Digital Humanities and Intelligent Computing

INTELLIGENT COMPUTING FOR CULTURAL HERITAGE

GLOBAL ACHIEVEMENTS AND CHINA'S INNOVATIONS

Edited by
Xiaoguang Wang, Marcia Lei Zeng,
Jin Gao, and Ke Zhao

ROUTLEDGE

Intelligent Computing for Cultural Heritage

This book offers a global perspective on the latest advancements and trends in digital humanities and intelligent computing of cultural heritage, covering both academic research and case studies within cultural institutions.

This edited volume brings together views and practices from different regions, including Asia, Europe, Africa, North America, and Australia. It offers innovative approaches and case studies related to humanities data and digital methods, with a focus on digital humanities research and pedagogy and cultural heritage organisation and preservation, in particular the development of digital knowledge repositories and methods for digital intelligence in cultural heritage. Each case study highlights unique cultural characteristics and academic histories, resulting in diverse development priorities and thematic directions. However, this diversity can also lead to imbalances and isolation within the field. To gain a better understanding of the complex trends in the development of the digital humanities, this book offers valuable insights from case studies and research practices, showcasing global contributions from scholars and institutions.

This title will appeal to scholars and students of digital humanities and information science, particularly those studying heritage management and intelligent computing. Professionals working at the intersection of technology and cultural heritage will also find this book of great interest.

Xiaoguang Wang is Professor and Vice Dean of the School of Information Management and Executive Vice Dean of the Big Data Institute, Director of the Intellectual Computing Laboratory for Cultural Heritage, and Director of the Centre for Digital Humanities at Wuhan University, China. His research interests are digital asset management, knowledge organisation, semantic publishing, and digital humanities.

Marcia Lei Zeng is Professor of Information Science at Kent State University (USA), with a PhD from the University of Pittsburgh (USA). Her research interests include knowledge organisation systems, metadata, semantic technologies, and digital humanities.

Jin Gao is Lecturer in Digital Archives at University College London, UK, and Research Fellow at the Victoria and Albert Museum. Gao's research interests focus on Digital Humanities history, network analysis, and digital cultural heritage.

Ke Zhao is a PhD researcher in Publishing Science at the School of Information Management, Wuhan University, China. Her research interests include digital humanities, digital publishing, digital storytelling, and human-centred design.

Digital Humanities and Intelligent Computing

The biennial book series, *Digital Humanities and Intelligent Computing*, provides a global perspective on the latest advancements and trends in digital humanities and intelligent computing related to cultural heritage. It covers both theoretical foundations and practical applications, offering a global view with a specific focus on China. As the digital humanities continue to flourish, research in different countries and regions prioritises their distinct cultural characteristics and scholarly traditions, resulting in a rich diversity of developments and thematic directions within the field, while underlining the need to further advance digital humanities research in cultural heritage on a global scale.

Digital Humanities and Intelligent Computing offers a unique opportunity to explore new frontiers and pioneer innovative research. The series will feature edited volumes that not only track global progress, explore innovations in technologies, methods, and tools, but also capture trends in the development of digital humanities and intelligent computing. Over the coming years, *Digital Humanities and Intelligent Computing* will provide digital humanities researchers as well as scholars from diverse disciplines (such as library and information sciences, arts and humanities, computer science, cultural studies, music, etc.) with access to current advances and further research in fields related to digital humanities and intelligent computing.

Strands of book titles

Global achievements and China's innovations will be the perspectives of each book.

Strands of book titles to be included in the series (examples only, non-exhaustive):

- Intelligent Computing for Cultural Heritage
- AI and Smart Data for Cultural Heritage
- Digital Storytelling and Creative Curation
- Digital Methods for Humanities
- Inclusive and Sustainable Laboratories for Humanities

Please contact the series editors, Xiaoguang Wang (Wuhan University, China), Marcia Lei Zeng (Kent State University, US), Jin Gao (University College London, UK), Ke Zhao (Wuhan University, China), and Taylor & Francis Publisher Lian Sun (Lian.Sun@informa.com), if you have an idea for a book for the series.

Titles in the series currently include:

Intelligent Computing for Cultural Heritage
Global Achievements and China's Innovations
Edited by Xiaoguang Wang, Marcia Lei Zeng, Jin Gao, and Ke Zhao

Intelligent Computing for Cultural Heritage

Global Achievements and China's Innovations

Edited by
Xiaoguang Wang, Marcia Lei Zeng, Jin Gao, and Ke Zhao

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

This book is funded by the major project “Resources Development and Service of Cultural Heritage Smart Data” (Grant No. 21&ZD334) supported by the National Social Science Foundation of the People’s Republic of China, and the national key research and development program “Research and Development of Key Technologies and Platforms for Digital Services of National Archives of Publications and Culture Collection Resources (Typical Application Demonstration)” (Grant No. 2022YFF0901902) supported by the Ministry of Science and Technology of the People’s Republic of China.

First published 2025
by Routledge
4 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
605 Third Avenue, New York, NY 10158

Routledge is an imprint of the Taylor & Francis Group, an informa business

The Open Access version of this book is funded by the Intellectual Computing Laboratory for Cultural Heritage, Wuhan University (Laboratory of Philosophy and Social Sciences, Ministry of Education, China)

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-032-70540-8 (hbk)
ISBN: 978-1-032-78358-1 (pbk)
ISBN: 978-1-032-70721-1 (ebk)

DOI: 10.4324/9781032707211

Typeset in Times New Roman
by codeMantra

Contents

Figures

Tables

Contributors

Anton Anikin is a PhD and an Associate Professor in the Software Engineering Department at Volgograd State Technical University. His primary research interests include ontology learning, natural language processing (NLP), intelligent tutoring systems, and decision-making processes. Since 2004, Dr. Anikin has been actively engaged in e-learning research. His latest work explores the application of ontologies within Intelligent Tutoring Systems and the NLP-based automatic generation of ontologies across various domains.

Adam Crymble is an Associate Professor of Digital Humanities in the Department of Information Studies at UCL. He also serves as the Deputy Director of the UCL Centre for Digital Humanities. His research interests are in the study of migration and community, and the ways that technology provides new problems or opportunities for the pursuit of that knowledge.

Huiling Feng is a National First-class Professor at Renmin University of China's School of Information Resource Management and Ex-Executive Vice-President of Renmin University of China. She serves as Deputy Director of the School Council and Director of Research Centres for Humanistic Beijing and Digital Humanities, leading the "Beijing Memory" digital project since 2013.

Jin Gao is a Lecturer in Digital Archives in the Department of Information Studies at UCL and a Visiting Research Fellow at the Victoria and Albert Museum. Her research interests lie primarily in the history of the Digital Humanities communities, social network analysis, digital cultural heritage, provenance archive studies, and data standards in GLAM.

Steven Gray is a Professor of Spatial Computation at UCL's Centre for Advanced Spatial Analysis, a Director of UCL Digital Humanities, and a Fellow at the Alan Turing Institute (The UK's National Centre for Data Science and AI) where he co-leads Turing's Tools, Practices and Systems Programme. His research interests include large-scale data collection and analysis, human-computer interaction, ubiquitous computing, and on-device machine learning for mobile systems. Through his software engineering outreach, he is a recognised Google Developer Expert for Google Cloud and Maps, Google Cloud Champion Innovator for Data Analytics, and Modern Architecture and Machine Learning.

Xilong Hou is an Associate Professor at the School of Communication, Qufu Normal University. He completed his postdoctoral research at the School of Information Management, Wuhan University, and holds a PhD in management from Central China Normal University. His research interests include knowledge organisation systems, linked data, digital humanities, and cultural heritage.

Xiao Hu is an Associate Professor in the Faculty of Education at the University of Hong Kong (HKU). Her research expertise is in learning analytics and cultural informatics. With an award-winning teaching philosophy, Dr. Hu inspires students to promote cultural heritage digitally. Dr. Hu has led publicly funded research projects, receiving multiple awards in international conferences.

Eero Hyvönen is a Professor in the Department of Computer Science at the Aalto University and a Director of the Helsinki Centre for Digital Humanities (HELDIG) at the University of Helsinki, Finland. A major theme since 2001 in his research has been the development of the national-level Semantic Web infrastructure and its applications in different domains.

Humphrey Keah is an Information & Knowledge Management Consultant at Rightpoint Information Services Ltd., Nairobi, Kenya. His professional experience spans international organisations including the UN, public and private sector organisations, and teaching experience at the university.

Joseph Kiplang'at is the Deputy Vice-Chancellor Academic Affairs and Research, Africa International University, Karen, Kenya. Prof. Kiplang'at is a Professor of Information Science with over 27 years of teaching and research experience in the university.

Jueying Lei is a PhD candidate at the School of Information Management, Wuhan University. Her research interests lie primarily in digital publishing, digital humanities, and digital cultural heritage.

Chengyong Liu is a Member of the 14th National Committee of the Chinese People's Political Consultative Conference (CPPCC), Secretary of the Party Committee, Curator (Director) of the China National Archives of Publications and Culture (National Archives of Publications and Culture Data Centre), and Vice-Chairman and Secretary-General of the National Press and Publication Information.

Ying-Hsang Liu is a Researcher in the Department of ALM at Uppsala University in Sweden and a Visiting Scholar at Professorship of Predictive Analytics, Chemnitz University of Technology in Germany. He holds a PhD in Information Science from Rutgers University, and his research lies at the intersections of knowledge organisation, interactive information retrieval, and human information behaviour in various domains, including digital humanities.

Linqing Ma is an Associate Professor at Renmin University of China's School of Information Resource Management, with an MA and a PhD in Archival Science from Renmin University of China. She is a Visiting Scholar at UBC and a

Research Fellow at the Key Laboratory of Data Engineering and Knowledge Engineering, MOE, China, focusing on digital records management and archival science theories.

Rongqian Ma is an Assistant Professor of Information and Library Science at the Luddy School of Informatics, Computing, and Engineering of Indiana University Bloomington. Her research focuses on digital humanities, visualisation, and scholarly communication. Previously, she earned her PhD in Library and Information Science from the University of Pittsburgh, USA.

Simon Mahony is a Professor of Digital Humanities at Beijing Normal University, Emeritus Professor at University College London, and Visiting Professor at Peking University. His research focus is on digital humanities with specific interests in education, information studies, equality, diversity, inclusion, and the open agenda.

Tianjiao Qi is a Lecturer at Renmin University of China's School of Information Resource Management and Visiting Scholar at UCLA and also a Member of the Digital Humanities Research Center at Renmin University of China. She leads youth and postdoctoral projects, and has published over 20 papers in digital resource organisation and digital humanities.

Yi Su is the Director of the Digital and Information Department and the Institute for Digital Application of Cultural Assets at the Palace Museum. Her research interests focus on the application of technologies related to digital museums and smart museums.

Shigeo Sugimoto is a Professor Emeritus of the University of Tsukuba. He served as a Faculty Member at Faculty of Library, Information and Media Science (iSchool@Tsukuba) over 30 years. He received his PhD in Information Science from Kyoto University in 1985. He is currently a governing board member of the Dublin Core Metadata Initiative and an executive director of Japan Library Association.

Jeremy Tzi Dong Ng is a PhD Candidate at the Faculty of Education, the University of Hong Kong (HKU). His research interest lies in technology-enhanced humanities education, and his doctoral research project focuses on how analytic-supported virtual reality creation facilitates cultural heritage education. He aspires to democratise digital content creation to under-resourced and non-tech-savvy learners.

Xiaoguang Wang is a Professor and Vice Dean of the School of Information Management and Executive Vice Dean of the Big Data Institute, Director of the Intellectual Computing Laboratory for Cultural Heritage, and Director of the Centre for Digital Humanities at Wuhan University, China. His research interests are digital asset management, knowledge organisation, semantic publishing, and digital humanities.

Yujue Wang is a Professor at the School of Information Management, Wuhan University. She also holds the position of Director of the Sino-French Humanities

Research Centre at Wuhan University and serves as a Consultant to the National Commission for UNESCO of the People's Republic of China. Her research interests include digital heritage, digital memory, archival memory theory, world heritage and documentary heritage, and Sino-French humanistic exchanges.

Claire Warwick is a Professor of Digital Humanities in the Department of English at Durham University, having been Pro-Vice-Chancellor: Research 2014–2019. Before coming to Durham, she was Head of UCL Department of Information Studies and Director of the UCL Centre for Digital Humanities, and began her academic career as a Lecturer at Sheffield University's iSchool. Her research is concerned with the way that digital resources are used in the humanities and cultural heritage and in reading behaviour in physical and digital spaces.

Mengjuan Weng is a PhD student in Information Science at the School of Information Management, Wuhan University. Her research interests include designing knowledge organisation system, digital preservation of ancient books, semantic publishing, and exploratory visual analytics.

Chiranthi Wijesundara serves as an Academic Librarian at the University of Colombo and holds a position as a Visiting Lecturer at the National Institute of Library and Information Sciences (NILIS), University of Colombo, Sri Lanka. In 2019, she earned her Doctorate in Informatics from the Graduate School of Library, Information, and Media Studies at the University of Tsukuba in Japan. Additionally, Chiranthi possesses two Master's degrees, one in Geo-Informatics and another in Library and Information Science. Her research interests encompass metadata modelling for digital archives, linked data, heritage informatics, and geo-informatics.

Marcia Lei Zeng is a Professor of Information Science at Kent State University (USA), with a PhD from the University of Pittsburgh (USA). Her research interests include knowledge organisation systems, metadata, semantic technologies, and digital humanities. She has authored six books and over 100 research papers.

Qi Zhang is the Director of the Digital Resources Department, China National Archives of Publications and Culture, and General Manager of National New Book Information Magazine Limited Liability Company.

Ke Zhao is a PhD researcher at the School of Information Management, Wuhan University, China, and a Visiting Postgraduate Research Fellow at the University of London. She holds an MSc in Digital Humanities from UCL, and previously earned a BFA in Film Studies and a BEng in Digital Production from China and South Korea. Her research interests preliminary focus on digital humanities, digital storytelling, digital publishing, and human-centred design.

Liang Zhao is an Associate Researcher at the School of Information Management, Wuhan University. Her research interests include Big Data, computational social science, social media and cultural computing, and data science.

Wenhong Zhou is an Associate Professor at Renmin University of China's School of Information Resource Management, and holds a PhD in Archival Science from Renmin University of China, including studies at UBC. He specialises in participatory recordkeeping on online platforms and leads a project funded by the National Social Science Fund of China, with over 100 publications in various languages.

Ziqi Zhou is a Master's student at the School of Information Management, Wuhan University. Her research interests focus on digital humanities, cultural heritage digitisation, and digital twins of tangible cultural heritage.

Preface

Digital humanities (DH) is an interdisciplinary field that uses computational methods and technologies to conduct humanities research while also applying humanistic lens to study the digital applications (Berry & Fagerjord, 2017). Globally, an increasing number of scholars have been embracing technologies and methodologies such as computing, geographic information systems (GIS), data visualisation, digital media, and artificial intelligence (AI) to address research questions and critically evaluate and debate such interdisciplinary efforts in the digital era. In China, significant developments have been made in the field of DH. The first DH research centre in the mainland China was established at Wuhan University in 2011. By December 2023, over 20 research centres dedicated to DH had been established across the region. The 5th Chinese Digital Humanities (CDH2023) Conference, convened in December 2023 in Wuhan, was themed "Symbiosis of Digital and Physical: Envisioning the Future of Digital Humanities". CDH2023 covered a diverse range of topics, encompassing smart data, digital publishing, revitalisation of ancient texts, digitisation of cultural heritage, AI, and digital gaming, as well as studies in linguistics, literature, music, art, history, and archaeology. The discussions and presentations vividly demonstrated the latest developments and achievements in China's DH research, highlighting its growing influence and significance in the academic field.

The continued flourishing of DH has also infused innovative thinking towards the digital revitalisation of cultural heritage, providing traditional humanities scholars with new application scenarios, digital tools, and research perspectives (Kenderdine, 2016, p. 23). This promotes the extensive application and advancement from digital cultural heritage research to DH research in computational social sciences (Münster et al., 2019; Gefen et al., 2021). Research from different nations and regions prioritises their unique cultural characteristics and academic histories, which generates diverse developments and thematic directions (Mahony, 2018; Münster & Terras, 2020). On the rich foundation of cultural heritage digitisation and advancement of computational methods, it is important to consider how to further promote DH research of cultural heritage using computing technologies and critical analysis to support future practices with DH thinking and contribute to the cultural and spiritual connotations of cultural heritage.

In view of this, this book proposes the concept of intelligent computing of cultural heritage (ICCH). It integrates DH and advanced technologies such as Big Data, AI, and virtual reality to capture, record, analyse, and represent the information and knowledge inherent within cultural heritage. The ultimate goal is to digitally revitalise cultural heritage with sustainable approaches, facilitating its transition from the physical world to digital spaces. It enables extraction of historical and cultural meanings from cultural heritage (Zeng, 2017), thereby creating a foundation for its broader dissemination, immersive viewing experiences, and critical engagement (Burdick, 2016, p. 44). This concept aims to advance the digital transformation of human civilisation and ensure the sustainable continuity of traditional culture in digital environments, opening new avenues for multidisciplinary research and education.

To support the argument for the importance of DH and ICCH, this "Digital Humanities and Intelligent Computing for Cultural Heritage: Global Developments and China Solutions" is planned to be a book series organised by the Intelligent Computing Laboratory for Cultural Heritage (ICLCH) at Wuhan University, originated from the Centre for Digital Humanities at Wuhan University. This book series aims to provide a comprehensive global perspective on the latest advancements and trends in these fields. The first volume brings together views and practices from various regions, including Asia-Pacific, Europe, Africa, and North America, offering a truly international outlook on the subject matter. It promotes cross-cultural dialogue, knowledge exchange, and collaboration, ultimately contributing to the future interdisciplinary research of DH, DH research on cultural heritage, and the development of ICCH.

This volume is divided into two parts, Global Creative Approaches (Part I) and Innovative Practice in China (Part II). It consists of 11 chapters that contribute to the DH and ICCH, providing valuable insights from case studies and research practices, highlighting global contributions from scholars and institutions, with a specific emphasis on China in the second part.

Part I. Global creative approaches

The first part of this book comprises six chapters that collectively examine the global developments and challenges in the field of DH and cultural heritage. It explores various perspectives, including open humanities data in cultural heritage sectors in the United States, data modelling for digital archiving of intangible and experiential entities, publishing and studying military history on the Semantic Web, global trends in DH education, the history of DH development and collaboration within a university, and the digital readiness of cultural heritage institutions in a country. These chapters provide a comprehensive overview of the state of DH on a global scale, emphasising the importance of collaborative efforts and creative solutions.

Chapter 1, "Towards an Open Humanities Data: Current States, Challenges, and Cases," by Rongqian Ma from Indiana University Bloomington, USA, discusses practices of open humanities data, including data access, curation, modelling, and

sharing, drawing upon research cases in the cultural heritage institution settings in the United States. The chapter also draws upon two case studies informed by the author's own research expertise to further reflect on the challenges and prospects of facilitating practices of open humanities data.

Chapter 2, "Data Modelling for Digital Archiving of Intangible and Experiential Entities," by Shigeo Sugimoto from the University of Tsukuba, Japan, and Chiranthi Wijesundara from the University of Colombo, Sri Lanka, discusses data modelling for the digital archiving of intangible and experiential entities such as intangible cultural heritage (ICH), performing arts, and disasters. The chapter introduces several key terms and previous models, and presents a data model for the digital archiving of intangible and experiential entities, which is a domain-agnostic conceptual model.

Chapter 3, "Military History on the Semantic Web: Lessons Learned from Developing Three In-use Linked Open Data Services and Semantic Portals for Digital Humanities," by Eero Hyvönen from Aalto University and the University of Helsinki, Finland, reviews three in-use systems for publishing and studying military history on the Semantic Web, namely, WarSampo, WarVictimSampo 1914–1922, and WarMemoirSampo. The chapter discusses the lessons learnt from the development of these systems based on the Sampo model, which is a set of principles developed in the process of creating and publishing Knowledge Graphs (KG) as LOD services with semantic portal User Interfaces (UI).

Chapter 4, "Identifying Main Topics of Digital Humanities Courses Across Countries: A Topic Modelling BERTopic Technique," by Ying-Hsang Liu from Uppsala University and Chemnitz University of Technology, and Anton Anikin from Volgograd State Technical University, uses the BERTopic (based on the transformers approach and c-TF-IDF) topic modelling technique to analyse a corpus of 563 course descriptions in DH programmes globally. The chapter provides a nuanced view of the DH course topics across countries, aiming to refine the understanding of the evolution of the DH field as reflected in the courses offered in DH programmes.

Chapter 5, "Empowering Global Engagement: The Development of Digital Humanities Research and Pedagogy at University College London (UCL)," by Jin Gao, Adam Crymble, Simon Mahony and Steven Gray from UCL, and Claire Warwick from Durham University, examines the institutional history and evolution of DH at UCL and its engagement with Chinese DH, positioning itself within the broader context of this edited volume. It emphasises UCL's global engagement with DH communities, arguing its transformative impact on research and teaching in this field.

Chapter 6, "Digitalisation-preparedness of Cultural Heritage Institutions – To What Extent Is Africa Ready?", by Joseph Kiplang'at from Africa International University and Humphrey Keah from Rightpoint Information Services Ltd in Kenya, explores the digitisation landscape of Cultural Heritage Institutions in Kenya with a view of determining their preparedness for digitalisation. Through questionnaires, interviews, and literature review, the chapter reveals some positive nascent effort towards AI-readiness in Kenya with government involvement in

acquisition of world-class digitisation infrastructure to realise the national aspiration of digitising 5 billion records as spelt out in the National Digital Master Plan. The study proposes a way forward regarding digitisation projects in Kenya and Africa at large.

Part II. Innovative practices in China

The second part of this book stands out by presenting China's contributions to DH and ICCH. It covers digitising efforts, digital memory construction, the use of digital twin technology for tangible heritage, experimental methodologies for digital revitalisation, and advanced digital reconstruction of ancient texts, offering a comprehensive and valuable view of the field.

Chapter 7, "Digitising Cultural Heritage in Hong Kong: An Overview," by Jeremy Tzi Dong Ng and Xiao Hu from the University of Hong Kong, surveys cultural heritage digitisation projects in Hong Kong both in the literature and in practice, summarising in particular the approaches adopted in these projects and reporting on their development trends. The chapter discusses the multifaceted facilitating factors, limitations, and challenges of these projects, and recommends best practices for future projects in Hong Kong and beyond.

Chapter 8, "Digital Memory Construction for Cultural Heritage: Methodology and Applications in China," by Huiling Feng, Linqing Ma, Tianjiao Qi, and Wenhong Zhou from Renmin University of China, elucidates the understanding of digital memory, introduces a theoretical framework for constructing digital memory oriented towards cultural heritage, and presents an example of its application in the 'Beijing Memory' project. The chapter discusses prospective developments and challenges in the digital memory field, aiming to furnish a scalable solution for an exhaustive exploration, in-depth interpretation, digital representation, and wide-ranging dissemination of cultural heritage in the digital era.

Chapter 9, "China's Exploration and Experience in Digital Twinning of Tangible Cultural Heritage," by Yi Su from the Palace Museum, Ziqi Zhou, Liang Zhao, and Yujue Wang from Wuhan University, discusses progress and experiences in applying digital twin technology to tangible cultural heritage sites in China. The chapter analyses three typical digital twin applications of cultural heritage for the Shanghai History Museum, the Longmen Grottoes and the Yellow Emperor's Native Place on the three typical carriers of tangible cultural heritage: museums, heritage sites, and scenic spots, as well as the pre-practice exploration of the digital twin for the large-scale comprehensive museum – the Palace Museum. The chapter also discusses the future development paths of tangible cultural heritage digital twin from four perspectives: application scenarios, data models, relevant technologies, and cultural value.

Chapter 10, "Digital Deduction Theatre: An Experimental Methodological Framework for the Digital Intelligence Revitalisation of Cultural Heritage," by Ke Zhao and Xiaoguang Wang from Wuhan University, Qi Zhang and Chengyong Liu from the China National Archives of Publications and Culture, examines an experimental methodological framework for digital intelligence revitalisation

of cultural heritage by exploring the digital deduction theatre (DDT) driven by digital intelligence and DH. The chapter captures the changes in humanities research powered by digital intelligence, analyses the unique features of DH laboratories, reflects on their development and characteristics, and aims to present a new paradigm for humanities experimentation driven by digital intelligence and DH that support cultural heritage research infused with historical, cultural, scientific, and artistic knowledge.

Chapter 11, "Combining Ontology and Nanopublication Models to Reconstruct Digital Commentaries on Ancient Chinese Books," by Mengjuan Weng, Xiaoguang Wang, and Jueying Lei from Wuhan University, along with Xilong Hou from Qufu Normal University, proposes a path for reconstructing commentaries, with a particular focus on refining the representation of semantic data within commentaries, and illustrates the practical application of digital reconstruction techniques for commentaries through a case study of Wen Fu and its commentaries. The case not only serves as a pilot application of methods and techniques for digitally reconstructing ancient books but also offers guidance for building a knowledge base and semantically publishing ancient texts.

Bibliography

Berry, D. M., & Fagerjord, A. (2017). *Digital humanities: Knowledge and critique in a digital age*. John Wiley & Sons.

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnapp, J. (2016). *Digital_Humanities*. MIT Press.

Gefen, A., Saint-Raymond, L., & Venturini, T. (2021). AI for digital humanities and computational social sciences. In B. Braunschweig & M. Ghallab (Eds.), *Reflections on artificial intelligence for humanity* (pp. 191–202). Springer. https://hal.science/hal-03043393

Kenderdine, S. (2016). Embodiment, entanglement, and immersion in digital cultural heritage. In S. Schreibman, R. Siemens & J. Unsworth (Eds.), *A new companion to digital humanities* (pp. 22–41). John Wiley & Sons.

Mahony, S. (2018). Cultural diversity and the digital humanities. *Fudan Journal of the Humanities and Social Sciences, 11*(3), 371–388. https://doi.org/10.1007/s40647-018-0216-0

Münster, S., Apollonio, F. I., Bell, P., Kuroczynski, P., Di Lenardo, I., Rinaudo, F., & Tamborrino, R. (2019). Digital cultural heritage meets digital humanities. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42*(2/W15), 813–820. https://doi.org/10.5194/isprs-archives-XLII-2-W15-813-2019

Münster, S., & Terras, M. (2020). The visual side of digital humanities: A survey on topics, researchers, and epistemic cultures. *Digital Scholarship in the Humanities, 35*(2), 366–389. https://doi:10.1093/llc/fqz022

Zeng, M. L. (2017). Smart data for digital humanities. *Journal of Data and Information Science, 2*(1), 1–12. https://doi.org/10.1515/jdis-2017-0001

Acknowledgements

We express our profound gratitude to the numerous contributors from around the world, whose diverse perspectives and scholarly expertise have greatly enriched the pages of this book. We also sincerely appreciate the time and effort they have devoted to this book. Special acknowledgment is due to Humphrey Keah, whose assistance in English editing has been invaluable. We also extend our sincere thanks to the editors at Routledge. Your insightful feedback and professional guidance have been instrumental in improving this book.

We would also like to thank the funding of this publication by the major project "Resources Development and Service of Cultural Heritage Smart Data" (Project code: 21&ZD334) supported by the National Social Science Foundation of China and the National Key Research and Development Program "Research and Development of Key Technologies and Platforms for Digital Services of National Archives of Publications and Culture Collection Resources (Typical Application Demonstration)" (Project code: 2022YFF0901902) supported by the Ministry of Science and Technology of China.

Additionally, we are excited to announce the book's launch at University College London (UCL) in 2024, a fitting venue that symbolises the spirit of this publication. UCL's distinguished contributions to digital humanities research and teaching, specifically through its Department of Information Studies (DIS) and the UCL Centre for Digital Humanities (UCLDH), have significantly influenced this volume. The institution's commitment to global engagement and digital innovation resonates with the themes of our work, which will not only celebrate the current achievements in the field but also mark a stepping stone towards a more connected future of digital humanities.

Xiaoguang Wang, Marcia Lei Zeng, Jin Gao, Ke Zhao

Part I

Global creative approaches

1 Toward an open humanities data

Current states, challenges, and cases

Rongqian Ma

1 The open approaches: Toward the open humanities data

The collective idea of open approaches, including concepts such as open data, open-access, and open sciences, has been widely discussed and promoted as a contributing factor to various aspects of society (Corrall & Pinfield, 2014). Corrall (2016) discussed three major typologies of the open approaches, which include open content, open infrastructure, and open process. Following this framework, open initiatives such as open data and open access movements provide opportunities for better accessibility and transparency of data and content, while practices such as open research and open sciences explore how distinct procedures and knowledge infrastructures and constructs address disciplinary challenges and lead to more transparent, equitable, and accessible research environment. The open data movement—a paradigm shift emphasizing transparency, reproducibility, and widespread accessibility of data and research—has demonstrated tremendous benefits to fields such as scientific research, government, journalism, and the general aspects of society (Baack, 2015; Janssen et al., 2012; Kitchin, 2014; Attard et al., 2015). The adoption of open data principles, however, involves addressing unique challenges, "myths," and prospects of various fields (Janssen et al., 2012).

An important component for promoting open data is providing open-access to resources. Originating in the late 20th century as a response to increasing journal subscription costs, the open access movement with significant milestones like the Budapest Open Access Initiative in 2002 and the Berlin Declaration in 2003 has greatly accelerated the progress of open data across various domains (Suber, 2012). Particularly, the open access movement promotes free and unrestricted access to scholarly research through various models and has revolutionized science and scholarly communication by accelerating the dissemination of research, fostering global collaboration and innovation, and promoting transparency of information among researchers and the general public (Mann et al., 2009; Suber, 2002; Björk et al., 2014). In congruent with the open access movement, *open science* has been a concept that is widely acknowledged and embraced by universities, government, information institutions, and the broader public (McKiernan et al., 2016; Lyon, 2009). Sometimes used interchangeably with open research in the literature, open science typically refers to the "practices and norms of more open and transparent

DOI: 10.4324/9781032707211-2

communication and research in scientific disciplines and the discourse on these practices and norms" (Knöchelmann, 2019). The openness discussed here involves various aspects of research, such as the infrastructures that drive the production, as well as the principles and practices of communicating, sharing, and evaluating scientific scholarship (Fecher & Friesike, 2014). Such practices, as illustrated by research literature, also have the potential to promote scientific knowledge communication as well as advance researchers' careers (McKiernan et al., 2016). For example, open science research practices help early career researchers gain more citations and exposure for their work, and establish collaborative work with other scholars (Eysenbach, 2006; McKiernan et al., 2016; Allen et al., 2019). A wide variety of practices and infrastructures implemented have also facilitated open science initiatives, such as preprints, open peer review practices, and open licenses (Knöchelmann, 2019).

By contrast, the discourse of open research and scholarship in the humanities has become a recent development (Knöchelmann, 2019). Multiple reasons and issues might have led to such a result. For example, Arthur and Hearn (2021) argued that humanities research often focuses on in-depth analysis of well-defined topical issues, such as "studies in local regional contexts and language-specific communities," and therefore, is less generalizable and applicable compared to scientific fields. In addition, the priority placed on books over other forms of publications such as journal articles and conference papers contributes as an important barrier to the openness of the humanities (Severin et al., 2018). Cost can be another issue. Compared to open-access scientific publications that usually require article processing charges (APCs) of $2,000–$3,000, open-access humanities monographs and publications may draw the APCs up to $130,000 (Maron et al., 2016). Despite the challenges and barriers, new approaches to research and emerging areas of inquiry have created opportunities to accelerate open humanities research. Digital humanities is such an example. The digital tools, archives, and platforms created by researchers in this broadly defined area have offered ways to communicate humanities knowledge and scholarship among wider audiences, facilitating openness in research (Arthur & Bode, 2014; Schreibman et al., 2016). With the rapid growth of digital humanities as an interdisciplinary inquiry and an increasing scale of adopting digital methods and data-driven research in the humanities, concerns about data accessibility and reuse have become a hotspot of scholarly discussion.

Given the inherent peculiarities in humanities and cultural data, research methodologies, and interpretative analyses predominant in humanities, the integration of open data principles prompts a unique set of opportunities and impediments. This chapter endeavors to provide an investigation of the prevailing status quo of open humanities data practices, following the framework of open content, process, and infrastructure. More specifically, this chapter focuses on the US context and aims to elucidate the multifaceted challenges confronting scholars and information professionals working with such datasets in research and cultural heritage institutions, before drawing on two empirical case studies to further reflect on the collective efforts to build open, accessible, and reusable humanities data.

2 Humanities and cultural data

With the development of the open data movement, its benefits, as well as concerns, issues, and challenges have been widely discussed, particularly within scientific research contexts. For example, the availability of reusable and high-quality open datasets is perceived to have the potential to accelerate research, especially in supporting large-scale, data-driven, and computational research. Scholars in information science and data curation domains have particularly discussed that open data have the potential to make methods more transparent and accessible, maintain accountability for both researchers and funders, promote replication and novel use of shared data corpora, and facilitate interdisciplinary research collaborations (Uhlir & Schröder, 2007; Janssen et al., 2012). However, the adoption of open data practices in the humanities research context bears unique challenges. For example, the varying notions and understandings of "data" in the humanities research context, the interpretive nature of humanities research, properties associated with data such as data format and type, copyright, and quality, as well as the lack and quality of infrastructural support contribute to the difficulty of implementing open data practices in the humanities research context.

While the notion of data is readily embraced by scientific disciplines and communities, its role in humanities research has perpetuated ongoing discourse. Humanities researchers seem to be less straightforward in using the word "data" to refer to their primary sources and research materials; rather, they tend to be more specific in identifying their data resources as documents, manuscripts, or images, for instance (Schöch, 2013). As Borgman (2017) highlighted, humanities and cultural data present an "innumerable" range and possess "fuzzy" boundaries with "publications." Humanities scholars draw on documents, artifacts, and materials like newspapers, correspondence, manuscripts, images, and archaeological objects as essential data sources. In addition, these humanities and cultural datasets usually focus on very particular, specific topics, and researchers usually need to craft and apply unique data curation practices according to the contexts of individual studies. The identification and use of such source materials as research data for humanistic inquiry are also considered as an esteemed form of scholarly endeavor and contribution (Borgman, 2017).

The uniqueness of humanities and cultural data also emanates from their detachment and contextual dissociation: cultural artifacts are subject to acquisition, displacement due to conflict, and relocation to museums or private collections. The dispersion of these artifacts is mirrored in the dispersion of records that document them, often held by their possessors, creating a distinctive landscape. This relationship between data and their contexts demands additional nuanced interpretation and evaluation of the datasets before reusing them in other contexts. In addition, since the use and discussion of humanities and cultural data usually involve documents displaced within contexts and the analysis of publications, copyright is an issue when it comes to making data widely available and accessible. Research literature has identified copyright as a potential barrier to facilitating open humanities research data at a large scale, particularly given that the process involves

stakeholders of institutions, publishers, as well as the constraints of infrastructural support (Arthur & Hearn, 2021).

While individualized approaches to data curation often work well to afford nuances for projects tackling humanities and cultural data, more infrastructural effort could be done from a library and information science perspective to expand research horizons in digital humanities. An increasing volume of scholarship has attempted to explore the evolution of digital humanities as a community of practice and as a dynamically shifting intellectual domain from its early development (Luhmann & Burghardt, 2021; Sula & Hill, 2019). More specifically, studies of this kind utilise bibliographic data and records about digital humanities and perform quantitative analyses of the datasets (Spinaci et al., 2022; Su et al., 2021; Gao et al., 2023; Wang, 2018; Tang et al., 2017). The bibliodata can also be valuable humanities data in this context. In addition, numerous digital humanities work that apply computational methods to study cultural phenomena and patterns demands more open, machine-readable, and reusable data in volumes (e.g., Underwood, 2019; Hu et al., 2023; Walsh & Antoniak, 2021). As illustrated in these studies, book reviews data and social media data, for example, also make emerging, valuable humanities and cultural datasets. These new forms of humanities datasets and quantitative approaches in digital humanities research raise challenges to researchers as to where and how to find quality data and reuse them effectively to address the identified questions. One barrier, as illustrated in the literature and the pervious work of the author, may come from the lack of centralized databases and digital infrastructures (Aschenbrenner et al., 2013; Li et al., 2023). In the following sections, I will introduce the major, existing infrastructural support to locate and reuse quality data in the broad field of humanities and cultural research, including but not limited to the publication of data papers, open data repositories, as well as semantic web technologies.

3 Infrastructural support for open and reusable humanities research data

Within the digital humanities communities, researchers have been working to provide infrastructural support to promote open humanities data and more effective data reuse for research. One of the major infrastructural supports for open humanities data comes from journals that focus on publishing *data papers* to make humanities datasets openly available. A data paper is a unique publication genre that focuses on providing detailed descriptions and contexts for a dataset as well as its possible uses (Schöpfel et al., 2020; Li et al., 2019). Compared to research articles that focus primarily on theoretical insight and analysis of the research subjects, data papers emphasize *data* as an important object as well as factor in research, illustrating "research workflows [that are] often undervalued, namely the creation, management, processing, access and usage issues around datasets" (McGillivray et al., 2022).

Data papers play a crucial role in promoting open, reliable, and reusable empirical data for scholarly purposes. More specifically, it promotes transparency of

research data, facilitates reproducibility of empirical research, while at the same time ensuring that credit is given to those who contribute to the creation and sharing of datasets as an invaluable part of scholarly activities. Data papers put data within specific research contexts, making datasets more usable and beneficial among researchers. Among scientific research communities, journals that specialize in data papers provide a major venue for researchers to publish their datasets as research outcomes. *Scientific Data* is one notable example. Launched in 2014 and published by the Nature group, *Scientific Data* is a peer-reviewed, open-access journal that focuses on publishing data descriptors and facilitating the sharing and reuse of scientific data. It covers all areas of natural sciences, medicine, engineering, and social sciences. As of October 14, 2023, the journal has published 2,936 data descriptors. An examination of the sample publications reveals the key elements for the data descriptors published in this journal, including the background, methodology of collecting, cleaning, modeling, and constructing data, descriptive statistics showcasing the shape and key characteristics of the data, the use cases, and contribution to research topics and fields. The structured data descriptors increase dataset reuse and reproducibility by highlighting the details and innerworkings of data processing and curation.

In the humanities research community, several open-access journals have committed to publishing data papers introducing humanities and cultural data. Table 1.1 gives an overview of the existing journals that publish data papers for humanities datasets. The open-access *Journal of Open Humanities Data* is one typical example. *Journal of Open Humanities Data* facilitates open humanities data by "featur[ing] peer-reviewed publications describing humanities data or techniques with high potential for reuse."[1] Through data papers and full research articles, researchers can publish the dataset to complement their research on a specific topic or subject matter, with detailed illustrations of the process of creating and curating the dataset. The journal has thus far published 63 data papers introducing humanities and cultural datasets of various subjects, which emphasize multiple elements: (1) the release of all datasets in open-access repositories, such as Harvard Dataverse and Zenodo; (2) a context section introducing the subject matter of the dataset, as well as the background for the research project; (3) the detailed illustration and explanation of methodologies used for data collection, processing, curation, and quality control; (4) a comprehensive data description which includes dataset format, dataset accreditation, and detailed metadata for the dataset; (5) and statements on reuse potential where the authors lay out the potential use cases for the dataset, the main contribution of the dataset to research topics and fields, and the broader impact the dataset can generate.

Journal of Cultural Analytics is another example of a digital humanities journal that publishes open humanities and cultural datasets as well as data papers. Among other types of publications such as research articles and debates, the journal has published multiple data papers introducing specific humanities datasets. Detailing computational methods of data creation and curation make ups most of the data papers in this journal. *Journal of Cultural Analytics*'s policy also details the data sharing policy and requires all data and code to be made publicly available in Dataverse

Table 1.1 List of journals that publish data papers and open datasets for humanities and cultural research

Journal title	*Description*
Journal of Open Humanities Data	The journal publishes peer-reviewed publications describing humanities data or techniques with high potential for reuse. The journal currently publishes two types of papers: Short data papers: contain a concise description of a humanities research object with high reuse potential. These are short (1,000 words) highly structured narratives and must conform to the Metapaper template. A data paper does not replace a traditional research article but rather complements it. Full-length research papers discuss and illustrate methods, challenges, and limitations in the creation, collection, management, access, processing, or analysis of data in humanities research, including standards and formats. These are intended to be longer narratives (3,000–5,000 words), which give authors the ability to contribute to a broader discussion. (https://openhumanitiesdata.metajnl.com/about)
Research Data Journal for the Humanities and Social Sciences (RDJ)	RDJ is a peer-reviewed, Diamond Open-Access e-journal designed to comprehensively document and publish deposited datasets and facilitate their online exploration. *RDJ* contains data papers: scholarly publications of medium length (with a maximum of 2,500 words) containing a description of a dataset and putting the data in a research context. (https://brill.com/view/journals/rdj/rdj-overview.xml)
Journal of Cultural Analytics	The *Journal of Cultural Analytics* is an open-access, peer-reviewed journal dedicated to the computational study of culture. It represents an intersection between the humanities and data science, focusing on the application of computational methods to analyse and interpret cultural texts and practices. According to its website, the journal features three major sections: "Articles profile peer-reviewed scholarship; Data Sets offer lengthy discussions of curatorial choices associated with new data sets relevant to cultural study; and Debates offer shorter more timely interventions into key discussions surrounding the computational analysis of culture." (https://culturalanalytics.org/)
Post45 Data Collective	The *Post45 Data Collective* peer reviews and houses post-1945 literary data on an open-access website designed, hosted, and maintained by Emory University's Center for Digital Scholarship. As detailed in the journal's submission guidelines, "each dataset must be accompanied by a contextual statement or 'data essay' that addresses (in 4,000 words or fewer) the significance of the data; the social and historical context of the data; how the data was collected, cleaned, and organized; and possible ethical concerns or misuses of the data. Data essays are published alongside datasets on the Post45 Data Collective website." (https://data.post45.org/submissions-2/)

to ensure access, reuse, as well as reproducibility of the identical dataset.[2] Other typical journals publishing open humanities data also include, for example, the Diamond Open Access journal *Research Data Journal for the Humanities and Social*

Sciences (RDJ) and specialized journals such as the literary journal *Post45 Data Collectives*, which focuses on post-1945 American literature data.

3.1 Data sharing repositories

Accompanied by the journal publications are the open data sharing infrastructures that help researchers facilitate research communication. The landscape of data sharing infrastructures for open humanities data has evolved significantly, influenced by advances in digital technologies and the growing emphasis on open-access and collaboration in academic research (Burrows, 2011; Kansa & Kansa, 2011). These infrastructures, ranging from digital libraries to specialized repositories, play a crucial role in facilitating the accessibility, preservation, and reusability of humanities data. The humanities benefit from a repositories and data sharing infrastructures like HathiTrust, Harvard Dataverse, and JSTOR Data for Research (DfR), to name a few. These platforms offer access to a vast array of materials including books, journals, manuscripts, and multimedia content, catering to diverse research needs. At the same time, these data sharing infrastructures ensure that researchers get the credit for publishing and sharing datasets as a research product and benefiting the scholarly communities. Table 1.2 provides a more detailed overview of a few existing infrastructures for humanities and cultural sharing.

The existing data infrastructures have offered numerous research opportunities for digital humanities, and the resources and tools made available by them have been increasingly adopted in individual research projects. For example, Downie et al. (2020) examined the potential of using HathiTrust Digital Library collections in musicology research. Bauder (2019) also highlighted the use of HathiTrust as a valuable data source to research early 19th-century library collections. Despite the fruitful results of the existing infrastructures, however, the variety of humanities and cultural data are still scattered across different resources, posing challenges to humanities researchers to collect and utilise available datasets. The lack of central, comprehensive data sharing infrastructures has also created problems for researchers who want to conduct large-scale, quantitative analyses of digital humanities as a research domain. For example, the digital humanities publication data, such as journal articles and conference papers, are not well indexed in existing databases such as the Web of Science or Scopus. Therefore, the analysis using digital humanities bibliodata has remained at a relatively small scale compared to other quantitative science studies, making it challenging to draw generalizable conclusions.

Information science researchers have made an effort to address this infrastructural challenge. For example, Weingart and Eichmann-Kalwara (2020) have developed the Index of Digital Humanities Conferences database to provide access to publications at the Digital Humanities annual conferences. Spinaci et al.'s (2022) work was another example where efforts were made to map available bibliographic records for digital humanities and create a comprehensive corpus for research. In the author's work with a collaborator, we have been working to establish a representative bibliographic data corpus to quantitatively examine the ecology of digital humanities as a cross-domain field and research conventions such as visual practices

Table 1.2 Available infrastructures that support open data in the humanities

Journal title	*Description*
Dataverse	The Harvard Dataverse Repository offers an open, cost-free platform for researchers of any field to store, share, reference, obtain, and examine research data. This customizable platform features individual Dataverse collections, serving as virtual repositories for dataset organization, management, and display. Researchers have the option to make their data publicly available or to limit access, with the ability to set specific usage terms. Published data is assigned a standardized citation with a DOI, ensuring the discoverability of metadata through search engines, even for restricted data. Harvard Dataverse stands as a pivotal archive for sharing data within the arts and humanities scholarly community. As of October 21, 2023, it houses approximately 36,166 entries in the arts and humanities category, encompassing data since 2008. Each entry provides vital metadata including unique identifiers, publication dates, titles, author details, subjects, keywords, and information about the depositor. (https://dataverse.harvard.edu/)
HathiTrust	HathiTrust is a collaborative digital library initiative that provides an expansive open data repository, primarily catering to the needs of humanities research. It was established as a partnership among academic and research institutions to preserve and provide access to a vast collection of digitised works. This repository is notably rich in historical and cultural texts, making it an invaluable resource for scholars, researchers, and students in the humanities. One of the key features of HathiTrust is its massive scale, encompassing millions of volumes of digitised books, periodicals, and manuscripts. This extensive collection includes works from numerous libraries and archives, offering a wide range of materials that span various periods, regions, and languages. Such diversity allows researchers to conduct comprehensive studies in fields like literature, history, arts, and linguistics. Moreover, the repository offers robust search and discovery tools, facilitating easy access to its contents. Users can perform full-text searches across the entire collection, a feature particularly beneficial for in-depth humanities research (York, 2012). (https://www.hathitrust.org/)
JSTOR Data for Research (DfR)	JSTOR Data for Research (DfR) is another infrastructure that offers technologies to facilitate research in the humanities and social sciences. Particularly, "JSTOR's Data for Research (DfR) program accommodates text analysis and digital humanities research by providing datasets for the journals, books, research reports, and pamphlets in the digital library." Powered by Constellate, a project of JSTOR labs, DfR also provides multiple data mining and text analysis features, custom dataset building, topic modeling methods, and data visualisation capabilities. Such exploratory features and tools contribute to the platform by enhancing usability of the available data. JSTOR's DfR has great impact on facilitating data-driven humanities research. (https://about.jstor.org/whats-in-jstor/text-mining-support/)

(Ma & Li, 2021; Li et al., 2023). In an attempt to expand our previous research on digital humanities visual use patterns, we apply computational techniques to extract figures and figure-related information from digital humanities publications to build a large-scale dataset for further analysis. Although we recognize the value of such

a dataset for research, copyright issues among many journals, articles, manuscripts, and archives may create obstacles to provide open-access to the dataset. This challenge also applies to broader contexts where records and documents are used as primary sources, as will be illustrated in the case studies in the following sections.

3.2 Linked open data and applications for the humanities research

Going beyond data papers and open data sharing repositories, Sematic Web technologies have approved to support access and use of large-scale humanities and cultural data. Linked Open Data (LOD) is one typical application in the humanities that have largely improved how researchers manage, share, and utilise data (Heath & Bizer, 2022). LOD connects data from various sources, allowing for a richer, more interconnected dataset. LOD typically uses RDF (Resource Description Framework) as a standard format for data interchange, enabling the description of relationships among data and providing a consistent yet flexible structure. LOD has been widely embraced as a best practice for publishing data on the web, and researchers have developed common criteria and standards to evaluate the quality of linked open data online (Nurmikko-Fuller, 2023). Figure 1.1 showcases the five-star scheme to evaluate LOD, which provides a framework for more interconnected, dynamic, and context-rich humanities and cultural research. Introduced by Tim Berners-Lee, the five-star scheme highlights a series of steps to publish data in a way that maximizes its utility and ease of integration on the web (Lebo et al., 2017):

- One Star (*): Make your data available on the web in any format, under an open license, ensuring the data is publicly accessible and can be used by anyone.

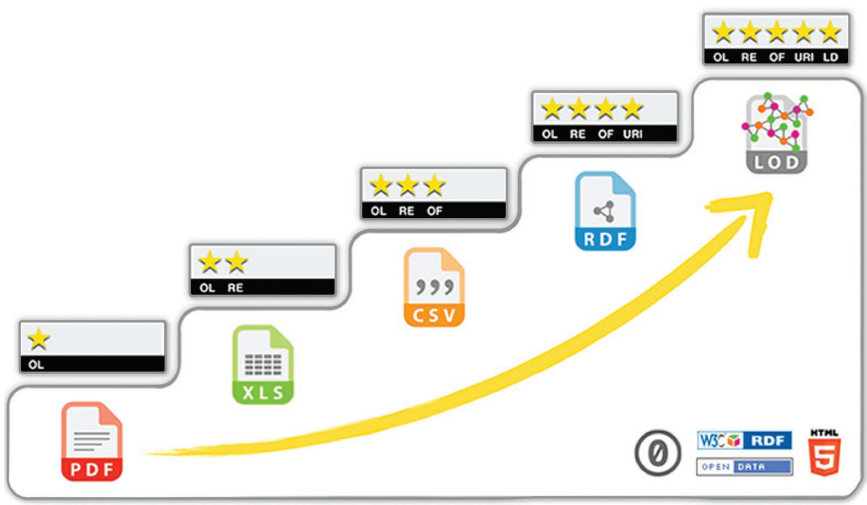

Figure 1.1 Five-star standard of Linked Open Data. Image from Wikimedia Commons. Creative Commons CC0 1.0 Universal Public Domain Dedication.

- Two Stars (**): Make the data available as structured data, in formats such as Excel or CSV.
- Three Stars (***): Use non-proprietary formats for structured data, such as CSV, which can be read by a wider variety of tools.
- Four Stars (****): Use URIs to denote things, which involves using a standardized way of addressing data, so that it can be referenced and linked.
- Five Stars (*****): Link your data to other data to provide context. This is the most advanced level, where the data is not only accessible, structured, and standardized but also connected to other relevant data on the web, making it part of a broader, interconnected data ecosystem.

One typical application of LOD can be the large-scale knowledge graphs (KGs) underpinning numerous databases and systems. KG is a structured way of representing knowledge in graphs, where entities (such as people, places, and events) are nodes, and relationships between them are edges. Such KGs offer more effective data integration from distributed resources, provide richer contextualization for datasets—and, therefore, usually leading to better reuse potential, and enhance the discoverability of humanities and cultural data. One concrete example of a large knowledge graph used in the humanities domain is DBpedia (https://www.dbpedia.org/), which represents a crowd-sourced community effort to extract structured content from the information created as part of the Wikipedia project. DBpedia aggregates and interlinks a vast amount of data from Wikipedia according to the RDF (Resource Description Framework) standards, covering topics as diverse as geography, history, science, and the arts. This makes it an incredibly rich resource for humanities research. In addition, DBpedia provides data in multiple languages, making it a global resource accessible to a wide range of researchers. Users from around the globe can retrieve information from the DBpedia KG using SPARQL query language, providing diverse and flexible approaches to exploring the data. The large-scale KG infrastructures applying LOD technologies unveil new insights and foster a deeper understanding of the complex tapestry of human culture and history.

3.3 Evaluate the value of open data infrastructures

Evaluating the potential impact of these open data infrastructures, however, is not a simple task. Traditional metrics for data infrastructure assessment, such as the usage statistics, citations, technical performance metrics, may not suffice to fairly evaluate the value as well as reflecting the meaningful use of humanities data and infrastructures (Mayernik et al., 2017; Ell & Hughes, 2013). More specifically, the use measured by quantities may not reflect the depth of use as well as the context in which the impact is generated. Alternative methods may need to be developed to better reflect the actual impact and use of such resources may be helpful to evaluate digital humanities infrastructures (Ma et al., 2023). For example, how should researchers or project curators decide if the chosen metadata scheme contributes to the best representation of the collection that suits the subject of representation

as well as its target audiences? How do data curators evaluate if the infrastructures (e.g., the interface, repository, exploratory tools) have added value to the data and collections? To address these problems that oftentimes need more individualized solutions, teams working with humanities and cultural data have continued to develop their own infrastructures and curating practices, to better address the specific needs of individual datasets as well as the challenges posed by the unique traits and diverse range of the humanities and cultural data.

4 Facilitating open and reusable humanities research data: Practices, workflows, and challenges

In the following section, I will use two research case studies to discuss data curation and modeling practices, which leveraged digital infrastructures creatively to augment users' interaction with humanities and cultural data, ultimately increasing their value and potential use. They also demonstrate the efforts undertaken by cultural heritage institutions—particularly academic and research libraries—to foster the creation, sharing, and reuse of open data for scholarly research in cultural studies and humanities domains. I will also discuss how *data modeling* and interaction augmented by the design and creation of *exploratory tools* and applications contribute to the better reuse of openly accessible humanities datasets. Exploratory tools refer to the interactive functions and features of an interface that enables users to explore the data in real time. Many such exploratory tools are embedded in existing digital humanities knowledge infrastructures, such as databases, digital scholarly editions, and collections (Ma et al., 2023). Such exploratory features have great potential to increase the value and use of humanities and cultural data. More specifically, we can evaluate how different types of exploratory tools and features facilitate interaction with data among various research communities, and how the exploratory tools along with data modeling and curation practices help shape storytelling and narratives of a humanities subject.

4.1 Contemporary Chinese village gazetteer (CCVG) data project

The Contemporary Chinese Village Gazetteer (CCVG) Data Project[3] was initiated and led by the East Asian Library (EAL) within the University of Pittsburgh Library System (ULS) in 2018. It aims to create a dataset that holds significant value in the fields of humanities and social sciences, utilizing EAL's extensive collection of village gazetteers (Zhang et al., 2020; He et al., 2022). These village gazetteers are invaluable primary resources and offer both quantitative and qualitative insights into China's most fundamental administrative units, encompassing a wide range of subjects such as local history, genealogy, economics, education, politics, management, public health, and more. The purpose of CCVG is to create a database that specializes in contemporary Chinese village gazetteers and assist social science and humanities researchers of relevant fields in better locating available data and facilitating potential work with data.

As He et al. (2022) demonstrated in a previous work, CCVG data have been systematically housed within a MySQL database. Within this database, 38 tables have been meticulously identified to accommodate the data pertinent to the villages' 12 thematic topics, encompassing areas such as gazetteer information, village basis, natural environment, natural disasters, ethnic groups, economy, and education. The online interface supports two ways of exploring CCVG data: a single village search for basic inquiries, and a multiple village search for more advanced explorations. The search platform, which is currently under construction, facilitates open-access as well as user-friendly interaction. Our initial assessments of these search interfaces underscore their utility and efficacy in catering to scholarly requirements, indicating their alignment with the needs and expectations of the academic community.

The massive data collected in the CCVG project were also enhanced by the design and implementation of various exploratory tools and features. For example, visualisation features have been embedded in the interactive interface to facilitate research discovery and enable researchers to perform live analysis with searchable data. Such existing visual analytic features include (1) an interactive map highlighting the geographic distribution of villages in the dataset and their specific details; (2) multiple interactive charts providing a broad overview of the data. Figures 1.2 and 1.3 provide examples of the interactive features. Figure 1.2

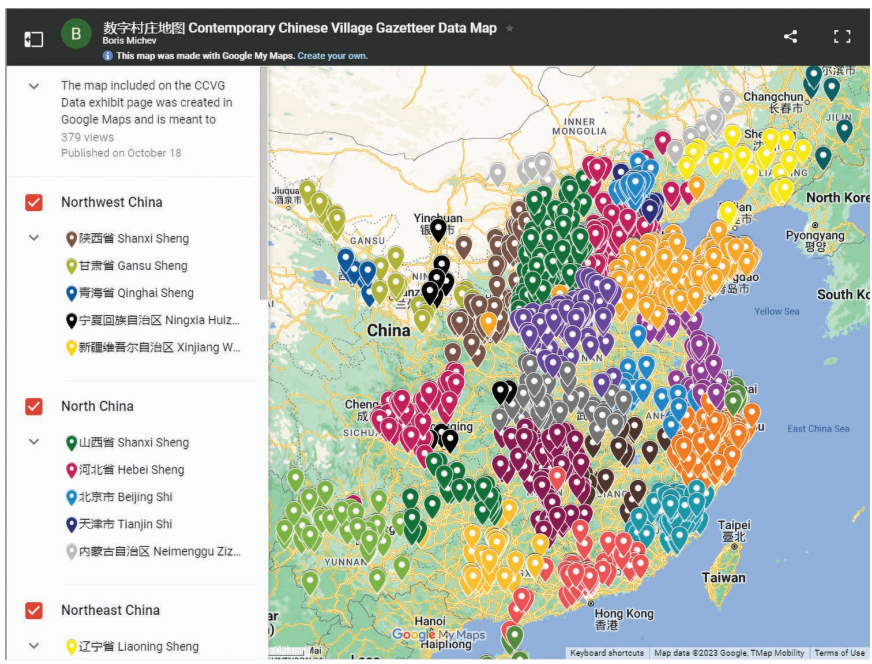

Figure 1.2 Interactive map of the CCVG data platform. East Asian Library, University Library System, University of Pittsburgh (2019). https://www.chinesevillagedata.library.pitt.edu/.

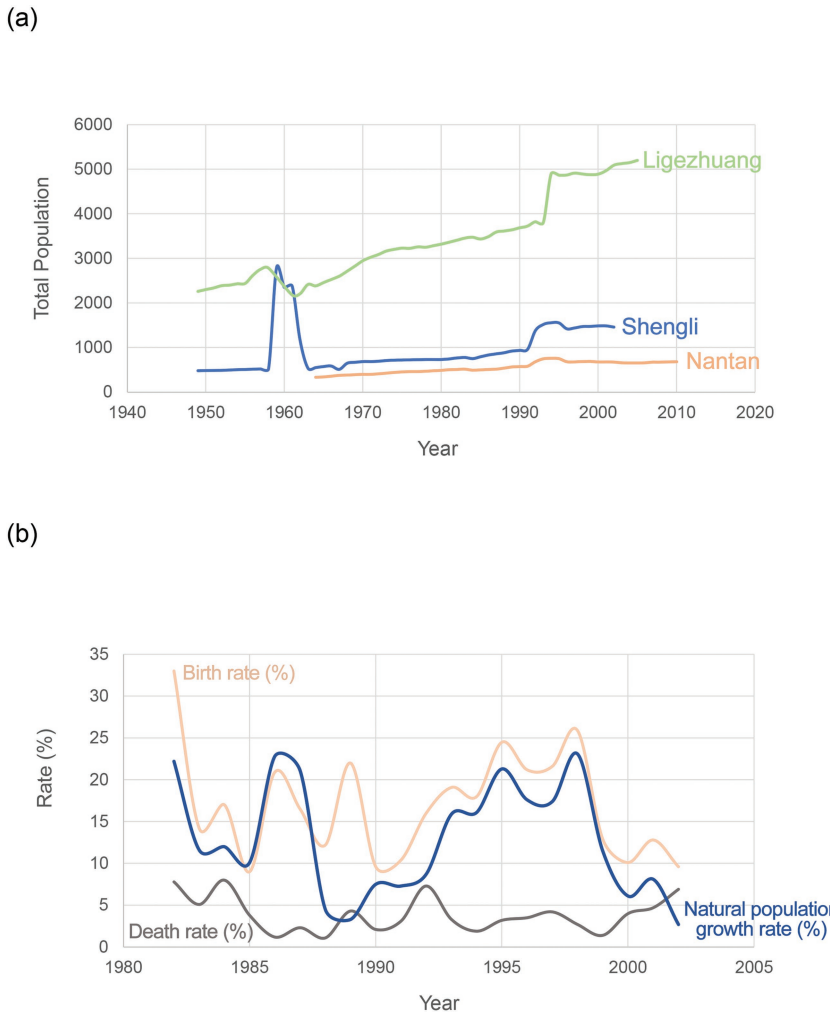

Figure 1.3 Examples of the interactive charts providing data overview on the CCVG data project. (a) shows total population change across three villages in Shandong Province, China, from 1940 to 2020; (b) shows the shifts of birth rate, death rate, and natural population growth rate across the dataset from 1980 to 2005 (Ma, 2022).

showcases the interactive map function, which utilises the contemporary map of China as a base map and marks all the villages covered in the data platform. Built with Google Maps, this interactive map enables users to develop an overview of the geographic distribution of the villages and explore further details of specific villages (e.g., see Figure 1.2). Another important visualisation form to highlight data content is graphs and charts (e.g., line graphs, bar charts) reactive to users' data search on the platform, providing a visual tool to explore and analyse

the massive numeric data. CCVG demonstrates a case where the openness and availability of cultural data is one but not the only goal. The increased usability, generated along with the creation of the interactive interface and exploratory features, makes this project an invaluable data project.

4.2 Challenges and solutions

Working on the CCVG data project presents several unique challenges that require thoughtful solutions. One of the primary challenges is the necessity for researchers to have access to context data while navigating copyright issues. The curating team discovered during the process that historians working with the CCVG data, for example, were particularly interested in understanding the context in which the numeric data were produced, collected, and possibly also "selected" in the gazetteer publications. However, satisfying scholars' needs to access full-text context data in the published Chinese gazetteers was difficult due to copyright constraints. It should not be a unique problem faced only by the CCVG project, but rather, it is probably a common issue confronted by many digital humanities projects that work with documents, archives, physical and digital artworks, to name a few. This challenge demands that curators come up with a delicate balance between utilizing essential historical and sociopolitical context for a comprehensive understanding of the data and respecting the intellectual property rights associated with the source materials. This complexity is even amplified by the diverse nature of the materials, which may include text, images, and maps, each with its own set of copyright considerations.

Another significant challenge is the requirement for human supervision to address inconsistencies in the original dataset. The process of transforming print materials into a structured digital database is not straightforward, as it involves standardizing diverse data formats and rectifying errors or omissions in the original sources. This task demands a high level of expertise and attention to detail, often requiring the involvement of skilled researchers, data specialists, and detail-oriented assistants. This also leads to the last, but not the least, aspect of the workforce for the project. The main CCVG project workforce encompasses librarians at the University of Pittsburgh's East Asian Library, researchers from the School of Computing and Information, and a student team across the campus, including master's students in programs such as information science, East Asian studies, education, and accounting. During the four-year period, the student team contributed significantly to data collection, transcription, cleaning, and processing. In addition, the labor-intensive process of designing and implementing the CCVG interface, as well as the exploratory visualisation features in the interface, relied heavily on students' efforts and their initiatives. The workforce composition has created both benefits and limitations for the project. On one hand, CCVG has provided valuable educational resources and training opportunities for students to improve their skills and knowledge with hands-on experience. But on the other, the short turnaround rate among master's student workers further complicated the project's working process. To address the corresponding challenges and risks, a detailed

documentation practice was implemented, ensuring that all students on the team were aware of previous practices and could collaboratively contribute to improve the practices without disrupting the progress of the project.

4.3 Cultural revolution 10 (CR/10) project

Compared to the CCVG data project, the Cultural Revolution 10 (i.e., CR/10) project[4] focused on creating a video oral history collection that focus on the collective cultural memories of China's Cultural Revolution (1966–1976). Initiated by the University of Pittsburgh in 2016, CR/10 aims to explore the shift of memory accounts on China's Cultural Revolution through hundreds of ten-minute video interviews with participants across generations and backgrounds, who have either experienced the historical incident or learned about it from external resources. One of my previous works illustrated the specific practices of curating the CR/10 project (Ma, 2022). Using a snowball sampling technique for interviewee selection, the CR/10 project has thus far featured more than 300 interviews with people from different generations, geographies, occupations, and social backgrounds who have either experienced the incident themselves or only learned about it from family, school, digital media, or other circumstances. Each interviewee of the project was given about ten minutes to speak freely of their impressions, memories, and thoughts of the historical incident, although, in actual practice, some interviews went longer than ten minutes. The rationale behind the "ten-minute" grammar was to capture the most distinct recollections of the Cultural Revolution. All the interviews were video-recorded and made openly accessible with complete transcriptions through the interactive CR/10 website.

In addition to enabling open-access to data collections, the website also aims to enhance the utility of the materials by facilitating user interaction via data modeling. The available oral history data are modeled on the interface from multiple aspects. Major metadata include the title, creator, date of creation, identifier, collection, contributor, and type. This allows the project to follow the institutional guidelines and standards. To further complement the metadata, the curators of the project also added aspects of descriptions and themes—such as the participants' occupation, education background, and generation—to enrich the oral history video collection and its potential value for research and teaching. Such additional, more user-oriented and flexible metadata categories are also implemented in the CR/10 interface to facilitate users' exploration with the collection, particularly through two exploratory functions, namely, (1) an interactive timeline that enables a chronological exploration of the videos in the collection (Figure 1.4); and (2) a map illustrating the geo-distribution of interviews and showcasing "typical" memories and experiences on a geographic landscape (Figure 1.5).

The interactive timeline and map create two distinct pathways to explore and interpret the oral history collections. The timeline creates an "atlas of forgetting," where the user of the timeline would find the collective memories of China's Cultural Revolution gradually fade across generations. Similarly, the interactive map highlights the regional differences in terms of the impact of the Cultural

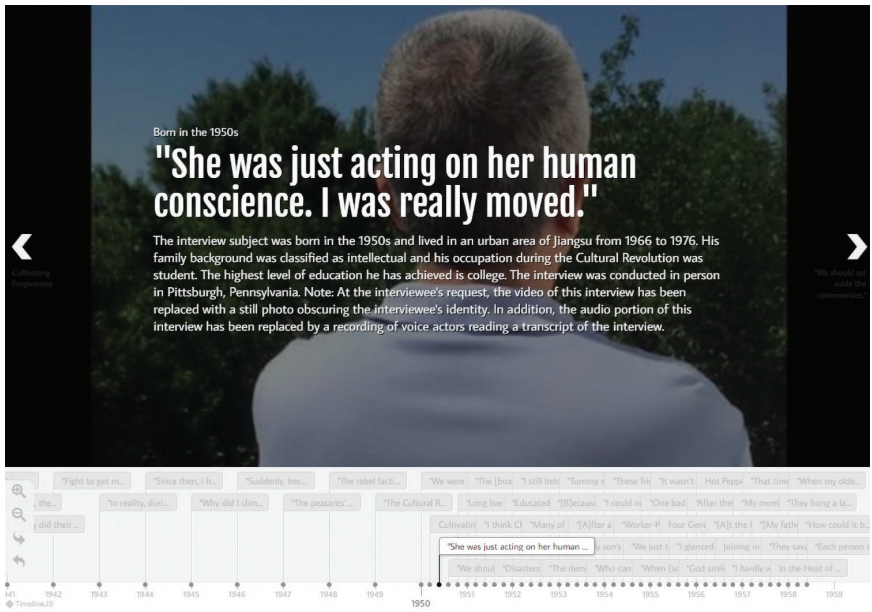

Figure 1.4 Interactive timeline of the CR/10 interface. East Asian Library, University Library System (ULS), University of Pittsburgh. (2017, September 25). https://culturalrevolution.pitt.edu/.

Figure 1.5 Interactive map of the CR/10 interface. East Asian Library, University Library System (ULS), University of Pittsburgh. (2017, September 25). https://culturalrevolution.pitt.edu/.

Revolution. These pathways generate powerful, multifaceted narratives of the historical incident, supporting users to better explore and reflect on its implications. To gain insights into how different groups interact with interactive features on the CR/10 platform, I carried out several user studies, which are reported in detail in a previous work (Ma, 2022).

The user studies aimed to assess the effectiveness of exploratory tools and data usage for two primary user groups: an academic cohort and a general user group. The academic group, consisting of Chinese studies scholars, graduate students, and academic librarians with a deep interest in China's Cultural Revolution, was recruited at the Association for Asian Studies (AAS) annual conference, a key event for Asian studies research in North America. In contrast, the general user group, composed of individuals fascinated by China's Cultural Revolution but without extensive knowledge of it, was gathered from an online book group in 2018.

The user research included two short-term ethnographic studies (one in-person and one online) and six semi-structured interviews. The in-person study took place at the 2018 AAS conference during a roundtable discussion on CR/10, featuring five scholarly presentations and a Q&A session. Complementing the on-site user study, a virtual ethnography was conducted with non-academic users. I joined a ProBoards online discussion group on the Chinese Cultural Revolution, facilitated by the National Consortium for Teaching about Asia (NCTA) at the University of Pittsburgh. This group included 25 K–12 educators and library professionals globally, interested in China and the Cultural Revolution. While with the book group, I monitored weekly online interactions and designed discussion prompts about CR/10 to gather reflections on its website and how participants engaged with its collections. These responses helped select six individuals for more detailed semi-structured interviews, aiming to explore general user navigation preferences on CR/10. Each interviewee spent 15 minutes navigating three different scenarios on the CR/10 website, while I observed their processes and decision-making without interference. Afterward, interviewees were asked to evaluate their navigation experience, focusing on their preferred pathways and the strengths and weaknesses of each. They also rated the website's usability on a scale of 1–5 and offered suggestions for improving the CR/10 interface and features.

The studies revealed differing information needs and navigation preferences between the academic and general users. Academics particularly valued the complexity and flexibility of CR/10's multiple pathways, especially the timeline and map, for shaping narratives of the Cultural Revolution. These features were seen as academically valuable for the breadth of options they provided. The studies also showed that the visual pathways aided users, especially those new to the topic, in understanding individual videos and the narrative of the CR/10 project.

4.4 Modeling and curation of data

The two cases demonstrate the importance of data curation and modeling practices as they pertain to the FAIR (Findable, Accessible, Interoperable, Reusable) standards to improve sharing and reuse of humanities data in research (Wilkinson et al., 2016). More specifically, during the process of curating the data collections for the two cases, we worked to increase data findability by providing more detailed, richer metadata for the collections, while improving interoperability of the collections by offering data dictionaries, guides, and use cases of the data. To further facilitate effective and creative use of the collections, curators of the two

collections developed interactive, exploratory tools to model data and inspire users to develop multi-faceted accounts of the data collections. As Flanders and Jannidis (2019) discussed in their recent book on data modeling in digital humanities, models demonstrate the *choice* of analysis and argumentation, thus comprising an important component of humanities research. As an increasing number of cultural heritage institutions start to create open-access data collections, either in the form of digitisation or born-digital collections and datasets, data modeling and curation practices remain of great importance to ensure that humanities and cultural data collections are utilised to their most potential and value.

Enhancing the CR/10 project's data modeling and curation strategies involves addressing several intricate challenges. Safeguarding the confidentiality of interview recordings and the privacy of participants was the first and foremost. All the interviews were conducted in a confidential setting, accompanied with a transparent consent process, to maintain trust and integrity between the participant and the research team. In addition to strictly following the data security protocols indicated in the Institutional Review Board, the curating team of CR/10 implemented further practices to address the ethical challenges posed by the project. For example, we instructed all the participants not to specify any personally indefinable information while discussing their experiences during the interview, such as individuals' names or working institutions. Given the sensitive nature of the topic of discussion, we also offered options of covering participants' faces, removing identifiable information, or distorting their voices. The interview recordings were also reviewed before being released to ensure the confidentiality and the use of ethical curation practices in the project.

When it comes to video editing and curating the collection, the project faces the dual challenge of preserving the authenticity of original interviews while making them accessible and engaging for a diverse audience. The curators and student assistants reviewed all the interview recordings and curate them collaboratively to create a balance between editorial cuts for clarity and conciseness and preserving the original context and emotional depth of the interviews. More specifically, we strived to consistently keep all the videos under ten minutes without hurting the essence of each narrative and storytelling. However, the curating team also kept a few lengthier videos with minimal edits, to remain respectful to participants' personal experiences and emotions, and preserve the unique stories and accounts of the Cultural Revolution. Additionally, as a number of participants spoke only mandarin Chinese or regional dialects, the CR/10 team faced translation challenges when creating transcriptions for the collection. To preserve the nuances in the participants' accounts in the video transcriptions, we implemented a collaboration model among narrative English speakers, China experts, and personnel who speak regional dialects during the process. Continuous review and feedback mechanisms were also placed among team members to enhance the accuracy of translations and transcriptions. Finally, curating the website and its features to enhance data reuse for the diverse user communities is a continuous process. As discussed previously, the CR/10 team has been working to improve user interface design for better navigation based on findings from user research, and to refine the interactive, exploratory features like search filters, thematic tagging, and multimedia presentations.

Engaging with user feedback and analytics have contributed to the ongoing refinement of the CR/10 website, making it a more open, responsive, and valuable resource for research and teaching.

5 Conclusion

The practices of open humanities data have been catalyzed by the overarching open initiatives, reflecting a paradigm shift toward transparency, reproducibility, and accessibility within the academic landscape. This shift reveals unique challenges when applied to the humanities due to the discipline's intricate nature, the specificity of data, and the interpretive traditions. This chapter focused on the topic of open data access and reuse for the humanities and cultural research, highlighting the key issues, challenges, as well as opportunities that a more accessible, reusable, and shared data environment can create. I discussed how existing digital infrastructures, such as data paper publications and journals, open humanities data sharing repositories, and semantic web technologies, can facilitate better reuse of humanities and cultural data. To particularly contextualize the general issues within the setting of cultural heritage institutions, I drew upon two case studies from my research: the Contemporary Chinese Village Gazetteer (CCVG) Data Project and the Cultural Revolution 10 (CR/10) Project, which illustrate the added value and potential of such data when coupled with effective curation practices and interactive tools. These projects underscore the importance of modeling data and designing data interfaces to ensure that data can be well represented, and therefore, findable, accessible, interoperable, and reusable, maximizing the impact of humanities research. In summary, the integration of open data practices within the humanities and cultural domains is a process filled with both opportunities and complexities. As the field continues to evolve, it is imperative that scholars, librarians, and technologists work collaboratively to further refine the infrastructures that support open data. By doing so, the humanities can fully embrace the transformative potential of open data initiatives, ensuring that data is not merely a record of human culture but also as a bridge to future explorations and understandings.

Notes

1 https://openhumanitiesdata.metajnl.com/.
2 https://culturalanalytics.org/.
3 https://www.chinesevillagedata.library.pitt.edu/.
4 https://culturalrevolution.pitt.edu/.

References

Allen, C., & Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PLoS Biology, 17*(5), e3000246. https://doi.org/10.1371/journal.pbio.3000246

Arthur, P. L., & Bode, K. (2014). *Advancing digital humanities: Research, methods, theories*. Palgrave Macmillan.

Arthur, P. L., & Hearn, L. (2021). Toward open research: A narrative review of the challenges and opportunities for open humanities. *Journal of Communication, 71*(5), 827–853. https://doi.org/10.1093/joc/jqab028

Aschenbrenner, A., Blanke, T., Fritze, C., & Pempe, W. (2013). Data-driven research in the humanities—the DARIAH research infrastructure. In Atkinson, M., Baxter, R., Brezany, P., Corcho, O., Galea, M., Parsons, M., Snelling, D., & van Hemert, J. (Eds.), *The DATA bonanza: Improving knowledge discovery in science, engineering, and business* (pp. 417–430). https://doi.org/10.1002/9781118540343.ch20

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly, 32*(4), 399–418. https://doi.org/10.1016/j.giq.2015.07.006

Baack, S. (2015). Datafication and empowerment: How the open data movement rearticulates notions of democracy, participation, and journalism. *Big Data & Society, 2*(2), 2053951715594634. https://doi.org/10.1177/2053951715594634

Bauder, J. (2019). HathiTrust as a data source for researching early nineteenth-century library collections. *Information Technology and Libraries, 38*(4), 14–24. https://doi.org/10.6017/ital.v38i4.11251

Björk, B. C., Laakso, M., Welling, P., & Paetau, P. (2014). Anatomy of green open access. *Journal of the Association for Information Science and Technology, 65*(2), 237–250. https://doi.org/10.1002/asi.22963

Borgman, C. L. (2017). *Big data, little data, no data: Scholarship in the networked world*. MIT Press.

Burrows, T. (2011). Sharing humanities data for e-research: Conceptual and technical issues. *Sustainable data from digital research: Humanities perspectives on digital scholarship*. Proceedings of the conference held at the University of Melbourne, 12–14 December 2011. https://ses.library.usyd.edu.au/handle/2123/7938

Corrall, S. (2016). *The open movement: What libraries can do*. D-scholarship.pitt.edu. https://d-scholarship.pitt.edu/28249/1/Corrall_Charleston_2015_proceedings_eprint.pdf

Corrall, S., & Pinfield, S. (2014, March). Coherence of "open" initiatives in higher education and research: Framing a policy agenda. In *Breaking Down Walls: Culture–Context–Computing, iConference 2014 Proceedings* (pp. 293–313). https://doi.org/10.9776/14085.

Downie, J. S., Bhattacharyya, S., Giannetti, F., Koehl, E. D., & Organisciak, P. (2020). The HathiTrust Digital Library's potential for musicology research. *International Journal on Digital Libraries, 21*, 343–358.

Ell, P. S., & Hughes, L. M. (2013). E-Infrastructure in the humanities. *International Journal of Humanities and Arts Computing, 7*(1–2), 24–40.

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology, 4*(5), e157. https://doi.org/10.1371/journal.pbio.0040157

Fecher, B., & Friesike, S. (2014). *Open science: One term, five schools of thought* (pp. 17–47). Springer International Publishing.

Gao, J., Nyhan, J., Duke-Williams, O., & Mahony, S. (2023). Exploring international collaboration and language dynamics in digital humanities: Insights from co-authorship networks in canonical journals. *Journal of Documentation, 79*(7), 240–260.

He, D., Ma, R., Zheng, R., & Zhang, H. (2022). CCVG data: A unique, curated, and searchable Chinese village dataset for Chinese Study scholars. In *Proceedings of ADHO digital humanities conference*. https://dh-abstracts.library.virginia.edu/works/11897

Heath, T., & Bizer, C. (2022). *Linked data: Evolving the web into a global data space*. Springer Nature.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management, 29*(4), 258–268.

Kansa, S. W., & Kansa, E. C. (2011). Enhancing humanities research productivity in a collaborative data sharing environment. *White Paper NEH Division of Preservation and Access*.

Kitchin, R. (2014). *The data revolution: big data, open data, data infrastructures & their consequences*. London: Sage Publications Ltd.

Knöchelmann, M. (2019). Open Science in the humanities, or: Open humanities? *Publications, 7*(4), 65.

Flanders, J., & Jannidis, F. (2019). *The shape of data in digital humanities: Modelling texts and text-based resources*. Routledge.

Hu, Y., LeBlanc, Z., Diesner, J., Underwood, T., Layne-Worthey, G., & Downie, J. S. (2023). Complexities of leveraging user-generated book reviews for scholarly research: Transiency, power dynamics, and cultural dependency. *International Journal on Digital Libraries*, 1–24. https://doi.org/10.1007/s00799-023-00376-z

Lebo, T., Del Rio, N., Fisher, P., & Salisbury, C. (2017). A five-star rating scheme to assess application seamlessness. *Semantic Web, 8*(1), 43–63.

Li, K., Greenberg, J., & Dunic, J. (2019). Data objects and documenting scientific processes: An analysis of data events in biodiversity data papers. *Journal of the Association for Information Science and Technology,71*(2), 172–182.

Li, K., Ma, R., & Fang, Z. (2023). Explicit or implicit digital humanities? An examination of search strategies to retrieve digital humanities publications from large-scale scholarly databases. SocArXiv.

Luhmann, J., & Burghardt, M. (2021). Digital humanities—A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape. *Journal of the Association for Information Science and Technology, 73*(2), 148–171.

Lyon, L. (2009). Open science at web-scale: Optimising participation and predictive potential. *Consultative Report*. JISC.

Ma, R. (2022). Curating China's Cultural Revolution (1966–1976): CR/10 as a Warburgian memory atlas and digital humanities interface. *Information & Culture, 57*(1), 27–45.

Ma, R. (2022). Boundaries, extensions, and challenges of visualization for humanities data: Reflections on three cases. *IEEE 7th Workshop on Visualization for the Digital Humanities (VIS4DH)*. https://doi.org/10.48550/arXiv.2210.03630.

Ma, R., Chen, A., Poole, A., Chassanoff, A., & Wingate, A. (2023). Evaluating the value of exploratory tools in digital humanities collections and scholarly projects: Discussions from researchers, developers, and users' perspectives (Panel). *Proceedings of the Annual Meeting of the Association for Information Science and Technology, 70*(1), 804–807.

Ma, R., & Li, K. (2021). Digital humanities as a cross-disciplinary battleground: An examination of inscriptions in journal publications. *Journal of the Association for Information Science and Technology, 73*(2). https://doi.org/10.1002/asi.24534

Mann, F., von Walter, B., Hess, T., & Wigand, R. T. (2009). Open access publishing in science. *Communications of the ACM, 52*(3), 135–139.

Maron, N. L., Mulhern, C., Rossman, D., & Schmelzinger, K. (2016). The costs of publishing monographs: Toward a transparent methodology. *Journal of Electronic Publishing*. https://doi.org/10.3998/3336451.0019.103

Mayernik, M. S., Hart, D. L., Maull, K. E., & Weber, N. M. (2017). Assessing and tracing the outcomes and impact of research infrastructures. *Journal of the Association for Information Science and Technology, 68*(6), 1341–1359.

McGillivray, B., Marongiu, P., Pedrazzini, N., Ribary, M., Wigdorowitz, M., & Zordan, E. (2022). Deep Impact: A study on the impact of data papers and datasets in the humanities and social sciences. *Publications, 10*(4), 39.

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., … Yarkoni, T. (2016). How open science helps researchers succeed. *elife, 5*, e16800.

Nurmikko-Fuller, T. (2023). *Linked open data for digital humanities*. Taylor & Francis.

Severin, A., Egger, M., Eve, M. P., & HuÅNrlimann, D. (2018). Discipline-specific open access publishing practices and barriers to change: An evidence-based review. *F1000 Research, 7*, 1925–1925. https://doi.org/10.12688/f1000research.17328.2

Schreibman, S., Siemens, R. G., & Unsworth, J. (2016). *A new companion to digital humanities*. John Wiley & Sons, Ltd.

Schöch, C. (2013). Big? Smart? Clean? Messy? Data in the humanities? *Journal of the Digital Humanities, 2*(3). https://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/

Schöpfel, J., Farace, D., Prost, H., & Zane, A. (2020). Data papers as a new form of knowledge organization in the field of research data. *Knowledge Organization, 46*(8), 622–638. https://doi.org/10.5771/0943-7444-2019-8

Spinaci, G., Colavizza, G., & Peroni, S. (2022). A map of digital humanities research across bibliographic data sources. *Digital Scholarship in the Humanities, 37*(4), 1254–1268. https://doi.org/10.1093/llc/fqac016

Su, F., Zhang, Y., & Immel, Z. (2021). Digital humanities research: Interdisciplinary collaborations, themes and implications to library and information science. *Journal of Documentation, 77*(1), 143–161. https://doi.org/10.1108/JD-05-2020-0072

Suber, P. (2012). *Open access*. The MIT Press.

Suber, P. (2002). Open access to the scientific journal literature. *Journal of Biology, 1*, article 3. https://doi.org/10.1186/1475-4924-1-3

Sula, C. A., & Hill, H. V. (2019). The early history of digital humanities: An analysis of computers and the humanities (1966–2004) and literary and linguistic computing (1986–2004). *Digital Scholarship in the Humanities, 34*(Supplement_1), i190–i206.

Tang, M. C., Cheng, Y. J., & Chen, K. H. (2017). A longitudinal study of intellectual cohesion in digital humanities using bibliometric analyses. *Scientometrics, 113*, 985–1008.

Uhlir, P. F., & Schröder, P. (2007). Open data for global science. *Data Science Journal, 6*, OD36–OD53.

Underwood, T. (2019). *Distant horizons: Digital evidence and literary change*. University of Chicago Press.

Wang, Q. (2018). Distribution features and intellectual structures of digital humanities: A bibliometric analysis. *Journal of Documentation, 74*(1), 223–246.

Walsh, M., & Antoniak, M. (2021). The goodreads "Classics": A computational study of readers, amazon, and crowdsourced amateur criticism. *Journal of Cultural Analytics, 6*(2), 243–287.

Weingart, S. B., Eichmann-Kalwara, N., Lincoln, M., et al. (2020). *The index of digital humanities conferences*. Carnegie Mellon University. https://dh-abstracts.library.cmu.edu. https://doi.org/10.34666/k1de-j489

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*, 160018. https://doi.org/10.1038/sdata.2016.18.

York, J. (2012). HathiTrust: The elephant in the library. *Library Issues, 32*(3), 1–4.

Zhang, H., Zhang, Y., Xin, Q., & Xiao, F. (2020). Contemporary Chinese Village Gazetteer Data project: From books to data. In *Proceedings of the ACM/IEEE joint conference on digital libraries*. Association for Computing Machinery, New York.

2 Data modeling for digital archiving of intangible and experiential entities

Shigeo Sugimoto and Chiranthi Wijesundara

1 Introduction

Since early 1990s, Libraries, Archives and Museums (LAMs) – the so-called memory institutions – have curated a huge number of digital resources in various cultural and historical domains and made them accessible on the Internet, such as American Memory and Making of America. Those memory institutions select and collect valuable resources and organize them as unique digital collections for use over the Internet and over time. Those digital collections are called cultural digital archives, or simply digital archives in this chapter as they are aimed at keeping important resources consistently accessible on the Internet and for posterity. This chapter aims at providing an overview of the digital archiving process of cultural resources and discussing metadata models for digital archives, covering both tangible and intangible cultural entities.

Early activities for building digital collections at memory institutions started as digitisation of physical objects selected from their institutional collections, such as books, manuscripts, paintings, sculptures, photographs, and audio-visual recordings on analog media. Since then, the domains and types of cultural resources have drastically expanded. Nowadays, cultural digital collections are created not only by digitizing physical/analog objects but also by copying/converting those objects originally created as digital objects (born-digital objects). Generally, institutional cultural collections, which may be digital or non-digital, are organized in accordance with the policies and conventions of the hosting institutions. Metadata is a very basic component to organize the collected items and make them findable and accessible. There are several metadata standards developed in accordance with the requirements and conventions of the memory institutions. A common underlying feature of those standards is a one-to-one correspondence between a collected item and its metadata. This item-oriented feature of metadata is common among those collections regardless of the type of institution.

This chapter focuses on the digital archiving of intangible and experiential cultural entities and metadata models for digital archiving. It discusses them in contrast with the digital archiving of tangible cultural objects. Intangible cultural entities are known as the knowledge, skills, and customs such as dance,

DOI: 10.4324/9781032707211-3

music, theater plays, craftsmanship, rituals, and services, which are maintained by individuals and communities. We can see and listen to those intangible cultural entities through the practices and performances. In other words, those intangible cultural entities are instantiated in the real world as events that we can experience. There are other types of events which are often archived such as natural and man-made disasters, exhibitions, sports, and games. A fundamental issue for archiving is that we can collect only tangible entities as an item which may be physical, analog, or digital. In the case of intangible entities, we can collect recordings created in a particular event of an intangible entity as well as tangible objects related to the intangible entity. There are, for instance, many digital archives of intangible cultural heritage (ICH). Those digital archives digitally curate recordings of events of an ICH as well as materials used in the event(s) and related to the ICH. Documentations of the ICH and event(s), which may be textual and/or audio-visual, are also collected.

An important aspect of this research involves the fundamental changes in the metadata models for digital collections caused by the inclusion of intangible entities in the digital collections. All collected materials and their metadata are digital objects that are linkable and will form a large network of digital objects. This feature may bring fundamental changes in the metadata models for the cultural collections. As we can flexibly create a complex entity composed of several collected items using any entities linkable on the Internet, the metadata schemas of digital archives do not need to be item-centric like those metadata standards used in existing institutional digital collections.

2 Backgrounds

2.1 Basic concepts

This section presents definitions of some important terms used in this chapter.

i Digital Archiving and Digital Archives

Many digital cultural collections have been developed by memory institutions and research institutions in the cultural domains. Those institutions curate digital items, organize them for use and maintain them for current and future users. This chapter refers to this function as *digital archiving*. In the 1990s when digital collection development became popular among memory institutions, many libraries and museums developed digital collections and provided them as a part of their services which were often called *digital libraries* and *digital museums*. The term *digital archives* may be used in a similar meaning, with respect to services and functions at archival institutions based on their digital collections. On the other hand, this term is sometimes used to mean any digital collections which are created by curating and organizing valuable resources mainly in the cultural and historical domains provided on the Internet over time. This chapter uses the term *digital archive* based on the latter meaning.

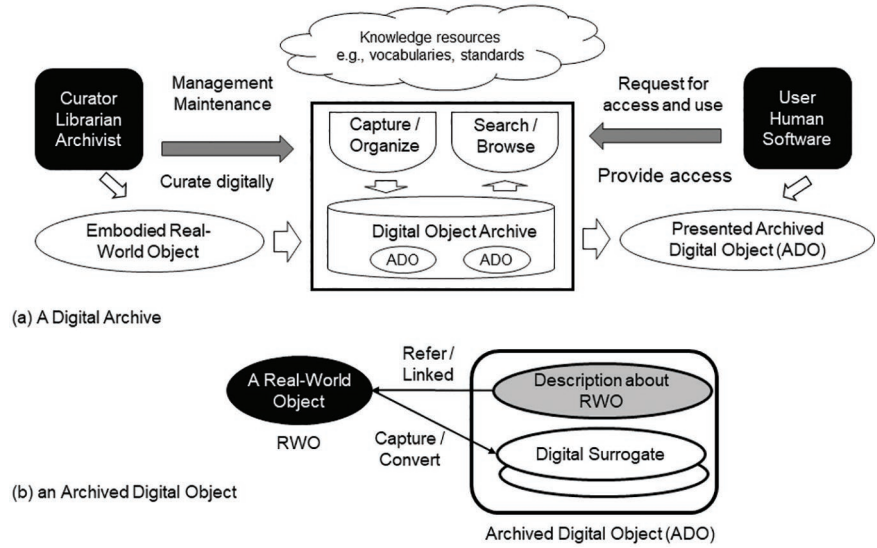

Figure 2.1 A generalized digital archive model (ADO model).

ii A Generalized Structural Model of Digital Archives (ADO Model)

In this chapter, an item curated in a digital archive is referred to as an Archived Digital Object (ADO). Figure 2.1 shows an outline of a digital archive and an ADO. An ADO is composed of one or more digital surrogates which are digital objects created from a real-world object and some descriptions of the real-world object. Thus, an ADO is a composite digital object, which is a basic concept shared with other models such as Metadata Encoding and Transmission Schema (METS) and the Europeana Data Model (EDM).

A digital surrogate may be created by capturing/digitizing an original object or converting/copying an original object in accordance with the type of the original object, namely, digital or non-digital. A real-world object from which a digital surrogate is created may be an original entity or a recording of an original entity as shown in Figure 2.2. This aspect is crucial in identifying objects which are referred to in the digital archiving process as discussed in the following paragraphs.

iii Tangible Entities

This chapter defines a tangible entity as one which can be perceived through sight, touch, and hearing, (un)aided by electronic, optical, or mechanical media such as video displays, audio speakers, tele/microscopes, film projectors, and 3D printers. Based on this definition, the following are considered as *tangible*: (1) physical objects such as paintings, sculptures, and monuments; (2) content carriers such as tapes and disks; and (3) the contents stored in a carrier. Digital objects such as photographs and audio-visual data are *tangible entities* embodied in a digital space that is a part of the real world. Tangible entities may be called objects in this chapter.

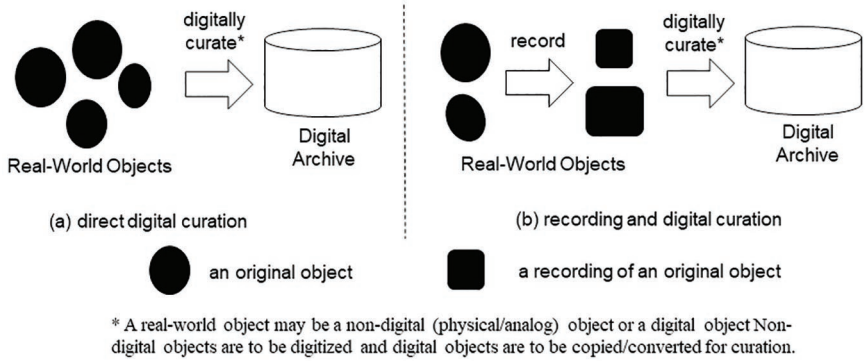

Figure 2.2 Basic scheme of digital archiving.

iv Intangible and Experiential Entities

UNESCO states in the definition of Intangible Cultural Heritage

> Cultural heritage does not end at monuments and collections of objects. It also includes traditions or living expressions inherited from our ancestors and passed on to our descendants, such as oral traditions, performing arts, social practices, rituals, festive events, knowledge and practices concerning nature and the universe or the knowledge and skills to produce traditional crafts.
>
> (UNESCO, 2023a)

Since the scope of this chapter is not limited to cultural heritage, it defines an intangible entity as an entity that we can experience once it is instantiated in the real world, for example, rituals, performing arts, games, craftsmanship, events, and so forth. Those entities which we can experience are called *experiential entities* or *experientials* in this chapter. Generally, experiential entities are events in the real world that may be intended or not. Most cultural events such as performances and exhibitions are intended. Natural events such as earthquakes and tsunamis are not intended. Both intended and non-intended events are important domains of digital archiving. For example, a theater play, which is an intellectual and artistic creative work, is intangible. A performance of the play at a theater which is an event can be recorded, and the recordings, which are tangible, are to be archived.

v Metadata Models and Standards for Digital Archiving

Data modeling is a fundamental issue in designing the organization of digital archives and their metadata. There are some basic aspects in the data modeling – identification of the entities that constitute the whole data model for the target domain, structural features of the entities such as relationships among the entities,

and definitions of classes and attributes of the entities, i.e., ontologies. An example is the FRBR model developed by IFLA, which defines the entities for bibliographic descriptions and the relationships among the entities (IFLA, 2009). IFLA LRM (Library Reference Model) formally defines the classes and attributes of those entities (Riva et al., 2017). In addition, standards for sharing and exchanging those entities over the Internet are also crucial. Some metadata models and standards related to this study are highlighted in the proceeding section.

2.2 Metadata models and standards for cultural digital archiving

There are many data models and standards related to metadata and digital archiving of cultural resources. The paragraphs below briefly introduce some of them which are chosen from different viewpoints – a metadata standard for visual resources (VRA Core), a standard data model designed for metadata aggregation in Europeana Data Model (EDM), an ontology standard for museum resources (CIDOC CRM), a system model which uses heterogeneous linked open data resources to build cultural portals in various domains (Sampo model and portals) and some Internet standards for interchange and sharing of metadata, vocabularies, and image resources over the Internet.

i VRA Core

VRA Core is a metadata standard for visual resources developed by the Visual Resource Association. The literature says "VRA Core is a data standard for description of works of visual culture as well as images that document them" (Visual Resource Association, 2014). The data model of VRA Core is composed of three types of entities: Work, Image, and Collection as shown in Figure 2.3. A Work may be a tangible or intangible entity. A Work may be presented by one or more images and an Image may present one or more Works. A Collection is a collection of Works or Images.

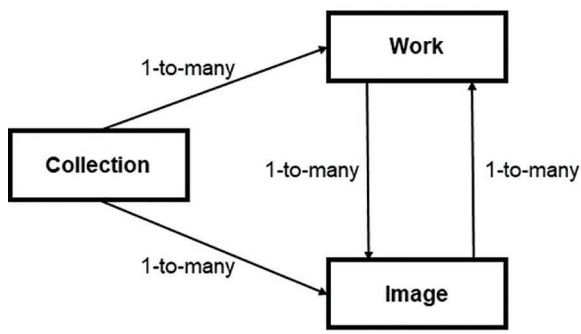

Figure 2.3 VRA core data model (Visual Resource Association, 2014, p. 1).

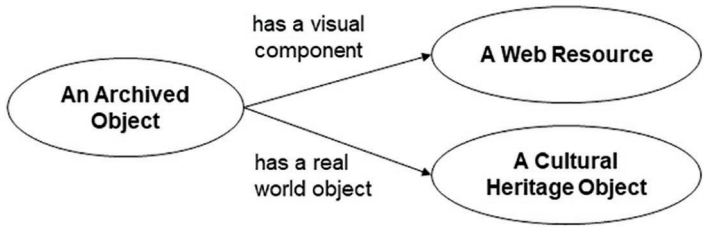

Figure 2.4 Relationships among objects of three core classes of EDM (c.f. Isaac, 2013, p. 10, fig. 4).

ii Europeana Data Model (EDM)

Europeana is a large-scale portal for digital cultural collections provided by the memory institutions in Europe. It collects metadata from participant institutions and aggregates the metadata to provide archived resources across the institutions. EDM defines a data model to define the structure of the curated digital resources and their aggregation (Isaac, 2013). Figure 2.4 shows a simple example adopted from the EDM data model.

iii CIDOC Conceptual Reference Model (CRM)

CIDOC CRM is a standard developed by the museum community and defines a set of classes and properties of the museum resources as an extensible ontology for concepts and information in cultural heritage and museum documentation. There are several activities to develop domain-specific ontologies such as bibliographic description and spatio-temporal resources (CIDOC CRM, 2023). CIDOC CRM is standardized as an international standard (ISO 21127). This standard specifies all information required for the exchange and integration of heterogeneous scientific and scholarly documentation about the past at a human scale and the available documented and empirical evidence for this (International Organization for Standardization [ISO], 2023).

iv Sampo Model

The Sampo portal series, which is developed using the data resources provided in various domains and based on the Linked Open Data technologies, provides rich cultural resources for the Digital Humanities (DH) researchers. The Sampo model, which is the underlying model for the Sampo series, is built based on a set of principles of data and technological aspects important for collaborative development and sharing of data as well as ontologies and for building the services for digital humanities research (Hyvönen, 2023). Each portal of the Sampo series uses several ontologies and data resources which are made available for users through user interfaces. The shared open ontology infrastructure, ontologies in various cultural domains, as well as the General Finnish Ontology (Finto, 2023) are used as the semantic basis and are provided as linked open data resources.

v Technological Standards for Metadata and Digital Archiving on the Internet

Resource Description Framework (RDF), Simple Knowledge Organization System (SKOS), and Web Ontology Language (OWL) are an infrastructure to share metadata and ontologies on the Internet. These are used in the metadata standards and models highlighted above. International Image Interoperability Framework (IIIF) provides APIs and metadata schemas to share image objects such as image and audio-visual files of digitised objects, over the Internet.

2.3 Conceptual Models for Cultural Digital Archiving – CHDE and CEDA

The authors proposed two conceptual models CHDE (Cultural Heritage in Digital Environments) and CEDA (Concepts, Embodiment, Digital Archives) in their earlier articles (Wijesundara et al., 2017; Wijesundara & Sugimoto, 2019; Sugimoto et al., 2021, 2022).

CHDE is proposed as a process model of digital archiving for both tangible and intangible cultural heritage. As mentioned previously, cultural digital collections are composed of archived digital objects (ADOs), each of which is composed of one or more digital surrogates and descriptions of original cultural entities. There is need to clarify from what objects the digital surrogates are created. Intangible Cultural Heritage (ICH) may be composed of intellectual and tactical entities such as skills, knowledge, and customs, so that what can be seen, heard, or touched is an instantiation of an ICH, for instance, dance and music performances, and craftsmanship. However, as we cannot archive those instantiations as they are, we collect recordings of the instantiations and tangible objects related to the ICH and performances to create an ADO. Figure 2.5 shows a process model proposed in the CHDE model.

The CEDA model was proposed from the viewpoint of shifting from item-centric description to content-oriented description of metadata. Cultural entities are expressed as a conceptual entity and an embodied entity. This is similar to FRBR, a well-known standard model for bibliographic description and now consolidated into the IFLA LRM model; FRBR has Work and Item entities which represent a conceptual and an embodied entity of a book, respectively, while the data model of VRA Core makes a distinction between Work, Image, and Collection. Figure 2.6 shows a structural model of CEDA. An elementary item for an ICH is composed of an instantiation and a set of recordings created for the instantiation. More than one elementary item may be aggregated to form a composite item for instance, an event composed of sub-events and a series of events. An item may be linked to one or more conceptual entities which represent the intellectual contents of the item.

3 Models for digital archiving of intangible and experiential entities

3.1 Basic characteristics of intangible and experiential entities

This section first discusses the basic characteristics of intangible and experiential entities from the viewpoint of digital archiving, and then it shows some examples to discuss aspects crucial for the digital archiving of intangible and experiential entities.

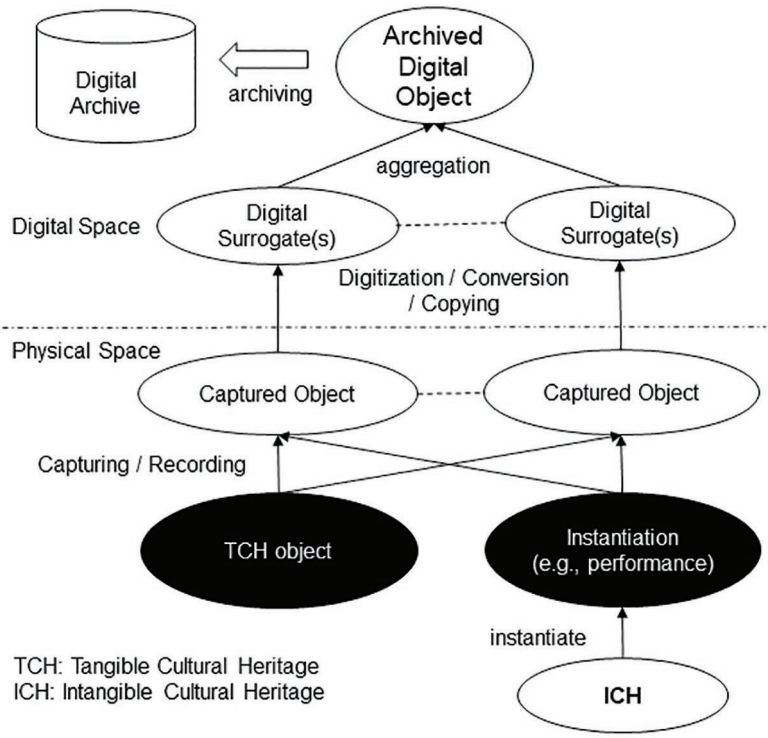

Figure 2.5 CHDE model.

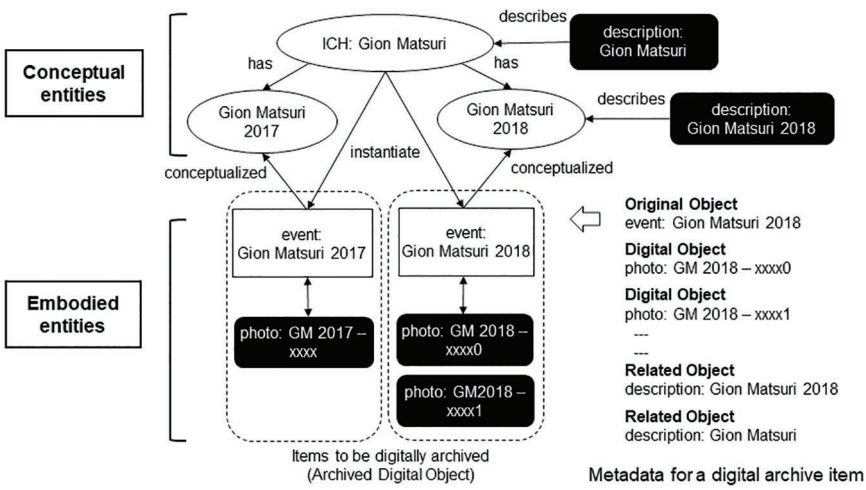

Figure 2.6 CEDA instances – digital archive item, conceptual entities, and metadata.

ICH is composed of conceptual and intellectual entities such as knowledge, skills, customs among others maintained by the people and their communities. They are things that we know and share but we need to express them in some forms through performances and practices in order to archive them. Those expressions are embodiments of ICH in our real world. Games and sports have similar features, in the sense that players have knowledge and skills to play, and their plays and performances are the embodiments in the real world. Those embodiments, which may be called cultural events, or simply events, are to be digitally archived.

There are various artworks such as interactive digital media arts which have features determined by the time, place, and audience of their exhibits. This means that the audience may see different features of the artwork depending on a particular exhibit. In other words, those artworks can be recognized as an exhibition event but not as a tangible object. Similarly, ephemeral artworks such as fireworks and ice sculptures may be recognized as an exhibition event.

Social and historical events such as natural and man-made disasters and elections are important domains of digital archiving. Cultural and social events have different features depending on the specific domains. For example, cultural and artistic performances are made based on the intentions of people and communities, but there are no such intentions in natural disasters.

The following paragraphs show some examples of digital archiving from different viewpoints.

i Datasets of Individual Archived Resources

It is quite common to organize a digital archive of ICH as a collection of ADOs created from individual resources such as photographs, motion pictures, postcards, and news articles. Similar organization strategies are used for digital archives of disasters, performing arts, and so forth.

For example, there are many digital archives developed to record the disasters caused by the Great East Japan Earthquake that happened on March 11, 2011, which collect and archive various materials such as photographs, video images, and documents. There is a portal named Hinagiku developed by the National Diet Library, Japan (National Diet Library, 2013), which harvests metadata from those disaster archives and provides search and browsing functions. Every collected item is obviously important as a record of the earthquake. However, the authors have learned that their simple and item-centric metadata may not be sufficient for the audience who want to know contextual information about an item such as a set of photographs that shows a place before and after the disaster, the recovery process of a place and so forth.

ii Documentations of Intangible Entities – Books and Audio-Visual Programs

Books and audio-visual programs would be typical resources useful to record intangible and experiential entities. Audio-visual recordings are commonly used to archive performing arts such as dance, music, and theater plays. An example is the National Museum of Ethnology, Japan, which provides a video collection

created at the museum, Videotheque, which started in the early 1970s. Another example is UNESCO which provides several video programs of ICH for the public (UNESCO, 2023b).

iii Content Access Using Spatiotemporal Information and Narratives

Textual and oral explanations about a historical event used with archived materials help the audience learn about the event. For example, digital archives of the atomic bombing at Hiroshima and Nagasaki and the Great East Japan Earthquake developed at the University of Tokyo use geospatial information and narratives about the events in the interfaces (University of Tokyo, 2023). Those archives collect information about the geospatial moves of people and narratives during and after the disasters and provide the resources through designed portals which help the audience to understand what happened in the disasters and the regions by connecting the archived items, people, location, and time.

iv Metadata Models for Event-oriented Artworks

There are digital archives and databases for event-oriented artworks such as performing arts and interactive arts. Their metadata models reflect the features of the domains.

The Tsubouchi Memorial Theatre Museum at Waseda University has developed the Japan Digital Theatre Archives (JDTA) (Theatre Museum of Waseda University, 2023). In the JDTA data model, a theater performance program is an entity linked to one or more performances, where every performance is a performance event of a theater performance program with a particular set of staff, theater goods, and environments. A theater performance program represents an intellectual entity, and every performance is an embodiment of the theater performance program.

The Agency for Cultural Affairs of the Japanese government has developed a database named Media Art Database (MADB) for media artworks of Manga (Japanese comics), video games (Game), animations (Anime), and digital media works (Art) (Agency for Cultural Affairs, 2023). MADB has four component DBs, where each component DB has its own metadata schema which is defined considering interoperability across the domains because the types of individual items in these domains are diverse. A comic book as an Item of the FRBR model is obviously a tangible object. On the other hand, we often recognize comic books by a series title and/or as a Work entity of FRBR. A video game could be collected from different viewpoints such as a computer program, a set of functions that interact with players, or a specific play/match by a player(s). Animations may be identified as a video content packaged in a video media or in a broadcasting program. In the latter case, animations could be modeled as a broadcasting event. Every artwork in the Art domain may be recognized as an event where the artwork is exhibited, which means that tangible objects such as devices and computer programs used in the artwork may be recognized as a component of the event.

3.2 A conceptual model of intangible and experiential entities for digital archiving

This section aims to present a conceptual model for the digital archiving of intangible and experiential entities. The model is named CEDA-X as it is defined based on the CEDA model shown earlier in this chapter. Figure 2.7 illustrates the conceptual model of CEDA-X, which includes component models for intangible cultural heritage (ICH), disaster, and tangible objects.

Figure 2.7(a) shows a model for ICH where an ICH is instantiated as an event. Events are created based on the rules and customs of the ICH. Those objects to be collected and archived are the recordings created for the event and the objects related to the event. Some objects that are related to the ICH but not to any particular event may be archived as well. An ICH event entity in the real world may be associated with a conceptual entity which represents the ICH event.

Social and historical events are obviously a major domain for digital archives. Their features are slightly different from those of cultural events. Figure 2.7(b) shows a model for a disaster, both natural and man-made disasters. Disasters are not an embodiment of knowledge or skills, and they are to be recognized as a fact after the disasters have happened. The fact may be modeled as a conceptual entity which is associated with a conceptual entity. This modeling scheme is similar to that for ICH events described above. The recordings of the events together with objects related to the events are digitally collected and archived.

In both Figure 2.7(a) and (b), the ADOs, which are created from the recordings of an event and related objects, are grouped shown by dotted circles and linked to each other. As presented in Figure 2.1(b), an ADO is an object which is composed of one or more digital surrogates and a description about an original real-world

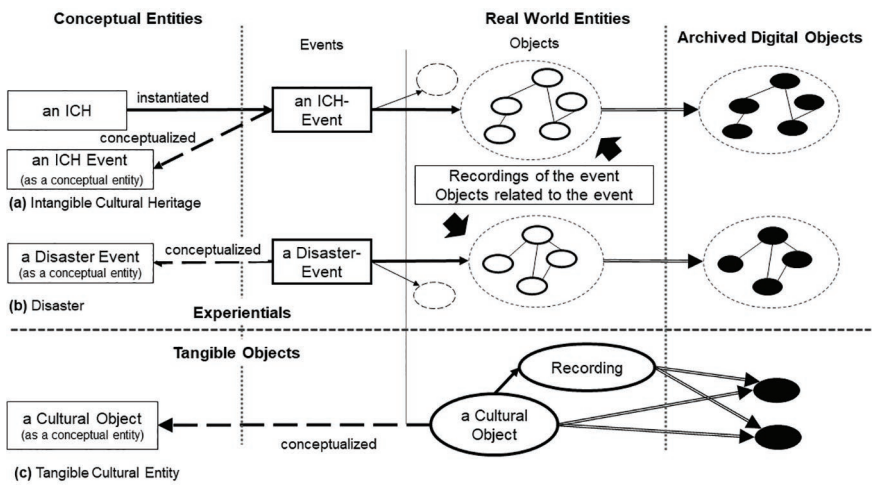

Figure 2.7 CEDA-X: A simple conceptual model of digital archiving.

object. In Figure 2.7, a dotted circle could be recognized as a digital surrogate of an event which may be used in an ADO.

A simple model for a tangible cultural object is shown in Figure 2.7(c). A cultural object may have one or more ADOs and may be linked to a conceptual entity which can be used as a symbolic entity by which the object is widely recognized.

Each event in Figure 2.7 means a real-world entity that we can experience. However, since we cannot directly convert an event to an ADO, we need to use recordings created for the event and/or tangible objects related to the event as the original object from which ADOs are created. An event may be archived as a set of those ADOs. Thus, *Event* is a key entity to differentiate intangible entities from tangible entities. As an event means something which we can experience, ICH and disasters are included in the Experientials category in Figure 2.7.

Figure 2.8 shows another view of Experientials. "A Conceptual Entity" on the left may be used as a symbolic instance to identify those conceptual entities such as ICH, ICH Events, Disaster Events, and Cultural Objects. "Descriptions/Metadata" beneath "a Conceptual Entity" and "an Event" are to be realized as a digital object and collected in a digital archive together with the ADOs on the right. The ADOs are created from various types of recordings and documentation as well as various types of original objects.

Thus, *Objects* and *Experientials* are the two major categories of the entities which are to be digitally archived in CEDA-X.

- *Objects* may be physical, digital, analog, or hybrid.
- *Experientials* are embodied as an event, which may be cultural events such as dance performances, practices of craftmanship, game plays, or social events such as disasters and incidents.

Those events are recorded, which means objectification of events, and the recordings together with various objects related to the events are digitally archived. Some types of objects such as ephemeral objects, machineries, and digital media arts, may be modeled as Experientials if their digital archiving focuses on their dynamic features which can be modeled as an event.

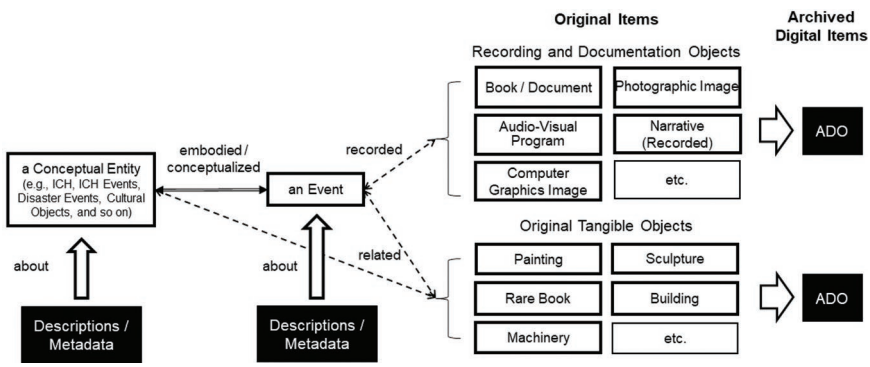

Figure 2.8 CEDA-X: A structural view of digitally archived experientials.

3.3 Summary

Figure 2.7 shows ICH and disasters as two cases of Experientials. As discussed in Section 2.1, there are various intangible and experiential entities that can be modeled as Experientials. Figure 2.7 shows two categories of Experientials: ICH and disasters. As roughly discussed in Section 2.1, a fundamental aspect to differentiate them would be intentions to make an event happen. On one hand, ICH, performing arts, and exhibitions are intentionally embodied by a person and/or community; on the other hand, man-made/natural events such as wars, elections, and earthquakes are not. Figure 2.7 uses ICH and disasters as two typical cases of these two categories. In both cases, events are the real-world entities from which we can create digital objects for archives via recordings and related tangible objects. The authors consider that these models can be applied to various genres such as performing arts, interactive arts, sports, games, and, among others.

In the Tangible Objects category at the bottom of Figure 2.7, no event entity is included, which is the basic difference between the Tangible Objects and Experientials categories. However, the border is not very crisp because there are tangible objects which have dynamic features namely, features with temporal dependency such as ephemeral objects and objects in motion.

There are many digital archives developed for intangible and experiential entities such as ICH and disasters. The underlying metadata models used in those digital archives are oriented to description of collected items (item-centric), and those conceptual entities and events defined in CEDA-X may not be explicitly defined in those models. The authors consider that item-centric description is quite natural from the viewpoint of organization of collections of physical items, although it does not satisfy the requirements to explicitly include those conceptual entities and events in the organization of the collections. The data models of CEDA-X are proposed to include those conceptual entities and events as first-class entities as well as archived digital objects to make them connectable to each other in the networked information environment.

4 Concluding remarks

This chapter aimed at discussing metadata models for the digital archiving of intangible and experiential entities. In this chapter, we first defined several basic concepts required for the discussion. Metadata models of existing digital archives are mostly defined based on the metadata models designed for conventional institutional collections. Those models are primarily designed to describe the collected resources item by item. It is quite natural for memory institutions to apply the conventional item-oriented description schemes not only to their collections of physical items but also to their digital collections because they have rich experiences and resources obtained during their extensive experiences. On the other hand, the information environment of digital archives has been greatly transformed in the recent decades by the progress of the Internet, in particular the Linked Data technologies, and by the broad acceptance of those technologies by memory institutions and their communities. Any entities identifiable on the Internet may be linked to, regardless

of their types or sizes, which enables us to develop new functions for digital archives such as the combination and aggregation of objects across digital archives and genres of the collections.

This chapter provides only a conceptual view of CEDA-X and it does not include definitions of the classes or properties of metadata. The reason for this is that CEDA-X is defined as a domain-agnostic model and any domain-specific metadata schemas and ontologies may be used on top of CEDA-X. However, the authors consider that mappings to metadata standards and standard ontologies commonly used in the cultural domains are required to provide a more formal definition of CEDA-X such as CIDOC CRM, which is beyond the scope of this chapter, but is left as work for the future.

References

Agency for Cultural Affairs. (2023). *Media arts database* (in Japanese). Retrieved from https://mediaarts-db.bunka.go.jp/

CIDOC CRM. (2023). *What is the CIDOC CRM?* ICOM/CIDOC Documentation Standards Group. Retrieved October 7, 2023, from https://cidoc-crm.org

Finto. (2023). *Finnish National Thesaurus and Ontology Service*. Retrieved November 8, 2023, from https://finto.fi/yso/en/

Hyvönen, E. (2023). Digital humanities on the semantic web: Sampo model and portal series. *Semantic Web, 14*(4), 729–744. https://doi.org/10.3233/SW-223034

IFLA. (2009). *Functional requirements for bibliographic records*. IFLA Study Group on the Functional Requirements for Bibliographic Records. Retrieved October 23, 2023, from https://cdn.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr/frbr_2008.pdf

Isaac, I. (Ed.). (2013). *Europeana data model primer*. Europeana Professional. Retrieved October 7, 2023, from https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf

International Organization for Standardization (ISO). (2023). *ISO 21127:2023 Information and documentation: A reference ontology for the interchange of cultural heritage information*. ISO. Retrieved November 10, 2023, from https://www.iso.org/standard/85100.html

National Diet Library. (2013). *National diet library great east Japan earthquake archive*. Retrieved October 7, 2023, from https://kn.ndl.go.jp/en/#/

Riva, P., Le Bœuf, P., & Žumer, M. (2017). *IFLA library reference model: A conceptual model for bibliographic information*. Consolidation Editorial Group of the IFLA FRBR Review Group. IFLA. Retrieved October 23, 2023, from https://repository.ifla.org/bitstream/123456789/40/1/ifla-lrm-august-2017_rev201712.pdf

Sugimoto, S., Wijesundara, C., Mihara, T., & Fukuda, K. (2021). Modelling cultural entities in diverse domains for digital archives. In Golub, K., & Liu, Y-H. (Eds.), *Information and knowledge organisation in digital humanities* (pp. 25–42). Facet Publishing. https://doi.org/10.4324/9781003131816-2

Sugimoto, S., Wijesundara, C., Mihara, T., & Nagamori, M. (2022). A generalized data model for digital archiving in cultural and historical domains. In Calzada-Prado, F. J. (Ed.), *Boosting the knowledge economy* (pp. 121–139). Elsevier. https://doi.org/10.1016/B978-1-84334-772-9.00003-8

Theatre Museum of Waseda University. (2023). *Japan digital theatre archives*. Retrieved from https://enpaku-jdta.jp/en

UNESCO. (2023a). *UNESCO intangible cultural heritage*. Retrieved October 7, 2023, from https://ich.unesco.org/en/what-is-intangible-heritage-00003

UNESCO. (2023b). *Outline of digital archiving project*. UNESCO Video and Sound Collections. Retrieved October 7, 2023, from https://www.unesco.org/archives/multimedia/?pg=14

University of Tokyo. (2023). *Archive series*. Hidenori Watanabe Laboratory. Retrieved October 7, 2023, from https://labo.wtnv.jp/p/blog-page_29.html

Visual Resource Association. (2014). *An introduction to VRA Core*. LOC. Network Development and MARC Standards Office of the Library of Congress. Retrieved October 7, 2023, from https://www.loc.gov/standards/vracore/VRA_Core4_Intro.pdf

Wijesundara, C., Monika W., & Sugimoto, S. (2017). A metadata model to organize cultural heritage resources in heterogeneous information environments. In S. Choemprayong, F. Crestani, & S. Cunningham (Eds.), Lecture Notes in Computer Science (LNCS): Vol. 10647. *International Conference on Asian Digital Libraries (ICADL)* (pp. 81–94). Springer. https://doi.org/10.1007/978-3-319-70232-2_7

Wijesundara, C., & Sugimoto, S. (2019). Shifting from item-centric to content-oriented: A metadata framework for digital archives in the cultural and historical domains. *9th Asia-Pacific Conference on Library and Information Education and Practice (A-LIEP)* (pp. 310–324).

3 Military history on the Semantic Web

Lessons learned from developing three in-use linked open data services and semantic portals for digital humanities

Eero Hyvönen

1 Introduction

Georg Friedrich Hegel (1770–1831) maintains that *we learn from history that we learn nothing from history*. He seems to be right, given the current raging war in the Ukraine happening not so long after the Second World War (WW2), which is arguably the most devastating catastrophe of human history, and the First World War only 20 years before that. In spite of this, the discussion in this chapter maintains that *the more we learn about the price of a war, the less there will be wars*. It is argued that the evolving Semantic Web (Bizer et al., 2009; Hitzler, 2021) and Linked Open Data (LOD) technologies (Heath & Bizer, 2011; Hyvönen, 2012) offer a promising way to expose war data, such as deaths, to researchers and the general public for study and learning.

As a proof of concept, a series of three LOD services and semantic portals listed in Table 3.1 have been designed, implemented, and published for public use in Finland on the Semantic Web.

1 WarSampo aggregates and publishes data about the Second World War (WW2) in Finland from some 20 data sources and several collaborating organizations.
2 WarVictimSampo 1914–1922 is a related system based on the death records and battles of the Civil War in Finland and Kindred Wars during 1914–1922, with a focus on data analytics for Digital Humanities research.
3 WarMemoirSampo demonstrates a novel way of publishing and watching videos on the Semantic Web, with a focus on publishing memoirs of WW2 veterans on the Semantic Web.

These prototype portals have become quite popular in Finland suggesting feasibility of the approach. For example, WarSampo has had 1.2 million distinct users in a small country like Finland of only 5.5 million inhabitants. WarSampo and WarVictimSampo 1914–1922 are based on death records of their topical wars that were originally provided by the National Archives of Finland via separate online services. In 2022, the provision of these legacy systems was terminated, and their users were redirected to the corresponding Sampo systems.

DOI: 10.4324/9781032707211-4

Table 3.1 Three Sampo portals and LOD services for digital humanities on military history. Distinct user counts, based on site visits, using Google Analytics by July 2023

Portal	*Published*	*Domain*	*# Users*	*# Triples*	*Data owners*
WarSampo[a]	2015–2019	World War II	1,200,000	14M	National Archives, Defense Forces, and others, Finland
WarVictimSampo 1914–1922[b]	2019	Finnish Civil and Kindred wars	53,000	10M	National Archives of Finland
WarMemoirSampo[c]	2021	War memoirs on videos	4,400	0.32M	Tammenlehvän Perinneiitto ry and National Archives, Finland

[a] Portal: https://sotasampo.fi; project: https://seco.cs.aalto.fi/projects/sotasampo/
[b] Portal: https://sotasurmat.narc.fi/; project: https://seco.cs.aalto.fi/projects/sotasurmat-1914-1922/
[c] Portal: https://sotamuistot.arkisto.fi; project: https://seco.cs.aalto.fi/projects/war-memoirs/

The three tabulated portals share not only their domain of discourse (military history), but also the underlying semantic infrastructure, the model of aggregating and publishing LOD as a service, and the user interface (UI) logic of the portals. These general commonalities, based on the Sampo Model (Hyvönen, 2023a) and the Sampo-UI Framework (Ikkala et al., 2022; Rantala et al., 2023), as well as on the Finnish LOD infrastructure (Hyvönen, 2024), are the topics of this chapter. It is suggested that the model and the tools presented could be of use for similar new applications locally and globally.

In the proceeding sections, these three systems on military history are first briefly introduced (Section 2). Afterwards, lessons learned in creating and publishing LOD on the Semantic Web are discussed (Section 3), followed by issues related to UI design (Section 4). Application of the principles of the Sampo Model to the three portals is demonstrated. In conclusion (Section 5), related works are discussed, contributions summarized, and directions for further research suggested.

2 Three LOD services and semantic portals for military history

The applications of Table 3.1 demonstrate how to aggregate and enrich heterogeneous, distributed datasets into harmonized Knowledge Graphs (KG), based on a shared ontology infrastructure. In this chapter, a knowledge graph is defined as a semantic network that represents real-world entities (objects, concepts, events, etc.) based on an ontological data model. The KGs can be published as LOD services and then be used for (1) data analyses in Digital Humanities (DH) with tools such as the Yasgui editor[1] (Rietveld & Hoekstra, 2017), Google Colab,[2] and Jupyter Notebooks,[3] and for (2) developing ready-to-use applications. The Sampo portals,

where faceted search and browsing are integrated seamlessly with data analytic tools, test and demonstrate the applications of the LOD services.

2.1 WarSampo – Finnish Second World War on the Semantic Web

WarSampo (Hyvönen et al., 2016) is a system for publishing and studying collections of heterogeneous, distributed data about the Second World War in Finland on the Semantic Web. Its Knowledge Graph (KG) (Koho et al., 2021) is based on aggregating and harmonizing some 20 war-related data sources, which makes it possible to enrich the datasets semantically with each other's contents. The key dataset originates from a database of some 95,000 WW2 death records of the National Archives of Finland; rich metadata about everybody killed in action during the war in Finland is included in the database.

As customary in Sampo systems, WarSampo has two components: first, there is a LOD service for DH research and for creating web applications. Secondly, a semantic portal has been created to test and demonstrate the usability of the data service. The portal allows both historians and laymen to study military history and destinies of the soldiers in the war from different interlinked perspectives. The data model used is event-based extending the CIDOC CRM standard[4] to event types of war, such as battle, bombardment, and personal histories like getting wounded, getting killed, and promotion in rank.

WarSampo was published first in 2015 with six application perspectives into the data: events, persons, army units, places, casualties, and memoir articles. After this, new datasets and perspectives have been added into the system, including a search view for 160,000 authentic photographs provided by the Defense Forces of Finland, a perspective of 672 war cemeteries (Ikkala et al., 2017), and yet another one about the 4,200 Finnish prisoners of war in the Soviet Union between 1939 and 1945 (Koho et al., 2020).

A key innovation of WarSampo is to try to re-assemble the life histories of the soldiers automatically by data linking. For example, if one's relative was killed in action, he can be found in the death records of the KG showing the army unit in which he served. Data on battles, movements, and other activities of army units can be found and, by extension, the actions of an individual as part of the unit, assuming that the individual was serving in the unit. The events related to a person can be illustrated on maps and timelines. Additional data, such as photographs, can be linked not only to persons and units mentioned, but also to events based on places and times mentioned in the metadata using Named Entity Linking (Heino et al., 2017). Furthermore, links to the actual war diaries can be provided based on the army unit data. In addition to reassembling personal war histories, the data has also been used for data analyses (Koho & Hyvönen, 2023; Koho et al., 2017).

For example, Figure 3.1 illustrates data about a unit selected on the left (a). Related events are shown on a timeline (c), and those in the particular time window in the middle are visualised on the map (b) with a heat map about the casualties within the currently visible time window. On the right, information regarding the unit is

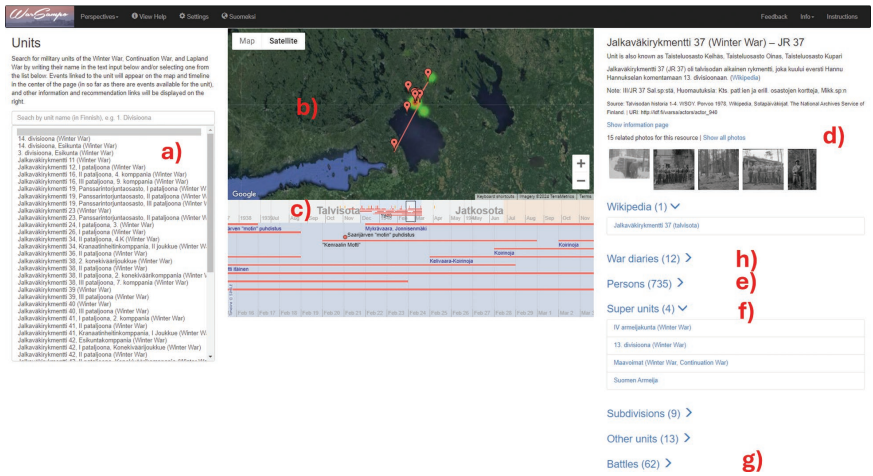

Figure 3.1 A view from the Army Units application perspective of WarSampo showing aggregated spatio-temporal data about the history of the Infantry Regiment 38.

shown, including photographs related to it (d). Additional links are shown to 445 persons in the unit (e), related units (f), battles (g), and authentic war diaries (h) of the unit for primary source data.

2.2 WarVictimSampo – Casualties of Finnish wars 1914–1922

WarVictimSampo 1914–1922 is a follow-up system of WarSampo, a semantic portal and LOD service about the war victims, battles, and prisoner camps in the Finnish Civil War in 1918 and casualties of the Kindred Wars (Rantala et al., 2020, 2022). The system contains detailed information about some 40,000 deaths extracted from several data sources, data about over 1,000 battles of the Civil War between the "Reds" and "Whites", and some additional related data created, compiled, and linked during the project.

A key novelty of WarVictimSampo 1914–1922 is the integration of ready-to-use Digital Humanities visualisations and data analysis tooling with semantic faceted search and data exploration. This facilitates the study of data about wider prosopographical groups in addition to individual war victims.

The portal includes two application perspectives, one for studying the victims of war and one for the Civil War battles. As customary in the Sampo model, these perspectives correspond to classes in the underlying KG and allow searching, browsing, and analyzing their individuals, here victims and battles. Figure 3.2 illustrates how the victims can be studied. On the left there are 21 facets for filtering out the casualties; only four first ones are visible in the figure due to limited space, i.e., Name, Birth date, Death date, and Party. The user has made two selections: Party = "Reds", and the Place of registry = "Viipuri Province". The result set can then be studied and visualised on the right using five tabs that can be selected from the top:

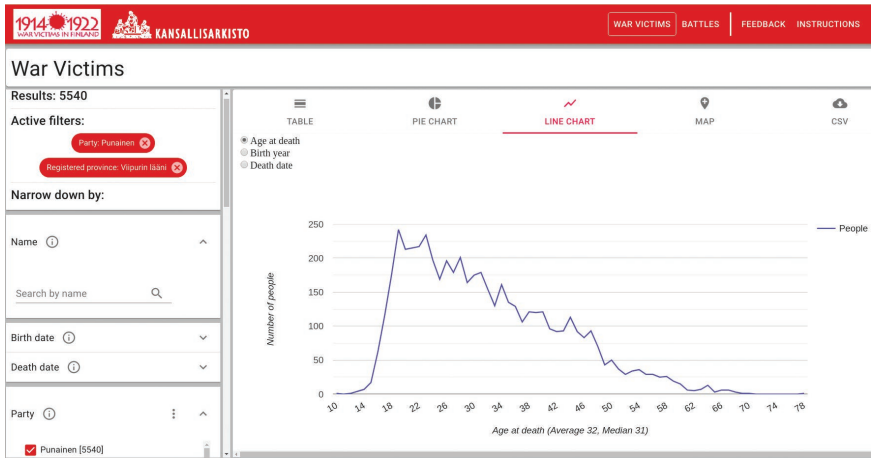

Figure 3.2 The age distribution of people who supported the Reds side in the Finnish Civil War and who were from the Viipuri Province as shown in WarVictimSampo 1914–1922 portal. The LINE CHART tab can be used for visualizing the birth and death years, too.

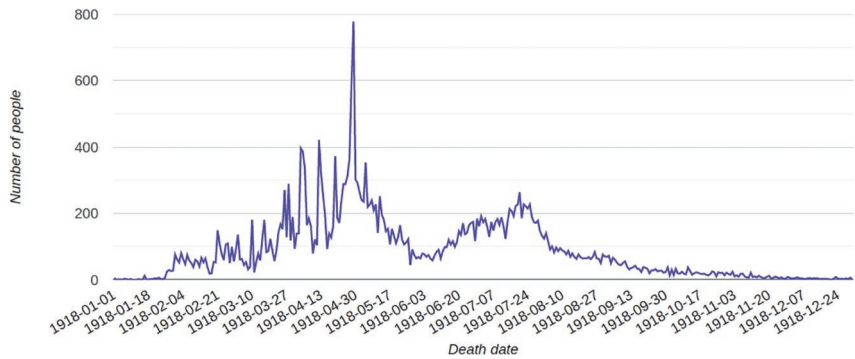

Figure 3.3 Distribution of death dates in 1918 in the knowledge graph of WarVictimSampo 1914–1922.

TABLE lists the filtered people, PIE CHART shows their distributions along different facets, LINE CHART visualises the people using a few numerical property values, the MAP tab shows the deaths of the selected people on a map, the CSV tab makes it possible to download the results in CSV format. The last option was considered important by history researchers as it allows re-use of the data in various other tools, such as spreadsheet programs.

In Figure 3.3, the line chart visualisation was selected for the filtered victims. Currently, there are three different options for the *x*-axes of this visualisation, namely, Age at death, Birth year, and Death date. The example given shows a

selection of the Age at death option automatically plotting a line chart with the *x*-axis representing age in years and *y*-axis representing the number of victims.

The average and median values are also shown under the visualisation graph. The user can then easily and quickly compare distributions and average values between certain subsets of people in the data. For example, in Figure 3.2 we have used the facets to select victims who supported the Reds side in the Civil War and who were registered to the Viipuri Province. Comparing this distribution to people from other provinces shows that war victims who supported the Reds side of the civil war from the Viipuri Province tended to be older than those from other provinces according to the data, with the median age of over 30 years. Explaining this would require more detailed analysis and close reading of the data, but this demonstrates how faceted search combined with data analytic tools can be effective at discovering interesting phenomena in the data.

The MAP tab view shows the death places of the victims on a map. This map is clustered so that nearby places are grouped together depending on the zoom level. Each cluster shows the number of victims that died in that area. The death places are shown on a municipality level. A victim's information is not shown on the map if there is no death geo-data available about them.

The other perspective of the portal is the Battles of the Finnish Civil War. This application perspective works in a way similar to the war victims' perspective. The user can search and filter the battles using facets and the results can be visualised in different ways on tabs; in this case as a table, a map, or an animation. The temporal animation visualisation is unique to the battles. It shows the battle sites at different times on a map as markers. A marker appears on the map as red when the animation reaches the starting date of the battle and then stays on the map as gray turning into black when the animation progresses in time. Figure 3.4 shows the situation on March 10, 1918. You can see how a front line has been formed across Finland from east to west, and how the battles are raging along it.

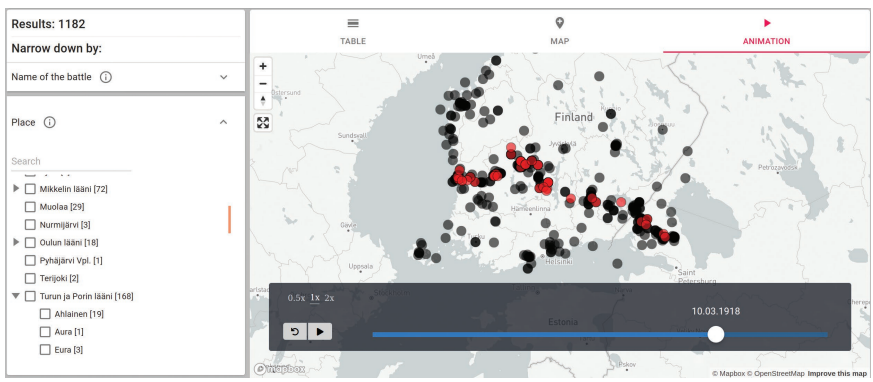

Figure 3.4 The battles animation view of the WarVictimSampo 1914–1922 portal, stopped on March 10, 1918. A front line can be seen as red markers from East to West. The battles can be filtered using the facets on the left.

2.3 WarMemoirSampo – Memoir interviews of WW2 veterans

WarMemoirSampo focuses on preserving and publishing memoirs of the WW2 veterans for future generations (Hyvönen et al., 2022). The dataset here is a set of video interviews created and stored by the Finnish WW2 Veteran Association *Tammenlehvän perinneliitto* and the National Archives of Finland.

A technical novelty of the system is to enable scene segments in videos to be searched by their semantic textual content (Koho et al., 2022; Leal et al., 2022). This makes it possible, for example, to find particular points in long videos, where a person, place, or other entity is mentioned. Another key idea of WarMemoirSampo is to enhance video-watching experience by data linking, whereby while watching a video, additional contextual information is provided dynamically based on the underlying LOD (Hyvönen et al., 2022).

The system is based on and demonstrates the idea of re-using the WarSampo infrastructure and KG in other applications. The LOD for WarMemoirSampo has been extracted automatically from time-stamped textual natural language descriptions of the video contents; the data is interlinked not only internally but also externally with the WarSampo KG.

Based on the Sampo model and Sampo-UI framework, the landing page of WarMemoirSampo portal provides three application perspectives to the underlying KG with faceted semantic search as follows:

1 *Interviews perspective* – used for searching whole videos based on their nine key properties, namely, interview content (text facet for traditional search), interviewee, interviewee gender, and mentioned places, persons, military units, organizations, events, other entities, and topic.
2 *Scenes perspective* – used for searching video scenes inside interviews using the same facets as in the interviews perspective.
3 *Directory perspective* – contains a total of about 3,000 entities mentioned in the texts with direct links to the scenes where the entities were mentioned. It is a kind of semantic index of the underlying dataset.

Figure 3.5 depicts the *Scenes perspective*, where the user has selected "Carl Gustav Emil Mannerheim" on the facet *Person*. The 40 scenes mentioning this marshal are shown on the right-hand side with metadata links for further information. By clicking on a video, it is opened for dynamic viewing as depicted in Figure 3.6. Links to additional information are provided on the fly. Selecting the tab *Map* on top shows the places mentioned on a map; Figure 3.7 shows all 4,566 of them. A click on a marker on the map opens a pop-up window with links to all related scenes. Finally, the tab *Word Cloud* summarizes the topics of the video interview as a word cloud based on the extracted subject matter concepts in the KG. The concepts were extracted by an automatic annotation tool (Leal et al., 2022) using the subject matter ontologies of the Finnish national ontology infrastructure (Hyvönen, 2024).

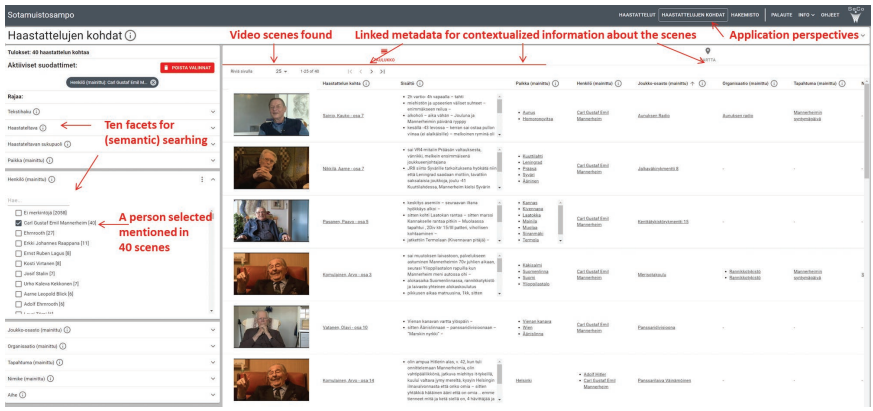

Figure 3.5 Faceted search for scenes inside videos.

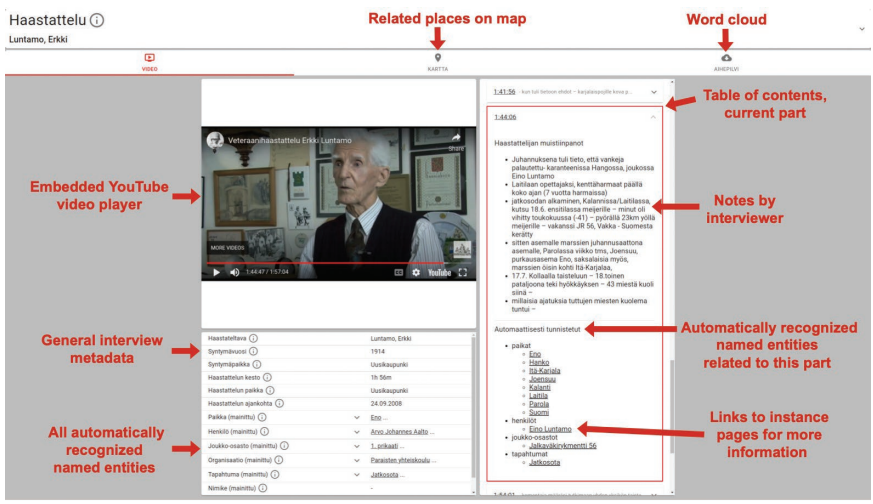

Figure 3.6 Video viewing page with a dynamic table of contents for contextual linked data.

3 Lessons learned: Creating and publishing cultural heritage LOD

Since 2002, the Semantic Computing Research Group (SeCo)[5] at the Aalto University and University of Helsinki has been involved in creating a national semantic web infrastructure. To test and demonstrate its usability, over 20 LOD services and semantic portals based on them have been created, mostly in the domain of Cultural Heritage (CH) and DH. During this work, the experiences have gradually evolved into the Sampo model, a set of general principles on (1) how to create LOD services and (2) user interfaces on top of them.

Regarding LOD creation there are three major principles P1–P3 in the model.

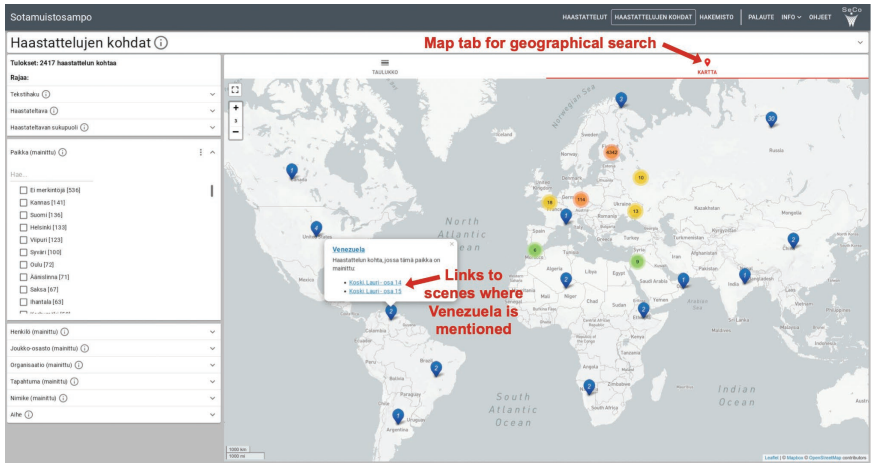

Figure 3.7 Map view to access scenes that mention a place.

3.1 Support collaborative data creation and publishing (PI)

Leonardo da Vinci said: *Learn how to see. Realize that everything connects to everything else*. This wisdom applies well to the general idea of LOD where mutually interlinked aggregated datasets are used to enrich each other. When creating search and data exploration systems based on data aggregation, there are two basic approaches available.

1 *Distributed strategy*: Federated search, where the traditional approach is to take the user's query, send it to distributed local data services hosting the data to be aggregated, collect the answers, and present them to the user.
2 *Centralized strategy*: Aggregating global data, where the approach is to aggregate and harmonize the distributed heterogeneous datasets first into a global database or KG, and apply the query to the centralized data service.

In the distributed strategy, the burden of figuring out what the user wants can be distributed to the local data providers that transform the query for their local databases. Also, the burden of actually executing the query can be distributed. However, it is difficult to transform the query and present results in a semantically interoperable way in local services whose data models and vocabularies[6] used in the metadata are different. This deteriorates precision and recall, and makes data analyses challenging. For example, entities, such as persons and places, are typically represented in different ways locally and, therefore, confused with each other. Furthermore, not having simultaneous access to the global data is a severe restriction on what can be analyzed globally. For example, finding out relations between entities in local datasets is hard. In the Sampo model, the centralized strategy was therefore selected, introduced already in the first Sampo system MuseumFinland (Hyvönen et al., 2005) (online since 2004). However, using the global strategy

brings in its own challenges. These challenges are of concern to data model harmonization of the local datasets and disambiguating and linking the data instances for semantic interoperability. However, these challenges are not due to the centralized strategy, but to the heterogeneity of the local datasets and the ways they are created, and have to be addressed in any case when dealing with local data in a semantically proper interoperable way.

3.2 Use a shared open ontology infrastructure (P2)

Albert Einstein said: *Intellectuals solve problems. Geniuses prevent them*. This wisdom applies well to the idea of developing and using an infrastructure in creating CH and DH applications (Hyvönen, 2024): It is arguably better to prevent interoperability problems in advance during data creation, than fixing problems afterwards when aggregating data (Hyvönen, 2010). According to our experiences, most of the time is "wasted" in projects like the Sampo systems in cleaning, aligning, and harmonizing distributed local datasets, which could have been avoided by using shared data models and vocabularies in populating them in databases in the first place.

Our work on developing the Sampos has, therefore, been supported by a systematic effort on creating a national semantic web infrastructure for vocabularies and by using standard data models, such as CIDOC CRM and Dublin Core. For example, ontologies developed from national keyword thesauri as well and ontologies for historical places, actors, and occupations have been re-used and developed further step-by-step in the Sampo systems. This work started in 2003 as the series of FinnONTO projects[7] in which, for instance, a LOD cloud called KOKO of interlinked domain-specific ontologies was created (Frosterus et al., 2015; Hyvönen et al., 2008) and published as a national ontology service ONKI.fi (Tuominen et al., 2009), deployed by the National Library of Finland in 2014 as the contemporary Finto.fi service (Suominen et al., 2014). This work is now continuing as part of the national FIN-CLARIAH initiative for DH research infrastructures in Finland,[8] with a goal to combine the European CLARIN[9] and DARIAH[10] infrastructures on a national level.

3.3 Make clear distinction between the LOD service and the user interface (UI) (P3)

This principle was tested first when developing the ontology service ONKI Light for SKOS vocabularies (Suominen et al., 2012): is it possible to re-implement the original ONKI.fi ontology services (Tuominen et al., 2009; Viljanen et al., 2009) by using SPARQL queries only for data access? The answer was "yes", and the Finto.fi ontology service was deployed based on ONKI Light. Another related step was to test whether it makes sense to apply this idea to implement faceted semantic search, too, used in the early Sampo systems starting from Museum Finland. The answer was "yes" again, and this development led to developing the tools SPAQRL Faceter (Koho et al., 2016), used in WarSampo, and later Sampo-UI in 2018. By 2023, Sampo-UI has been used in some 15 Sampos including WarVictimSampo 1914–1922 and WarMemoirSampo.

3.4 Maintaining linked open data and data services

According to Heraclitus (fl. 500 BC), *everything changes and nothing remains still, and you cannot step twice into the same stream*. An important issue of using LOD is maintaining changes in the KG as time goes by and software evolves. However, the Sampo principles discussed above focus only on how to create and publish a LOD service.

A piece of good news regarding the challenges of change is that linked data formats are open, standardized by W3C recommendations, and are based on text. The data is, therefore, pretty sustainable and re-usable, but tools, such as triple stores and UI frameworks that use the data change more often and may support and extend the standards, such as the SPARQL query language, in different ways.

A more severe challenge is what to do, when either the metadata models (Zeng & Qin, 2022), or vocabularies used in populating the models, and the data itself evolves. For example, the WarSampo KG has been extended with new data as new application perspectives were developed. Then new vocabularies or entries in old ones may be needed, and the entities in the new data need to be aligned with pre-existing ones in the underlying infrastructure. This problem is discussed in (Koho et al., 2018a, 2018b).

There are two basic approaches depending on how the primary data is managed. If the data is maintained in a legacy system using traditional formats, it makes sense to design the LOD transformation in such a way that it can be re-run automatically from scratch. This means that there should preferably be no intermediate manual phases in the process, as their results would be overridden when the KG is reconstructed. The challenge in this approach is that the new data is likely to contain typos and linking textual descriptions may need subsequent manual work and fixes. In order to determine quality issues in the linked data, semantic validation languages and frameworks, such as SHACL[11] and ShEx[12] can be used.

Another better approach would be managing the KG in its native linked data form. This would keep the data automatically consistent and ready to be uploaded into a LOD service. Unfortunately, there are still few tools for editing and managing RDF data. An exception to this are ontology editors, such as Protegé[13] and Topbraid Composer.[14] In the case of the Sampo systems, the SPARQL SAHA editor (Mäkelä & Hyvönen, 2014) was, for instance, developed and has been used in maintaining the BookSampo KG (Mäkelä et al., 2013) by the data owners, namely, the public libraries of Finland. SAHA is also used for maintaining the RDF data of OperaSampo on historical opera and musical performances (Ahola et al., 2023) by its data owner, the Sibelius Academy.

4 Lessons learned: User interfaces for digital humanities research

After a LOD service has been established, its data can be used in the following two ways:

1 *Using Application Programming Interfaces (API)*. The LOD publication methodology provides different ways to access the data: (1) The data can be downloaded from the service as data dumps. (2) The data can be browsed in a human

readable way using a linked data browser.[15] (3) The LOD service provides content negotiation where URIs can be resolved and either data for the machine or HTML for the human user can be returned.[16] (4) Most importantly, the data service can be queried in flexible ways using the SPARQL query language[17] and endpoint. There are easy-to-use tools, such as Yasgui (Rietveld & Hoekstra, 2017), for editing and executing SPARQL queries with some built-in visualisation options for the results. The SPARQL endpoint can be accessed from any programming environment, such as Jupyter notebooks and Python scripting for querying and analyzing data.

2 *Using portals and other applications*. Ready-to-use applications for accessing and using the data without programming skills can be developed on top of the LOD service, as exemplified by the Sampo portal series.

In the following section, some lessons learned are discussed in developing UIs for the three portals of this paper, as formulated in the Sampo model and the Sampo-UI framework.

Figure 3.8 illustrates the navigational structure of using a Sampo-UI-based portal. The user first lands on the *landing page* with several *application perspectives* to the data. The perspectives are based on classes of the underlying KG, such as Artefacts, Persons, and Places. The usage cycle of each perspective can be divided into two steps: (1) filter and (2) analyze. The user first filters the data by using the faceted semantic search (Hearst, 2006; Tunkelang, 2009) tools provided by the portal. The results, as well as the facet option hit counts are updated after each category selection on a facet, making it possible for the user to precisely filter the end-result entities by different aspects, for instance, filtering by party and

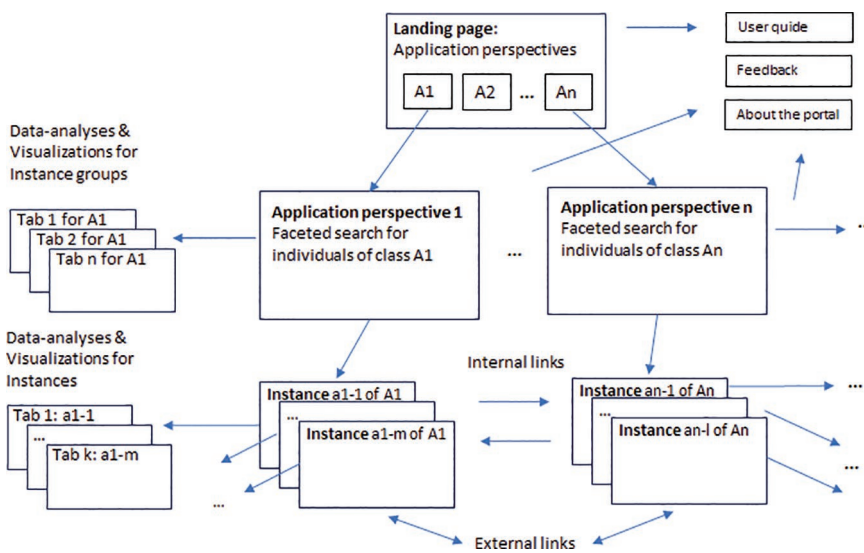

Figure 3.8 Navigational page structure of a portal based on Sampo-UI.

registration municipality as shown in Figure 3.2. In faceted search, the hit counts direct the search and prevent arriving at dead-end situations where no results are found. Faceted search was developed already in the 1990s and early 2000s but under the name "view-based search" (Hyvönen et al., 2004; Pollitt, 1998) and also as "dynamic taxonomies" (Sacco, 2005).

After filtering the data to the wanted subset, the *target group*, the user can analyze the results set – a set of instances of the class corresponding to the application perspective, with integrated data analytic tools available as tabs on the application perspective page. An example of a visualisation can be seen in Figure 3.2.

It is also possible to select a particular instance of the result set for a closer look: each instance has an *instance page* that provides aggregated information about the individual with internal and external links for further information to browse. In addition, instance pages may also have a set of tabs that provide contextualized data analyses of the individuals in the same way as for target groups. For example, BiographySampo (Hyvönen et al., 2019) and AcademySampo (Leskinen et al., 2022) contain an application perspective corresponding to the class *Person*. Visualisations for studying prosopographical target groups, such as persons with the same occupation or place of birth, are available on different tabs of the application perspective page, and on instance pages of each particular person there are tabs such as the one for studying ego-centric networks of individuals.

This *filter-analyze* two-step usage cycle allows an iterative approach to exploring the data (Marchionini, 2006; Tzitzikas et al., 2017). It is possible to find potentially interesting subsets and individuals in the data without having to be already familiar with the content. By providing a text facet, it is also possible to support use cases where the user is looking for a specific instance, say a person with a known name, thereby formulating the search query easily.

The UI logic above is based on the three principles P4–P6 articulated in the Sampo model:

1 *Provide multiple perspectives to the same data (P4)* The idea here is the same as in the FAIR principles,[18] but adapted to UI design – reusing the data even within one UI. The class structure of KGs provides a natural approach for this; classes such as *Person, Place,* and *Battle,* can be used as a basis for searching their individuals (people places, battles, etc.) in the application perspectives. In our case, WarSampo has nine application perspectives, WarVictimSampo 1914–1922 two, and WarMemoirSampo three.
2 *Standardize portal usage by a simple filter-analyze two-step cycle (P5)* This idea, discussed above, was inspired by the prosopographical research method on groups of people (Verboven et al., 2007), where a target group of people sharing some common features is first filtered out and then analyzed in more detail.
3 *Support data analysis and knowledge discovery in addition to data exploration (P6)* Finally, in addition to semantic faceted search and data exploration, one should consider providing the user with intelligent tools for analyzing the data, or even with intelligent agents for establishing interesting knowledge patterns inherent in the data, solving research problems, and explaining the results to the user, leading to "third generation" systems in DH (Hyvönen, 2020).

5 Contributions, related work, and discussion

In this section, the overall contributions of using LOD in the three Sampo systems presented are summarized, followed by a review on related works. Finally, some critical issues on using LOD are discussed and a view into the future is given.

5.1 Contributions

There are three main technical reasons for using LOD in publishing and using contents about Military History, and more generally about Cultural Heritage:

1 It is possible to enrich everybody's data collaboratively from separate data silos. In other words, *everybody can win through collaboration.*
2 By creating Findable, Accessible, Interoperable, Re-usable data, as suggested by the FAIR principles for scientific data management and stewardship, the value of data increases in the four FAIR dimensions.
3 By using Semantic Web Linked Data semantics (Hitzler et al., 2010), based on first-order logic, the machine can "understand" the data and, for example, enrich the data by reasoning and consequently, solve problems. This means that creating more intelligent applications for the public, curators, and researchers is possible and more cost-efficient.

From a DH perspective the idea of using LOD services and semantic portals on top of them is promising in filtering out patterns of possibly interesting phenomena in Big Data using distant reading (Moretti, 2013). However, also close reading can be supported through verifying and interpreting the data analytic results.

5.2 Related works

There are several projects that have published linked data about wars, such as World War I: Europeana Collections 1914–1918,[19] 1914–1918 Online,[20] WW1 http://www.1914-1918-online.net, Discovery,[21] Out of the Trenches,[22] CENDARI,[23] and Muninn.[24] Our work on WarSampo was inspired by our earlier system WW1LOD on WW1 data that was created in collaboration with the University of Colorado, Boulder (Mäkelä et al., 2017).

In addition to WarSampo, there are a few works that have used the Linked Data approach to WW2, such as (Collins et al., 2005; de Boer et al., 2013) Open Memory Project[25] on holocaust victims, and the Dutch project Netwerk Orloogsbronnen.[26] In Wang (2023), knowledge graph technologies are studied for integrating heterogeneous data related to wars, in this case the Second Sino-Japanese war 1937–1945 as a case study.

The ideas behind the Sampo model have been explored and developed before in different contexts. A case in point is the notion of collaborative content creation by data linking, which is a fundamental idea behind the Linked Open Data Cloud movement[27] and has also been developed in various settings such as ResearchSpace.[28] The idea of providing multiple analyses and visualisations to a set of filtered search results has been employed in other portals, such as the

ePistolarium[29] (Ravenek et al., 2017) for epistolary data, and using multiple perspectives have been studied as an approach to decision making (Linstone, 1989). Faceted search (Hearst, 2006; Tunkelang, 2009), also known as view-based search (Hyvönen et al., 2004; Pollitt, 1998) and dynamic taxonomies (Sacco, 2005), are well-known paradigms for explorative search and browsing (Marchionini, 2006) in computer science and information retrieval, based on S. R. Ranganathan's original ideas of faceted classification in Library Science in the 1920s (Ferreira et al., 2017). The two-step study model used in our work has been used, for instance, in prosopographical research (Verboven et al., 2007) (without the faceted search component). The novelty of the Sampo Model lies in combining several ideas and operationalizing them for developing LOD services and applications in Digital Humanities.

The research area of video indexing is surveyed in Hu et al. (2011). Indexing can be done by analyzing the frames and/or audio of the recording to locate specific spots, for instance, where goals are made in a football match. Another option is to use the textual subtitles (dialogues, commentaries) of the video. In some cases, like historical film archives, manually curated textual descriptions or commentaries of the videos may be available for preserving cultural heritage – they can be used for annotations and indexing too, as in our case study. Various methods and tools are available for extracting linked data from texts (Martinez-Rodriguez et al., 2020). Providing contextual information and ads while watching videos has been suggested already in the 1980s in systems such as Hypersoap.[30] Examples of works on enriching video watching experience using linked data-based recommendations include (Nixon et al., 2013).

5.3 Discussion

We have learned that even in the rural northern parts of Europe, massive amounts of WW2 data can be found and opened for public use. In WarSampo some 100,000 people, mostly casualties of the WW2 events are considered. However, there are also data available about hundreds of thousands of soldiers who survived the war only in Finland. On a European level, these numbers are counted in tens of millions. Managing the data, and providing it for different user groups, suggests serious challenges when dealing with aspects such as the war events in the central parts of Europe, where the amount of data is in orders of magnitude larger than that in Finland, is multilingual, and distributed in different countries. For example, solving entity resolution problems regarding historical place names and person names can be difficult. However, it seems that using Linked Data is a promising way to tackle these challenges.

A major challenge in creating data analyses like the ones shown in this chapter is related to the quality of the data produced. Historical (meta)data is typically incomplete and our knowledge about it is uncertain. Also using more or less automatic means for transforming and linking the data leads to problems of incomplete, skewed, and erroneous data (Mäkelä et al., 2020). In general, more data literacy (Koltay, 2015) is usually needed from the end-user when using data analytic tools.

The methods of network analysis, for instance, can be very sensitive to even small errors in the data or biases in the sampling schemes.

This as well as conceptual difficulties in modeling complex real-world ontologies, such as historical geographical gazetteers, become sometimes embarrassingly visible when using and exposing the knowledge structures to end-users. The same problems exist in traditional systems but are often hidden in the non-structured presentations of the data and UIs.

Arguably, three successive generations of semantic portals for cultural heritage can be identified (Hyvönen, 2020): at the beginning of the millennium, the research focus in semantic portal development was on data harmonization, aggregation, search, and browsing ("first generation systems"). The rise of Digital Humanities research then started to shift the focus to providing the user with tools for solving data analytic research problems in interactive ways ("second-generation systems"). The portals presented in this chapter are examples of this generation. The next step ahead to "third-generation systems" is based on Artificial Intelligence (AI): future portals not only provide tools for the human to solve problems but are used for finding research problems in the first place, for addressing them, and even for solving them automatically under the constraints set by the human researcher. Such systems should preferably be able to explain their reasoning, which is an important aspect in the source-critical humanities research tradition. In the novel *Hitchhiker's Guide to the Galaxy* by Douglas Adams, the supercomputer came to the conclusion that the meaning of life, the universe, and everything is "42". In addition to knowing that it would be nice to know *why so*.

Acknowledgments

Lots of people at the Semantic Computing Research Group (SeCo)[31] have contributed to the work presented in this chapter, funded by several organizations in Finland. Thanks to Marcia Zeng for reviewing an earlier version of this chapter. Thanks also to the Finnish Cultural Foundation for an Eminentia Grant for reflecting our work. CSC – IT Center for Science has provided us with computational resources.

Notes

1 https://yasgui.triply.cc.
2 https://colab.research.google.com/notebooks/intro.ipynb.
3 https://jupyter.org.
4 https://www.cidoc-crm.org/.
5 SeCo homepage: https://seco.cs.aalto.fi.
6 In this paper the term *vocabulary* is used to refer to (hierarchical) knowledge organization systems, such as thesauri, authority files, and geographical gazetteers, whose entries are used to fill in metadata element (property) values.
7 FinnONTO project series: https://seco.cs.aalto.fi/projects/finnonto/.
8 FIN-CLARIAH initiative LOD work package: https://seco.cs.aalto.fi/projects/fin-clariah/.
9 CLARIN infrastructure: https://www.clarin.eu/.
10 DARIAH infrastructure: https://www.dariah.eu/.

11 https://www.w3.org/TR/shacl/.
12 https://shex.io/.
13 https://protege.stanford.edu/.
14 https://allegrograph.com/topbraid-composer/.
15 See, e.g., the browser for DBpedia: https://dbpedia.org/ontology/Browser/.
16 Content Negotiation by Profile: https://www.w3.org/TR/dx-prof-conneg/.
17 SPARQL 1.1 Query Language: https://www.w3.org/TR/sparql11-query/.
18 https://www.go-fair.org/fair-principles/.
19 http://www.europeana-collections-1914-1918.eu.
20 http://www.1914-1918-online.net.
21 http://ww1.discovery.ac.uk.
22 http://www.canadiana.ca/en/pcdhn-lod/.
23 http://www.cendari.eu/research/first-world-war-studies/.
24 http://blog.muninn-project.org.
25 http://www.bygle.net/wp-content/uploads/2015/04/Open-Memory-Project 3-1.pdf.
26 https://www.oorlogsbronnen.nl.
27 https://lod-cloud.net.
28 https://www.researchspace.org.
29 http://ckcc.huygens.knaw.nl.
30 www.media.mit.edu/hypersoap/.
31 https://seco.cs.aalto.fi/.

References

Ahola, A., Hyvönen, E., Rantala, H., & Kauppala, A. (2023). Publishing and studying historical opera and music theatre performances on the Semantic Web: Case OperaSampo 1830–1960. *Proceedings of SWODCH 2023. Semantic Web and Ontology Design for Cultural Heritage. Co-located with the 22nd International Semantic Web Conference (ISWC 2023) in Athens, Greece.* CEUR Workshop Proceedings, vol. 3540. Retrieved from https://ceur-ws.org/Vol-3540/paper8.pdf

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data – The story so far. *International Journal on Semantic Web and Information Systems (IJSWIS), 5*(3), 1–22. https://doi.org/10.4018/jswis.2009081901

Collins, T., Mulholland, P., & Zdrahal, Z. (2005). Semantic browsing of digital collections. In Y. Gil, E. Motta, V. Richard Benjamins, & M. A. Musen (Eds.), *The Semantic Web – ISWC 2005* (vol. 3729, pp. 127–141). Springer. https://doi.org/10.1007/11574620_12

de Boer, V., van Doornik, J., Buitinck, L., Marx, M., & Veken, T. (2013). Linking the kingdom: Enriched access to a historiographical text. In *Proceedings of the Seventh International Conference on Knowledge Capture (K-CAP 2013)* (pp. 17–24). Association for Computing Machinery, New York. https://doi.org/10.1145/2479832.2479849

Ferreira, A. C., dos Santos Maculan, B. C. M., & Naves, M. M. L. (2017). Ranganathan and the faceted classification theory. *TransInformacao, Campinas, 29*(3), 279–295. https://doi.org/10.1590/2318-08892017000300006

Frosterus, M., Tuominen, J., Pessala, S., & Hyvönen, E. (2015). Linked open ontology cloud: Managing a system of interlinked cross-domain light-weight ontologies. *International Journal of Metadata, Semantics and Ontologies, 10*(3), 189–201. http://doi.org/10.1504/IJMSO.2015.073879

Hearst, M. (2006). Design recommendations for hierarchical faceted search interfaces. *ACM SIGIR workshop on faceted search* (pp. 1–5). Retrieved from https://demo.kerko.whiskyechobravo.com/bibliography/F8FUWSY5

Heath, T., & Bizer, C. (2011). *Linked data: Evolving the web into a global data space* (1st ed.). Morgan & Claypool. https://doi.org/10.1007/978-3-031-79432-2

Heino, E., Tamper, M., Mäkelä, E., Leskinen, P., Ikkala, E., Tuominen, J., … & Hyvönen, E. (2017). Named entity linking in a complex domain: Case Second World War history. In J. Gracia, F. Bond, J. P. McCrae, P. Buitelaar, C. Chiarcos, & S. Hellmann (Eds.), *Language, data, and knowledge: First international conference, LDK 2017* (Vol. 10318, pp. 120–133). Springer. https://doi.org/10.1007/978-3-319-59888-8_10

Hitzler, P. (2021, January). A review of the semantic web field. *Communications of the ACM, 64*(2), 76–83. https://doi.org/10.1145/3397512

Hitzler, P., Krötzsch, M., & Rudolph, S. (2010). *Foundations of semantic web technologies*. Springer.

Hu, W., Xie, N., Li, L., Zeng, X., & Maybank, S. (2011). A survey on visual content-based video indexing and retrieval. *IEEE Transactions on Systems, Man, and Cybernetics, 41*(6), 797–819. https://doi.org/10.1109/TSMCC.2011.2109710

Hyvönen, E. (2010). Preventing interoperability problems instead of solving them. *Semantic Web – Interoperability, Usability, Applicability, 1*(1–2), 33–37. https://doi.org/10.3233/SW-2010-0014

Hyvönen, E. (2012). *Publishing and using cultural heritage linked data on the semantic web*. Morgan & Claypool. https://doi.org/10.1007/978-3-031-79438-4

Hyvönen, E. (2020). Using the semantic web in digital humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery. *Semantic Web – Interoperability, Usability, Applicability, 11*(1), 187–193. https://doi.org/10.3233/SW-190386

Hyvönen, E. (2023a). Digital humanities on the semantic web: Sampo model and portal series. *Semantic Web – Interoperability, Usability, Applicability, 14* (4), 729–744. https://doi.org/10.3233/SW-223034

Hyvönen, E. (2024). How to create a national cross-domain ontology and linked data infrastructure and use it on the semantic web. *Semantic Web – Interoperability, Usability, Applicability* (pp. 1–15). Pre-press. https://doi.org/10.3233/SW-243468

Hyvönen, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M., Tamper, M., … & Mäkelä, E. (2016). WarSampo data service and semantic portal for publishing linked open data about the Second World War history. In H. Sack, E. Blomqvist, M. d'Aquin, C. Ghidini, S. P. Ponzetto, & C. Lange (Eds.), *The Semantic Web. Latest Advances and New Domains: 13th International Conference, ESWC 2016* (pp. 758–773). Springer. https://doi.org/10.1007/978-3-319-34129-3_46

Hyvönen, E., Ikkala, E., Koho, M., Leal, R., Rantala, H., & Tamper, M. (2022). How to search and contextualize scenes inside videos for enriched watching experience: Case stories of the Second World War veterans. In *The Semantic Web: ESWC 2022 Satellite Events* (pp. 163–167). Springer. https://doi.org/10.1007/978-3-031-11609-4_31

Hyvönen, E., Leskinen, P., Tamper, M., Rantala, H., Ikkala, E., Tuominen, J., & Keravuori, K. (2019). BiographySampo – Publishing and enriching biographies on the Semantic Web for digital humanities research. In *The Semantic Web. 16th International Conference, ESWC 2019* (pp. 574–589). Springer. https://doi.org/10.1007/978-3-030-21348-0_37

Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M., & Kettula, S. (2005). MuseumFinland—Finnish museums on the semantic web. *Journal of Web Semantics, 3*(2), 224–241. https://doi.org/10.1016/j.websem.2005.05.008

Hyvönen, E., Saarela, S., & Viljanen, K. (2004). Application of ontology techniques to view-based semantic search and browsing. In *The Semantic Web: Research and Applications. Proceedings of the First European Semantic Web Symposium* (ESWS 2004). Springer. https://doi.org/10.1007/978-3-540-25956-5_7

Hyvönen, E., Viljanen, K., Tuominen, J., & Seppälä, K. (2008). Building a national semantic web ontology and ontology service infrastructure – The FinnONTO approach. In *The Semantic Web: Research and Applications. ESWC 2008* (pp. 95–109). Springer. doi: 10.1007/978-3-540-68234-9_10

Ikkala, E., Hyvönen, E., Rantala, H., & Koho, M. (2022, January). Sampo-UI: A full stack JavaScript framework for developing semantic portal user interfaces. *Semantic Web – Interoperability, Usability, Applicability, 13*(1), 69–84. https://doi.org/10.3233/SW-210428

Ikkala, E., Koho, M., Heino, E., Leskinen, P., Hyvönen, E., & Ahoranta, T. (2017). Prosopographical views to Finnish WW2 casualties through cemeteries and linked open data. In A. Adamou, E. Daga, & L. Isaksen (Eds.), *Proceedings of the Second Workshop on Humanities in the Semantic Web (WHiSe II)* (pp. 45–56). CEUR Workshop Proceedings, Vol. 2014. Retrieved from http://ceur-ws.org/Vol-2014/

Koho, M., Heino, E., & Hyvönen, E. (2016). SPARQL Faceter – client-side faceted search based on SPARQL. In *Joint Proceedings of the 4th International Workshop on Linked Media and the 3rd Developers Hackshop*. CEUR Workshop Proceedings, Vol. 1615. Retrieved from http://ceur-ws.org/Vol-1615/semdevPaper5.pdf

Koho, M., & Hyvönen, E. (2023). Studying occupations and social measures of perished soldiers in WarSampo linked open data. In *Biographical Data in a Digital World 2022 (BD 2022)*, Tokyo. Retrieved from https://seco.cs.aalto.fi/publications/2022/koho-et-al-warsampo-bd-2022.pdf (Forth-coming)

Koho, M., Hyvönen, E., Heino, E., Tuominen, J., Leskinen, P., & Mäkelä, E. (2017). Linked death — representing, publishing, and using Second World War death records as linked open data. In E. Blomqvist, K. Hose, H. Paulheim, A. Lawrynowicz, F. Ciravegna, & O. Hartig (Eds.), *The Semantic Web: ESWC 2017 Satellite Events* (vol. 10577, pp. 369–383). Springer. https://doi.org/10.1007/978-3-319-70407-4_45

Koho, M., Ikkala, E., Heino, E., & Hyvönen, E. (2018a). Maintaining a linked data cloud and data service for Second World War history. In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*. 7th International Conference, EuroMed 2018 (vol. 11196). Springer. https://doi.org/10.1007/978-3-030-01762-0_12

Koho, M., Ikkala, E., & Hyvönen, E. (2018b). How to maintain a linked data cloud in a deployed semantic portal. In *Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks*. CEUR Workshop Proceedings, Vol. 2180. Retrieved from http://ceur-ws.org/Vol-2180/

Koho, M., Ikkala, E., & Hyvönen, E. (2020, September). Reassembling the lives of Finnish prisoners of the Second World War on the Semantic Web. In *Proceedings of the Third Conference on Biographical Data in the Digital Age (BD 2019)*. CEUR Workshop Proceedings, Vol. 3152. Retrieved from https://ceur-ws.org/Vol-3152/BD2019_paper_5.pdf

Koho, M., Ikkala, E., Leskinen, P., Tamper, M., Tuominen, J., & Hyvönen, E. (2021). WarSampo knowledge graph: Finland in the Second World War as linked open data. *Semantic Web – Interoperability, Usability, Applicability, 12*(2), 265–278. https://doi.org/10.3233/SW-200392

Koho, M., Leal, R., Ikkala, E., Tamper, M., Rantala, H., & Hyvönen, E. (2022). Building lightweight ontologies for faceted search with named entity recognition: Case WarMemoirSampo. In S. Tiwari, N. Mihindukulasooriya, F. Osborne, D. Kontokostas, J. D'Souza, & M. Kejriwal (Eds.), *Proceedings of the 1st International Workshop on Knowledge Graph Generation from Text and the 1st International Workshop on Modular Knowledge Co-located with 19th Extended Semantic Conference (ESWC 2022)* (pp. 19–35). CEUR Workshop Proceedings, vol. 3184. Retrieved from http://ceur-ws.org/Vol-3184/TEXT2KG Paper 2.pdf

Koltay, T. (2015). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science, 49*(1), 3–14. https://doi.org/10.1177/0961000615616450

Leal, R., Rantala, H., Koho, M., Ikkala, E., Merenmies, M., & Hyvönen, E. (2022, March). WarMemoirSampo: A semantic portal for war veteran interview videos. In *DHNB 2022 the 6th Digital Humanities in Nordic and Baltic Countries Conference*. CEUR Workshop Proceedings, long papers, vol. 3232. Retrieved from http://ceur-ws.org/Vol-3232/paper30.pdf

Leskinen, P., Rantala, H., & Hyvönen, E. (2022). Analyzing the lives of Finnish academic people 1640–1899 in Nordic and Baltic countries: AcademySampo data service and portal. In *DHNB 2022 the 6th Digital Humanities in Nordic and Baltic Countries Conference*. CEUR Workshop Proceedings, long papers, vol. 3232. Retrieved from http://ceur-ws.org/Vol-3232/paper07.pdf

Linstone, H. A. (1989). Multiple perspectives: Concept, applications, and user guidelines. *Systems Practice, 2*(3), 307–331. https://doi.org/10.1007/BF01059977

Mäkelä, E., Hypén, K., & Hyvönen, E. (2013). Fiction literature as linked open data – The BookSampo dataset. *Semantic Web*, *4*(3). https://doi.org/10.3233/SW-120093

Mäkelä, E., & Hyvönen, E. (2014). SPARQL SAHA, a configurable linked data editor and browser as a service. In V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, & A. Tordai (Eds.), *The Semantic Web: ESWC 2014 satellite events. ESWC 2014* (pp. 434–438). Springer. https://doi.org/10.1007/978-3-319-11955-7_62

Mäkelä, E., Lagus, K., Lahti, L., Säily, T., Tolonen, M., Hämäläinen, M., … & Nevalainen, T. (2020). Wrangling with non-standard data. In *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference* (pp. 81–96). CEUR Workshop Proceedings, vol. 2612. Retrieved from http://ceur-ws.org/Vol-2612/paper6.pdf

Mäkelä, E., Törnroos, J., Lindquist, T., & Hyvönen, E. (2017). WW1LOD: An application of CIDOC-CRM to World War 1 Linked Data. *International Journal on Digital Libraries, 18*(4), 333–343.

Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM, 49*(4), 41–46. https://doi.org/10.1145/1121949.1121979

Martinez-Rodriguez, J. L., Hogan, A., & Lopez-Arevalo, I. (2020). Information extraction meets the semantic web: A survey. *Semantic Web – Interoperability, Usability, Applicability, 11*(2), 255–335. https://doi.org/10.3233/SW-180333

Moretti, F. (2013). *Distant reading*. Verso Books. https://doi.org/10.1093/llc/fqu010

Nixon, L., Bauer, M., & Bara, C. (2013). Connected media experiences: Web based interactive video using linked data. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 309–312). ACM, New York. https://doi.org/10.1145/2487788.2487931

Pollitt, A. S. (1998). *The key role of classification and indexing in view-based searching* (Tech. Rep.). University of Huddersfield. Retrieved from http://www.ifla.org/IV/ifla63/63polst.pdf

Rantala, H., Ahola, A., Ikkala, E., & Hyvönen, E. (2023). How to create easily a data analytic semantic portal on top of a SPARQL endpoint: Introducing the configurable Sampo-UI framework. In *Proceedings of VOILA 2023, Co-locatd with ISWC 2023, Athens*. CEUR Workshop Proceedings (vol. 3508). Retrieved from https://ceur-ws.org/Vol-3508/paper3.pdf

Rantala, H., Ikkala, E., Jokipii, I., Koho, M., Tuominen, J., & Hyvönen, E. (2020). WarVictimSampo 1914–1922: A semantic portal and linked data service for digital humanities research on war history. In *The Semantic Web: ESWC 2020 Satellite Events* (pp. 191–196). Springer. https://doi.org/10.1007/978-3-030-62327-2_33

Rantala, H., Jokipii, I., Ikkala, E., & Hyvönen, E. (2022). WarVictimSampo 1914–1922: A national war memorial on the semantic web for digital humanities research and

applications. *ACM Journal on Computing and Cultural Heritage, 15*(1). https://doi.org/10.1145/3477606

Ravenek, W., van den Heuvel, C., & Gerritsen, G. (2017). The ePistolarium: Origins and techniques. In A. van Hessen & J. Odijk (Eds.), *CLARIN in the low countries* (pp. 317–323). Ubiquity Press. https://doi.org/10.5334/bbi

Rietveld, L., & Hoekstra, R. (2017). The YASGUI family of SPARQL clients. *Semantic Web – Interoperability, Usability, Applicability, 8*(3), 373–383. https://doi.org/10.3233/SW-150197

Sacco, G. M. (2005). Dynamic taxonomies: Guided interactive diagnostic assistance. In N. Wickramasinghe (Ed.), *Encyclopedia of healthcare information systems* (pp. 631–635). Idea Group.

Suominen, O., Johansson, A., Ylikotila, H., Tuominen, J., & Hyvönen, E. (2012). Vocabulary services based on SPARQL endpoints: ONKI Light on SPARQL. In *Poster Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2012)*. Retrieved from https://seco.cs.aalto.fi/publications/2012/suominen-et-al-onkilight-2012.pdf

Suominen, O., Pessala, S., Tuominen, J., Lappalainen, M., Nykyri, S., Ylikotila, H., … & Hyvönen, E. (2014). Deploying national ontology services: From ONKI to Finto. In *Proceedings of the Industry Track at the International Semantic Web Conference 2014. CEUR Workshop Proceedings* (vol. 1383). Retrieved from http://www.ceur-ws.org/Vol-1383

Tunkelang, D. (2009). *Faceted search*. Morgan & Claypool. https://doi.org/10.1007/978-3-031-02262-3

Tuominen, J., Frosterus, M., Viljanen, K., & Hyvönen, E. (2009). ONKI SKOS server for publishing and utilizing SKOS vocabularies and ontologies as services. In *Proceedings of the 6th European Semantic Web Conference (ESWC 2009)* (pp. 768–780). Springer. https://doi.org/10.1007/978-3-642-02121-3_56

Tzitzikas, Y., Manolis, N., & Papadakos, P. (2017). Faceted exploration of RDF/S datasets: A survey. *Journal of Intelligent Information Systems, 48*(2), 329–364. https://doi.org/10.1007/s10844-016-0413-8

Verboven, K., Carlier, M., & Dumolyn, J. (2007). A short manual to the art of prosopography. In Keats-Rohan, K. S. B. (Ed.), *Prosopography approaches and applications. A handbook* (pp. 35–70). Unit for Prosopographical Research (Linacre College). https://doi.org/1854/8212

Viljanen, K., Tuominen, J., & Hyvönen, E. (2009). Ontology libraries for production use: The Finnish ontology library service ONKI. In *Proceedings of the 6th European Semantic Web Conference (ESWC 2009)* (pp. 781–795). Springer. https://doi.org/10.1007/978-3-642-02121-3_57

Wang, Q. (2023). *Forging a path to understanding: Integrating and analyzing war history data through knowledge graph technologies*. PhD Thesis manuscript submitted, The Australian National University, Canberra.

Zeng, M., & Qin, J. (2022). *Metadata* (3rd ed.). ALA Neal-Schuman.

4 Identifying main topics of digital humanities courses across countries

A topic modelling BERTopic technique

Ying-Hsang Liu and Anton Anikin

1 Introduction

The field of Digital Humanities (DH) has emerged from the earlier field of Humanities Computing and is concerned with the use of digital and computational methods in humanities research, as well as applying humanities theories in studying digital media and culture (e.g., Drucker, 2021; Golub & Liu, 2021). The knowledge structure within DH has been characterised as multidisciplinary, interdisciplinary and transdisciplinary, based on the three forms of supradisciplinary activities (Rosenfield, 1992, p. 1351). A recent critical analysis of the intellectual foundations of DH has demonstrated that DH and Humanities Computing are analytically distinguished since the name change has reflected the substantial changes in the field (Isemonger, 2018). This thread of research has been extensively studied, using bibliometric techniques (e.g., Tang et al., 2017; Su & Zhang, 2022; Wang, 2018; Yang et al., 2020), based on the corpora of research publications, with particular references to the peer-reviewed journal articles. These studies have revealed the broad and unbalanced coverage of DH disciplinary fields (Yang et al., 2020), interdisciplinarity of DH by diversity of origins and coherence (interconnected topics, concepts and tools) (Tang et al., 2017), close associations of DH with the field of history, literary and cultural heritage, and information and library science (Wang, 2018), and topics and themes of DH research in which collaboration, interdisciplinarity and support are included (Su & Zhang, 2022). However, these studies have not reflected the practice of DH research in learning and teaching activities, such as curriculum development and course offerings.

As revealed in a recent special issue on DH in information science, it is crucial to outline a research agenda that "identifies critical points of intersection and gaps in knowledge that require collaboration" (Zeng et al., 2022, p. 143). At the level of disciplines, there are strong ties between DH and the disciplines of history, literary and cultural heritage, as well as information and library science (Wang, 2018). A detailed analysis of the three established English-language DH journals, using latent Dirichlet allocation topic modelling, has demonstrated that DH is a discipline in itself and the disciplines of computational linguistics

DOI: 10.4324/9781032707211-5

and information science are connectors to neighbouring disciplines (Luhmann & Burghardt, 2022). A study of author keywords from a sample of 1,924 articles in DH has identified four major clusters of research topics, including (1) text and analysis techniques; (2) data and collection; (3) project, people and infrastructure; and (4) humanities theory and practice (Su & Zhang, 2022). In another study that aims to identify the research agenda in DH, using a text mining approach, it was found that popular topics in the recent decade include cultural heritage, geographic information, semantic web, linked data and digital media; research trends identified by using structural topic models included linked open data, text mining, semantic web and ontology, text digitisation and social network analysis (Joo et al., 2022). However, the level of interdisciplinary research in DH and international collaboration can be constrained by disciplinary and geographical boundaries (Lee & Wang, 2018; Tang et al., 2017). Nonetheless, these studies have demonstrated that due to advances in computational techniques for text analysis, the adoption of text mining approaches and visualisation techniques have made it possible to uncover the research topics in DH more accurately. The unit of analysis of disciplinary relationships has evolved into the level of research topics and themes.

Since DH research is generally considered interdisciplinary, digital research methods and tools in DH have been introduced to students to become aware of the disciplinary differences in research-based learning (Cosgrave, 2021). In an analysis of keywords in DH course syllabi, using co-occurrence of terms, it was found that the key themes from courses include data work ("Big Data," "data visualisation," "data projects"), text analysis and introductions to the digital humanities (Sula & Berger, 2023, p. 810). In another study that analyses DH course descriptions, using structural topic modelling, the core components of DH courses and their hierarchical relationships have been identified, which results in five clusters of topics: (1) management of language resources and interactive data visualisation; (2) statistical data analysis and text, geospatial data analysis and modelling; (3) software programming and development and web applications, architecture and metadata; (4) critical history of cultural heritage issues and public communication and production; and (5) socio-political theory and concepts and critical digital media and socio-cultural impact (Walsh et al., 2022). These studies have provided a more detailed picture about the DH research as reflected in the topics covered in the courses. However, a more in-depth understanding of the similarities and differences in DH courses across the level of education, programmes and country is needed for research-based learning and curriculum development globally.

To that end, this chapter has used the topic modelling technique of BERTopic (Grootendorst, 2022) to analyse a corpus of course descriptions. Our findings have revealed an emerging core of DH courses and the similarities and differences of the course offerings, varied by level of education, programmes and country. Visualisations of the results have been produced to assist our understanding of the identified topics and themes in course descriptions. Overall, our findings have painted a nuanced picture of the main DH course topics across countries, which may be useful for the global curriculum development of DH programmes.

2 Methods

To identify the main topics covered in DH courses, the state-of-the-art topic modelling technique of BERTopic (Grootendorst, 2022) was implemented, together with hierarchical clustering and visualisation, to analyse the course descriptions in programmes from any school, department or discipline. A snowball sampling method was adopted to include the DH programme registry and the known DH programmes. The course description dataset has been expanded from a previous study (Walsh et al., 2022) by including additional courses in the Digital Humanities Course Registry[1] in August 2023.

In this study, a total of 563 DH course descriptions (comprising a corpus of 8,214 terms and 87,666 tokens) were collected and analysed. The data was extracted from extensive searching and browsing of the DH programme websites, which includes the country, level of study (e.g., undergraduate and graduate), type of programme (e.g., Bachelor of Arts, Master of Science, certificate and major/minor), university, course title, field of study and course description.

To enhance the reproducibility of the study, the Google Collaboratory platform,[2] a hosted Jupyter Notebook service to combine Python executable code and annotations in a single document, was operated and visualisation results were produced. Specifically, after installing the required packages for running BERTopic and uploading the text file of course descriptions, the universal-sentence-encoder[3] that encodes the text into high-dimensional vectors at the sentence level for document embeddings was applied. The ClassTfidfTransformer was employed to reduce the impact of frequent words on the topic model. These steps constitute our global topic model in this study.

To perform additional analysis on the topics, the fitted model was then generated and visualised (see Figure 4.1). Since each topic was represented by a set of words, a bar chart visualisation of the terms of each topic ranked by term score was generated to assist with the interpretation of topics (see Figure 4.2). The hierarchical clustering technique was used to provide an overview of the relationships among the identified topics (see Figure 4.3). Additionally, we look into the categories across the course topics to investigate the similarities and differences by academic level, programme and country. The distribution of topic categories was then visualised across each category (see Figures 4.4–4.6). By doing so, we were able to gain a nuanced picture of the major course topics varied by academic level, programme and country.

3 Results

Taking a topic modelling approach to finding about the main topics in DH course descriptions, this section presents the results from our visualised outputs.

3.1 What are the main topics in DH course descriptions?

The results of our topic modelling suggest that the programming languages for data analysis was the most common topic in DH course descriptions (see Figure 4.1, topic 0). Topic 1 was concerned with the design of interactive web applications,

such as 3D visualisations, whereas Topic 2 was about the DH tools and methods. Topic 3 dealt with the digital media culture, whereas Topic 4 was concerned with the cultural heritage in museum and archaeology. Topic 5 was devoted to the placement, research projects and master-level dissertations in DH. Topic 6 represented the topic of metadata standards and preservation of archival materials.

3.2 What are the relationships among the DH course topics?

In addition to the individual topics identified in Figures 4.1 and 4.2, our results of hierarchical clustering of topics (see Figure 4.3) revealed that Topic 6 metadata standards of archival materials has a unique place in DH courses, since it formed a main cluster. Other topics belonged to another big cluster, in which Topics 2 and 3 formed a small cluster of DH tools and digital media culture and Topics 0 and 1 were another small cluster, dealing with the design of interactive web applications, and programming languages for data analysis. Topics 2, 0 and 1 joined together represented the technical side of DH courses. Topic 5 project-based learning courses joined with other clusters in this cluster.

Overall, the derived topics revealed the interdisciplinary nature of the DH courses, ranging from the technical abilities of programming languages for data analysis, the design of interactive web applications to the cultural heritage collections and digital media culture. From a course design perspective, it's not surprising that DH methods and project-based learning were well represented in the courses. Importantly, metadata standards are one of the core areas of interests in DH courses since they are the building blocks of DH infrastructures and applications (see Figure 4.3).

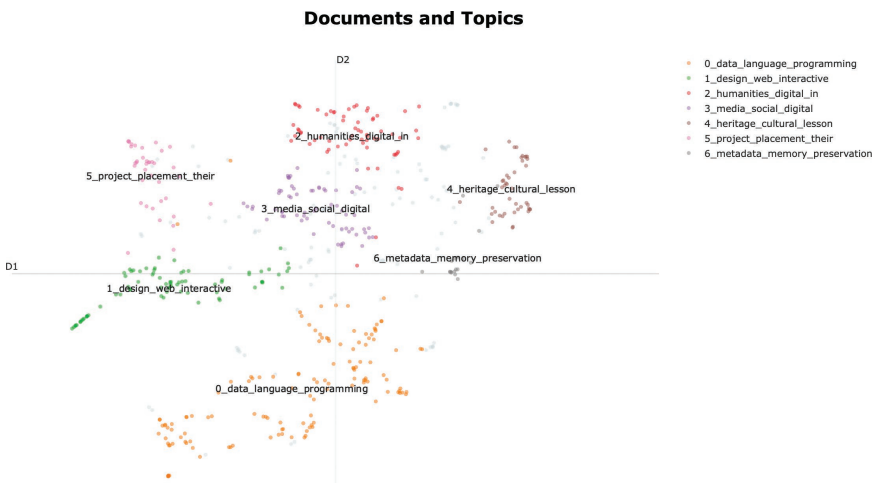

Figure 4.1 Topic models of DH course descriptions represented by a set of words using BERTopic.

Figure in colour available at https://github.com/ruyhliu/DHIC-24/blob/14a96c5fa15b1a3d6f796ae7adca20c7c5f839b0/Figure%204.1.png

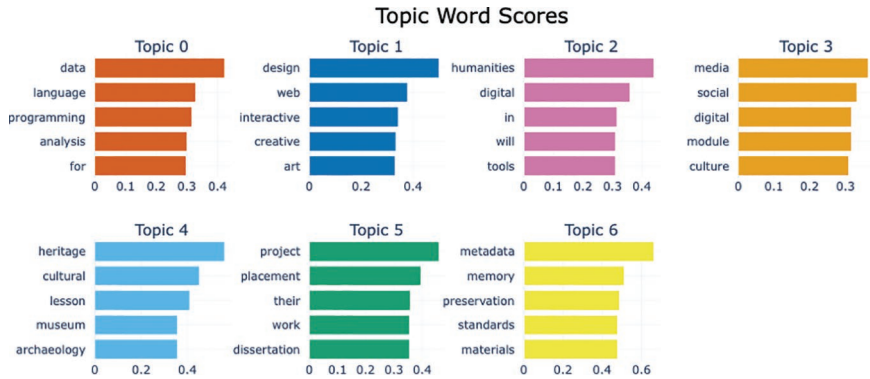

Figure 4.2 Bar chart visualisation of topics ranked by word scores.

Figure in colour available at https://github.com/ruyhliu/DHIC-24/blob/14a96c5fa15b1a3d6f796ae7adca20c7c5f839b0/Figure%204.2.png

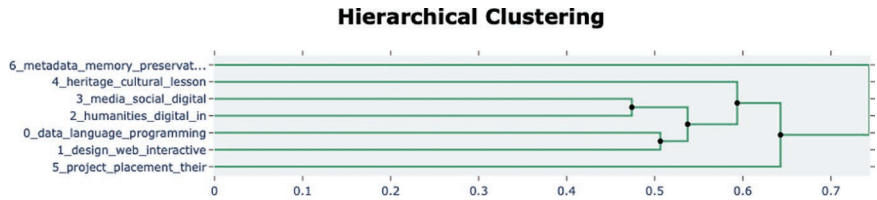

Figure 4.3 Hierarchical clustering of DH course descriptions topics.

Figure in colour available at https://github.com/ruyhliu/DHIC-24/blob/14a96c5fa15b1a3d6f796ae7adca20c7c5f839b0/Figure%204.3.png

3.3 What are the similarities and differences among the DH course topics by academic level, programme and country?

Our results revealed that the dominating DH course topic is programming languages for data analysis and digital media course at the postgraduate and undergraduate level, respectively (see Figure 4.4). At the postgraduate level, the top three DH course topics were (1) programming languages for data analysis; (2) design of interactive web applications; and (3) DH methods and tools. At the undergraduate level, the top three DH course topics were (1) programming languages for data analysis, (2) digital media culture and (3) design of interactive web applications. As such, the shared top DH course topics were programming languages for data analysis, as well as design of interactive web applications. In addition, project-based learning and cultural heritage topics were also well represented. As such, these observations suggested the core elements of DH courses are concerned with the DH infrastructure topics of programming languages for data analysis and the design of interactive web applications. The topic of DH methods and tools was emphasised at the postgraduate level, while digital media culture was more prominent at the undergraduate level.

Since the DH field is multidisciplinary, there were some variations in the name of programmes (see Figure 4.5). The terms digital humanities and digital culture(s) were the most frequently used, followed by the digital studies in specific domains,

such as arts and humanities and art history. Other programme names revealed the interdisciplinary nature of the DH programme, such as digital scholarship and digital heritage. Our results revealed that the topics of programming languages for data analysis, DH methods and tools and digital culture were most prominent in the DH programme. The course topic of cultural heritage and archaeology was strongly represented in the Digital Heritage programme, whereas the topic of digital media culture and social aspects was highlighted in the Digital Culture programme. Overall, these results suggested that there are emerging core courses in DH, with a few other variations in programme names and particular emphases on the course topics.

Regarding the course topics varied by country (see Figure 4.6), the results revealed some differences across countries. Specifically, the DH course topics were most represented by Australia, UK and the USA. In Australia, the top three course topics included (1) programming languages for data analysis, (2) design of interactive web applications and (3) cultural heritage and museums. In the UK, the top three course topics included (1) digital media culture, (2) programming languages for data analysis and (3) placement, research projects and master-level dissertations. In the USA, the top three course topics were (1) programming languages for data analysis, (2) DH methods and tools and (3) design of interactive web applications (tied with metadata and preservation standards). The course topic DH methods and tools was also well represented in Ireland and Sweden. The course

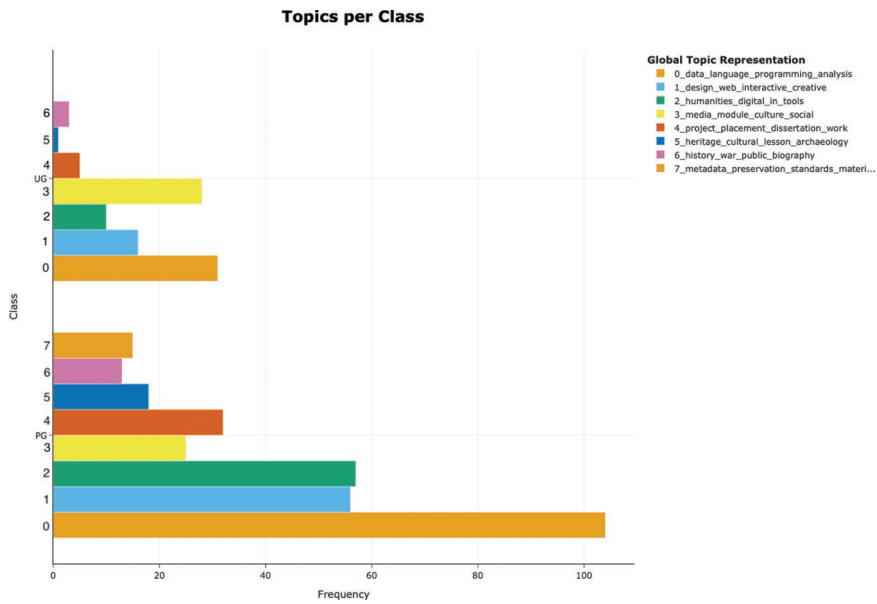

Figure 4.4 Distribution of DH course descriptions topics by academic level (UG = undergraduate; PG = postgraduate).

Figure in colour available at https://github.com/ruyhliu/DHIC-24/blob/14a96c5fa15b1a3d6f796ae7adca20c7c5f839b0/Figure%204.4.png

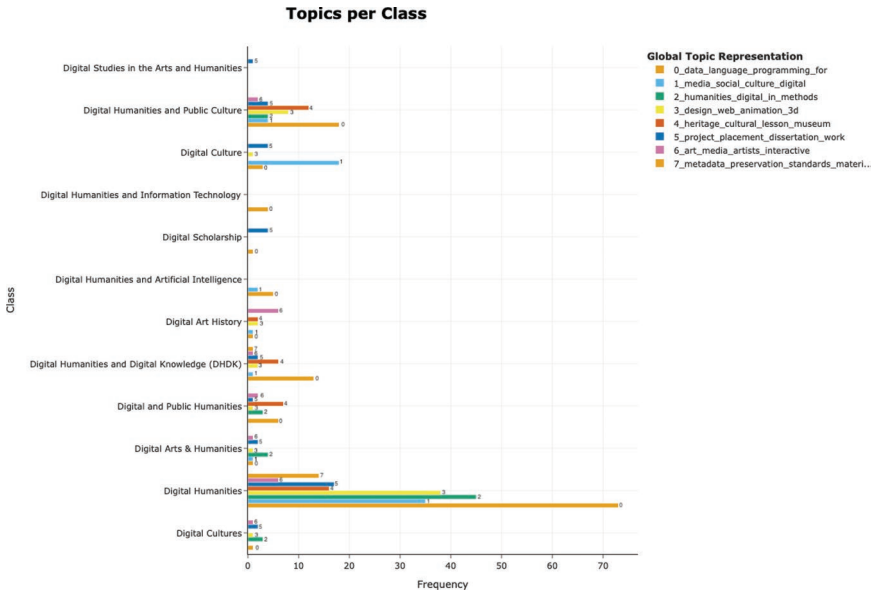

Figure 4.5 Distribution of DH course descriptions topics by programme.

Figure in colour available at https://github.com/ruyhliu/DHIC-24/blob/14a96c5fa15b1a3d6f796ae7adca20c7c5f839b0/Figure%204.5.png

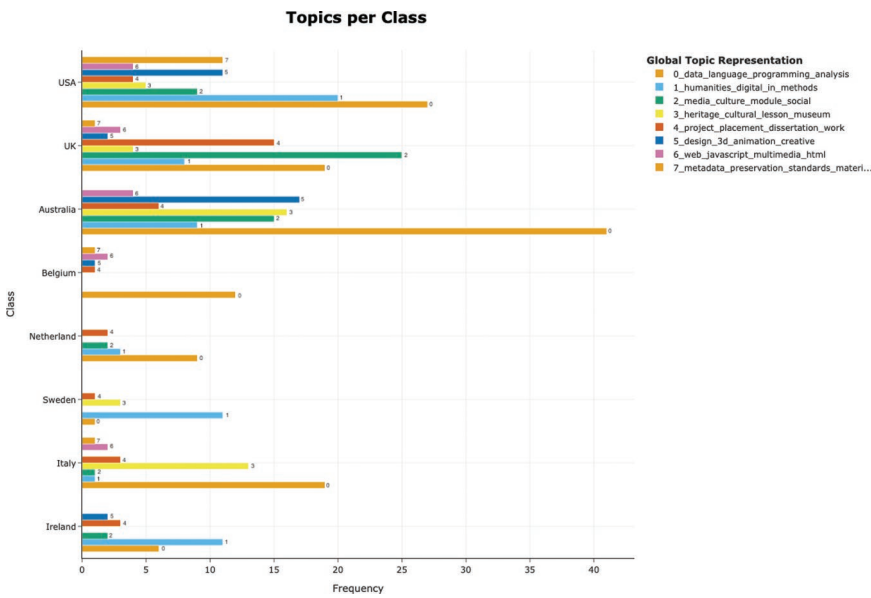

Figure 4.6 Distribution of DH course descriptions topics by country.

Figure in colour available at https://github.com/ruyhliu/DHIC-24/blob/14a96c5fa15b1a3d6f796ae7adca20c7c5f839b0/Figure%204.6.png

topic programming languages for data analysis was also highlighted in Belgium and Italy. The course topic of culture heritage and archaeology was prominent in Italy and Sweden. Overall, the course topic of programming languages for data analysis was probably the core across the countries, with different emphases on the topics of design of interactive web applications, DH methods and tools, design of interactive web applications and project-based learning.

4 Summary and conclusion

This chapter has used the topic modelling technique of BERTopic (based on transformers approach and c-TF-IDF) to analyse a corpus of 563 course descriptions in DH programmes globally. The corpus was a purposive and convenience sample of the DH programmes, biased towards the English-speaking countries, and those countries with online presence and English version of the course descriptions. Nonetheless, it provided a snapshot of the development of DH programmes across countries. This chapter has demonstrated that the usefulness of the state-of-the-art topic modelling technique of BERTopic to uncover the latent relationships among the identified main topics covered in the global DH programmes.

The research findings have revealed an emerging core of DH courses covering the topics of programming languages for data analysis, design of interactive web applications, cultural heritage resources and collections, DH methods, digital media culture, placement, research projects and master-level dissertations in DH and metadata standards of archival materials. These findings are in line with the previous studies that investigate the intellectual foundations and knowledge structures of DH, in which there are strong associations between DH and the disciplines of history, literary and cultural heritage, as well as information and library science (Wang, 2018). Additionally, digital tools and DH methods have been emphasised in the courses, which reflects the interdisciplinary nature of DH programmes (Cosgrave, 2021; Sula & Berger, 2023; Walsh et al., 2022) and the geographical differences in the offering of academic degrees and credentials (Cobb & Golub, 2022). The disciplinary differences in DH course offerings are reflected in the programme titles where Digital Humanities is the most used. Overall, the core course topic of programming languages for data analysis seems to be consistent across countries, although there are varying emphases on specific areas such as the design of interactive web applications, DH methods and tools and project-based learning.

With the recent rapid development of AI (Artificial Intelligence) technologies, this chapter has shown the potential of using the topic modelling technique (part of text mining) for extracting semantic information from a large corpus, which is consistent with a recent study that employed the structural topic models in the DH research domain (Joo et al., 2022). This approach has the advantage of producing more interpretable topics than previous topic modelling techniques, partly due to the new representation model, such as weighting with c-TF-IDF that has been adopted in this study. As noted in Grootendorst (2022), BERTopic can be fine-tuned by utilising text generation Large Language Models (LLMs), such as ChatGPT and GPT-4. However, this approach is still limited by the availability of

digitised text (which should be representative of the focus of a study) and requires human interpretation for the generated topics. Importantly, to enhance the transparency and fairness of the systems, the language biases in the internal knowledge of LLMs need to be further considered (Salinas et al., 2023) since they can affect downstream applications, such as the topic modelling technique of BERTopic. AI technologies like LLMs have opened up new opportunities for computational DH research, such as sentiment analysis for understanding narrative (Chun & Elkins, 2023) and facilitating new perspectives with data stories in literary studies (Mischke & Ohge, 2023). While the course topic of programming languages for data analysis has been dominant in DH courses across countries, it is expected that DH methods and tools will become more prominent shortly.

Notes

1 https://dhcr.clarin-dariah.eu.
2 https://colab.google/.
3 https://tfhub.dev/google/universal-sentence-encoder/4.

References

Chun, J., & Elkins, K. (2023). eXplainable AI with GPT4 for story analysis and generation: A novel framework for diachronic sentiment analysis. *International Journal of Digital Humanities, 5*(2), 507–532. https://doi.org/10.1007/s42803-023-00069-8

Cobb, P. J., & Golub, K. (2022). Digital humanities degrees and supplemental credentials in Information Schools (iSchools). *Education for Information, 38*(1), 67–92. https://doi.org/10.3233/EFI-200452

Cosgrave, M. (2021). Digital humanities methods as a gateway to inter and transdisciplinarity. *Global Intellectual History, 6*(1), 24–33. https://doi.org/10.1080/23801883.2019.1657639

Drucker, J. (2021). *The Digital Humanities coursebook: An introduction to digital methods for research and scholarship*. Routledge.

Golub, K., & Liu, Y.-H. (2021). Information and knowledge organisation in digital humanities: Global perspectives. https://doi.org/10.4324/9781003131816

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. ArXiv Preprint ArXiv:2203.05794.

Isemonger, I. (2018). Digital humanities and transdisciplinary practice: Towards a rigorous conversation. *Transdisciplinary Journal of Engineering & Science, 9*, 116–138. https://doi.org/10.22545/2018/00105

Joo, S., Hootman, J., & Katsurai, M. (2022). Exploring the digital humanities research agenda: A text mining approach. *Journal of Documentation, 78*(4), 853–870. https://doi.org/10.1108/JD-03-2021-0066

Lee, H.-L., & Wang, S. (2018). Investigating digital humanities: A domain analysis of conference proceedings published in Taiwan, 2009–2016. *Journal of Library and Information Studies, 2*, 1–23. https://doi.org/10.6182/jlis.201812_16(2).001

Luhmann, J., & Burghardt, M. (2022). Digital humanities—A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape. *Journal of the Association for Information Science and Technology, 73*(2), 148–171. https://doi.org/10.1002/asi.24533

Mischke, D., & Ohge, C. (2023). Digital Melville and computational methods in literary studies. *Leviathan,* 25(2), 35–60. https://doi.org/10.1353/lvn.2023.a904374

Rosenfield, P. L. (1992). The potential of transdisciplinary research for sustaining and extending linkages between the health and social sciences. *Social Science and Medicine, 35*(11), 1343–1357. https://doi.org/10.1016/0277-9536(92)90038-R

Salinas, A., Penafiel, L., McCormack, R., & Morstatter, F. (2023). "Im not racist but...": Discovering bias in the internal knowledge of Large Language Models. https://arxiv.org/abs/2310.08780

Su, F., & Zhang, Y. (2022). Research output, intellectual structures and contributors of digital humanities research: A longitudinal analysis 2005–2020. *Journal of Documentation, 78*(3), 673–695. https://doi.org/10.1108/JD-11-2020-0199

Sula, C. A., & Berger, C. (2023). Training information professionals in the digital humanities: An analysis of DH courses in LIS Education. *College and Research Libraries, 84*(5), 802–823. https://doi.org/10.5860/crl.84.5.802

Tang, M. C., Cheng, Y. J., & Chen, K. H. (2017). A longitudinal study of intellectual cohesion in digital humanities using bibliometric analyses. *Scientometrics, 113*(2), 985–1008. https://doi.org/10.1007/s11192-017-2496-6

Walsh, J. A., Cobb, P. J., Fremery, W., Golub, K., Keah, H., Kim, J., Kiplang'at, J., Liu, Y.-H., Mahony, S., Oh, S. G., Sula, C. A., Underwood, T., & Wang, X. (2022). Digital humanities in the iSchool. *Journal of the Association for Information Science and Technology, 73*(2), 188–203. https://doi.org/10.1002/asi.24535

Wang, Q. (2018). Distribution features and intellectual structures of digital humanities: A bibliometric analysis. *Journal of Documentation, 74*(1), 223–246. https://doi.org/10.1108/JD-05-2017-0076

Yang, M., Wang, M., Wang, H., Yang, G., & Liu, H. (2020). Exploring the transdisciplinary nature of digital humanities. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries* (pp. 553–554). https://doi.org/10.1145/3383583.3398552

Zeng, M. L., Sula, C. A., Gracy, K. F., Hyvönen, E., & Alves Lima, V. M. (2022). JASIST special issue on digital humanities (DH). Journal of the Association for Information Science and Technology, 73(2), 143–147. https://doi.org/10.1002/asi.24584

5 Empowering global engagement

The development of digital humanities research and pedagogy at UCL[1]

Jin Gao, Adam Crymble, Simon Mahony, Steven Gray and Claire Warwick

1 Introduction

The discipline now known as Digital Humanities (DH) has origins in several different fields. The most well-known story is its origin in linguistics and English studies, where the use of computational techniques and methodologies in the humanities is usually linked to the work of Roberto Busa from the 1940s, and the work of Josephine Miles, although far less known (Miles, 1946; Busa, 1950). Many in the field look to Busa as its progenitor, through his collaboration with IBM and his creation of an *index variorum* of the combined works of Thomas Aquinas. Yet, alternative foundation narratives also exist, both in linguistics and in other disciplines such as history, which place emphasis on other key moments that better reflect issues like disciplinary differences and the key projects that inspired much of the work we now celebrate (Rockwell, 2007; Nyhan & Flinn, 2016a; Crymble, 2021). Indeed, some argue that DH as a field only emerged at the beginning of the 2000s (Berry & Fagerjord, 2017, p. 10), while others claim that it started in the 19th century (Brandeis Library, 2013; Hayes, 2017). Whichever view we take, the history of DH is gaining increasing traction (Nyhan, 2023), and DH itself has grown out of earlier movements within scholarship, which can be traced in the Anglophone sphere through the changing nomenclature: "applied computing in the humanities", to "humanities computing", and then to "digital humanities" (Mahony & Gao, 2019). Nyhan and Flinn argue that this history 'once neglected, is now emerging and is absolutely necessary' (Nyhan & Flinn, 2016a, p. 14). This chapter seeks to contribute to the broad conversation about the history of DH that has opened up in the last decade (Crymble, 2021; Kemman, 2021).

Within that conversation, various treatments of DH at the national level have emerged, e.g., the Netherlands (Zaagsma, 2013), Germany (Thaller, 2017), and France (Le Deuff, 2018). Why zoom in further, to write institutional histories of DH, such as that set out here? Institutional histories can play an important role in acknowledging and addressing the "role of difference" within the knowledge production enterprise (Jones, 2013, p. 31). Recognising the significance of diversity in terms of gender, race, geo-linguistic background, and intersectionality, DH scholars and organisations are increasingly focalising the need for inclusive representation and community diversity (Bailey, 2011; Gil & Ortega, 2016; Liu, 2018; Mahony,

DOI: 10.4324/9781032707211-6

2018; Risam, 2015). In recent years, there has been a growing recognition of the need to move beyond traditional Anglophone and male-dominated perspectives, to embrace a more inclusive and representative DH landscape (ADHO, 2022; Estill et al., 2022; Gao et al., 2022; Nyhan, 2023). These efforts have sought to dismantle historical hierarchies, disrupt exclusionary practices, and amplify marginalised voices within the DH community.

Thus, understanding the historical trajectory of DH at institutions like University College London (UCL) through retrospective studies, such as the one presented in this chapter, is a necessary part of the wider whole. Though presenting a necessarily partial view, the institutional view is one that can be profitably co-joined, or indeed contrasted, with other available historical lenses, such as the history of digital technology in DH. By delving into institutional pasts, we can gain not only a better understanding of "who we are" and "where we have come from" as a community, and at a high level of local resolution, but also gain valuable insights that may be applicable to the challenges and opportunities that lie ahead for DH at UCL and comparable institutions. Ideally, then, the lessons learnt from institutional histories might enable us to build upon existing strengths, confront past inequalities, and effectively address the evolving needs of the DH field, and the multiplicity of actors and agents that constitute it, at individual, institutional, national, trans-national, and global levels. The limitations of institutional histories written, as this one is, by those who participated in the events at hand should be acknowledged too, and this chapter does lay an emphasis on successful events, and milestones, in a way that a more critical outsider, less hindered by institutional politics and ongoing professional relationships may not (Chowdhury, 2013). We acknowledge that as a world-leading and well-funded university, the discussion presented in this chapter is of a variant of DH that could only be pursued in highly privileged circumstances.

In the following sections, we present discussions that explore the development of DH at UCL (Section 2), the co-authorship networks at UCL (Section 3), DH pedagogy (Section 4), the widening of engagement (Section 5), and conclude with the key insights gained from these investigations.

2 The DH development at UCL

2.1 The early works (1970–1994)

University College London (UCL), founded in 1826, holds a significant place in the history of higher education for its progressive stance on gender equality. In 1878, UCL became the first university in the UK to admit women on equal terms with men, demonstrating its pioneering role in promoting gender inclusivity in academia. This early commitment to coeducation contributed to UCL's reputation as a leader in advancing educational opportunities for women and fostering a more inclusive and equitable learning environment. Over the years, UCL has emerged as an important institution in the global field of DH with innovative research, teaching, and interdisciplinary collaborations. At UCL, the most well-known DH

milestone was probably the establishment of UCL Centre for Digital Humanities (UCLDH), formally launched in 2010, but active from 2009 onwards. At the same time, other research centres and departments at UCL were also actively engaged with DH work, e.g., the UCL Centre for Advanced Spatial Analysis (CASA); the UCL Interaction Centre (UCLIC); the UCL International Centre for Chinese Heritage and Archaeology (ICCHA); the UCL Centre for Medical Image Computing (CMIC), and more. Some examples will follow later in this chapter. UCLDH did not seek to 'own' or control DH activities at UCL but rather to provide a central hub to facilitate exchange and coordination, cooperation, and collaboration (Warwick et al., 2011). For a significant time before the foundation of these centres, UCL scholars had been making substantial contributions in the relevant fields and participating at related events.

UCL's role in the development and application of computational techniques and methodologies to the Humanities can be traced to the 1970s at least. This early phase was driven by pioneering scholars in computational linguistics (e.g., Arthur Colin Day, David W. Packard), Geographic Information System (GIS) (e.g., Roger Tomlinson), who built a global network of collaborators (e.g., Sidney Greenbaum, Andrew Rosta).

According to the Index of DH Conferences (Weingart et al., 2023),[2] the first UCL scholar to present at a major DH conference was Arthur Colin Day in March 1970. He presented at the ALLC/EADH (Association for Literary and Linguistic Computing) *Symposium on the Uses of Computers in Literary Research*, hosted by Cambridge University (23–26 March 1970). There, Day presented the paper *FORTRAN as a language for linguists* with the affiliation of 'Computer Centre – University College London' (Farringdon, 1970; Nyhan & Flinn, 2016b, p. 237). This symposium is one of two early conferences held by the ALLC before its formal inauguration in 1973 and its first so-titled conference in 1974. This association later changed its name to the European Association for Digital Humanities (EADH) and continues to be a co-host of a major, annual international DH conference (EADH, 2023a).

Day was awarded his PhD in General Linguistics in 1966 from the School of Oriental and African Studies (SOAS), University of London, and he worked at the UCL Computer Centre from 1967 to 1992, starting as a programmer and later promoted to the Head of Applications (Day, 2004). According to his biography, after giving a talk to the School of Library, Archive and Information Studies (now, the UCL Department of Information Studies, UCLDIS) on how computers could be used for non-numerical work, the then Head of Department suggested that he write a book about it, which resulted in the *Fortran Techniques, with Special Reference to Non-numerical Applications* (Day, 1972b).

This book quickly became 'the one to have' for studying and applying the Fortran programming language and it sold tens of thousands of copies (Barron, 1973). Day started to run the Fortran course for "all-comers" at UCL, and later this popular course was videotaped and published by Athlone Press, with the title *A London Fortran Course* (Day, 1972a). Day also actively contributed to the review of the COCOA (COunt and COncordance Generation on Atlas) project (Russell,

1965), which was a concordance programme developed by UCL and the Atlas Computer Laboratory. Following this, he published the article *Software Reviews: COCOA: A Word Count and Concordance Generator* in the journal *Computers and the Humanities* (*CHum*) with Ian Marriott, who was doing his PhD in Latin at UCL, and they became the first UCL scholars to publish in what many consider to have been one of the canonical DH journals – *Computers and the Humanities* (Day & Marriott, 1976). COCOA played a significant role in the emergence of digital humanities (then known as applied computing in the humanities), offering word-counting and concordance-building functionalities. While successor formats like TEI XML gained wider adoption, COCOA's legacy and influence remain noteworthy (Hockey, 2004).

While discussing COCOA and the Atlas Computer Laboratory, one key person who was related to these early works and who also presented at the same 1970 Cambridge symposium was Susan Hockey (Hardesty & Mann, 1973). Although a much more well-known scholar to the UCL DH community, she was not affiliated with UCL until 2000 (EADH, 2023b). At the 1970 symposium, together with Robert F. Churchhouse from Cardiff University, Hockey presented their research *The use of an SC4020 for output of a concordance program* under her pre-marriage name of 'Miss S. M. Petty' with her affiliation as 'the Atlas Computer Laboratory' (Farringdon, 1970, p. 315; Nyhan & Flinn, 2016c, p. 87).

Hockey played a crucial role as a founding member of the ALLC in 1973 and she joined the Oxford University Computing Services in 1975. Her first formal association with UCL is recorded in the *Index* as her presentation at the ACH/ALLC 1999 conference in Virginia, when she was listed as one of the ten authors of *Can a Team Tag Consistently? Experiences on the Orlando Project* (Butler et al., 1999). In 2000, Hockey joined UCL as Professor of Library and Information Studies, and later became the director of the School of Library, Archive, and Information Studies (now UCLDIS) from 2001. Upon her retirement in 2004, she was awarded the Roberto Busa Prize by ADHO (the Alliance of Digital Humanities Organizations), recognising her as a leader in the field of humanities computing (EADH, 2023b). Later, to celebrate the fifth anniversary of UCLDH in 2015, Hockey gave the inaugural lecture in what became the annual Susan Hockey Lecture at UCLDIS, commemorating her contributions to the field.

Roger Tomlinson, widely recognised as the "Father of GIS" completing his PhD at the UCL Department of Geography in 1974 (Tomlinson, 1974). He went on to pursue innovative work in digital mapping, a legacy that endures today in the fundamental principles underlying modern computer-generated cartography. It was during his time at UCL that Tomlinson laid the groundwork for the *Canada Geographic Information System*, influential in the field of geographical information science to this day. Tomlinson's dissertation, titled *Geographical Information Systems, Spatial Data Analysis, and Decision Making in Government*, remains relevant with his forward-looking perspective on the flow and types of spatial data elements, their governance and impact on the real world (Cheshire, 2017). Tomlinson's groundbreaking maps, developed during his time at UCL, were among the earliest to demonstrate the value of integrating spatial data for insightful analysis,

which exemplified their potential for identifying land areas with favourable prospects for specific land uses, based on a set of prioritised criteria.

Apart from the key projects and efforts mentioned, growing scholarly collaborations and exchanges across different countries and regions can also be detected during this early time. After the foundation of ALLC in 1973, the first two official ALLC conferences were held at Cardiff University in April 1974 (Jones & Churchhouse, 1976) and at King's College London in December 1974 (Butler, 1974). Looking through the two proceedings, along with Day and Hockey, many DH pioneers also attended the conference, e.g., Roberto Busa, Andrew Morton, Roy Wisbey, and Joseph Raben. One attendee who recorded a UCL affiliation was David W. Packard, a former professor of Greek and Latin at the University of California, Los Angeles (UCLA) (Kenneth, 2016). He gained his PhD in classical philology from Harvard University and taught Classics at UCLA and other universities in the USA. There are no recorded ties for Packard with UCL apart from these proceedings. Through personal correspondence with the authors, Packard confirmed that he was a visiting professor, and active member of the research community at UCL in 1974–1975. For example, he discussed his own system for the automatic morphological analysis of ancient Greek titled *Metrical and Grammatical Patterns in the Greek Hexamater* at the 1974 ALLC conference in Cardiff, and he introduced his scheme for using computers in the teaching of ancient Greek at the 1974 conference at King's College London (Butler, 1974; Jones & Churchhouse, 1976). Packard is currently the president of the Packard Humanities Institute, and he is also the son of Hewlett-Packard (HP)'s co-founder David Packard.

This time also saw UCL collaborations at a wider international scale. The International Corpus of English (ICE) founded by Sidney Greenbaum at UCL in 1988 is a large-scale corpus linguistics project that aims to collect, analyse, and compare written and spoken English from different countries and regions around the world (The ICE, 2021). The project's primary goal is to provide a comprehensive and representative sample of global English usage, which considers the linguistic variations and differences that exist among different English-speaking communities, including the United States, Canada, Australia, New Zealand, South Africa, and several other regions where English is a dominant or significant language, such as India and Nigeria. This project was established on the success of the Survey of English Usage (SEU) project, founded by Randolph Quirk in 1959 at UCL, which collected and analysed written and spoken English from Great Britain from 1955 to 1985, creating a substantial and influential corpus of British English, which served as a model and inspiration for the later ICE project.

Many well-known scholars worldwide took part in the development of SEU and ICE projects, which compiled a million-word corpus of their respective English variety, and it had a common design and annotation scheme for comparability (UCL, 2022). The two projects have very close connections, which demonstrated not only a clear academic genealogy with scholars at UCL, but also an expanding global network with significant international impact. The ICE project lead, Greenbaum, completed his PhD at UCL within the SEU project under Quirk's supervision in 1967. Greenbaum succeeded Quirk as the Quain Professor of English Language

and Literature, as well as director of the SEU at UCL in 1981. From 1986 to 1988, Greenbaum served as the Dean of the Faculty of Arts and Humanities at UCL, and supervised many PhD students including Andrew Rosta, who presented the ICE project at the 1992 ALLC conference. Rosta continued Greenbaum's work and joined the SEU project in 1987. Although ICE is not affiliated with UCLDH, through Rosta's email correspondence with us, he communicated his belief that his time at UCL saw the beginnings of the field now known as DH, and he also witnessed the forming of the ICE project's global collaborative network spanning Singapore, South Africa, New Zealand, India, Nigeria, Pakistan, Jamaica, Kenya, and beyond. When working at the ICE project, Rosta took the responsibility for developing the markup scheme which led to him joining the Text Encoding Initiative (TEI) Council as part of a four-person working group on the encoding of spoken texts, along with Jane Edwards (University of California at Berkeley), Stig Johansson (University of Oslo), and Lou Burnard (Oxford University).

2.2 The structural development (from 1995)

From this initial take-up of computational techniques and methodologies by individual scholars and research projects, by the 1990s DH appears to have gained structural foothold at UCL, as evidenced by the establishment of many DH-related research centres and projects across different departments and faculties. Additionally, the wider context of the evolving digital and communication landscape driven by the Internet and the general adoption of computational applications also contributed to this expansion, which, in turn, helped foster greater collaboration and international engagement between UCL and the wider world.

UCL's expertise in the spatial realm was fostered by the Centre for Advanced Spatial Analysis (CASA) founded in 1995 within the Bartlett Faculty of the Built Environment at UCL. Led by Michael Batty as the first centre director, CASA aimed to pioneer the advancement of comprehensive urban science by leveraging techniques in modelling, urban environment sensing, visualisation, and computation (Shiode et al., 1998). It was the first among UK universities to employ virtual reality (VR) systems for studying multi-user GIS for London and creating a virtual world in which users as avatars could manipulate urban designs (Batty et al., 1998). Over the years, CASA and the Bartlett Faculty have made their contributions to the field through pioneering projects and collaborations around the world, e.g., Africa Centre/I-Sense project (Manley et al., 2016), GNOMEs project (Milton et al., 2018), Harmony project (Batty & Evans, 2022), City of Women project (Watson et al., 2022). CASA also has close research overlap and connections with DH. Apart from similar research topics and methods, such as feminist studies (Sheppard et al., 2023) and relevant history studies (Cheshire, 2017), scholars in CASA and the Bartlett have collaborated on projects with UCLDH, such as QRator (Hudson-Smith et al., 2012), Textal (Terras et al., 2013), NFTs (Non-Fungible Tokens) related research (Valeonti et al., 2021), and high-performance computing on digitised collections (Terras et al., 2018), and have also jointly supervised PhD research.

The field of Human-Computer Interaction (HCI) is also relevant to research in DH, and the foundation of the UCL Interaction Centre (UCLIC) in 2001 represented another DH milestone at UCL. The centre sits between the Department of Computer Science and the Psychology and Language Sciences Division at UCL. While its roots lie in cognitive psychology, ergonomics, and computer science, under the directorships of Harold Thimbleby, Ann Blandford (from 2004), and Yvonne Rogers (from 2011), its research later embraced the design and social dimensions of HCI, and physical computing (Blandford, 2011; Harrison & Rogers, 2013). Researchers from UCLIC collaborated with UCL Information Studies, before UCLDH was established, on topics including digital libraries (Buchanan et al., 2006) and information-seeking behaviour (Makri & Warwick, 2010). UCLIC is now a globally recognised hub for HCI, and its collaborations extend widely through projects such as the EnTimeMent project, which involves a number of institutions and industries across Europe (Ceccaldi et al., 2020), and the Global Disability Innovation (GDI) Hub, which operates in 61 countries and is the world's first and only WHO Collaborating Centre for Assistive Technology (Holloway & Barbareschi, 2021).

Medical Physics and Computer Vision may seem distant from core DH interests but specialist imaging, such as at the UCL Computer Science Department, the Centre for Doctoral Training in Computer Vision, the Centre of Science and Engineering in Arts, Heritage and Archaeology (SEAHA), and the UCL Centre for Medical Image Computing (CMIC), play important parts in DH research practices at UCL. Collaborations for both research and teaching have resulted in projects such as The Great Parchment Book of the Honourable the Irish Society led by Tim Weyrich and Melissa Terras (Pal et al., 2016), general heritage and manuscript imaging (Giacometti et al., 2014), and a collaborative doctoral project on multi-spectral images of parchment (Giacometti, 2014). Imaging techniques and expertise at UCL have been central to the development and success of the UCL Multi-Modal Digitisation Suite which is coordinated by UCLDH and located in UCL Library Services, Special Collections archival space.

UCL's connection with the development of Chinese DH has been long established through the UCL Institute of Archaeology (IoA) and the International Centre for Chinese Heritage and Archaeology (ICCHA) (Fuller & Pang, 2015). ICCHA was established in 2003 as a collaborative research centre between UCL IoA and the School for Archaeology and Museology of Peking University. The centre is dedicated to the study and preservation of China's rich cultural heritage, with a particular focus on integrating DH techniques and building collaborations beyond the UK and China. For example, the Central Asian Archaeological Landscapes (CAAL) project (Nebbia et al., 2021), launched by UCL IoA and led by Tim Williams, has been digitising and digitally preserving the intricate archaeological legacy spanning Central Asia from the Caspian Sea to Western China. Collaborating with institutions from Kazakhstan, Kyrgyz Republic, Tajikistan, Turkmenistan, Uzbekistan, and China's Xinjiang Uyghur Autonomous Region, as well as involving the International Institute for Central Asian Studies and the ICOMOS International Conservation Centre Xi'an, China, CAAL unites a multinational and multidisciplinary team. This team is entrusted with digitising existing archival materials and

integrating information from regional entities into a unified multilingual repository. The project's holistic approach blends advanced imaging techniques with on-site exploration to uncover new sites, enrich documentation, encourage scholarly engagement, and facilitate informed conservation policies. The CAAL project also employs the open-source ARCHES inventory package, co-developed by the Getty Conservation Institute and the World Monuments Fund, to comprehensively catalogue and manage the diverse archaeological heritage dispersed throughout Central Asia (CAAL, 2019).

Other ICCHA digital humanities projects include the Early Rice project, which pursues linked research on the origins and early development of rice cultivation, as well as the spread of agriculture across China; the Terracotta Army project, which collaborates with the Museum of Emperor Qin Shihuang's Mausoleum and investigates craft specialisation, interactions, and social cohesion in the emerging imperial systems. These project not only facilitate research but also contribute to the interdisciplinary approach, combining archaeology, history, and digital methodologies to understand the cultural exchange in China and other relevant countries. Through its engagement in these projects, IoA and ICCHA have significantly contributed to China's digital cultural heritage as key players in the global efforts to study and protect cultural treasures.

2.3 The UCL Centre for Digital Humanities (UCLDH) (from 2010)

The UCL Centre for Digital Humanities (UCLDH) was established in 2010 by a group of scholars including Claire Warwick and Melissa Terras, aiming to form a vibrant interdisciplinary research hub dedicated to exploring the intersection of digital technology and the humanities (Warwick et al., 2011). The UCL Grand Challenges were launched in 2008 (UCL, 2008a), with the aim of fostering interdisciplinary research, an approach to which DH's collaborative working methods were ideally suited. After a successful application to the Provost's Strategic Development fund, in 2009, UCLDH was granted funding for two years, which allowed for a high-profile launch with much media coverage and, more importantly, additional new staff to be hired for teaching, administration, and project work. With support from the Faculties of Engineering, most particularly Computer Science, and the Arts and Humanities, the centre grew and since 2021 is now a part of the UCL Institute of Advanced Studies.

Before the founding of UCLDH, Warwick and Terras had already worked at UCL for several years, having been appointed as lecturers in Electronic Communication and Publishing in 2002 and in 2003, respectively. They taught on, and successively directed, the MA in Electronic Communication and Publishing, which had been established in the late 1990s. This included modules on topics of immediate relevance to DH, such as XML, Internet Technologies, and Digital Resources in the Humanities. This work, combined with their joint efforts in securing grant funding, from Jisc, RIN, EPSRC, and AHRC, demonstrated their commitment to building a strong foundation in digital humanities at UCL. Their work on projects like LAIRAH (UCL, 2005) and VERA (UCL, 2008b) was evidence of the growth

of the field and made the establishment of UCLDH a logical progression. However, it is important to note that their journey was not without challenges: as Terras emphasised with us through personal correspondence, their successes were achieved in spite of adversity. As Kirschenbaum found, in the 2010s, the development of DH was not always welcomed by more traditional humanities scholars (Kirschenbaum, 2014), and UCL was not immune to such attitudes.

In addition to the publication of conference papers and journal articles, which we discuss below, members of UCLDH also collaborated on the publication of books which were significant in demonstrating the breadth of the UCL impact on the digital humanities landscape. *Digital Humanities in Practice*, edited by Warwick, Terras, and Nyhan, with chapters and case studies authored by numerous UCLDH members, presented the works of the Centre, both in terms of research and teaching. *Defining Digital Humanities: A Reader*, edited by Terras, Nyhan, and Vanhoutte (an affiliate member of UCLDH) has become an essential reference for anyone wishing to understand the theoretical foundations of the field.

As above, UCLDH did not plan to take ownership of the many DH activities across the college but acted as a central hub to bring people together to share experiences and expertise. Part of this sharing was to run a series of public events and seminars with speakers from across UCL and beyond; this included the Susan Hockey Lecture in Digital Humanities from 2015 to celebrate the fifth anniversary and continued success of the centre. This was the first, and as far as the authors are aware the only, named lecture series in DH. This outreach also placed UCL firmly within the wider DH context as London had by then become a significant point of DH focus, with our close neighbours at the Department of Digital Humanities, King's College London, and the School of Advanced Study, University of London, to further develop collaboration and the community aspect of our work. In addition to departments and faculty across UCL, from its foundation onwards, the centre has collaborated and worked closely with UCL Museums and Collections (now UCL Culture), UCL Library, and particularly UCL Library Services including Special Collections, one of the foremost university collections of rare books and manuscripts in the UK. Building on these connections, Terras, together with support from Computer Science, developed the multi-modal digitisation suite which is used for teaching and significant research into non-destructive and heritage imaging of bespoke and high value research objects.

UCLDH has members across all parts of UCL, but, nevertheless, there is no building, no dedicated office space, no sign over a door saying, "Digital Humanities Centre". UCLDH is a virtual centre; it is made up of people and hence "People" is the uppermost link on the home webpage menu. The structure has changed over the years. On launch, there was Claire Warwick as Director and Melissa Terras as Deputy Director (both at UCLDIS), Tim Weyrich (Department of Computer Science) and Ulrich Tiedau (Department of Dutch) as Associate Directors, and a small immediate team who invited colleagues and DH practitioners to come together as part of the centre's wider team and affiliates. With the subsequent growth and increase in activities, a more robust management structure developed leading to the one in place today. There is now a management team with director, deputy

and associate directors as dictated by the centres' constitution. The website lists the extensive UCLDH Team with members across the wide range of UCL Faculties, Honorary Members listing the former directors, Affiliated PhD students, and other UCLDH Affiliates. In addition to this, there is the UCLDH Industry Advisory Panel with the centre's major contacts and partners in industry, such as the British Library, the V&A, IBM, Adobe, and Microsoft. One group not publicly displayed is the UCLDH Steering Committee which consists of the deans and representatives from the faculties and departments that work closely with UCLDH and who act in an advisory compacity for strategic and tactical planning. It is this committee that formally appoints the directors and ratifies any changes to the constitution.

UCLDH is unique within the usual university structures. Initially it was situated within UCLDIS and thus the Faculty of Humanities. However, as the centre continued to develop, it became clear that, to better foster interdisciplinary collaboration, a new organisational model was required. In 2013, therefore, it was agreed that the centre should be situated between faculties, so that no one faculty could lay claim to represent DH within the institution. The inaugural and three subsequent directors of UCLDH, Claire Warwick, Melissa Terras, Simon Mahony, and Julianne Nyhan, have all been staff members at DIS, while the current director, Steven Gray, is based at CASA, and Tim Weyrich, one of the first two associate directors is based in Computer Science, which shows this cross-faculty-led structure of the centre. However, the centre is closely linked with the Department of Information Studies (DIS) through research and teaching. The teaching programme was always planned as part of UCLDH but due to various institutional structures and rationale, the programme is owned and offered by DIS. The department offers both Master of Arts and Master of Science degrees in DH as well as a full range of other options such as Short Courses, Diplomas and Certificates, Master of Research, and full MPhil/PhD study. The master's programme has a strong emphasis on cultural heritage and the application of innovative digital methods to the study of the humanities more widely. The students develop strong technical as well as theoretical and critical skills. Whenever appropriate we involve the students in the centre and, where possible, in our research too; this is covered more broadly below.

Overall, UCLDH has played an important role in advancing the DH field, and fostered many projects that span diverse areas within humanities research, digital methodologies, and cultural heritage preservation. Many innovative projects are listed on the centre's website notable for its distinctive graphic design by Rudolf Ammann, produced when he was a PhD student at DIS. The research page features award-winning projects such as the Great Parchment Book, QRator, and Transcribe Bentham. It also highlights notable collaborations between the UCL Digitisation Suite and UCL Library Services and Special Collections.

Faculty affiliated with UCLDH have published books and articles that are highly cited by the wider field, e.g., (Warwick et al., 2012; Terras et al., 2013; Nyhan & Flinn, 2016a). In the third section of this chapter, therefore, we present a co-authorship network analysis, to allow us to further investigate UCL's collaborative engagement in the field of digital humanities. This complements our background study in the earlier section and enhances our understanding of this

historical context. By looking through the collaborations at UCL from a quantitative perspective, we can identify prominent individuals, influential collaborations, and notable patterns of UCL scholarly exchange.

3 The co-authorship networks at UCL

Various methods have been employed to study the development of DH and its communities, such as comprehensive literature reviews (McCarty, 2003), interviews and oral histories (Nyhan & Flinn, 2016a; Nyhan & Passarotti, 2019), statistical and infographic analyses (Terras, 2012; Nyhan & Duke-Williams, 2014), bibliometric analyses (Wang & Inaba, 2009; Weingart & Eichmann-Kalwara, 2017), and social network analyses (Grandjean, 2016; Gao et al., 2017). Among these methodologies, co-authorship network visualisation has emerged as a valuable tool for helping comprehend the collaborative dynamics and scholarly influences within the field (Gao et al., 2022).

This section presents a quantitative exploration of the collaboration networks formed by scholars affiliated with UCL who participated in DH and proto-DH conferences from 1970 to 2023. Through the use of co-authorship network analysis, we offer a distinct lens for examining the intellectual and collaborative history of digital humanities at UCL.

Data has been collected from the Index of DH Conferences for the years 1970–2023; note that this index is a growing record constructed entirely by volunteers and so cannot be considered complete (Weingart et al., 2023). The ADHO conference seems to be fully ingested, but many others are not (for example, ALLC, EADH, DHC are included, and at the time of writing in 2023, up until 2018). Consequently, this does not represent the entirely of UCL's (or any other institutional) representation at DH conferences during this period. Nor does it include DH contributions to discipline-specific conferences not led by DH organisations. We also acknowledge the lack of inclusiveness and diversity of topics and scholars among DH conference proceedings. Nevertheless, this dataset provides a useful snapshot of participation in conferences between 1970 and 2023, and so is used as the basis for the network analysis in this study.

In general, it seems that DH at UCL has taken the lead in significant areas – one of which is in the number of co-authored conference papers, and another is the number of female scholars when compared to other UK universities with prominent DH contingents during the same period, such as King's College London (KCL) and the University of Cambridge, UK.

In total, UCL scholars presented 97 papers at DH conferences held between 1970 and 2023 and included in this dataset: 77 were co-authored papers and 20 were single-authored. The multi-authored papers account for 79% of the total from UCL; in comparison, during the same period, 65% of papers by KCL scholars were multi-authored and 47% of University of Cambridge scholars. Figure 5.1 shows the number of single-authored and multi-authored papers presented by UCL scholars each year at DH conferences based on the data held in the Index of DH Conferences (Weingart et al., 2023).

Figure 5.1 The number of single-authored and multi-authored papers presented by UCL scholars each year at DH conferences, data from (Weingart et al., 2023) 1970 to 2023.

There are many aspects to unpack and compare, but our research here mainly focuses on investigating the collaborative network of UCL scholars. Additionally, when counting scholars' affiliations, we are aware that there are many cases where scholars move and change institutions, but we count only the papers where author(s) declare an affiliation with UCL.

We can see from Figure 5.1 that there is overall growth in the total numbers of papers and multi-authored papers at UCL, peaking in 2013 (a clear anomaly that may have resulted from the relaxation of restrictions on the travel budgets that followed the banking crisis). From 2014 to 2019, the number of papers given by UCL scholars remained similar to 2012, while the number of multi-authored papers increased. After 2019, the impact of the pandemic on individual staff members with caring responsibilities, the scholars moving to other universities, or indeed the changing conference locations and their move online in response to the global pandemic might have influenced the numbers. Staff mobility is likewise relevant, with individuals moving from one institution to another or having papers accepted to conferences not included in the *Index* (iSchools, for example), which echoes the lack of inclusiveness in topics and scholars from the ADHO conference proceedings acknowledged earlier. The sharp decline in 2020 reflects the impact of the global pandemic, not only restricting travel but also significantly increasing staff workloads as all teaching needed to be moved online and staff with caring responsibilities were often directly impacted.

In total, 246 authors contributed to papers that listed UCL affiliations; 77 of them were UCL scholars, with 169 affiliated elsewhere. On average, one UCL author has 4.4 co-authors in the dataset, which is relatively high compared to other institutions during the same period. For example, based on the Index of DH Conferences (Weingart et al., 2023), authors affiliated with KCL have on average 3.9 co-authors, and Cambridge authors have 3.0 co-authors. Table 5.1 shows the top ten UCL authors ranked by number of papers for DH conferences in our dataset.

Table 5.1 The top ten UCL scholars by the number of papers for DH conferences, data from (Weingart et al., 2023) 1970 to 2023

	Scholars	*Papers*	*Average year*	*Co-authors*	*Presumed gender*
1	Melissa Terras	38	2012	99	F
2	Claire Warwick	24	2009	62	F
3	Julianne Nyhan	13	2017	28	F
4	Simon Mahony	8	2016	13	M
5	Ann Blandford	5	2008	26	F
6	Oliver Duke-Williams	5	2015	13	M
7	Alejandro Giacometti	4	2013	33	M
8	Anne Welsh	4	2013	14	F
9	Claire Ross	4	2011	9	F
10	George Buchanan	3	2006	4	M

As shown in Table 5.1, we can see the total number of papers each author with UCL as an affiliation presented at the DH conferences in our dataset, their average year of attendance,[3] the number of unique co-authors they have worked with, and their presumed gender. Given that authors of different ages, and with varying lengths of institutional affiliation with UCL are included in Table 5.1, the "average year of attendance" provides a useful temporal context for this data. We have identified the author gender based on a well-tested name-gender assignment method (Sugimoto et al., 2015) and our knowledge of these authors. If we have mis-gendered any individuals through this method of classification, we apologise. It is acknowledged that some people are gender diverse, but the sources for that information are very limited, so this study follows the previous binary gender category convention (Rørstad & Aksnes, 2015). Among the most prolific ten presenters who were affiliated with UCL, six of them are female scholars, including the three most productive, with 88 papers in total – Terras, Warwick, and Nyhan. In comparison, four male scholars contributed 20 papers in total. Compared to other DH centres in the UK, UCL had a relatively higher female acceptance rate at conferences, led by three prominent female scholars who were active during the time covered in this dataset, although they have since this moved to other institutions. For example, within our current dataset, female scholars account for 48% of those recorded with a UCL affiliation, whereas previous scholarship shows that, based on both publications in major DH journals from 1966 to 2017 and conference papers DH2004 to DH2016, female scholars accounted for 30% of the total authors (Weingart & Eichmann-Kalwara, 2017; Gao et al., 2022).

In the visualisation of the co-authorship network that follows, scholars are the nodes, while the affiliative co-authorship association serves as the defining edge. Thus, the edges within a designated co-authorship network are established through the collaborative efforts of two scholars who cooperated on a scholarly paper. Specifically, this study counts the total number of papers to weigh the author node and the number of co-authored papers for the edge. The network's size increases as more scholars collaborate, so it results in a more clear structure that offers an

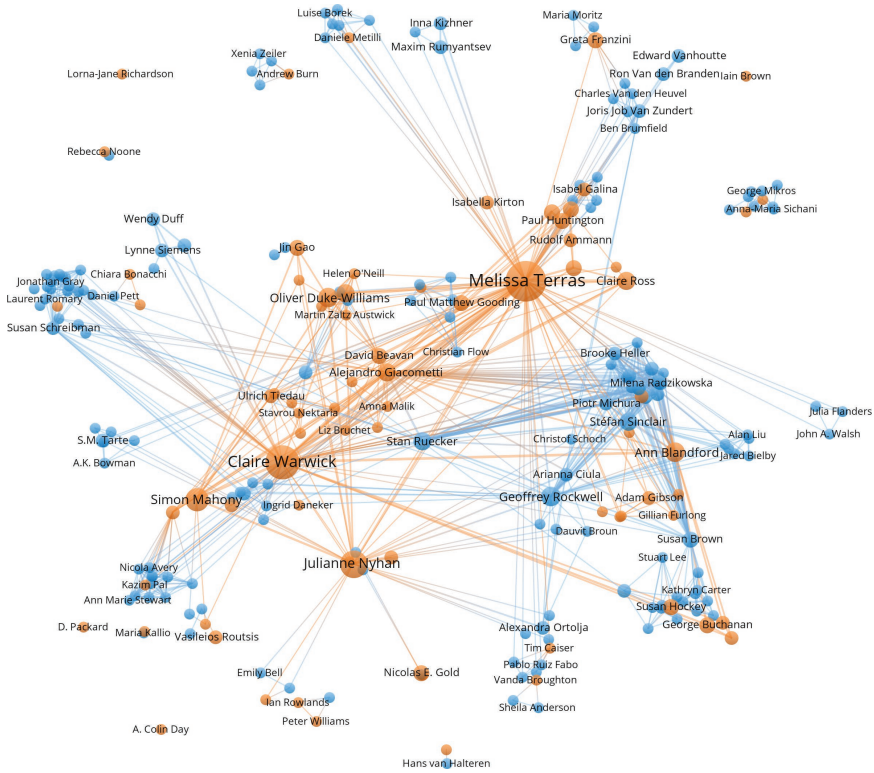

Figure 5.2 Co-authorship network of UCL authors and their collaborators at DH conferences, data from (Weingart et al., 2023) 1970 to 2023. [For printed version, light grey – UCL scholars and dark grey – non-UCL scholars; for online coloured version, orange – UCL scholars and blue – non-UCL scholars.]

output-centric view of characteristics such as connectivity levels and the prominence of scholars. VOSviewer 1.6.7 and Gephi 0.9.2 have been used for network construction and centrality measures, and the default disperse method was applied in VOSviewer. Figure 5.2 Co-authorship network of UCL authors and their collaborators at DH conferences, data from (Weingart et al., 2023) 1970 to 2023 [for printed version, light grey – UCL scholars and dark grey – non-UCL scholars; for online coloured version, orange – UCL scholars and blue – non-UCL scholars] and Figure 5.3 present the visualised co-authorship networks with UCL scholars highlighted in orange, and average publication year colour-coded in a heatmap timeline.

As illustrated in Figure 5.2, Co-authorship network of UCL authors and their collaborators at DH conferences, data from (Weingart et al., 2023) 1970 to 2023 [for printed version, light grey – UCL scholars and dark grey – non-UCL scholars; for online coloured version, orange – UCL scholars and blue – non-UCL scholars] and Figure 5.3, specific nodes (scholars) within the network act as crucial

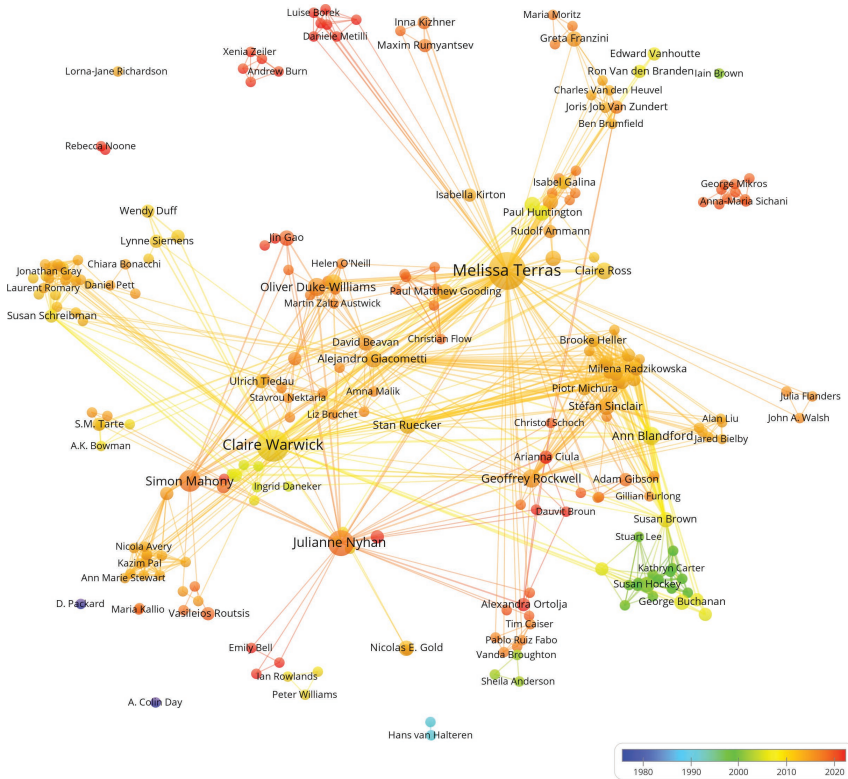

Figure 5.3 The co-authorship network of UCL authors and their collaborators at DH conference, heatmap timeline version, data from (Weingart et al., 2023) 1970 to 2023.

connectors, evident through higher betweenness centrality values computed using Gephi's measurements. Notably, a significant portion of these influential connectors are female scholars, indicating their substantial contributions to the establishment and development of collaborative DH networks at UCL. As we know from previous studies (Gao et al., 2022), female scholars in DH are often the main forces maintaining scholarly connections, and they are frequently the icebreakers that bridge isolated groups. Whether or not women are better communicators, or have been socialised in this way (Eagly, 2013), female scholars in digital humanities at UCL have different patterns of collaboration which helped UCL build more collaborative networks and achievements. These findings prompt a consideration of wider academic publication patterns that indicate men often have a tendency to work alone or cite and collaborate more frequently with male authors (Bozeman & Gaughan, 2011). This phenomenon invites reflection on the implications of such gender-related biases in the context of DH collaborations. However, it is important to recognise that UCL is unusual in its approach to gender inclusivity, perhaps owing to its long history of female participation and early coeducation. It also boasts a larger proportion of female faculty than many other universities in the UK

(Times, 2022). These factors suggest that UCL's DH network represents a distinctive and forward-thinking model for promoting diverse and collaborative scholarly connections, which challenges assumptions about gender imbalances often observed in academic collaboration patterns.

One particular factor to notice is that the interdisciplinarity of the DH scholars, especially the female ones, helps to bring new ways of working to the field. It is necessary to emphasise that these scholars, such as Terras, Warwick, and Hockey, have championed inclusion of everyone involved in the research, especially research students, as authors of conference papers and publications stressed the importance of collaboration and collaborative culture in reaction to traditional humanities. This is consistent with the collegial, inclusive, approach, which Hockey and other digital humanities pioneers always strove to promote, from the infancy of the field.[4]

The analysis of the dataset reveals new patterns in the gender distribution of presenters affiliated with UCL at DH conferences. The considerable representation of female scholars, particularly within the top ten most prolific presenters, suggests that women at UCL made significant progress towards greater gender diversity, a pattern that challenges the historical gender disparities documented in previous studies (Weingart & Eichmann-Kalwara, 2017). UCL's female faculty contributed to the progress made in fostering inclusivity and gender balance within academic conferences, and these findings not only provide insights into the changing landscape of DH research but also highlight the role of prominent female scholars in driving this transformation.

Additionally, there is substantial evidence within the dataset indicating that both early-career scholars and PhD students frequently presented their work with the assistance and supervision of their mentors. This practice further highlights UCL's proactive approach in providing financial support at departmental and faculty level for the participation of PhD students at conferences. From a quantitative perspective, we have provided a more visualised image of the UCL scholarly co-authorship and collaborations in digital humanities, and below we continue our discussion about teaching DH at UCL, with in-depth examples that reflects on the collaborative patterns demonstrated here.

4 DH pedagogy at UCL

Fostering collaboration not only within research networks but also within the classroom has been a feature of digital humanities at UCL, and it has come in two phases. The first was from the launch of the MA/MSc Digital Humanities in 2011 (Terras, 2010). The innovative dual designation of the programme as MA and MSc was intended to allow students to follow paths through the programme with different degrees of technical, computer scientific, content. It was also designed to attract students from a range of backgrounds, including STEM disciplines, although in practice most students had a first degree in the humanities.

This programme was originally developed with desktop and laptop computing in mind, and in an era when getting material online was somewhat less straightforward than it is in 2024. The first iPhone smartphone by Apple launched in the UK

in November 2007 (Apple, 2007) and took time to have its full impact on the way of life of students and DH practitioners (Crymble, 2021, pp. 72–78), nevertheless, mobile computing was introduced to the DH module Internet Technologies from 2012. However, in general, students tended to focus on textual content in the classroom, as was typical of much DH work circa 2010, while digital images and videos remained difficult formats to work with at scale until later in the decade.

Some of the modules included learning the practicalities of digitising images and objects to get them into the computer, for example, Digital Resources in the Humanities at UCLDIS, started by Hockey and developed further by Terras in 2003. Later, case studies that discussed the teaching for this module revealed the integration of object-based learning within the DH programme (Nyhan et al., 2014; Kador et al., 2018, p. 164), exemplified by the Galton collection,[5] which highlighted the student connections with seemingly disparate subjects within the collections and their challenges to problematise their knowledge and apply it to novel situations. Discussions on creating a digital collection from this material not only indicate technological aspects but also emphasise the necessity of sensitive and ethical approaches in making digital collections universally available. These case studies raise essential questions about the role of DH in disrupting techno-triumphalism and preparing students for multifaceted careers in diverse industries. For this reason, teaching both on the MA/MSc in DH and the MA in Electronic Communication and Publishing which preceded it always included discussion of the social and ethical context of digital technologies.

The module Introduction to Digitisation has partnered with the Asia Department at the Victoria and Albert Museum (V&A), which demonstrates the transformative potential of digital tools in engaging with museum objects in storage (Gao & Hongxing, 2023). Through the Chinese Export Watercolours (CEW) project, our students contributed to digitising paintings made in the 18th and 19th centuries, fostering practical skills, and contributing to best practices in museum digitisation. The public nature of this learning, where students 'learn in public' expands beyond traditional research activities to encompass a broader, digitally iterative, and collaborative approach (Kador et al., 2018, p. 167). This project also reflects on how we foster collaborative pedagogy techniques that leverage historical materials, highlight the convergence of collections and digital spaces in a truly interdisciplinary manner.

Another module is XML, a textual markup format popular in the scholarly editing community to format and otherwise annotate interesting details in a text (all references to people, or places, or different parts of speech). This XML was taught alongside modules that introduced the basics of web publishing, with HTML, CSS, JavaScript, and PHP used to present that material on the live web, providing students with a broad technical aptitude that was not likely embedded in their first degree. The workflow was often about getting cultural relics (objects, texts) into the body of the computer, where they could be shared via the web. Students also learned about digital humanities tools and methods, both aimed at textual analysis (Voyant, Antconc, etc.) and at geospatial work (GIS), which could lead to research findings in traditional humanities disciplines, and which could act as a springboard into PhD study.

The focus on technology and skills development was an important part of the programme as feedback from both students and employers made it clear that skills made their CVs stand out against those of other humanities students and gave our graduates an advantage in the job market. They could speak the language of both the technologists and the non-technical staff (Mahony & Pierazzo, 2012). The specific skills they learnt were of course transient, as technology inevitably evolves; however, the ability to learn and teach oneself technology is a lifelong skill, as is the programme's focus on developing cognitive ability and new ways of thinking both with and against the machine. Students would often come with entrenched and unquestioned assumptions about the physical and digital being in an oppositional and hierarchical relationship; these assumptions needed to be challenged and disrupted. Both have a part to play in our understanding of artefacts and collections and the students are challenged to reflect on the ways in which knowledge is constructed, curated, and communicated (Mahony et al., 2016).

UCLDH has close relationships with UCL Library Special Collections as well as UCL Museums and Collections (now known as UCL Culture) which enabled both hands-on and object-based learning sessions as part of our modules. Indeed, an introduction and practical experience with specialists in UCL Special Collections and Museums was a feature of students' induction sessions. These, together with our work placement module, gave many students valuable experience within the museum and library sectors.

The second teaching phase of DH at UCL emerged later in the 2010s as the smartphone era progressed and new technologies such as more evolved social media ecosystems, virtual reality (VR) and augmented reality (AR) emerged and captured the public imagination; incoming students began to shift their interests and expectations about what they would hope to learn in a DH postgraduate degree. The programme always attracted an international cohort of students with the majority being non-native speakers of English. However, the 2016 Brexit referendum result had a significant impact on European student recruitment to British programmes, with European students facing a tripling of their tuition fees without the benefit of EU policy that allowed them access to 'home' tuition rates (HESA, 2023). This, coupled with a general feeling that they were perhaps not wanted in Britain, led to a dramatic decline of more than 50% in European student numbers, not only at UCL but across the sector, which is only starting to reverse nearly a decade later. In their place came a growing number of Chinese applicants, with significant increases year-on-year from 2013 onwards, who became the largest single group of students in short order.

Their welcome arrival necessarily changed the pedagogical needs of the classroom because it could no longer be taken for granted that most students in the class had a British or European education or preferred learning styles. The segregated nature of the Chinese and Western Internets meant that students also arrived having few digital or popular culture references in common. China also had invested heavily in some digital cultural technologies that the West had not, particularly in AR and VR. This is best exemplified by the International Dunhuang Project hosted at the British Library which launched in 1994 (British Library, 2022), and also the

Digital Dunhuang project in China, launched in 2016 (2017 in English), which had provided 3D digitised representations of the Dunhuang Caves, a UNESCO World Cultural Heritage site, and provided a means for visitors to explore the caves virtually on their phones (Dunhuang Academy, 2016). The Palace Museum in Beijing has also invested heavily in a beautiful app that provides an engaging walkthrough of the museum with an accompanying storyline, that could give a visitor the feeling of having a private tour in a virtual space (Zhang, 2019). Chinese applicants to the DH programme at UCL frequently referred to both the Dunhuang and Palace Museum apps in their personal statements. It was incredibly rare for Western-based students to mention AR or VR as an area of interest, in part because Western museums had not yet made much engaging use of these technologies, focusing instead on more nuts-and-bolts priorities such as collection management and building maintenance.

What these conditions created was a cohort of students with very diverse, but culturally specific interests. This diversity is of course very welcome for any educator. Where it posed a challenge was in the students' own self-awareness of how their interests were linked to their culture of origin, and what else was out there that they might like to explore. They generally did not realise that other students in the classroom had such different interests and previous experiences, or different hopes and dreams about where their degree could take them. The teaching team recognised this challenge, and wanted to consider how students could build this self-awareness of their interests and blind spots, and also how they could use their time at UCL to expand their horizons to understand how other cultures were using or planning to use technology in the cultural sectors.

To address these changing interests, our DH programme's emphasis shifted away from working with textual material, which remains available for those looking to specialise in that way, but is not the default starting point as it once was. The module on XML has been one casualty of that shift and is currently being phased out of the programme to make space for a new module on social media. Secondly, the team introduced a new module called Global Digital Humanities in 2021, which helps students explore how different languages, cultures, religions, geographies, and economics have sculpted the ways people in different parts of the world approach digital humanities (Fernández l'Hoeste & Rodríguez, 2020; Crymble & Afanador-Llach, 2021; Dodd, 2021; Risam & Josephs, 2021). This module built upon the expertise of one of the authors of this chapter, Adam Crymble, who chaired the multilingual *Programming Historian* project, whose tutorials offered free DH skills training in four languages (English, Spanish, French, and Portuguese). Through the evolution of that project, it became clear to Crymble and others working on it, that technological needs varied across space and in different cultures. The "Global Digital Humanities" module was designed to impart that knowledge to students. Each week focuses on a different country or region, with weeks on China, the UK, North Africa, Latin America, and a week on Indigenous digital humanities. Most students in the class will have one week focusing on the place they came from, providing them with the chance to be the temporary experts for their peers. The remaining weeks, they learn about assumptions they may have had

about how typical their own experiences with technology have been. For example, students in China are often surprised that Western students on the programme do not share their passion for the museum sector. Students in the West are often blind about how much work goes into multilingual technologies in the rest of the world, having grown up in an English-speaking bubble (Galina, 2014; Mahony & Gao, 2018; Spence & Brandao, 2021).

Students on the module are challenged to become aware of their digital positionality, and to be globally focused technology professionals. Their major assignment on the module is to write a report advising the government of a country they have never lived in, on how they can best develop their own digital humanities degrees that suit the local needs and culture best. This forces students to explore solutions across cultural boundaries, to provide culturally sensitive suggestions, and gives them the skills to challenge the status quo at home or to take their skills abroad in a productive manner. This has been a natural evolution of the programme, designed to empower students with skills that can help them make an impact on the world stage through a deeper understanding of technology's connections to local cultures.

5 Widening the engagement at UCL

The growth and development of UCLDH, together with the taught graduate programme, coincided significantly with the year-on-year increase in the number of students from China. UCL has paid increasing attention and allocated resources to the development of connections and collaborations in the Greater China area. UCL had a presence at the China Scholarship Council Graduate Fair (CSCGF) for many years (this is now the China Scholarship Council-UCL Joint Research Scholarship), as well as setting up a strategic partnership with Peking University (PKU). In addition to this, funds have been available under the UCL Global Engagement initiative for connections and collaborations in East Asia with specific Partner Funds which now, as well as Peking University, include Zhejiang University, and Shanghai Jiao Tong University. Simon Mahony, one of the authors of this chapter, was successful in attending the CSCGF Beijing event from 2014 to 2018 as well as being awarded funding grants for travel and networking in mainland China. The growth and development of DH in China has been covered elsewhere (Tsui, 2020; Chen & Tsui, 2020; Wang et al., 2020), and this growth together with Mahony's support from UCL Global for travel and research trips enabled UCLDH to set up many connections with Chinese researchers and research centres. Beyond this, Mahony received many invitations to speak at conferences and give guest lectures to faculty and students at universities and other institutions.

While the universities strive to construct strategic partnerships at a high level and set up joint degree programmes and other initiatives, individuals, however, set up connections and relationships at a personal level. This has been a strength of UCLDH's relationship with China and has given UCLDH a presence at all the major DH research centres and groups in China as well as at many conferences and symposia (Mahony & Gao, 2018). Chinese students considering studying DH in the UK would have several choices and could be guided beyond the university rankings with a simple web search which would find links to staff representing

UCL and UCLDH in China, speaking at universities and conferences, and perhaps (as has happened on several occasions) at their own university.

Research trips are often reciprocal and UCLDH was delighted to host Professor Wang from Wuhan (and editor of this volume) in 2016 as our first Chinese visitor and who was followed by a succession of others over the following years. In 2019, Terras, Vanhoutte, and Gao were invited to Nanjing University to launch the Chinese translation of *Defining Digital Humanities: A Reader*, and this trip also fostered the foundation of the Chinese Digital Humanities Alliance (CDH) in 2019. In addition, Mahony was an invited guest speaker at the first Chinese National Digital Humanities Conference (CDH), held at Dunhuang in 2019. In the following year, Terras gave an online keynote address at the second CDH conference hosted at Shanghai Library, with Mahony on the Expert Council for the event. The tenth anniversary of UCLDH was in 2020 and we had arranged a succession of events which was to have included a two-day symposium, 'Digital Humanities and the Library, Research Partners', in collaboration with UCLDH, UCLDIS, UCL Library, PKU DH and PKU Library. This was supported by the UCL Partner Funds and would have included the dean and the director of the PKU DH centre (PKUDH) as well as DH researchers from PKU Library. As with all UCLDH events, this would have been an opportunity to engage and involve our students. Unfortunately, this had to be cancelled due to the pandemic.

As mentioned, the increased UCLDH connection with China has coincided with the significant growth in the number of Chinese applicants and, particularly following Brexit and the consequent impact on applications from within the European Union, they have become the largest single group of students on our taught programme. This in turn helps to raise the profile of both the centre and programme as graduating students become our ambassadors on their return home. We are also seeing an increase in the number of Chinese PhD students within the department (one of whom, Jin Gao, is now permanent UCLDIS staff, lead-author of this article, and co-editor of this volume), giving us more opportunity to connect with cultural aspects and sensitivities with their guidance as staff and teaching assistants (TAs). As UCL, more generally, increases its links with China through the development of formal partnerships and making funds available for travel and networking, UCLDH will continue to nurture our Chinese connections. We shall continue to build bridges, reaching out beyond our own cultural echo-chamber to promote communication and understanding, teaching ourselves as well as our students, to share knowledge and learning, to overcome cultural barriers and other obstacles to achieve harmony and greater prosperity for all.

6 Conclusion

Reviewing the institutional history of DH holds significant importance in shaping the future trajectory of the field. Retrospective studies, like the one undertaken of UCL in this chapter, provide a critical lens through which we can reflect on our past, interrogate our present, and forge a path towards a more inclusive and sustainable future for DH. By examining the successes and setbacks, the triumphs and trials of DH at UCL, this chapter helps us learn from past experiences and ensure

that our collective efforts are informed and guided by a broader understanding of the field's evolution. This knowledge empowers us to actively work towards creating a DH community that not only celebrates diversity but also actively seeks to overcome the barriers that hinder equality, diversity, and inclusion.

Notes

1 We extend our gratitude to Prof. Julianne Nyhan, Prof. Melissa Terras, Prof. Oliver Duke-Williams, Prof. Andrew Rosta, Dr Scott B Weingart, Dr Pang Rui, and Dr Bonnie Buyuklieva, for their generous contributions of insights and comments, offered amidst their busy schedules, which significantly enriched this chapter.

2 We acknowledge that this dataset does not represent the entirely of UCL's (or any other institutional) representation at DH conferences during this period. Nor does it include DH contributions to discipline-specific conferences not spearheaded by DH organisations. We also acknowledge the lack of inclusiveness and diversity of topics and scholars among DH conference proceedings. Nevertheless, this dataset provides a useful snapshot of participation in conferences between 1970 and 2023, and will be used as the basis for the network analysis in this study.

3 The average year of attendance for scholars at DH conferences is calculated by finding the mean of the individual years in which the particular scholar attended these conferences, providing a measure of their collective historical engagement with the events. For example, if scholar A attended 2010 conference and 2012 conference, then the average year of attendance is (2010 + 2012)/2 = 2011.

4 It is striking that almost every individual interviewed by Nyhan and Flinn (2016a) remarks on how open and friendly they found DH to be, on being introduced to the field, as opposed to traditional humanities disciplines.

5 More information related to Galton collection can be found here (Langkjær-Bain, 2019). We acknowledge that the problematic aspects of the Galton collection were put front and centre in the teaching as indicated in (Nyhan et al., 2014).

References

ADHO. (2022). Response to "The circus we deserve? A front row look at the organization of the annual academic conference for the Digital Humanities". *Digital Humanities Quarterly, 16*(4). http://www.digitalhumanities.org/dhq/vol/16/4/000664/000664.html

Apple. (2007, January 9). *Apple reinvents the phone with iPhone*. Apple Newsroom (United Kingdom). https://www.apple.com/uk/newsroom/2007/01/09Apple-Reinvents-the-Phone-with-iPhone/

Bailey, M. Z. (2011). All the digital humanists are white, all the nerds are men, but some of us are brave. *Journal of Digital Humanities, 1*(1). http://journalofdigi talhumanities.org/1–1/all-the-digital-humanists-are-white-all-the-nerds-are-menbut- some-of-us-are-brave-by-moya-z-bailey/

Barron, D. W. (1973). FORTRAN techniques with special reference to non-numerical applications, A. C. Day, Cambridge University Press, 1972. No. of pages: 96. Price: £2·50. *Software: Practice and Experience, 3*(4), 404. https://doi.org/10.1002/spe.4380030415

Batty, M., & Evans, T. (2022, April 20). *HARMONY. The Bartlett Centre for advanced spatial analysis*. Retrieved from https://www.ucl.ac.uk/bartlett/casa/research/current-projects/harmony

Batty, M., Hudson-Smith, A., Dodge, M., & Doyle, S. (1998, January 1). *Modelling virtual urban environments*. CASA Working Paper 1, The Bartlett Centre for Advanced Spatial Analysis. https://www.ucl.ac.uk/bartlett/casa/publications/1998/jan/casa-working-paper-1

Berry, D. M., & Fagerjord, A. (Eds.). (2017). *Digital humanities: Knowledge and critique in a digital age*. Polity.

Blandford, A. (2011). UCLIC@10: The changing face of interaction design. *The Ergonomist*, 6–7.

Bozeman, B., & Gaughan, M. (2011). How do men and women differ in research collaborations? An analysis of the collaborative motives and strategies of academic researchers. *Research Policy, 40*(10), 1393–1402. https://doi.org/10.1016/j.respol.2011.07.002

Brandeis Library. (2013, May 17). Digital Humanities: From 1851? *Brandeis Library*. Retrieved from https://blogs.brandeis.edu/library/2013/05/17/digital-humanities-from-1851/

British Library. (2022, October 8). International Dunhuang Project. *The British Library*. Retrieved from https://www.bl.uk/projects/international-dunhuang-project

Buchanan, G., Gow, J., Blandford, A., Rimmer, J., & Warwick, C. (2006). Representing aggregate works in the digital library. *Lecture Notes in Computer Science* (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 4172 LNCS (pp. 532–535).

Busa, R. (1950). Complete index Verborum of works of St. Thomas. *Speculum: A Journal of Medieval Studies, XXV*(1), 424–425.

Butler, C. S. (1974, December 13). *International Meeting/Annual General Meeting of the Association for Literary and Linguistic Computing*, London.

Butler, T., Fisher, S., Hockey, S., Coulombe, G., Clements, P., Brown, S., Grundy, I., Carter, K., Harvey, K., & Wood, J. (1999). *Can a team tag consistently? Experiences on the Orlando Project*. ACH/ALLC/ACH/ICCH/ALLC/EADH 1999 – Virginia. http://www2.iath.virginia.edu/ach-allc.99/schedule.html

CAAL. (2019, April 10). *CAAL*. Institute of Archaeology. https://www.ucl.ac.uk/archaeology/research/caal

Ceccaldi, E., Bardy, B., Bianchi-Berthouze, N., Fadiga, L., Volpe, G., & Camurri, A. (2020). The first international workshop on multi-scale movement technologies. *Proceedings of the 2020 International Conference on Multimodal Interaction* (pp. 895–896). https://doi.org/10.1145/3382507.3420060

Cheshire, J. (2017, July 28). Roger Tomlinson's PhD: The first in GIS – James Cheshire. *James Cheshire's Blog*. https://jcheshire.com/resources/roger-tomlinsons-phd-the-first-in-gis/

Chowdhury, I. (2013). A historian among scientists: Reflections on archiving the history of science in postcolonial India. *Isis, 104*(2), 371–380. https://doi.org/10.1086/670955

Crymble, A. (2021). *Technology and the historian: Transformations in the digital age*. University of Illinois Press.

Crymble, A., & Afanador-Llach, M. J. (2021). Digital history: The globally unequal promise of digital tools for history: UK and Colombia's case study. In J. Clark & A. Nye (Eds.), *Teaching history for the contemporary world tensions, challenges and classroom experiences in higher education* (pp. 85–98). Springer (7; Issue 7). http://doi.org/10.1007/978-981-16-0247-4

Day, A. C. (1972a). *A London Fortran course*. The Athlone Press.

Day, A. C. (1972b). *Fortran techniques: With special reference to non-numerical applications (Reprinted)*. Cambridge University Press.

Day, A. C. (2004). *Biography of Dr A. Colin Day. Colin Day & Roget's Thesaurus of the Bible*. https://colinday.co.uk/biography/index.shtml

Day, C., & Marriott, I. (1976). Software reviews: COCOA: A word count and concordance generator. *Computers and the Humanities, 10*(1), 56–56. https://doi.org/10.1007/BF02399143

Dodd, M. (Ed.). (2021). *Exploring digital humanities in India: Pedagogies, practices, and institutional possibilities*. Routledge, Taylor & Francis Group.

Dunhuang Academy. (2016). 数字敦煌 [Research project website]. Digital Dunhuang (数字敦煌). Retrieved from https://www.e-dunhuang.com/

EADH. (2023a). *EADH – The European Association for Digital Humanities*. https://eadh.org/

EADH. (2023b). *Susan Hockey—Curriculum vitae & publications*. EADH – The European Association for Digital Humanities. https://eadh.org/susan-hockey-curriculum-vitae-publications

Eagly, A. H. (2013). *Sex differences in social behavior: A social-role interpretation* (1st ed.). Psychology Press. https://doi.org/10.4324/9780203781906

Estill, L., Guiliano, J., Ortega, É., Terras, M., Verhoeven, D., & Layne-Worthey, G. (2022). The circus we deserve? A front row look at the organization of the annual academic conference for the Digital Humanities. *Digital Humanities Quarterly, 16*(4). https://www.digitalhumanities.org/dhq/vol/16/4/000643/000643.html

Farringdon, M. G. (1970). Symposium on the uses of the computer in literary research: A conference report. *Computers and the Humanities, 4*(5), 315–317.

Fernández l'Hoeste, H. D., & Rodríguez, J. C. (Eds.). (2020). *Digital humanities in Latin America*. University of Florida Press.

Fuller, D. Q., & Pang, R. (2015). The International Centre for Chinese Heritage and Archaeology (ICCHA): After a decade of success. *Archaeology International, 18*(1). https://doi.org/10.5334/ai.1806

Galina, I. (2014). Geographical and linguistic diversity in the digital humanities. *Literary and Linguistic Computing, 29*(3), 307–316. https://doi.org/10.1093/llc/fqu005

Gao, J., Duke-Williams, O., Mahony, S., Bold, M. R., & Nyhan, J. (2017, August). The intellectual structure of digital humanities: An author co-citation analysis. *Digital Humanities 2017*, Montreal. https://dh2017.adho.org/abstracts/083/083.pdf

Gao, J., & Hongxing, Z. (2023, June 21). *Chinese Export Watercolours (CEW) Project*. UCL Information Studies. https://www.ucl.ac.uk/information-studies/chinese-export-watercolours-cew-project

Gao, J., Nyhan, J., Duke-Williams, O., & Mahony, S. (2022). Gender influences in digital humanities co-authorship networks. *Journal of Documentation, 78*(7), 327–350. https://doi.org/10.1108/JD-11-2021-0221

Giacometti, A. (2014). *Evaluating multispectral imaging processing methodologies for analysing cultural heritage documents*. UCL.

Giacometti, A., Gibson, A., Hess, M., Hindmarch, J., MacDonald, L., Pal, K., Robson, S., Terras, M., & Weyrich, T. (2014). Heritage imaging at UCL. *Presented at the Annual Conference of Association for Historical and Fine Art Photography (AHFAP)*.

Gil, A., & Ortega, É. (2016). Global outlooks in digital humanities: Multilingual practices and minimal computing. In C. Crompton, R. J. Lane, & R. Siemens (Eds.), *Doing digital humanities: Practice, training, research* (pp. 58–70). Taylor & Francis Group. https://www.taylorfrancis.com/books/e/9781315707860/chapters/10.4324/9781315707860-12

Grandjean, M. (2016). A social network analysis of Twitter: Mapping the digital humanities community. *Cogent Arts & Humanities, 3*(1), 1171458. https://doi.org/10.1080/23311983.2016.1171458

Hardesty, W. H., & Mann, D. D. (1973). Literary and linguistic research: The Cambridge symposium. *Computers and the Humanities, 7*(5), 287–291.

Harrison, D., & Rogers, Y. (2013). UCLIC. *Interactions, 20*(6), 84–87. https://doi.org/10.1145/2533768

Hayes, B. (2017, February 6). Belles lettres meets big data. *American Scientist*. https://www.americanscientist.org/article/belles-lettres-meets-big-data

HESA. (2023, January 19). *Higher Education Student Statistics: UK, 2021/22—Where students come from and go to study*. Higher Education Student Statistics. https://www.hesa.ac.uk/news/19-01-2023/sb265-higher-education-student-statistics/location

Hockey, S. (2004). The history of humanities computing. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A companion to digital humanities* (pp. 1–19). Blackwell Publishing Ltd. http://doi.wiley.com/10.1002/9780470999875.ch1

Holloway, C., & Barbareschi, G. (2021). *Disability interactions: Creating inclusive innovations*. Morgan & Claypool Publishers.

Hudson-Smith, A., Gray, S., Ross, C., Barthel, R., De Jode, M., Warwick, C., & Terras, M. (2012). Experiments with the internet of things in museum space: QRator. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 1183–1184). https://doi.org/10.1145/2370216.2370469

Jones, A., & Churchhouse, R. F. (Eds.). (1976). The computer in literary and linguistic studies. *Proceedings of the Third International Symposium*. University of Wales Press.

Jones, S. E. (2013). *The emergence of the digital humanities*. Routledge.

Kador, T., Hannan, L., Nyhan, J., Terras, M., Chatterjee, H. J., & Carnall, M. (2018). Object-based learning and research-based education: Case studies from UCL curricula. In J. P. Davies & N. Pachler (Eds.), *Teaching and learning in higher education: Perspectives from UCL* (pp. 157–176) (11; Issue 11). UCL IoE Press. https://www.ucl-ioe-press.com/ioe-content/uploads/2018/05/Teaching-and-Learning-in-Higher-Education.pdf

Kemman, M. (2021). *Trading zones of digital history*. De Gruyter Oldenbourg. https://doi.org/10.1515/9783110682106

Kenneth, T. (2016, March 5). Reels of classic films tend to melt into goo; philanthropist David W. Packard won't let that happen. *Los Angeles Times*. https://www.latimes.com/entertainment/movies/la-ca-mn-david-packard-film-preservation-20160306-column.html

Kirschenbaum, M. (2014). What is "Digital Humanities," and why are they saying such terrible things about it? *Differences, 25*(1), 46–63.

Langkjær-Bain, R. (2019). The troubling legacy of Francis Galton. *Significance, 16*(3), 16–21. https://doi.org/10.1111/j.1740-9713.2019.01275.x

Le Deuff, O. (2018). *Digital humanities: History and development*. ISTE Wiley.

Liu, A. (2018). *Digital humanities diversity as technical problem*. Alan Liu. https://doi.org/10.21972/g21t07

Mahony, S. (2018). Cultural diversity and the digital humanities. *Fudan Journal of the Humanities and Social Sciences*. https://doi.org/10.1007/s40647-018-0216-0

Mahony, S., & Gao, J. (2019). Linguistic and cultural hegemony in the digital humanities. *Proceedings of the Digital Humanities Congress 2018*. The Digital Humanities Institute. https://www.dhi.ac.uk/openbook/chapter/dhc2018-mahony

Mahony, S., Nyhan, J., Terras, M., & Tiedau, U. (2016). Digital humanities pedagogy: Integrative learning and new ways of thinking about studying the humanities. *Proceedings of the Digital Humanities Congress 2014*, The University of Sheffield. https://www.dhi.ac.uk/books/dhc2014/digital-humanities-pedagogy/

Mahony, S., & Pierazzo, E. (2012). Teaching skills or teaching methodology? In B. D. Hirsch (Ed.), *Digital humanities pedagogy: Practices, principles and politics* (pp. 215–225). Open Book Publishers. https://doi.org/10.11647/OBP.0024.09

Makri, S., & Warwick, C. (2010). Information for inspiration: Understanding architects' information seeking and use behaviors to inform design. *Journal of the American Society for Information Science and Technology, 61*(9), 1745–1770. https://doi.org/10.1002/asi.21338

Manley, E., Concannon, D., & Gray, S. (2016, November 28). *Africa Centre/I-Sense*. The Bartlett Centre for Advanced Spatial Analysis. https://www.ucl.ac.uk/bartlett/casa/research/current-projects/africa-centrei-sense

McCarty, W. (2003). *Humanities computing. In Encyclopedia of Library and Information Science*. Marcel Dekker.

Miles, J. (1946). *Major adjectives in English poetry from Wyatt to Auden*. University of California Press.

Milton, R., Buyuklieva, B., Hay, D., Hudson-Smith, A., & Gray, S. (2018). Talking to GNOMEs: Exploring privacy and trust around internet of things devices in a public space. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3170427.3188481

Nebbia, M., Cilio, F., & Bobomulloev, B. (2021). Spatial risk assessment and the protection of cultural heritage in southern Tajikistan. *Journal of Cultural Heritage, 49*, 183–196. https://doi.org/10.1016/j.culher.2021.01.012

Nyhan, J. (2023). *Hidden and devalued feminized labour in the digital humanities: On the index Thomisticus project 1965–67*. Routledge.

Nyhan, J., & Duke-Williams, O. (2014). Joint and multi-authored publication patterns in the digital humanities. *Literary and Linguistic Computing, 29*(3), 387–399. https://doi.org/10.1093/llc/fqu018

Nyhan, J., & Flinn, A. (2016a). *Computation and the humanities: Towards an oral history of digital humanities*. Springer.

Nyhan, J., & Flinn, A. (2016b). So, into the Chopper it went: Gabriel Egan and Julianne Nyhan. In J. Nyhan & A. Flinn (Eds.), *Computation and the humanities* (pp. 237–256). Springer International Publishing. https://doi.org/10.1007/978-3-319-20170-2_16

Nyhan, J., & Flinn, A. (2016c). They took a chance: Susan Hockey and Julianne Nyhan. In J. Nyhan & A. Flinn (Eds.), *Computation and the humanities* (pp. 87–97). Springer International Publishing. https://doi.org/10.1007/978-3-319-20170-2_6

Nyhan, J., & Passarotti, M. (2019). *One origin of digital humanities: Fr Roberto Busa in his own words*. https://doi.org/10.1007/978-3-030-18313-4

Nyhan, J., Terras, M., & Mahony, S. (2014). Digital humanities and integrative learning. In D. Blackshields, J. Cronin, B. Higgs, S. Kilcommins, M. McCarthy, & A. Ryan (Eds.), *Integrative learning: International research and practice* (pp. 235–247). Routledge (18; Issue 18). http://www.routledge.com/books/details/9780415711074/

Pal, K., Avery, N., Boston, P., Campagnolo, A., De Stefani, C., Matheson-Pollock, H., Panozzo, D., Payne, M., Schüller, C., Sanderson, C., Scott, C., Smith, P., Smither, R., Sorkine-Hornung, O., Stewart, A., Stewart, E., Stewart, P., Terras, M., Walsh, B., … Weyrich, T. (2016). Digitally reconstructing the Great Parchment Book: 3D recovery of fire-damaged historical documents. *Digital scholarship in the humanities*, fqw057. https://doi.org/10.1093/llc/fqw057

Risam, R. (2015). Beyond the margins: Intersectionality and the digital humanities. *Digital Humanities Quarterly, 9*(2). http://www.digitalhumanities.org/dhq/vol/9/2/000208/000208.html

Risam, R., & Josephs, K. B. (Eds.). (2021). *The Digital Black Atlantic*. University of Minnesota Press.

Rockwell, G. (2007, May 2). An alternate beginning to humanities computing? *Theoretica. Ca*. http://theoreti.ca/?p=1608

Rørstad, K., & Aksnes, D. W. (2015). Publication rate expressed by age, gender and academic position – A large-scale analysis of Norwegian academic staff. *Journal of Informetrics, 9*(2), 317–333. https://doi.org/10.1016/j.joi.2015.02.003

Russell, D. B. (1965). *COCOA — A word-count and concordance generator*. Chilton Computing and UKRI Science and Technology Facilities Council. http://www.chiltoncomputing.org.uk/acl/applications/cocoa/p001.htm

Sheppard, L., Reades, J., & Freeman, R. P. J. (2023). Gendering and diversifying the research pipeline: A quantitative feminist geographical approach to gender in higher education. *Annals of the American Association of Geographers, 113*(4), 817–833. https://doi.org/10.1080/24694452.2022.2160302

Shiode, N., Batty, M., Dodge, M., Jiang, B., & Thurstain-Goodwin, M. (1998). The centre for advanced spatial analysis and its projects on geographic information systems and urban analysis. *Theory and Applications of GIS, 6*(1), 81–86. https://doi.org/10.5638/thagis.6.81

Spence, P. J., & Brandao, R. (2021). Towards language sensitivity and diversity in the digital humanities. *Digital Studies / Le Champ Numérique, 11*(1). https://doi.org/10.16995/dscn.8098

Sugimoto, C. R., Ni, C., West, J. D., & Larivière, V. (2015). The academic advantage: Gender disparities in patenting. *PLoS ONE, 10*(5), e0128000. https://doi.org/10.1371/journal.pone.0128000

Terras, M. (2012, January 20). Infographic: Quantifying digital humanities. *UCL DH Blog*. http://blogs.ucl.ac.uk/dh/2012/01/20/infographic-quantifying-digital-humanities/

Terras, M., Amman, R., & Gray, S. (2013). Textal: A text analysis smartphone app for digital humanities. *Journal of Digital Humanities, 2*(3). https://journalofdigitalhumanities.org/2-3/textal-a-text-analysis-smartphone-app-for-digital-humanities/

Terras, M., Baker, J., Hetherington, J., Beavan, D., Zaltz Austwick, M., Welsh, A., O'Neill, H., Finley, W., Duke-Williams, O., & Farquhar, A. (2018). Enabling complex analysis of large-scale digital collections: Humanities research, high-performance computing, and transforming access to British Library digital collections. *Digital Scholarship in the Humanities, 33*(2), 456–466. https://doi.org/10.1093/llc/fqx020

Terras, M. M. (2010, July 30). Announcing the new MA/MSc in digital humanities at UCL. *UCLDH Blog*. https://blogs.ucl.ac.uk/dh/2010/07/30/announcing-the-new-mamsc-in-digital-humanities-at-ucl/

Terras, M. M., Nyhan, J., & Vanhoutte, E. (Eds.). (2013). *Defining digital humanities: A reader*. Ashgate Publishing Limited.

Thaller, M. (2017). Geschichte der digital humanities. In F. Jannidis, H. Kohle, & M. Rehbein (Eds.), *Digital humanities* (pp. 3–12). J.B. Metzler. https://doi.org/10.1007/978-3-476-05446-3_1

The ICE. (2021). *The International Corpus of English (ICE)*. Retrieved from https://www.ucl.ac.uk/english-usage/projects/ice.htm

Times. (2022, April 3). *Impact rankings: Gender equality*. Times Higher Education (THE). https://www.timeshighereducation.com/rankings/impact/2022/gender-equality

Tomlinson, R. F. (1974). *Geographical information systems, spatial data analysis and decision making in government* [Doctoral thesis, University of London]. https://discovery.ucl.ac.uk/id/eprint/1563584/

UCL. (2005). *LAIRAH: Log analysis of internet resources in the arts and humanities*. https://www.ucl.ac.uk/infostudies/LAIRAH/

UCL. (2008a). *About us*. UCL Grand Challenges. https://www.ucl.ac.uk/grand-challenges/about-us

UCL. (2008b). *(VERA) Virtual Environments for Research in Archaeology*. https://web.archive.org/web/20140728122248/http://www.jisc.ac.uk/whatwedo/programmes/vre2/vera.aspx

UCL. (2022). *UCL Survey of English Usage*. https://www.ucl.ac.uk/english-usage/

Valeonti, F., Bikakis, A., Terras, M., Speed, C., Hudson-Smith, A., & Chalkias, K. (2021). Crypto collectibles, museum funding and OpenGLAM: Challenges, opportunities and the potential of non-fungible tokens (NFTs). *Applied Sciences, 11*(21), 9931. https://doi.org/10.3390/app11219931

Wang, X., & Inaba, M. (2009). Analyzing structures and evolution of digital humanities based on correspondence analysis and co-word analysis. *Art Research, 9*, 123–134.

Warwick, C., Mahony, S., Nyhan, J., Ross, C., Terras, M., Tiedau, U., & Welsh, A. (2011). UCLDH: Big tent digital humanities in practice. *Digital Humanities 2011 Conference Abstracts*, 387–389.

Warwick, C., Terras, M., & Nyhan, J. (2012). *Digital humanities in practice*. Facet Publishing.

Watson, E., Eddo-Lodge, R., & Solnit, R. (2022, April 6). *City of Women London*. The Bartlett Centre for Advanced Spatial Analysis. https://www.ucl.ac.uk/bartlett/casa/research/current-projects/city-women-london

Weingart, S. B., & Eichmann-Kalwara, N. (2017). What's under the big tent?: A study of ADHO conference abstracts. *Digital Studies/Le Champ Numérique, 7*(1). https://doi.org/10.16995/dscn.284

Weingart, S. B., Eichmann-Kalwara, N., & Lincoln, M. (2023). *The Index of Digital Humanities Conferences Data* [dataset]. Carnegie Mellon University. https://doi.org/10.1184/R1/12987959.v4

Zaagsma, G. (2013). On digital history. *BMGN – Low Countries Historical Review, 128*(4), 3. https://doi.org/10.18352/bmgn-lchr.9344

Zhang, M. (2019, April 17). Palace Museum to launch new mobile app with informative features. *CGTN Culture China*. https://news.cgtn.com/news/3d3d414e3267444d34457a6333566d54/index.html

6 Digitalisation-preparedness of cultural heritage institutions

To what extent is Africa ready?

Joseph Kiplang'at and Humphrey Keah

1 Introduction

Digitisation refers to the process of converting information materials to binary representations. This process precedes digitalisation, which is the preparation of digitised information materials with various types of metadata for use by artificial intelligence (AI) agents for task automation and knowledge discovery. When this state is achieved, a digital transformation may be said to have been realised. AI implementations in cultural heritage institutions (Libraries, Archives and Museums) can be considered as a digital humanities implementation by virtue of the integration of computational methods into social science research. Some studies (Marciano, 2022) suggest the emergence of a new discipline of Computational Archival Science (CAS) from a convergence of the disciplines of computing and archival science. The Venn diagram (Figure 6.1) illustrates the convergence of the domains of information management and computing produce the intersection domain of AI and CAS. National Libraries, Archives and Museums are strategically placed to provide leadership on applications of emerging technologies for automation and knowledge discovery for the benefit of the country at large, hence our rationale for employing them as our primary information sources of information on the status of digitisation in Kenya.

2 Literature review

While some studies on various aspects of digitisation projects in Africa have been undertaken such as that on standards (Katuu, 2016), it is evident that most of the literature on digitisation available is fairly recent, pointing to a window of immense opportunity for information professionals in the wake of the fourth Industrial Revolution. Africa being home to nine out of 15 of the world's fastest-growing economies makes the continent an increasingly attractive environment for global business investment in the area of digitisation (Chambers, 2015). On a global scale, this observation is echoed in by the World Bank Digital Economy for Africa initiative (DE4A) which aims to ensure that every individual, business and government in Africa will be digitally enabled by 2030 in support of the African Union "Digital Transformation Strategy for Africa" (World Bank Digital Economy for Africa initiative [DE4A],

DOI: 10.4324/9781032707211-7

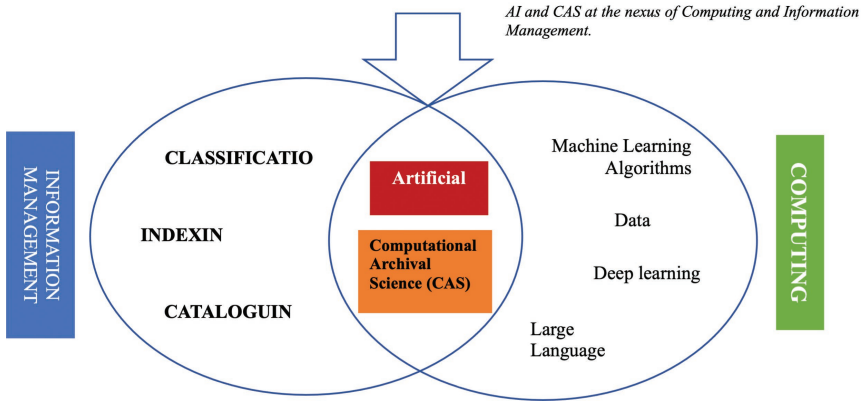

Figure 6.1 Convergence of computing and information management.

2023). It is important to note that Kenya ranks as one of the most-wired nations in Africa offering the right conditions for digital participation, including extensive Internet infrastructure and a well-established innovation ecosystem, with a growing list of start-up incubators. It is an environment that gave birth to the mobile money service known as M-Pesa, the crisis-mapping tool Ushahidi, and the iHub innovation space – all of which have put Kenyan creativity on the map (Nitsche, 2019). Kenya is also mentioned as the third-largest economy in sub-Saharan Africa, leading in the adoption of fourth Industrial Revolution (digital economy) in the region in the provision of essential citizen services, social protection services, health, energy, digital learning, tax administration, judiciary and land services, among others (Meru & Kinoti, 2022). The Kenya National Digital Master Plan is the guiding document for digitisation of information materials from a national perspective (Kenya-Digital Master Plan 2022–2032). According to Finsense (2023), there are challenges and opportunities for digitisation in Africa, some of which have also been cited by the institutions we surveyed for the purposes of this study.

3 Research methodology

The study was mainly qualitative with aspects of quantitative data. The study targeted information professionals in charge of digitisation projects in the surveyed institutions. The key informants were drawn from top management. A structured interview schedule was administered to respondents who were unable to give adequate time for face-to-face interviews. The respondents were cooperative in availing the necessary information for the study. However, a challenge was encountered through bureaucracy in some of the institutions which took too long to obtain the required data for the study. Consequently, one of the institutions under investigation had to be dropped from the study population. Permission to conduct the study was obtained through official communication channels by way of introductory letters to conduct the research.

3.1 Data collection procedures

Research data was gathered through face-to-face interview and structured interviews with information management professionals in charge of digitisation projects at national information agencies namely, the Kenya National Library Services (KNLS), Kenya National Archives & Documentation Services (KNADS) and the National Museums of Kenya. Other sources of information include the results of a baseline survey conducted on the status of Digital Humanities (DH) in sub-Saharan Africa by Walsh et al. (2022), a previous unpublished baseline survey on the status of legacy projects in international academic institutions in Kenya (AHS Legacy Centre Team, 2021), as well as the Anglican Church of Kenya (ACK), with precolonial archives dating back to the 1840s.

4 Data analysis and presentation of findings

Data was analysed qualitatively based on seven criteria under investigation, namely, the project background, project scope, project infrastructure, personnel, funding, challenges encountered and perceptions of staff on digitisation and digitalisation.

4.1 Background of digitisation projects

The study noted that digitisation projects in Kenya are at their nascent stages (1–2 years old); however, the National Archives of Kenya was one of the first institutions to digitise the records and archives in the years 2004 and 2005, respectively. The survey established that digitisation projects are generally anchored on corporate objectives and guided by the national strategy on digitisation. A good example is the Kenya National Library Services (KNLS) where the Government of Kenya (GoK) spearheaded digitisation efforts. At the National Museums of Kenya (NMK), the digitisation function is under the ambit of the Marketing Department, whereas at the National Archives (KNADS), there is a special project on digitisation in the research library services section. At one of the premier academic institutions in Kenya, the background to digitisation stems from a need by the alumni to establish a legacy dating back to 100 years since the school was founded (AHS Legacy Centre Team, 2021), whereas a mainstream church – the Anglican Church of Kenya (ACK) – envisions a multimedia archival gallery in a custom-built environment that will attract researchers from across the globe. Despite the specific backgrounds to the projects, the general objective is preservation of institutional memory. Table 6.1 summarises the backgrounds of digitisation projects in Kenya.

4.2 Project scope

The scope of information materials for the surveyed institutions is unlimited in terms of physical formats. The collections are historical in nature with archival value dating back to the 1840s and post-colonial times, serving a global clientele with some of the users visiting from Europe, Asia and the Americas. Economically speaking, the National Library digitisation facility doubles as an

Table 6.1 Summary of digitisation projects

Assessment criterion	*Aspects & notes*		
	Developmental stage	**Project basis**	**Project location**
Project background	I 1–2 years (Nascent) II Earliest started in 2004 but inconsistent	I Corporate objectives II High-level government involvement	I Fully-fledged digitisation department II Strategic placed in marketing department

Table 6.2 Scope of digitisation

Assessment criterion	*Aspect*	*Scope notes*
Project scope	**Geographic coverage**	I National
	a Collections	II East Africa region
	b Clientele (Users)	Global
	Subject coverage	I Unlimited
		II Mission-specific (for Church)
	Time-span	I Historical collections dating back to the 1840s
		II Post-colonial era
	Types of information materials digitised	Unlimited (Multimedia)
	Work accomplished	Up to 7 million records
	Maximum workload	680 million records
	Collaborators	I Institutional
		II Private
	Role of institution	I Advisory
		II Partnership

income-generating entity designed to digitise both public and private information collections. The National Archives estimates a total of 680 million records, with 7 million already digitised, while the premier academic institution estimates a total of 15 linear kilometres of documents dating back to the 1920s, representing 150 cubic metres of archival materials. It was established that some of the surveyed institutions play an advisory role while others play the role of partnerships in digitisation projects. Table 6.2 summarises the scope of digitisation projects in Kenya.

4.3 Project infrastructure

Courtesy of the Office of President, the Kenya National Library Services (KNLS) acquired an AI-enabled robotic scanner from Europe which is so far the most advanced technology in the region running on Q-Capture software, with a dedicated digitisation laboratory and controlled access. The robotic scanner has three operational modes, namely, manual, semi-auto and automatic (AI) modes. The AI mode

Table 6.3 Infrastructure of digitisation projects in Kenya

Assessment criterion	*Aspects/description*	
Project infrastructure	**Technology**	**Physical space**
	I Artificial Intelligence (Robotic) scanner II Q-Capture Software	I Dedicated digitisation lab II Satellite digitisation units

is fully automated and can digitise information materials of various sizes without human intervention. Being the only institution owning this equipment in East Africa, the Kenya National Library Services is able to extend its client coverage within the region. Table 6.3 is a summary of the project infrastructure of digitisation projects in Kenya.

4.4 Staffing level of digitisation project

Given the nascent stages of digitisation projects in the country, the staffing level is inadequate. This ranges between four and eight digitisation staff performing specialised tasks including, photography, content editing, selection, quality control, cataloguing and uploading to the virtual library. It is noted that the national library has a relatively large team of digitisation staff located at its satellite libraries who comprise of system librarians and information management staff. The surveyed institutions indicated that owing to the voluminous nature of archival collections for digitisation, more staff is required. One of the institutions employed casual labourers and interns but due to fiscal constraints they had to stop, pending an increase of recurrent budget from the exchequer or support of external grants. Capacity-building of existing staff and on-boarding of the right cadre of staff with a background in information sciences is key to running digitisation projects professionally. It was observed from the survey that although the national cultural heritage institutions should play an advisory and technical support role by virtue of their expertise, there is need for capacity-building of staff to effectively handle digitisation projects given the cutting-edge nature of innovations being experienced for the first time in Africa. Table 6.4 is a summary of staffing level.

4.5 Project funding

The sources of funding for digitisation projects range from government funding to external grants from international organisations such UNESCO, the International Council on Archives (ICA) and the British Library Endangered Archives Programme (BL-EAP), the Modern Endangered Archives Programme (MEAP), Electronic Information for Libraries (eIFL) and BookAid International. The surveyed institutions are always on the look out for potential collaborators and funders. Table 6.5 is a summary of project funding for digitisation projects in Kenya.

Table 6.4 Staffing level of digitisation projects

Assessment criterion	*Aspects & notes*	
Project personnel	**Staff population**	I 4–8 (minimum) II 10–20 (institutions with satellite digitisation staff)
	Division of labour	I Photography II Content editing III Selection IV Quality control V Cataloguing VI Uploads to virtual library
	Staff cadre	I Full-time II Casuals III Interns
	Designations	I System librarians II Information management staff
	Requirements	Additional staff

Table 6.5 Sources of funding of digitisation projects

Assessment criterion	*Aspects & notes*	
Source of funds	Corporate	Institution finances digitisation project
	State	High-level State intervention (Office of the President)
	International Organisations	UNESCO, British Library, eIFL, BookAid International
	Nature of funding	Equipment, salaries/stipends, travel, training

4.6 Project challenges

The surveyed institutions seemed to face similar challenges on account of their infancy stages, some of which have been mentioned by teamFinsense (2023).

4.6.1 Inadequate funding

The survey established that the problem of inadequate funding has a bearing on the following aspects:

1 Staffing Level. The right cadre and number of staff is required for digitisation projects. Due to the digital humanities nature, it was observed that often times a blend of ICT and information management staff is bound to occur in most projects, calling for a need to synchronise the working relations of the two categories of staff. One of the respondents gave an example of a digitisation project which was allocated to ICT department. However, the ICT staff had little regard for conservation aspects of original documents, with dire consequences for the

Table 6.6 Summary of challenges facing digitisation projects in Kenya

Assessment criterion	*Aspects & notes*	
Challenges of digitisation projects	Limited funding	Negatively impacts project execution.
	Staffing	Need to synchronise the working relations between ICT and information management staff
	Capacity-building	Inadequate funding to bring trainer for robotic scanner from abroad is often a major challenge
	Technical infrastructure	I Cost of imported digitisation equipment is prohibitive. II Some institutions still employ old technologies for information management.

project. In yet another instance, a digitisation project entrusted to ICT staff was completely abandoned upon completion of the scanning process due to waning excitement with the project as well as absence of information management expertise in handling the information management aspect of the project. Adequate funding will enable most of the institutions to scale up their human resource capacity in order to realise improved productivity for digitisation projects.

2 Capacity-building. It was observed that in some of the institutions, digitisation staff need sensitisation through relevant training in order to achieve desirable outcomes for digital archiving projects. Given that the technology employed for digitisation is foreign, inadequate funding to bring the trainer from abroad was cited as a major challenge which could have dire consequences for digitisation projects without instituting appropriate mitigation measures to build internal capacity of digitisation project staff.

3 Technical Infrastructure. It was observed that there is generally dedicated space for digitisation projects in the surveyed institutions equipped with computers and attendant equipment for researchers. However, the technologies for digitisation, such as the robotic scanner, are imported and the cost of the equipment is prohibitive. Some of the institutions still employ old technologies for information management as was reported by one of the respondents who observed that the information management system in use became unstable at 500,000 records, which required a more robust system that was never implemented on account of fiscal constraints. Without adequate funding digitisation projects run the risk of coming a cropper. Table 6.6 is a summary of challenges facing digitisation projects in Kenya.

4.7 Synopsis of digital humanities status in Sub-Saharan Africa

The phenomenon of digitisation being an aspect of digital humanities, we take this opportunity to highlight the results of a digital humanities infrastructure study by Walsh et al. (2022), in which the authors of this study conducted a baseline survey of

sub-Saharan African universities offering academic programmes in information sciences. It is expected that such institutions should be in a position to champion digitisation initiatives by virtue of their professional expertise in the information sciences.

The DH study was informed by the wider scientific domain in modeling interdisciplinary relationships between information sciences and humanities and research units (Rosenbloom, 2009). It was established that 79% of the surveyed LIS schools offer the following programs: Library and Information Studies, Archives and Records Management, Publishing and Media Studies, Information Technology and Computing and Informatics. Only 33% of the institutions surveyed are actively engaged in DH research and scholarship. Although 91% of LIS Schools in Africa have dedicated computer laboratories that support teaching, learning and research, it was established that the physical and technical infrastructure was inadequate, which corroborates with our observation in this study with regard to inadequate funding to support aspects such as infrastructure. On a more positive note, digital libraries (72.7%), digital archiving (54%) and institutional repositories (90.9%) are widely used to support DH activities in the universities. Eighty-two per cent (82%) of the respondents anticipated the integration of DH in their curricula. A significant number of respondents (91%) are collaborating with humanities departments in teaching and curricula design as well as in research.

This focused regional study concludes that DH initiatives are slowly taking root in universities in Africa. There is a need to create more awareness of DH through workshops and conferences tailored to DH thematic areas. More ambassadors of DH are required in all the regions of Africa as these could possibly champion digitisation projects in their countries to realise a digital transformation for the continent.

5 Discussion

Artificial intelligence is becoming an important agent for task automation and knowledge discovery in social science circles, despite eliciting mixed reactions from various quarters globally (Crew, 2020; Cushing & Osti, 2022; InterPARES, 2021; Seles, 2021). Although it was earlier noted that the twin processes of digitisation and digitalisation are important precursors to artificial intelligence implementations, this study has established a rudimentary level of preparedness on a national scale, with limited technical infrastructure and manpower to achieve the national goal of digitising 50 billion records by the year 2030 according to the National Digitisation Master Plan (Kenya-Digital Master Plan 2022–2032). It was evident from the face-to-face interviews that the subject of digitisation is often thought of as an ICT project with the information management perspective of digitisation being given less serious consideration. Cultural heritage institutions and information science professionals should be at the forefront in terms of ensuring integration of professional archival imperatives some of which Sarah (2018) elaborates in her study on archival principles and practices, and Katuu (2016) addresses in his study on global information management standards, in order for digitisation projects to be of impact to artificial intelligence implementations. Further, it was established in

this study that majority of the digitisation project staff find digitisation projects to be exciting and an excellent learning experience, although some of them did not originally possess an information management background. Such staff reckon their exposure to digitisation projects to be a great opportunity for cross-disciplinary collaboration with information management staff, for discovery and innovation through research on similar projects to study best practices on project workflows globally. Given the diversity of potential digitisation objects and the need to preserve cultural heritage for posterity, the question of funding becomes another major imperative, without which little can be achieved in terms of the digitisation work, as was established by the survey in one of the institutions which embarked on a digitisation project in the year 2004 and had to stop the project due to lack of funds. This aspect makes finances a top priority in determining the scope and successful execution of digitisation projects.

6 Conclusion

The conclusion to this study reverberates with the question at the title head – "to what extent is Africa ready?" A look at the existing literature on the digitisation landscape in Africa reveals a vast untapped potential (Chambers, 2015; teamFinsense, 2023), which is corroborated by this study as highlighted in the discussion section, and the previous study on the status of digital humanities in Africa. It can be confidently argued from the two studies that Africa's preparedness lies in its great potential resident in its digital humanities infrastructure, state agencies, cultural heritage institutions and information professionals. Institutions with digital humanities infrastructure could facilitate setting up of digitisation labs, cultural heritage institutions can play advisory and technical support roles, while the state can be a major player in digitisation infrastructure provision. Going forward, these key players, together with adequate funding, stand to animate the artificial intelligence dream of the continent to global standards.

References

AHS Legacy Centre Team. (2021). *Final Benchmarking Report*. AHS Old Boys Club, Nairobi.

Chambers, J. (2015, June 1). *Why digitization is an opportunity for Africa*. Retrieved November 10, 2023, from https://www.weforum.org/agenda/2015/06/why-digitization-is-an-opportunity-for-africa/

Crew, B. (2020). Artificial-intelligence research escalates amid calls for caution. *Nature Index, 588*(S101). https://doi.org/10.1038/d41586-020-03408-9

Cushing, A. L., & Osti, G. (2022). So how do we balance all of these needs?: How the concept of AI technology impacts digital archival expertise. *Journal of Documentation, 79*(7), 7–29. https://doi.org/10.1108/JD-08-2022-0170

Google AI. (2023, May 15). *Bard Chatbot*. Retrieved May 31, 2023, from https://bard.google.com/

Heaven, W. D. (2023, February 8). *ChatGPT is everywhere. Here's where it came from*. Retrieved March 15, 2023, from https://tinyurl.com/2p8kyjca

InterPARES. (2021). *InterPARES TRUST AI*. Retrieved March 19, 2023, from https://interparestrustai.org/trust/aboutus/partners

Katuu, S. (2016). Managing digital records in a global environment: A review of the landscape of international standards and good practice guidelines. *The Electronic Library, 34*(5), 869–894. http://doi.org/10.1108/EL-04-2015-0064

Kenya-Digital Master Plan 2022–2032. https//repository.kippra.or.ke/handle/123456789/3580

Marciano, R. (2022). AFTERWORD: Towards a new discipline of Computational Archival Science (CAS). In L. Jaillant (Ed.), *Archives, access and artificial intelligence: Working with born-digital and digitized archival collections* (Vol. 2, pp. 205–216). Bielefeld University Press.

Meru, A. K., & Kinoti, M. K. W. (2022, February 7). *Digitalisation and public sector service delivery in Kenya*. Retrieved November 8, 2023, from https://link.springer.com/chapter/10.1007/978-3-030-83909-3_10

Nitsche, L. (2019, January 23). *A tale of two Kenyas: Digitalization has brought change but not always for the better*. Retrieved November 8, 2023, from https://p.dw.com/p/3C2XW

OpenAI. (2023, March 7). *Introducing ChatGPT*. Retrieved March 14, 2023, from https://openai.com/blog/chatgpt

Rosenbloom, P. (2009). The great scientific domains and society: A metascience perspective from the domain of computing. *International Journal of Science and Society, 1*(1), 133–143. https://doi.org/10.18848/1836-6236/CGP/v01i01/51493

Sarah Welland. (2018). Archives principles and practices. *Journal of the Australian Library and Information Association, 67*(1), 64. https://doi.org/10.1080/24750158.2018.1429795

Seles, A. (2021). *Artificial intelligence and archives. West Nordic archives conference*. International Council on Archives (ICA) / Conseil International des Archives (CIA). Retrieved March 8, 2023, from https://www.youtube.com/watch?v=ZA3xqg9hW-s

teamFinsense. (2023, March 15). *Digitizing Africa 2023: The challenges, opportunities and solutions*. Retrieved November 10, 2023, from https://finsense.africa/blog/2023/03/15/digitizing-africa-2023-the-challenges-opportunities-and-solutions/

The World Bank. (2023, Fall Edition). *Digital Economy for Africa initiative (DE4A)*. Retrieved November 8, 2023, from https://www.worldbank.org/en/programs/all-africa-digital-transformation

Walsh, J. A., Cobb, P. J., de Fremery, W., Golub, K., Keah, H., Kim, J., … & Underwood, T. (2022). Digital humanities in the iSchool. *JASIST, Special Issue on Digital Humanities, 73*(2), 188–203. https://doi.org/10.1002/asi.24535

Part II

Innovative practices in China

7 Digitising cultural heritage in Hong Kong

An overview

Jeremy Tzi Dong Ng and Xiao Hu

1 Introduction

Cultural heritage is defined as the physical and intangible legacies passed down from previous generations, maintained and preserved in the present for the benefit of future generations (Macrì & Cristofaro, 2021). It encompasses a wide array of objects and artefacts (termed *movable cultural heritage*), buildings, structures, and environments (termed *immovable cultural heritage*), and phenomena and human activities (terms *intangible cultural heritage*), which embody and express knowledge, beliefs, and values (Dümcke & Gnedovsky, 2013; Pérez et al., 2010). Integrating the inheritance of our ancestors with the idea of shaping the future, cultural heritage acts as a vital pillar for sustainable development of the world (Kontiza et al., 2020), and is thus both explicitly part of and implicitly built into United Nation's Sustainable Development Goals (SDGs), such as Target 11.4[1] and Target 4.7,[2] respectively (Jagielska-Burduk et al., 2021; Koya & Chowdhury, 2020). The world has been experiencing globalisation, modernisation, and urbanisation at a rapid pace in recent decades, along with the fresh wounds from the COVID-19 pandemic and the Russian-Ukraine War. These worldwide phenomena and incidents have posed threats and opportunities to cultural heritages, leading to increased urgent demands for both global and local actions in protecting and preserving our valuable assets of humankind (Jagielska-Burduk et al., 2021; Quintero et al., 2020). The United Nations Education, Scientific, Cultural Organisation (UNESCO) also emphasises the importance of promoting the values and significance of cultural heritage to the younger generation, comprised of digital natives living in an ever-evolving, multicultural societies (İnanç & Liew, 2017). Due to the inherent multidisciplinary nature of cultural heritage (Koya & Chowdhury, 2020), the preservation and promotion of cultural heritage require concerted efforts from different stakeholders (Burri, 2010), ranging from government authorities and cultural heritage institutions (e.g., museums) to private (e.g., companies) and grassroots organisations (e.g., small-scale interest groups) as well as individuals (Corallo et al., 2019). Such efforts have also departed from being only traditional, one-size-fits-all methods and approaches to experimenting with innovative, localised perspectives and initiatives (Zhang et al., 2022), with an increasing emphasis on digitisation (Corallo et al., 2019).

DOI: 10.4324/9781032707211-9

As a special administrative region in China, Hong Kong is renowned for being not only an international financial centre but also a densely populated metropolitan city (Kee, 2019). Under the British rule for 156 years after thousands of years of immersion in the Chinese culture and before the handover back to China in 1997, Hong Kong experiences a long and unique history, which has led to the emergence and accumulation of a diversity of cultural heritage in this postcolonial, arguably multicultural city (Xue, 2016). As of the time of writing this chapter, in Hong Kong there are 1,444 statutorily recognised historic buildings and monuments,[3] 480 officially listed intangible cultural heritage,[4] along with an abundance of socio-historically significant objects and artefacts. Aiming to protect these valuable assets and resources and to heighten citizens' awareness of their importance, a variety of stakeholders have contributed their time and efforts towards the preservation and promotion of cultural heritage in Hong Kong. To name a few, these contributors include government bodies (e.g., Antiquities and Monuments Office) (Ho & Hou, 2019), cultural heritage institutions (e.g., Hong Kong Heritage Museum) (Meng et al., 2023), universities and schools (Allsop, 2022), non-governmental organisations (NGOs), and advocacy groups (Yung & Chan, 2011). Their endeavours range from systematically documenting and assessing every historical architecture in the city to engaging different members of society in learning about traditional craftsmanship inherited by indigenous communities (Kee, 2019; Meng et al., 2023). With their own objectives and agendas, these stakeholders of cultural heritage in Hong Kong have also initiated, supported, and participated in various types of digitisation projects.

To the best of our understanding, however, there is a lack of a comprehensive and up-to-date overview of Hong Kong-based cultural heritage digitisation projects. To fill the gap, this chapter aims to survey these projects in Hong Kong. First, it gives a brief account of related literature on cultural heritage digitisation around the world. Then, through our literature review and leveraging the document analysis method, we survey the various projects conducted in Hong Kong both in the literature and in practice. Based on these findings, we summarise the multidimensional approaches adopted in these projects and report the development trends across projects and over time. After identifying the facilitating factors, limitations and challenges of these projects, this chapter concludes with recommendations on best practices for future projects of cultural heritage digitisation in Hong Kong and beyond.

2 Related literature

2.1 Digitisation of cultural heritage: An overview

Digital technologies have been leveraged to record and disseminate cultural heritage content and information (Koukopoulos & Koukopoulos, 2019). By definition, *digitisation of cultural heritage* refers to the processes of systematically capturing, documenting, and re-creating cultural heritage in digital formats or media (Ocón, 2021), guided by standards and protocols (Hu et al., 2018) and producing audio-visual, structural, and contextual information (Zhang et al., 2022). In line with the

principles of preservation, digitisation of cultural heritage aims at ensuring the sustained accessibility of the digitised content (Ahmad et al., 2023), and very recently, is transitioning to a *participatory* paradigm that underscores the prominence of the audience's contributions to the digitisation process and outcomes (Tasker & Liew, 2020). Thus, this section summarises the pool of literature on cultural heritage digitisation projects that involve preservation of the cultural heritage content and access to the digitised materials, for example, in the form of an accessible database or information system (e.g., Pandey & Kumar, 2021), or a physical or virtual exhibition (e.g., Hu et al., 2021). Cases in which only digital preservation or documentation of cultural heritage were reported are excluded in this chapter.

There are several prime reasons for justifying the efforts on cultural heritage digitisation, among which the first and foremost lies in its provision of a "memory insurance policy" for cultural heritage (Ocón, 2021, p. 978). This rationale is especially important in times of tragic events, such as natural (e.g., weathering) or man-made causes (e.g., war), and unforeseen circumstances (e.g., pandemic) (Meng et al., 2023; Ginzarly & Srour, 2022). Related to the multidisciplinarity of the cultural heritage ecosystem, shifting to the realm of digitisation can bring together various stakeholders, as aforementioned, through innovative, experimental methods of preserving and disseminating cultural heritage content (Corallo et al., 2019; Kontiza et al., 2020). Next, digitisation helps record the essential aspects of cultural heritage, including its socio-historical information and sensorial characteristics (e.g., visual appearance), subsequently forming the basis for professional, academic, educational, and recreational purposes (De Bruycker & Girault, 2018; Hu et al., 2018). Thirdly, the ubiquity of digital technologies enables users of the digitised content to appreciate and experience cultural heritage anytime and anywhere at their convenience, transcending temporal and spatial boundaries (De Bruycker & Girault, 2018; Kużelewska & Tomaszuk, 2020). As the traditional preconditions such as transportation costs and opening hours hardly exist in the digital world (Ocón, 2021), digitisation can, therefore, create more equitable opportunities for accessing and experiencing cultural heritage (Pisoni et al., 2021). In the long run, the seamless availability of digitised cultural heritage will enhance the transmission and curation of collective knowledge and information among the global audiences (Koya & Chowdhury, 2020; Kużelewska & Tomaszuk, 2020). A much-needed intercultural dialogue will thence be developed, improving social inclusion and cohesion (Macrì & Cristofaro, 2021).

A recent scientometric analysis on 1,107 studies on the digitisation and datafication of cultural heritage concluded that the amount of publication in the arena of cultural heritage digitisation has surged exponentially in the past two decades of the millennium (Liao et al., 2020). Another up-to-date review (Mendoza et al., 2023) reported that, consistent with the multidisciplinary nature of cultural heritage, studies on digitising cultural heritage span across a variety of disciplines including not only Arts and Humanities (e.g., prior archaeological excavation) (Ocón, 2021), Hard Sciences (e.g., physical and chemical analyses of historic objects before 3D modelling) (Yang et al., 2022), and Information Science and Technology (e.g., creating metadata for cultural heritage information) (Hu et al., 2018),

but also Education (e.g., designing pedagogies for learning through digitising materials) (Hu et al., 2019), Management (e.g., museum operations and staffing for digitised materials) (Meng et al., 2023), and so on.

2.2 Flagship projects on digitising cultural heritage

In the scholarship of cultural heritage digitisation, a flagship project often cited is *Europeana*,[5] one of the largest portals to various digital collections of European cultural heritage (Koukopoulos & Koukopoulos, 2019). Launched in 2008, *Europeana* is an initiative of the European Union and co-funded by both the European Commission and other cultural heritage institutions, which are mostly publicly funded and operate as non-profit-making organisations (Macrì & Cristofaro, 2021). Further to offering public access to cultural heritage assets from and in European countries (Hou et al., 2022), this initiative also encourages participatory actions such as creative usage of the digitised content (Koukopoulos & Koukopoulos, 2019; Macrì & Cristofaro, 2021). On a related note, in 2018 the Association of Southeast Asian Nations (ASEAN) started the *ASEAN Cultural Heritage Digital Archive* project, aiming at raising public awareness of cultural heritage from Southeast Asia (Ocón, 2021). Another renowned project is *Open* Heritage,[6] a cross-sector collaborative effort from Google, a non-profit organisation *CyArk*, and the University of South Florida. This online platform showcases the digitised version of cultural heritage sites around the world, in the form of virtual tours and 3D models (Ocón, 2021). On the other hand, the International Council of Museums (ICOM) initiated the *Intangible Cultural Heritage and Museums* project[7] in 2017. With the objective of strengthening cultural diversity, this project leverages an information system to document and present how museums in Europe safeguard intangible cultural heritage (Hou et al., 2022; Mendoza et al., 2023). Different from these large scale, relatively well-funded digitisation projects, a notable grassroots initiative is the *Saving Ukrainian Cultural Heritage Online* (SUCHO) project,[8] which strives to digitally preserve Ukrainian cultural heritage amidst the Russian-Ukraine War (Dombrowski, 2022). These flagship projects illustrate the recent emphasis of cultural heritage digitisation projects on appreciating the diversity of cultural heritage, strategizing cross-sector collaboration, and empowering the audience's participation. In the next section, we will introduce and exemplify several common forms of cultural heritage digitisation around the world.

2.3 Forms of cultural heritage digitisation

2.3.1 3D and immersive technologies

Most of the current digitisation projects have involved the use of 3D and immersive technologies for preserving and promoting both tangible and intangible cultural heritage (Mendoza et al., 2023; Russo, 2021). For digitising immovable (e.g., monuments, buildings) and movable cultural heritage (e.g., objects, artefacts) in particular, specific technologies (e.g., laser scanning, photogrammetry, proximity sensors) are

adopted to obtain various types of data (e.g., spatial data, 3D point clouds, photos) for creating 3D models of heritage items (Yang et al., 2022). These 3D representations of cultural heritage have been applied to a variety of purposes, including historic preservation, education, and tourism (Champion & Rahaman, 2020), at times targeting lay users (e.g., tourists, elementary school students) and at other times being designated for professionals and experts (e.g., archaeologists). A well-known example is the *Digital Michelangelo* project[9] from Stanford University (Levoy et al., 2000), being one of the earliest to digitise cultural heritage items using 3D modelling and to distribute these 3D models for non-commercial use (Bruno et al., 2010).

As for immersive technologies, this chapter focuses on the applications of virtual reality (VR) and augmented reality (AR) for digitising cultural heritage, as their definitions have been much less debated when compared with mixed reality (MR) and extended reality (XR) (Bekele et al., 2018). VR is defined as a "three-dimensional, computer-generated environment" that a user can explore and interact with (Virtual Reality Society, 2017). In the cultural heritage domain, VR has been widely adopted to reproduce the heritage environment, mainly for exploration and education-related purposes (Boboc et al., 2022). In European countries, many funded projects on digitising cultural heritage involve VR technologies (Russo, 2021), for instance, the VR application named *Mona Lisa: Beyond the glass at the Louvre*[10] created by the Louvre in France (Li et al., 2022). A similar trend is also observed in Asia, where museums (e.g., China Digital Science and Technology Museum) and educators (e.g., Singapore Management University) have devoted efforts into digitising cultural heritage through VR (Ocón, 2023). On the other hand, AR is defined as the overlay of computer-generated information (e.g., text, graphics) on the real-world environment (Boboc et al., 2022). Through reliving historic events and reconstructing past states of cultural heritage, AR enhances users' in-situ experiences in the physical cultural heritage sites (Boboc et al., 2022; Russo, 2021). It has been commonly applied in museum exhibitions (e.g., National Museum of Natural History in France) (Mudička & Kapica, 2022), and also leveraged by governments and commercial organisations to promote cultural tourism through AR-enabled experience in heritage tourist attractions (Jung et al., 2018).

2.3.2 Digital galleries, libraries, archives, and museums

The advancements in digital technologies have given rise to more digitised and digital galleries, libraries, and museums (GLAMs), which realise a ubiquitous, location-aware, and personalised approach for enhancing visitor experience (Hijazi & Baharin, 2022). In addition to physical GLAMs establishing online exhibits (e.g., Metropolitan Museum of Art in New York) or digitising their collections (Li et al., 2022), there is also an increasing number of born-digital GLAMs showcasing digitised cultural heritage materials, such as the Digital Museums Canada[11] (Ocón, 2021). Some notable endeavours in Asia include Korean History Online[12] (Kim, 2019) and National Repository of Cultural Heritage in Taiwan[13] (Liu & Lin, 2021), where their interfaces are exclusive to their own native languages (i.e., Korean and Chinese, respectively).

Related to the notion of *participatory cultural heritage* (Tasker & Liew, 2020), community archives have emerged as grassroots initiatives of digital GLAMs, mostly with information work that departs from mainstream projects (Poole, 2020). Examples include *The ArQuives*,[14] an online collection of multimedia materials from and about the LGBTQ2+[15] people in Canada, the *South Asian American Digital Archive*[16] on the history and cultural heritage of South Asian Americans, and the *Singapore Memory Project*[17] that encourages citizens to contribute their heritage materials (Tasker & Liew, 2020; Poole, 2020). In the context of digitising cultural heritage, some projects are named otherwise but belong to digital GLAMs as well (Gireesh & Raman, 2022). For instance, the *South African Heritage Resources Information System*[18] offers resources on both digitised materials of cultural heritage and heritage management (Gireesh & Raman, 2022), and *HistoricalPlacesLA*[19] presents historic monuments and heritage sites in the city of Los Angeles in the United States (Barton et al., 2017).

2.3.3 Social media platforms and campaigns

Capitalising digitisation technologies, social media platforms make indispensable contributions towards online access to digitised cultural heritage content. Social media platforms are primarily designed for ubiquitous sharing of multimedia content, with features and functions for social interactions (Liang et al., 2021). Apart from social media accounts set up by GLAMs (Ginzarly, 2021; Liang et al., 2021), more grassroots initiatives on preserving cultural heritage, built on the notion of *collective memory* in particular, have proliferated (Ginzarly, 2021; Liang et al., 2021). Either way, social media has reinforced and facilitated digitally mediated practices of accessing and disseminating cultural heritage content (Ginzarly, 2021; Mendoza et al., 2023). A recent review showed that, for the purpose of crowdsourcing people-centred interpretation of cultural heritage, Facebook and Twitter have been the most popular platforms employed by cultural heritage institutions in European countries (Liang et al., 2021). Studies have also systematically analysed the grassroots, user-contributed efforts observed in social media platforms that are prevalent in mainland China, such as Douyin, Weibo, and Xiaohongshu (Liang et al., 2022). In addition, during the pandemic, the UNESCO also launched an international social media campaign[20] using the hashtags #ShareOurHeritage and #ShareCulture, in order to promote the importance of cultural heritage and encourage users of social media to share their own cultural heritage amidst the global lockdown (Ginzarly & Srour, 2022). These social media-based efforts, along with 3D and immersive technologies as well as digital GLAMs, have increasingly adopted the people-centred approach, which recognises the value of the mass public users' participation and interpretation of cultural heritage information.

In extant reviews of cultural heritage digitisation projects (e.g., Bekele et al., 2018; Mendoza et al., 2023; Yang et al., 2022), case studies and examples from China have often been reported. Nevertheless, their discussions largely focused on those in specific provinces and cities in mainland China. The situation in Hong Kong has rarely garnered interest from researchers on this topic.

3 Digitisation of cultural heritage in Hong Kong

This chapter aims to offer a summary of cultural heritage digitisation projects from various stakeholders in Hong Kong, both in the literature and in practice. To embrace the broad spectrum of initiatives and efforts, these projects are not limited to only those digitising cultural heritage in Hong Kong per se (e.g., temples located in Hong Kong). Digitisation projects that focus on cultural heritage from outside Hong Kong (e.g., mainland China) but take place predominately in local settings are also regarded as part of local efforts. Excluded from our survey are projects that are funded by local stakeholders but do not materialise in Hong Kong.

To fulfil our aim, projects and efforts reported in the literature were first identified. We conducted multiple iterative rounds of literature search on Google Scholar, using three groups of search terms connected by the Boolean operator AND: (1) "digitis*" OR "digitiz*", (2) "cultural heritage", and (3) "Hong Kong", followed by backward (i.e., references of each relevant publication) and forward (i.e., those citing each publication) citation search. Among the search results, we included English-written journal articles, conference papers, and book chapters with descriptions of Hong Kong-based cultural heritage digitisation projects. Then, for identifying these projects yet to be reported in scholarly publications, we employed the *document analysis* method that has been deemed suitable for digital humanities research (Bradley, 2016), particularly for surveying non-scholarly projects (Bowen, 2009; Finefter-Rosenbluh, 2017). Document analysis is defined as a "systematic procedure for reviewing or evaluating documents", especially those "produced, shared, and used in socially organized ways" (Bowen, 2009, p. 27). Examples of documents are wide-ranging and can be non-exhaustive, where the documents analyzed in this study included project-based websites, news articles, press releases, organisational and institutional reports, social media pages and posts, and other relevant publicly available records (e.g., official information system on public research grants) (Bowen, 2009; Sankofa, 2022). Our survey was also supplemented by the authors' professional experience in leading relevant projects on cultural heritage digitisation (e.g., Hu et al., 2018, 2020; Ng et al., 2022a), the repertoire of their pedagogical materials on digitising cultural heritage (e.g., Hu et al., 2017, 2019; Ng et al., 2022b), and personal experience. The grounded theory approach (Corbin & Strauss, 1990) was adopted to summarise information about these projects. For mitigating issues arising from interpretative subjectivity and potential biases, the authors conducted multiple rounds of reflexive discussions and maintained a neutral stance throughout the analysis (Bowen, 2009; Finefter-Rosenbluh, 2017). We acknowledge the reliance on a portion of publicly accessible documents might limit the survey and encourage further exploration and analysis of any other relevant data sources.

The following elaborates the cultural heritage digitisation projects in Hong Kong identified and analyzed in our survey. We categorise the projects based on the major contributors of cultural heritage preservation and promotion, namely government bodies, universities and schools, non-governmental and commercial organisations, cross-sector collaboration, and grassroots projects and efforts. This

categorisation aligns with the various stakeholders of cultural heritage management in a framework validated in the Chinese context (Du Cros & Yok-shiu, 2007). We also present the links to the projects' official websites or the deliverables created for users' access to the digitised materials.

3.1 Government-led projects

Table 7.1 summarises the projects led by governmental bodies and public museums. In Hong Kong, the Antiquities and Monuments Office[21] (AMO), which is administratively under the Development Bureau of the Hong Kong S.A.R. Government, takes charge of the preservation and promotion of immovable (e.g., historic buildings) and movable cultural heritage (e.g., excavated objects) (Ho & Hou, 2019). It has led three initiatives of cultural heritage digitisation. First is the *Hong Kong Archaeological Archive System* [1] (identifiers in tables, thereafter) that presents information of local archaeological discoveries (e.g., pottery, ceramics) online to the public. Second, with input from the Lands Department, the AMO also maintains the *Geographical Information System on Hong Kong Heritage* [2], which presents the computerised information on the local archaeological and built heritage for fostering public awareness and promoting cultural exchanges. Third, as a joint effort with the Commissioner for Heritage's Office, the AMO organised the public physical exhibition "In Virtual of Heritage: Exploring Monuments with Heritage Technology" [3] in 2022 where several historic buildings were digitally presented to visitors in immersive and interactive displays (Figure 7.1).

The Leisure and Cultural Services Department (LCSD), formerly administratively supervising AMO, has also led several different kinds of initiatives. For standardising and centralising the management of public museum collections, the LCSD has established the *LCSD Museum Collection Management System* [4] for enabling visitors to access to information of cultural heritage items across all public museums in Hong Kong. A similar initiative from the LCSD is the *Hong Kong Intangible Cultural Heritage Database* [5], which aims at allowing the public to appreciate local intangible cultural heritage items. A noteworthy effort, as part of a long-term exhibition titled "A City of Thousand Faces" [6], is an AR-enhanced exhibition inside the Former Yamen Buildings of Kowloon Walled City (Figure 7.2), which immerses visitors in the digitally re-created historic environment of the heritage site. In response to the COVID-19-induced lockdown measures in 2020, as part of the *Muse Fest HK 2020*, the LCSD also created a series of virtual tours of several cultural heritage sites (e.g., Old Tai Po Market Station) and institutions (e.g., Hong Kong Film Archive) [7], for engaging the audience in appreciating cultural heritage from home.

Interestingly, the *City in Time* project [8] led by the Tourism Commission,[22] whose mission is not primarily focused on cultural heritage preservation, was discussed in a series of case studies on how digitisation supported cultural heritage tourism in Asia (Ocón, 2022). In this project, aiming to "bolster the appeal [of post-pandemic Hong Kong] ahead of the return of foreign visitors" (Ocón, 2022, p. 173), augmented reality (AR) experience has been created for users to learn

Table 7.1 Overview of cultural heritage digitisation projects led by governmental bodies and public museums in Hong Kong

No.	*Name of project*	*Project leader(s)*	*Starting year*	*Digitised cultural heritage*	*Technologies adopted*	*Source*
[1]	*Hong Kong Archaeological Archive System* https://hkaas.amo.gov.hk/	Antiquities and Monuments Office	Unknown	Archaeological discoveries	Batch web publishing	Document analysis
[2]	*Geographical Information System on Hong Kong Heritage* https://gish.amo.gov.hk/internet/index.html?lang=en-us	AMO	2005	Local archaeological and built heritage	Geographical information system; Batch web publishing	Document analysis
[3]	*In Virtual of Heritage: Exploring Monuments with Heritage Technology*	Commissioner for Heritage's Office and AMO	2022	Historic buildings	Interactive display; Virtual reality	Document analysis
[4]	*LCSD Museum Collection Management System* https://mcms.lcsd.gov.hk/Search/search/enquire?&request_locale=en	Leisure and Cultural Services Department	2019	Public museum collections	Batch web publishing	Document analysis
[5]	*Hong Kong Intangible Cultural Heritage Database (HKICH Database)* https://www.hkichdb.gov.hk/en/index.html	LCSD	2014	Intangible cultural heritage items	Web publishing	Document analysis
[6]	*A City of Thousand Faces* https://www.lcsd.gov.hk/en/parks/kwcp/thousand.html#03	LCSD	2017	Former Yamen Buildings of Kowloon Walled City	Augmented reality	Document analysis
[7]	*Muse Fest HK 2020,* featuring 360-degree interactive virtual tours https://www.museums.gov.hk/en/web/portal/mf2020_virtualtour.html	LCSD and public museums in Hong Kong	2020	Several cultural heritage sites	Virtual reality; Web publishing	Document analysis

(*Continued*)

Table 7.1 (Continued)

No.	*Name of project*	*Project leader(s)*	*Starting year*	*Digitised cultural heritage*	*Technologies adopted*	*Source*
[8]	*City in Time* https://www.cityintime.hk/en/	Tourism Commission	2021 or earlier	Locations of historic interest	Augmented Reality (AR); Mobile app	(Ocón, 2022)
[9]	Digitised holdings from *Government Records Service* https://www.grs.gov.hk/en/online_holdings.html?date=8-2023	Government Records Service	Unknown	Physical holdings of documentary heritage	Web publishing	Document analysis
[10]	*Virtual Tour of Ex-Sham Shui Po Service Reservoir* https://www.wsd.gov.hk/VirtualTour/index.html	Leisure and Cultural Services Department	2021	Ex-Sham Shui Po Service Reservoir	Virtual reality; Web publishing	Document analysis
[11]	*Neonsigns.hk* https://www.neonsigns.hk/	M+ Museum	2014 or earlier	Neon signs and their designs	Web publishing	(Chen, 2020)

Figure 7.1 Interactive displays in the "In Virtual of Heritage: Exploring Monuments with Heritage Technology" exhibition (2022). Photograph from the authors.

Figure 7.2 AR exhibition in the Former Yamen Buildings of Kowloon Walled City. Photograph from the authors.

about and appreciate the past and present views of various designated locations of historic interest. There are two initiatives from other government bodies. The Government Records Service, striving to preserve and promote documentary heritage

of Hong Kong, has built a digital archive of part of its holdings [9], such as textual records and physical posters, aiming to cultivate public awareness, appreciation, and proper usage of these heritages. Another interesting case is an effort from the Water Supplies Department. Despite that it is not directly responsible for preserving and promoting cultural heritage, it took charge of producing the "Virtual Tour of Ex-Sham Shui Po Service Reservoir" [10], where the reservoir served as the main source of water supply to the district. The objective of this virtual tour was to facilitate the public consultation of a proposed grading decision of the said cultural heritage site, which was not open to the public during the consultation period.

Involving citizens' participation was also a strategy for government-funded projects. For illustrating the sustainability of digitised cultural heritage, Chen (2020) introduced the online, crowdsourced exhibition *neosigns.hk* [11] from the M+ Museum, which showcased user-submitted photos, videos, and information of neon signs and their designs in Hong Kong.

3.2 Projects led by universities and schools

Both the literature and results of our document analysis contain multiple university-led projects on digitising cultural heritage based in Hong Kong, as summarised in Table 7.2. Early on, two engineering researchers from the City University of Hong Kong (CityU) (Chow & Chan, 2009) reported the procedures of constructing 3D models of ceramic tea ware artefacts from the Flagstaff House Museum of Teaware[23] in Hong Kong [12], towards creating an "interactive virtual exhibition" (p. 161). A few cultural heritage digitisation initiatives stem from externally (e.g., public research grants) or internally funded research projects. As a major output of a study on the historical development of early Cantonese-written biblical content and Christian literature [13], a linguistics research team from the Education University of Hong Kong (EdUHK) built a database to document these local intangible cultural heritages for both professionals (e.g., historians, theologists) and amateurs. Also directly arising from a research study, a research centre specialised in classical Chinese poetry from the Chinese University of Hong Kong (CUHK) aimed to preserve the different schools and genres of Cantonese chanting, as an endangered intangible heritage, by building the *Archive of 20th Century Cantonese Chanting in Hong Kong* [14]. Another scholarly effort, presented as a digital humanities project led by the University of Hong Kong (HKU), produced the *Hong Kong Historical GIS (1900–1933)* [15]. It is a geographical information system (GIS) visualising the spatial distributions of commercial entities and activities in Hong Kong during a specific historic period of time. As briefly mentioned in the literature (Allsop, 2022), the *Hong Kong Memory* [16] is a project formerly from the Centre of Asian Studies at the HKU, commissioned by the Hong Kong Jockey Club[24] and now being managed by the LCSD. As a response to the UNESCO's Memory of the World Programme, its corresponding online platform *Hong Kong Memory* houses multiple digital collections and exhibitions of local cultural heritages. More recently, the Centre for Studies of Daoist Culture at the Chinese University of Hong Kong developed

Table 7.2 Overview of cultural heritage digitisation projects led by universities and schools in Hong Kong

No.	*Name of project*	*Project leader(s)*	*Starting year*	*Digitised cultural heritage*	*Technologies adopted*	*Source*
[12]	Untitled	City University of Hong Kong	2009 or earlier	Ceramic tea ware artefacts	3D modelling	Chow and Chan (2009)
[13]	*The Database of Early Cantonese Bible and Christian Literature* https://corpus.eduhk.hk/cantobible/index.html	The Education University of Hong Kong	2020	Literature	Web publishing	Document analysis
[14]	*Archive of 20th Century Cantonese Chanting in Hong Kong* https://dsprojects.lib.cuhk.edu.hk/en/projects/20th-cantonese-poetry-chanting/home/	The Chinese University of Hong Kong Library	2020	Cantonese chanting	Web publishing	Document analysis
[15]	*Hong Kong Historical GIS (1900–1933)* https://hkh-gis.lib.hku.hk/	The University of Hong Kong	Unknown	Commercial entities and activities in a historical period	Geographical information system; Batch web publishing	Document analysis
[16]	*Hong Kong Memory* http://hkmemory.hk/	University of Hong Kong	2011	Local cultural heritage items	Web publishing	Allsop (2022)
[17]	*Daoist Digital Museum* https://dao.crs.cuhk.edu.hk/digitalmuseum/CH/	Chinese University of Hong Kong	2015 or earlier	Chinese temples and Daoist literary works	Web publishing	Kim (2019)
[18]	*Buddhaverse: Virtual Dunhuang* https://arthistory.hku.hk/index.php/event/buddhaverse-virtual-dunhuang-hku-art-history/)	HKU	2023	Mogao caves	Virtual reality	Document analysis

(*Continued*)

Table 7.2 (Continued)

No.	*Name of project*	*Project leader(s)*	*Starting year*	*Digitised cultural heritage*	*Technologies adopted*	*Source*
[19]	*DigitalRepository@HKU* https://digitalrepository.lib.hku.hk/	HKU University Libraries	Unknown	Any kinds of cultural heritage	Batch web publishing	Document analysis
[20]	*Hong Kong Oral History Archives* https://sunzi.lib.hku.hk/hkoh/	HKU University Libraries and HKU Centre of Asian Studies	2001	Oral history	Batch web publishing	Document analysis
[21]	*Hong Kong Early Tabloid Newspapers* https://repository.lib.cuhk.edu.hk/en/collection/cuhk-hk-tabloid	CUHK Library	2014	Newspapers	Batch web publishing	Document analysis
[22]	*Hong Kong Education Image Database* https://imagedb.museum.eduhk.hk/	Hong Kong Museum of Education and EdUHK Library	Unknown	Educational heritage	Batch web publishing	Document analysis
[23]	Student-created 3D models of lighthouses	CityU University Library	2016	Lighthouses	3D modelling	Ching (2018)
[24a]	Student-created digital galleries of cultural heritage https://commoncore.hku.hk/slf2021/slf2021-ccch9051/	HKU	2021	Any kinds of cultural heritage	Web publishing	Ng et al. (2022b)
[24b]	Student-created virtual reality stories of cultural heritage https://lavrplatform.com/discover				VR; Web publishing	Hu et al. (2019)
[25]	Student-created 3D models of cultural heritage https://cchu9080.lib.hku.hk/	HKU	2021	Museum artefacts	3D modelling; Web publishing	Document analysis

[26]	*Catalogues of St. John's College Archives* https://archives-sjchku.com/	HKU St. John's College	2016	Relics and historic documents	Batch web publishing	Document analysis
[27]	*Rhizome of the Western Han* https://www.jeffreyshawcompendium.com/portfolio/rhizome-of-the-western-han/	City University of Hong Kong	2010	Rhizome of the Western Han	Interactive visualisation	Kenderdine et al. (2012)
[28]	*Pure Land AR* https://www.jeffreyshawcompendium.com/portfolio/pure-land-ar/		2012	Dunhuang caves	AR	Kim (2019)
[29]	*Gifts from Lanmama* https://lanmama.lib.hku.hk/	University of Hong Kong	2019	Embroidery from Guizhou	Web publishing	Hu et al. (2021)
[30]	*Habits and Haberdashery — Uncovering History and Heritage in the Hidden* https://foundation.mcs.hk/virtual-tour-habits-and-haberdashery-uncovering-history-and-heritage-in-the-hidden-attic/	Maryknoll Convent School	2020	Historic materials and architecture	Virtual reality; Web publishing	Document analysis

the *Daoist Digital Museum* [17], which provides "historical, epigraphic and geographic data" about 125 selected Chinese temples in Hong Kong, and also hosts a digital archive of classical Daoist literary works (Kim, 2019, p. 27). This year, HKU Art History organised a one-off exhibition on campus, *Buddhaverse: Virtual Dunhuang* [18], in which participants could explore four different Mogao caves using virtual reality headsets.

Apart from academic units, libraries and museums operating under the management of universities also contribute towards digitising cultural heritage. Academic departments in universities often seek collaboration and support from academic libraries and museums to transfer knowledge created in their research into permanent digital resources hosted and managed by these within-university GLAM units. The HKU University Libraries set up the *DigitalRepository@HKU* [19] for providing users with the access to digitised materials resulting from projects partnered with them, including both tangible (e.g., physical photos) and intangible (e.g., oral history) cultural heritage. Digitising a particular type of cultural heritage, the HKU University Libraries and its Centre of Asian Studies jointly created the *Hong Kong Oral History Archives* [20], as a systematic effort of archiving oral materials related to the history of Hong Kong. Focusing on journalistic literature, the CUHK Library houses a digital collection named *Hong Kong Early Tabloid Newspapers* [21] for opening the access to and promoting academic research of this less common form of cultural heritage. Also being a theme-based effort, the Hong Kong Museum of Education in EdUHK has collaborated with the EdUHK Library in curating the *Hong Kong Education Image Database* [22] for showcasing digitised collection of educational heritage in Hong Kong, ranging from covers of historic textbooks to photos of traditional classrooms.

In essence, universities also serve as institutions that empower and engage students in their learning and extra-curricular experience. Thus, there have also been initiatives involving university students as the contributors of cultural heritage digitisation. In a project from the University Library at CityU [23], with the aim to engage students with archival collections, they organised a service-learning programme in which students learned how to present lighthouses in Hong Kong, as a kind of immovable cultural heritage, through different media including 3D models (Ching, 2018). In a general education course titled "Digitising Cultural Heritage in Greater China" offered at HKU, students' capstone assignment is to create public-facing digital galleries [24a] (Ng et al., 2022b) and virtual reality (VR) stories [24b] (Hu et al., 2019) for presenting self-selected topics of cultural heritage, where some of these student works are featured in student festivals and online exhibitions (Figure 7.3). Notably, a self-developed VR authoring platform was developed for facilitating students' VR-based digitisation works (Wang et al., 2022). Adopting a similar rationale, another general education course themed archaeology at HKU involves students in creating 3D models of cultural heritage items from the University Museum and Art Gallery, then showcasing these student creations in a digital platform hosted by the University Libraries [25]. As part of a

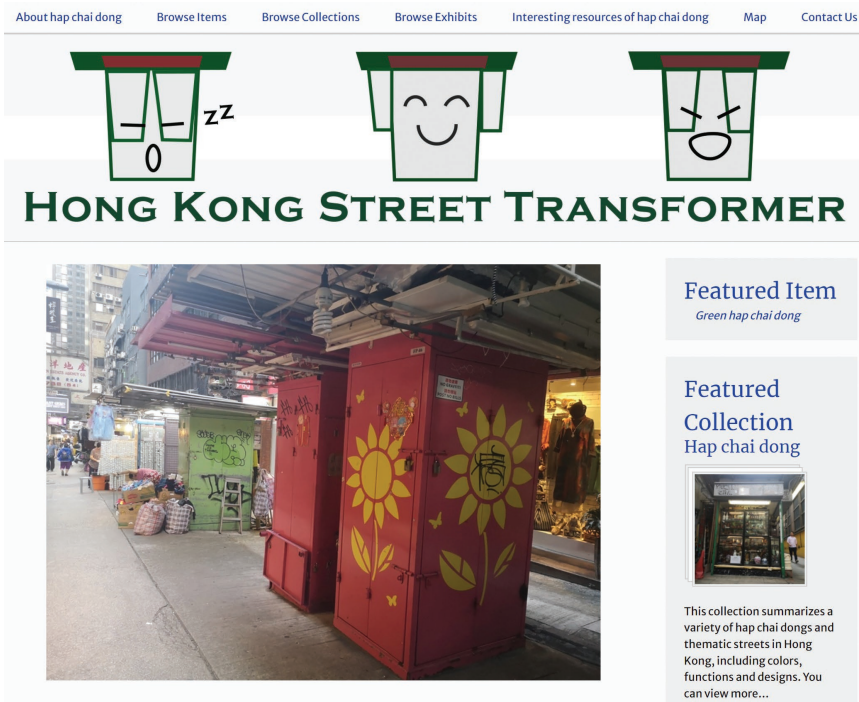

Figure 7.3 Screenshot of a HKU student-created digital gallery of local cultural heritage.

bigger project, "Retracing St. John's Project", students and alumni of the HKU St. John's College have been preserving relics and documents from and in the college in their physical archives, and then established the *Online Catalogues of St. John's College Archives* [26].

Notably, CityU has also been active in digitising cultural heritage from China as well, including the development of an interactive visualisation of the Rhizome of the Western Han [27] (Kenderdine et al., 2012) and an AR experience of Mogao Grottoes from Dunhuang [28] (Kim, 2019). Also contributing towards cultural heritage from mainland China, the Faculty of Education at the University of Hong Kong (HKU) collaborated with the HKU University Libraries and a private museum in Guizhou, China, on a knowledge exchange project, creating a digital museum named *Gifts from Lanmama* [29] for showcasing the unique intangible cultural heritage of embroidery from the ethnic minorities in Guizhou (Hu et al., 2021).

Apart from universities, in Hong Kong the sites and buildings where a number of K-12 schools are located have also been officially protected as immovable cultural heritage. One of these schools is the Maryknoll Convent School (Figure 7.4). This historic school, through its company, led the project "Habits and Haberdashery—Uncovering History and Heritage in the Hidden" [30], for conserving the

Figure 7.4 The historic architecture of the Maryknoll Convent School in Hong Kong. Photograph by the authors.

historic materials uncovered in an attic within the school's campus. A virtual tour was produced for guiding the public audience to appreciate the historical and architectural significance of the school and its related heritage.

3.3 Projects led by non-governmental and commercial organisations

Various non-governmental organisations (NGOs) in Hong Kong have been devoting their time and resources to digitising cultural heritage (Table 7.3). As one of the major NGOs dedicated to cultural heritage promotion in Hong Kong, the Conservancy Association Centre for Heritage (CACHe) conducted the project "Listen to Heritage – 100 Audio Stories of Hong Kong" [31], developing audio recordings for introducing 100 historic buildings in Hong Kong to the public through online means. Another large-scale NGO in Hong Kong, Caritas, led the project "A Glimpse Through Chinese Mosaic Windows: The Bridal Laments of the Last Walled Villages Brides" [32], aiming to preserve wedding-related cultural heritage in Lung Yeuk Tau, a historic local village in Hong Kong. In this project, they recorded and published online a series of oral history and singing from bridal laments, as an oral tradition. Independent of governmental support, the Hong Kong Museum of Medical Sciences (HKMMS) is run by the NGO named HKMMS Society. The Society exhibits the *Hong Kong Museum of Medical Sciences SARS Oral History Archive* [33], presenting online excerpts of oral history about the SARS epidemic in Hong Kong.

There are also NGOs operating at a small scale, which are often committed to particular groups of cultural heritage. For instance, the Pokfulam Village Cultural Landscape Conservation Limited, a company focused on heritage conservation in Pokfulam, a historic district in Hong Kong, worked on a digitisation project "Integrating and Sharing of Pokfulam Village Community Archives" [34]. They created a digital asset management system for the sustainable storage and use of cultural

Table 7.3 Overview of cultural heritage digitisation projects led by non-governmental and commercial organisations in Hong Kong

No.	*Name of project*	*Project leader(s)*	*Starting year*	*Digitised cultural heritage*	*Technologies adopted*	*Source*
[31]	Listen to Heritage – 100 Audio Stories of Hong Kong http://cache.org.hk/blog/heritage_100/	Conservancy Association Centre for Heritage	2020	Historic buildings	Web publishing	Document analysis
[32]	A Glimpse Through Chinese Mosaic Windows: The Bridal Laments of the Last Walled Villages Brides https://www.lordwilson-heritagetrust.org.hk/en/projects/project_listing/project_details/27-9-87.html	Caritas	2016	Cultural heritage in a local community	Web publishing	Document analysis
[33]	Hong Kong Museum of Medical Sciences SARS Oral History Archive https://www.hkmms.org.hk/en/sars-oral-history-archive/	Hong Kong Museum of Medical Sciences Society	2013	Oral history	Web publishing	Document analysis
[34]	Integrating and Sharing of Pokfulam Village Community Archives https://www.lordwilson-heritagetrust.org.hk/en/projects/project_listing/project_details/27-9-91-1.html	Pokfulam Village Cultural Landscape Conservation Limited	2018	Cultural heritage in a historic village	Web publishing	Document analysis
[35]	Documentation of general lexical items in four indigenous languages in Hong Kong http://www.hkilang.org/v2/%e7%99%bc%e9%9f%b3%e5%ad%97%e5%85%b8/	Association for Conservation of Hong Kong Indigenous Languages	2009	Indigenous languages	Batch web publishing	Document analysis
[36]	Promotion Project of the Historical Archives of Anti-Japanese War in Hong Kong https://joint2net.org/hkcf.html	Hong Kong Champion Family	2012	Oral history	Web publishing	Document analysis
[37]	Hong Kong Heritage Project https://www.hongkongheritage.org/	Kadoorie	2007	Business history	Batch web publishing	Allsop (2022)

heritage information about a historic village in this district. Next, the Association for Conservation of Hong Kong Indigenous Languages works on various indigenous languages in Hong Kong, some of which have been statutorily recognised as official intangible cultural heritage items. This association led the project "Documentation of general lexical items in four indigenous languages in Hong Kong" [35], compiling a database for preserving such intangible heritage and promoting them to the younger generation. Another NGO named the Hong Kong Champion Family, formed by a group of veterans who participated in the Anti-Japanese War in Hong Kong, initiated the "Promotion Project of the Historical Archives of Anti-Japanese War in Hong Kong" [36]. In this initiative, part of their deliverables was the establishment of an online archive of oral history relating to the project theme, aiming to let the public audience better understand this historic event.

As an initiative from a commercial organisation, the *Hong Kong Heritage Project* [37] was established in 2007 by Sir Michael Kadoorie, part of the Kadoorie family with wide-ranging business investments in Hong Kong (Allsop, 2022). In this project, a corporate archive of oral history was built to "preserve Kadoorie business history" (p. 363) and "inspire staff" (p. 364), with a portion of the catalogue available online for engaging external audiences (Allsop, 2022).

3.4 Cross-sector collaboration

It is also common for various parties, especially government bodies, GLAMs, and NGOs, to collaborate in running cultural heritage digitisation projects (Ocón, 2021), as listed in Table 7.4. Fung Ying Seen Koon, being one of the several massive Buddhist temples in Hong Kong, collaborated with the Intangible Cultural Heritage Office of the government and other NGOs including Hong Kong Taoist Association and Guangzhou Sanyuan Palace in mainland China on a cultural heritage digitisation project. Aiming for preserving heritage through modern technologies, this project built the *E-database of Hong Kong Quanzhen Temples Taoist Ritual Music* [38]. Another noteworthy example is the "HKACT! ACT 1 BeHere" project [39] commissioned by the Design District Hong Kong, organised by the Hong Kong Design Centre, and co-produced by the Osage Art Foundation. With the goals of promoting the communal history of Wanchai, a historic district in Hong Kong, this project encouraged public participants to explore the district through mobile app-enabled AR experience (Figure 7.5).

Another example is the 2016 exhibition, *300 Years of Hakka Kungfu – Digital Vision of Its Legacy and Future* [40], co-organised by the Intangible Cultural Heritage Office, International Guoshu Association, and CityU, at the Hong Kong Heritage Museum (Lo et al., 2019). This exhibition has been cited in the literature as an application of the digitisation and visualisation of intangible cultural heritage (Guidi & Frischer, 2020). Another exhibition [41], co-created by the Hong Kong Heritage Museum and the Dunhuang Academy, showcased the digital recreation of Dunhuang caves and murals (Figure 7.6), aiming to let visitors explore these cultural heritages and learn about their significance (Song, 2023).

Table 7.4 Overview of cross-sector collaborative projects on cultural heritage digitisation in Hong Kong

No.	*Name of project*	*Project leader(s)*	*Starting year*	*Digitised cultural heritage*	*Technologies adopted*	*Source*
[38]	*E-database of Hong Kong Quanzhen Temples Taoist Ritual Music* https://daoistmusichk.org/zh-hant/ritual-music-viewing	Fung Ying Seen Koon	2015	Music	Web publishing	Document analysis
[39]	*HKACT! ACT 1 BeHere* https://ddhk.designdistrict.hk/en/programme/	Hong Kong Design Centre	2018	Communal history	Augmented reality; Mobile App	Document analysis
[40]	*300 Years of Hakka Kungfu – Digital Vision of Its Legacy and Future* https://hk.heritage.museum/en/web/hm/exhibitions/data/exid238.html	Intangible Cultural Heritage Office, International Guoshu Association, and CityU	2016 or earlier	Kungfu	3D modeling	Ocón (2021)
[41]	*The Hong Kong Jockey Club Series: Digital Dunhuang*	Hong Kong Heritage Museum and the Dunhuang Academy	2018 or earlier	Dunhuang caves	3D animation; Interactive displays	Song (2023)

Figure 7.5 A participant with an AR-superimposed scenario of the old-styled haircut service in Wanchai, Hong Kong. Photograph by the authors.

3.5 Grassroots projects and efforts

Complementing the digitisation efforts from these more common stakeholders of cultural heritage in Hong Kong, the works from other co-contributors have also been indispensable. As shown in Table 7.5, there are several cases of grassroots projects from special interest groups (SIGs) and individuals. The Hong Kong Bird Watching Society conducted the project "Hong Kong's Got Fishpond – Aquaculture of Freshwater Fish in Hong Kong History" and digitised the oral history from fishermen who practiced freshwater fishing in Hong Kong [42]. A group of hobbyists, the Hong Kong Underwater Heritage Group, undertook the project "Surveying

Figure 7.6 Interactive displays of Dunhuang murals in Hong Kong Heritage Museum. Photograph by the authors.

and Documenting the Underwater Heritage of Hong Kong", compiling a database [43] that shows the locations and relevant information of shipwrecks and other underwater heritage sites in Hong Kong. Another group, with a team of doctoral researchers and undergraduate students, developed the *Bilingual Database and Annotated Bibliography of Cantonese Popular Periodicals of the Early Twentieth Century* [44], for showcasing such local cultural heritage materials stored in Hong Kong and other places.

Over the years, there have also been a number of projects from teams not affiliated with any organisations or societies. A team of independent researchers has established the digital platform *Hong Kong Stele Inscriptions* [45], curating photos and information about steles and their inscriptions in Hong Kong, specifically for supporting academic research and history education. Starting from a British expatriate residing in Hong Kong, *Gwulo.com* [46] is an online, crowd-sourced community archive of historic stories and information about Hong Kong (Poole, 2020). With a specific theme, the *Hong Kong War Diary* [47] documents photos and information about "the 1941 defence of Hong Kong by members of the garrison" (Affleck & Kvan, 2008, p. 270).

Another form of grassroots efforts on cultural heritage digitisation is the use of social media accounts and pages. Since it is commonplace for cultural heritage digitisation projects to employ social media for publicity purposes (Ginzarly, 2021; Liang et al., 2021), our survey focuses on stand-alone social media accounts and pages for preserving and promoting cultural heritage in Hong Kong. As shown in Table 7.6, among the popular social media platforms in Hong Kong, we identified a number of Facebook, Instagram, and Twitter groups, pages, and accounts to be related to cultural heritage preservation and promotion.

Table 7.5 Overview of grassroots projects and efforts on cultural heritage digitisation in Hong Kong, other than social media campaigns

No.	*Name of project*	*Project leader(s)*	*Starting year*	*Digitised cultural heritage*	*Technologies adopted*	*Source*
[42]	*Hong Kong's Got Fishpond - Aquaculture of Freshwater Fish in Hong Kong History* http://www.hkbws.org.hk/fishpondoralhistory/index.php/oral_history/	Hong Kong Bird Watching Society	2015	Oral history	Web publishing	Document analysis
[43]	*Surveying and Documenting the Underwater Heritage of Hong Kong* http://www.hkuhgroup.com/cms/index.php?option=com_content&view=category&id=18&Itemid=14&lang=en	Hong Kong Underwater Heritage Group	2009	Underwater heritage	Batch web publishing	Document analysis
[44]	*Bilingual Database and Annotated Bibliography of Cantonese Popular Periodicals of the Early Twentieth Century* https://cantonpp.com/en	A team of doctoral researchers and students	2020	Periodicals	Batch web publishing	Document analysis
[45]	*Hong Kong Stele Inscriptions* https://www.hkinscriptions.com/	A team of independent researchers	Unknown	Steles and inscriptions	Web publishing	Document analysis
[46]	*Gwulo.com* https://gwulo.com/	A British expatriate	2009	Community-based history	Web publishing	Poole (2020)
[47]	*Hong Kong War Diary* www.hongkongwardiary.com	Members of a garrison	Unknown	A historic war	Web publishing	Affleck and Kvan (2008)

Table 7.6 Overview of social media campaigns on cultural heritage preservation and promotion in Hong Kong

No.	*Name of group/page/account*	*Type*	*Link*
[48]	*@hongkongoldfans*	Instagram page	https://www.instagram.com/hongkongoldfans/
[49]	*Retro Gallery of Hong Kong in the 1950s, 1960s, and 1970s*	Facebook group	https://www.facebook.com/groups/454479074672446
[50]	Old Hong Kong Lovers	Facebook page	https://www.facebook.com/OldHongKongLovers/about
[51]	@OldHKinColour	Twitter account	https://twitter.com/OldHKinColour
[52]	Hong Kong Folk Memories Database	Instagram page	https://www.instagram.com/vcrbase/
[53]	Hong Kong &Macau old postcard, photo, slide, film	Facebook group	https://www.facebook.com/groups/391916297636331
[54]	@ancient_hk_stories	Instagram page	https://www.instagram.com/ancient_hk_stories/
[55]	@hk.localstoriesm	Instagram page	https://www.instagram.com/hk.localstoriesm/
[56]	Kongcept	Instagram page	https://www.instagram.com/kongcept852/
[57]	Built Heritage Hong Kong	Instagram page	https://www.instagram.com/built_heritage_hongkong/
[58]	Hong Kong Corner Houses	Instagram page	https://www.instagram.com/hkcornerhouses/
[59]	History of HK Shopping Malls	Facebook page	https://www.facebook.com/ShoppingMalls.hk
[60]	Hong Kong Heritage & Military Falls Endeavour	Facebook group	https://www.facebook.com/groups/HKHerMiliFallsEndeavor
[61]	Filing Archives and Architectural Research	Instagram page	https://www.instagram.com/hkfaar_/
[62]	Motoring Archive of Hong Kong	Facebook page	https://www.facebook.com/motoringarchivehk
[63]	Database of History of Canoeing	Facebook page	https://www.facebook.com/hongkongcanoeinghistory
[64]	History of Waterworks in Hong Kong	Facebook group	https://www.facebook.com/groups/1188839834474601/
[65]	HK Catholic History Research	Facebook group	https://www.facebook.com/groups/hkcatholichistory/
[66]	Hong Kong Community Arts Archive Catalogue	Facebook page	https://www.facebook.com/hkcommunityartarchive
[67]	@hkurbanfont	Instagram page	https://www.instagram.com/hkurbanfont/
[68]	@hkfont	Instagram page	https://www.instagram.com/hkfont/
[69]	Protecting HK Names	Instagram page	https://www.instagram.com/protecthknames/

A few of these pages are largely hinged on the notions of nostalgia and collective memories, including *@hongkongoldfans* [48] aiming to document traditional local heritage items in general, the era-based *Retro Gallery of Hong Kong in the 1950s, 1960s, and 1970s* [49], *Old Hong Kong Lovers* [50] focused on local cultural heritage items that can evoke Hongkongers' collective memories, and *Hong Kong Folk Memories Database* [52] that digitises Video Home System (VHS) tapes into videos. Interestingly, there is a Twitter account *@OldHKinColour* [51] presenting artificial intelligence (AI) colorised refurnished old photos of historic landscapes in Hong Kong. Also built upon nostalgia, a Facebook group titled *Hong Kong & Macau old postcard, photo, slide, film* [53] is aimed for participants to both share and exchange ephemeral items (i.e., movable cultural heritage) such as postcards and physical slides. More inclined towards the locality of cultural heritage, the Instagram pages *@ancient_hk_stories* [54] and *@hk.localstoriesm* [55] share cultural heritage contents through a story-telling narrative. A similar page, named *Kongcept* [56], specifies cultural heritage information related to architecture, music, craftsmanship, and community-based items.

The other social media initiatives are primarily focused on a particular theme of cultural heritage. For immovable cultural heritage, in addition to the titular Instagram pages *Built Heritage Hong Kong* [57] and *Hong Kong Corner Houses* [58], the Facebook page *History of HK Shopping Malls* [59] aims to document different aspects of historically significant shopping malls, whereas the Facebook group *Hong Kong Heritage & Military Falls Endeavour* [60] revolves around archaeological ruins of military sites. Related to these initiatives, the page *Filing Archives and Architectural Research* [61] presents archived materials about post-war architecture in Hong Kong. Interestingly, there are several pages about transportation and sports-related cultural heritage, namely *Motoring Archive of Hong Kong* [62] on the history and cultures of motoring and the *Database of History of Canoeing* [63]. Other unique groups include *History of Waterworks in Hong Kong* [64] focusing on heritages related to waterworks (e.g., water pumping stations), and *HK Catholic History Research* [65] for enthusiasts of Catholic history and heritage. As for movable and intangible cultural heritage, the *Hong Kong Community Arts Archive Catalogue* [66] page showcases community-based artworks, while the pages *@hkurbanfont* [67] and *@hkfont* [68] present fonts discovered in public venues (e.g., streets, shops). Another page *Protecting HK Names* [69] aims to preserve and revitalise Hong Kong toponymy.[25]

4 Discussion

In this section, we summarise the approaches adopted in these cultural heritage digitisation projects in Hong Kong, followed by the development trends across these projects. The facilitating factors and the limitations and challenges will then be discussed. For easy cross-referencing within this chapter, the identifying numbers (e.g., [1], [2]) in Tables 7.1–7.6 along with the abbreviation of project names (e.g., GIS on HK Heritage) will be used when the corresponding projects are mentioned as examples.

4.1 Approaches of the projects

4.1.1 Multi-layered objectives of digitising cultural heritage

While digitising cultural heritage serves the dual purposes of digitally preserving and promoting cultural heritage (Mendoza et al., 2023; Russo, 2021), multiple layers of objectives have been identified among the cultural heritage digitisation projects in Hong Kong. Many projects have explicitly articulated their goal or mission to preserve or document a theme or type of cultural heritage (e.g., [13] on Cantonese Christian literature, [30] on heritage from a local school), often attributed to the lack of systematic efforts in its preservation (e.g., canoeing [63], local names [69]) or it being endangered and at risk of disappearance (e.g., early Cantonese chanting [14], local stories [54]). While digitisation efforts are at times subsidiary to or only part of bigger projects on the compilation or documentation of cultural heritage information (e.g., [13], [30], [35 on local indigenous languages]), digitising cultural heritage or the use of digital technologies has also been strategized as one of the major goals (e.g., virtual tours [7], community archives [34], e-database [38], @ hongkongoldfans [48]).

Most of the projects have also shared their goals of promoting a theme of cultural heritage, through various ways of disseminating the digitised content, for example, enabling users' first-hand experience or appreciation of the cultural heritage items (e.g., HKICH Database [5], AR experience [6][8], audio stories [31]) and encouraging researchers, educators, and students to leverage the digitised materials for academic research, teaching, and learning (e.g., Government Records Services [9], [35], HKInscriptions [45]). A few projects have emphasised the need for improving access to specific themes (e.g., Muse Fest HK [7], Cantonese periodicals [44]), and one has highlighted their aim for fostering international cultural exchange (e.g., GIS on HK Heritage [2]). The majority of these projects are intended for the general public as their target audience, with one project directly aiming for the younger generation [35]. About half of the projects also have targeted academics and researchers (e.g., HKAAS [1], GIS [2] [15], [14], [44]) and amateurs (e.g., [21] on historic newspapers, [43] on underwater heritage). In particular, a few projects have specified more detailed target beneficiaries of their projects, such as linguists and theologists [13], professionals working in cultural heritage institutions [2], university staff and students (e.g., Buddhaverse [18], DigitalRepository@HKU [19]), and Chinese and international visitors [48].

In addition to directly aiming for cultural heritage preservation and promotion, numerous projects have spelled out alternative goals, which can be regarded as contributing towards this dual purpose indirectly. For example, several projects act as responses or solutions to international programmes (e.g., UNESCO Memory of the World [5]), public policies (e.g., [2], [5]), museum management strategies (e.g., LCSD Museum Collection Management System [4]), and preservation-related decision-making (e.g., virtual tour of a historic reservoir [10]) in Hong Kong, as well as temporary measures (e.g., lockdown) (e.g., Muse Fest HK [7]).

4.1.2 Funding sources and modes of collaboration

There are several funding sources for conducting cultural heritage digitisation projects in Hong Kong. As expected, projects led by government bodies and public museums have all been publicly funded. Next, around one-third of all surveyed projects have been financially supported by the *Lord Wilson Heritage Trust*, a yearly public grant established in 1992 for local cultural heritage preservation and conservation projects. The holders of these projects funded by this Trust are often academic researchers from universities, independent researchers, NGOs, schools, and SIGs. A few projects (e.g., Hong Kong Memory[16]) have been supported by the *Jockey Club Charities Trust*, where most of other projects (e.g., St. John's College Archives [26], HKMMS SARS Oral History Archive [33], HKInscriptions [45], Gwulo [46]), especially the social media campaigns, are self-financed or reliant on crowdfunding.

Several modes of collaboration among different contributors have been identified. The first mode, as portrayed in the literature (e.g., Ocón, 2021; Song, 2023), entails that the different contributors (e.g., a government department, LCSD, and public GLAMs, as in [4]) jointly take charge of the project work, and this mode prevails among contributors that are already related to each other prior to the collaboration (e.g., the museum and library of the same university HK Edu. Image Database [22]). Another mode of collaboration involves a leading unit which at the same time receives support (e.g., Lands Department providing information [2]; HKU Libraries as the host of digital platforms [25], [29]) from another contributor. Across several projects (e.g., Hong Kong Memory [16], an e-database of ritual music [38], the BeHere project [39]), there can be multiple co-contributors with various roles, such as commissioner of the project, producer or manager of a platform, consultant, etc. For several other projects, given the limited information available on the project-related websites (e.g., vision and mission) and webpages (e.g., acknowledgement), the mode of collaboration and role of contributors seem to be ambiguous or are not disclosed. It is also not uncommon that some projects seemingly do not involve any cross-sector contributions, i.e., only working within a unit: e.g., a single government department (e.g., HKAAS [1], [6]), a single research team (e.g., [13] from CUHK, [21] from the CUHK Libraru, [26] from HKU St. John's College), a single NGO [31], [32], [35], [36], or a single special interest group (e.g., [43], HKInscriptions [45]).

Irrespective of whether and how various stakeholders have collaborated with each other, the deliverables of these projects have often acknowledged the help and support of various cultural heritage-related governmental bodies (e.g., Antiquities and Monuments Office, Intangible Cultural Heritage Office), academic researchers, and unaffiliated individuals. Such acts imply the importance of considering and utilising both professional, authoritative as well as lay, grassroots knowledge and experience in conducting cultural heritage digitisation projects.

4.1.3 Selection of cultural heritage and theme-based projects

Following the UNESCO categorisation of cultural heritage (UNESCO Institute for Statistics, 2009), we categorise the surveyed projects into digitising immovable,

movable, and intangible heritage. Only a small subset of projects has focused exclusively on digitising immovable cultural heritage, i.e., heritage buildings and sites (e.g., various sites [2], a reservoir [10], [57]). A modest number of projects have digitised movable cultural heritage only, namely historic monuments (e.g., inscriptions [45]), archaeological objects [1], artefacts in GLAMs (e.g., museum collections [4], [25], periodicals [44]), and ephemeral items (e.g., postcards) [26], [53]. Occasionally, despite focusing on immovable cultural heritage, a few projects also present content and information about the movable cultural heritage discovered in or associated with the immovable heritage (e.g., artefacts from a school [30], underwater heritage [43], military heritage [60]). More than half of the surveyed projects have been devoted to digitising only intangible cultural heritage, covering a wide range of general (e.g., HKICH Database [5], @hongkongoldfans [48]) and specific intangible heritage items: folkways (e.g., old lifestyles [6], [39], aquaculture [42]), literature (e.g., Christian literature [13], names [69]), oral traditions (e.g., chanting [14], indigenous languages [35], ritual music [38]), oral history (e.g., [20], [32], [33]), and so on. There are also a few endeavours (e.g., HK Historical GIS [15], @hkurbanfont [67]) presenting both immovable and movable cultural heritage as the main types, while touching on elements of intangible cultural heritage as a peripheral part of the projects. Interestingly, one project [43] worked towards digitising underwater cultural heritage including both immovable and movable items.

Similar to the cultural heritage digitisation projects around the world, those in Hong Kong also come with their own themes. A common example is a sub-category of movable (e.g., archaeological finds [1]), immovable (e.g., corner houses [58]), or intangible (e.g., oral history [20]) cultural heritage. Other examples include themes specific to a cultural heritage site (e.g., [6], [10]) or the cultural heritage items within a district in Hong Kong (e.g., Lung Yeuk Tau [32], Pokfulam Village [34], Wanchai [39]), and those specific to a historical incident (e.g., SARS [33]) or a historical period in Hong Kong (e.g., [15], Anti-Japanese War [36], 1950s to 1970s [49]). Given a theme in each project, our survey has found that the mechanism (e.g., top-down, bottom-up) or criteria for selecting what cultural heritage items to digitise nearly always remain unknown, with a few exceptional cases that outline such criteria (e.g., areas of significance, urgency) [44], [45]. Notably, in the student-led digitisation projects (i.e., [24a], [24b], [25] from HKU), the students had the freedom and flexibility to determine which items of cultural heritage to digitise, whereas in the various social media-based campaigns, both the managers of the social media pages and the users can participate in the selection process.

4.1.4 Technological adoption and user involvement

From our survey, these projects in Hong Kong have adopted the common forms of cultural heritage digitisation as reported in the literature, including 3D and immersive experiences, digital GLAMs, and social media. The most frequently adopted technology has been web publishing, which gives users free access to various digital GLAMs. While the literature has shown the prevalence of immersive technologies in showcasing digitised cultural heritage contents (Mendoza et al., 2023; Russo,

2021), only a few projects in Hong Kong have adopted them, including VR (i.e., virtual tours in [7], [10], [18], [30]) and AR (i.e., AR-enabled displays in [6], [8], [28], [39]). As described, social media platforms including Facebook, Instagram, and Twitter have proliferated cultural heritage digitisation in Hong Kong. For cultural heritage digitisation projects, it is important to consider how and by whom the digitised materials are managed (Ahmad et al., 2023). Nevertheless, except that the systems and platforms arising from government-led projects have specified by which parties the systems are maintained (e.g., [1], [2], [4], [5]), information about the management strategies and protocols are unclear in other projects. On a related note, only a few projects have shared part of the digitisation process to the public (e.g., [25], [39], [52]).

All of these digitisation projects have produced either physical (e.g., an on-site technology-enhanced exhibition [3], [6], [18]) or digital outputs, among which most of them offer open access to the digitised materials. In a minority of cases on the digital archives and databases (e.g., [19]), part of the digitised contents is only accessible to users affiliated with the institutions leading the projects. On a related note, the content in a few social media campaigns (e.g., a Facebook group [60]) are also restricted to only group members. Regarding user involvement, apart from the social media pages being open to users' commenting, several of the online platforms (e.g., [13], [14], [21], [34], [45]) openly welcome enquiries and feedback from users, and especially giving the disclaimer on the potential inaccuracy of some presented information and reminding users to make their own judgment (e.g., [13]).

4.1.5 Development trends

This section discusses the development trends across cultural heritage digitisation projects in Hong Kong, in terms of various aspects. First, there have been an increasing number of projects whose primary objectives surround digitisation itself instead of preservation or promotion in general, for instance, building a digital archive, information system, or database for users' convenient access to cultural heritage contents (e.g., [4], [5], [13], [26], [35], [38], [44]). Concerning the types of cultural heritage being digitised, various contributors' attention on the three UNESCO types of cultural heritage (i.e., immovable, movable, intangible) remains fairly distributed. However, in recent years, the different stakeholders in society have started embracing "unofficial" cultural heritage items that have not yet been statutorily recognised (e.g., listed or assessed by the government), such as ruins (e.g., [60]), literature (e.g., Christian literature [13], historic newspapers [21]) and leisurely practices (e.g., motoring [62], canoeing [63]). The historic values of these groups of cultural heritage might not be as high as those digitised in governmental projects (e.g., officially graded historic sites, listed oral traditions), though they might hold remarkable social, commemorative, and iconic values (Bond & Worthing, 2016), especially in Hong Kong where the notions of locality and collective memories have gained more traction (Wong, 2019). Unsurprisingly, more recent projects have leveraged not only immersive technologies (Ocón, 2022) but also cutting-edge technologies, such as AI, in facilitating the presentation of cultural

heritage contents (Mendoza et al., 2023). Furthermore, the contributors of cultural heritage digitisation have been expanded from more organised or institutionalised bodies, e.g., governmental departments, GLAMs, universities, and NGOs, to individual, grassroots populations, e.g., hobbyists and volunteers (e.g., @hkurbanfont [67]), which is consistent with projects around the world (Ocón, 2022; Zhang et al., 2022). Related to the expertise in digitisation, aligned with the global trends in participatory cultural heritage (Tasker & Liew, 2020; Tsenova et al., 2020), not only experienced professionals but also novice users, upon training, have been empowered to contribute towards digitising cultural heritage in Hong Kong (e.g., HKU projects on student-created digital products [24a], [24b], [25]).

The target audience and beneficiaries of these projects have also gradually been shifted from only scholars and the general public audience to also including amateurs, educators, and specific groups of the public, such as the younger generation (e.g., [21], [33], [35], [43]). This shift is consistent with the increasing emphasis of UNESCO on educating the youth about the significance of cultural heritage (İnanç & Liew, 2017). Partly owing to the emergence of social media as an innovative channel for delivery of cultural heritage content (Liang et al., 2021), the style of language employed in project deliverables has also demonstrated less formality, for example, using colloquial expressions for engaging public users (e.g., @hongkongoldfans [48]). In relation to the general public, projects in the recent decade are more likely to encourage feedback and suggestions from the users, including not only social media (e.g., @hkurbanfont [67]) but other aforementioned digital GLAMs and databases as well. In addition to welcoming two-way interactions with users, the contents of the digitised materials have also evolved from being rich in text (e.g., metadata of cultural heritage items) to being predominately aural or visual (e.g., Neonsigns.hk [11], [21], social media posts), which might prompt more diverse interpretations from users (Civantos et al., 2016).

4.2 Facilitating factors

This section discusses the multiple factors that have facilitated the development of cultural heritage digitisation projects in Hong Kong.

4.2.1 Diversity of cultural heritage

The emergence of different kinds of cultural heritage digitisation projects can be attributed to the diversity of cultural heritage in Hong Kong. Due to historical and socio-cultural reasons, Hong Kong has accumulated a variety of immovable and intangible cultural heritage (Kee, 2019; Wong, 2019). Examples of immovable cultural heritage include Chinese temples, neoclassical architectures, and old-styled corner houses (commonly known as "Tong Laus" [Figure 7.7]), to name a few, where intangible cultural heritage can be exemplified by oral traditions (e.g., Cantonese chanting) and various traditional craftsmanship (e.g., mooncake making and the Hungry Ghost Festival [Figure 7.8]). The large numbers of diverse cultural heritage items have thus motivated efforts from different stakeholders, with their

Figure 7.7 Old-styled corner houses as a unique kind of cultural heritage in Hong Kong. Photograph by the authors.

Figure 7.8 The stage set-up for the Hungry Ghost Festival in Hong Kong. Photograph by the authors.

own areas of interest, expertise, and experience, in systematically documenting and presenting them to the public audience. Another strength of these projects lies in that the groups and items of cultural heritage being digitised have not been limited to only those in urgent need of preservation, and have also included uncommon, less recognised ones (e.g., educational heritage [22]). This inclination towards *preventive digitisation* efforts (Ocón, 2021) can be more effective in preserving

and promoting cultural heritage proactively at earlier opportunities. Additionally, albeit that most projects have been mainly Hong Kong based, the frequent and ever-growing exchanges between Hong Kong and mainland China, along with various resources and funding opportunities from the Greater Bay Area, have also incentivised contributors from Hong Kong to start initiatives towards digitising cultural heritages in not only Hong Kong but also mainland China (e.g., (Hu et al., 2021 [29]; Kim, 2019 [28]; DigitalRepository@HKU [19]).

4.2.2 Multi-perspective contributors

In this dynamic, multivocal Hong Kong society, the success of cultural heritage preservation and promotion highly depends on harnessing the contributions from a plethora of stakeholders. Multiple projects have been financially supported by grants from the government (e.g., Lord Wilson Heritage Trust), large-scale NGOs (e.g., Jockey Club Charities Trust), and university grants. Aligning well with the higher-level missions of these funding bodies, the objectives and efforts of these projects have still been focused on preserving and promoting cultural heritage, rather than being distracted by other agendas. It is also an encouraging sign that, regardless of the modes of collaboration, governmental bodies, academic researchers, GLAMs, and the community often rely on each other's knowledge, skills, and experience. When digitising cultural heritage, especially those with local significance (Bond & Worthing, 2016), considering the plurality of perspectives in society not only can improve the audience's experience in learning about and interpreting the contents but also might widen the audience base (Corallo et al., 2019; Kontiza et al., 2020). This strategy is distinctively applicable to such contexts as Hong Kong where the locality of cultures and heritage has been increasingly valued (Wong, 2019). It is also noteworthy that some projects have considered how the digitised materials can be leveraged for teaching and learning (e.g., HKMMS SARS Oral History Archive [33]), especially those with tailor-made educational materials (e.g., LCSD Museum Collection Management System [4]) and educational outreach activities (e.g., [13]). This implies the educational potential of these projects in Hong Kong.

4.2.3 High accessibility

Except a few instances where part of the access is restricted, the majority of the project outputs, ranging from digital archives to immersive exhibitions, target the general public as both the audience and the beneficiaries. These projects often utilised high-tech digitisation technologies (e.g., drone-based VR production [7]) through offering access to digitised cultural heritage in a low tech-barrier way. For instance, the digital platforms arising from the projects can all be accessed using computer web browsers, with the more recent ones including the social media pages also supporting mobile access (e.g., HKInscriptions [45]). The databases, digital archives, and geographical information systems in particular also have largely user-friendly interfaces and facilities for supporting users' searching

and browsing of cultural heritage information (Gireesh & Raman, 2022), such as clear instructions on navigation [9], [19], browsing at different levels of granularity (e.g., collections, items) [13], simple search function (e.g., [31], [37], [38]), and the availability of hashtags [44]. The platforms that enable users' immersive experience also provide useful guidance, as illustrated by the details in navigating the virtual heritage environment [10] as a signature example. The multimodality of presented cultural heritage information, including text (e.g., metadata of cultural heritage items [2]), images (e.g., [5], [11]), audio (e.g., recordings of oral history) (e.g., [33]), and videos (e.g., introduction of cultural heritage [36]), can also cater to the needs of audience with wide-ranging preferences (Ng et al., 2022a). The free-of-charge access to these digital platforms and physical exhibitions showcasing digitised cultural heritage (e.g., [3], [6], [18]) also indicates the potential for low-cost participation in technology-enhanced cultural heritage experience among individuals from different socio-economic backgrounds.

4.2.4 Availability of practical information

In addition to user guides or instructions, many platforms have also included the sources of the digitised contents and also disclaimers to remind users about the potential inaccuracy and lack of comprehensiveness in the presented cultural heritage information (e.g., [13], [14], [45]). Notably, on the social media pages it is always remarked that the input from users is purely their personal opinions and do not represent the page itself. Disclaiming these details can manage the users' expectations on the content, educate them in information literacy, and protect the project leaders from possible legal or ethical liabilities. Moreover, the platforms also often emphasise that the digitised materials should only be used for academic, educational, and other non-commercial purposes (e.g., virtual tour of a reservoir [10], HK Oral History Archives [20], HKInscriptions [45]). For this rationale, in addition to showing clear statements on intellectual property rights, some platforms even supply recommendations on how the contents can be properly used (e.g., [60]) and how to cite the presented information (e.g., GIS on HK Heritage[2], the SARS Oral History Archive [33], [45]), which can encourage more exposure and a higher mobility of the digitised contents and can maximise their academic and educational values.

4.3 Limitations and challenges

The cultural heritage digitisation projects in Hong Kong are not without their own limitations and challenges from external sources (e.g., society, funding opportunities).

4.3.1 Inadequacy of designated policies and absence of standards

A prevalent challenge, which has also affected other countries and regions, is the absence of direct, effective policies on preserving and promoting cultural heritage (Pandey & Kumar, 2020). Despite the enforcement of the Antiquities and

Monuments Ordinance and the establishment of the Intangible Cultural Heritage Office for preserving tangible and intangible cultural heritage, respectively, in Hong Kong, graded historic buildings are not necessarily protected by legislation from being repurposed or demolished (Ho & Hou, 2019; Kee, 2019). In turn, it is even less likely for a policy on digitising cultural heritage to be proposed in the short term. This gap might have led to the lack of a standardised set of procedures on proposing, implementing, and evaluating cultural heritage digitisation projects in Hong Kong. As a consequence, in spite of the emergence of projects led by various stakeholders, the digitisation efforts still seem shattered, without much coherence or any connections among the different projects. More specifically, for databases and digital archives, perhaps owing to the lack of protocols for digitising cultural heritage, projects led by parties other than governmental bodies and GLAMs might not have strictly followed existing standards on creating metadata and performing cataloguing for cultural heritage items (Hu et al., 2018). This issue has also been reported as a common impediment in cultural heritage digitisation projects outside Hong Kong (Pandey & Kumar, 2020). In the long run, the lack of interoperability among platforms of digitised cultural heritage might affect the quality of cultural heritage information and thereby users' experience.

4.3.2 Lack of extrinsic incentives

Another practical challenge is the overall lack of extrinsic incentives for the various non-governmental stakeholders to devote time and resources into digitising cultural heritage. Even for public GLAMs in Hong Kong under the management of governmental bodies, their staff might still perceive the digitisation projects would bring about a heavier workload in addition to their current tasks (Meng et al., 2023). For NGOs and universities, though they might possess the intrinsic motivation to work towards preserving and promoting cultural heritage (e.g., to realise their social responsibilities), the existing public grants are not designated to technological projects, nor do they mandate or encourage the use of technologies for cultural heritage preservation or promotion. Regarding social media campaigns that might require less resources (e.g., funding, staffing) to start and operate, it can in turn be more challenging for the creators or managers to scale up and sustain their digitisation efforts.

4.3.3 Insufficient publicity and user evaluation

Notwithstanding the endeavours on digitising cultural heritage, insufficient publicity might be a salient limitation impeding the extent to which these projects have fulfilled their intended objectives. In our document analysis, we iteratively searched for and conducted in-depth inspection into various publicly available online materials, supplemented by our personal knowledge (e.g., physical advertisements in public). First, most of the projects have not been extensively promoted through various possible channels (e.g., government news and announcements, university mass emails, social media posts). Next, for users accessing the platforms

or audience participating in experiencing the digitised materials, very few of the projects have developed any evaluation mechanisms (e.g., post-experience survey) for soliciting users' feedback on the effectiveness of the projects and their deliverables. The project teams would miss the valuable opportunities to identify the strengths and weaknesses of their projects and their deliverables, which would be detrimental to the long-term development of these digitisation efforts.

4.3.4 Low international awareness

An evident limitation of many projects is their lack of international awareness. While most projects have clearly articulated their objectives of promoting a theme of cultural heritage to the public audience, it can be inferred that, based on the display languages, many projects seem to be targeting only the Hong Kong Chinese population. Among the projects, only the websites and platforms of government-led projects and one university-led project (Hu et al., 2021) show Traditional Chinese, Simplified Chinese, and English in their interface and also for the metadata of cultural heritage items. The availability of these three options on a website or platform implies catering to the needs of Hong Kong Chinese, mainland Chinese, and part of the non-Chinese-speaking audience. There is one exceptional case from the government-led project [7], where the audio guide of virtual tours is only available in spoken Cantonese. Many projects either do not provide English information (e.g., [32], [45]) and metadata or inconsistently provide English translation for only part of the Chinese content (e.g., [20], [21], [22]). Neglecting the needs of the international audience might limit the potential of these projects in promoting local cultural heritages in Hong Kong to the international audience.

4.3.5 Uncertainty about sustainability

Last but not least, as a possible consequence of the aforementioned limitations and challenges, it remains largely unknown the extent to which these projects are sustainable in the long run. As opposed to being one-off, short-term projects, cultural heritage digitisation projects require constant and consistent supplies of resources and efforts from not only the contributors but the involvement of different stakeholders (Ahmad et al., 2023; Pandey & Kumar, 2020). However, our analysis has discovered that, some projects are inherently temporary (e.g., non-permanent exhibitions of digitised materials [3], [7], [18]), and in projects presenting the digitised contents online, few have explicitly presented their long-term digital preservation strategies (e.g., Gifts from Lanmama (Hu et al., 2021) [29]). These observations demonstrate the risk of the projects lacking sustainability, not to mention any future development and extensions.

4.4 Recommendations for future projects

When proposing a cultural heritage digitisation project, the formulation of objectives should align with both international agendas (e.g., United Nation's Sustainable

Development Goals) and regional public policies on cultural heritage preservation and promotion. In addition to systematic documentation and dissemination of cultural heritage contents and information, future endeavours should amplify digitisation as one of the major objectives. The target audience should also be identified and spelled out early on for the sake of clarity and for upcoming implementation work (e.g., designing metadata, seeking dissemination channels). As an important consideration for planning the project timeline and budgeting, a needs analysis should be conducted in advance, which can also prepare the project team for managing the audience's expectations.

Future teams and individuals should strive for more cross-sector collaborative projects, such that the knowledge, skills, and experience of various stakeholders in society can be synergised for effective preservation and promotion of cultural heritage. Effective collaboration may also improve the rigor and efficiency of existing, in-progress digitisation projects in various organisations, in which a large portion of materials are waiting to be digitised. In particular, for realising the multivocality of participatory cultural heritage, members of the general public should be encouraged to offer their input from time to time. Given the increasing emphasis on digital humanities around the world, project teams can also seek collaboration opportunities with experts and communities outside Hong Kong.

Regarding the selection of cultural heritage to be digitised, future projects are suggested to place more attention on the rich pool of movable cultural heritage (e.g., museum artefacts, archaeological finds), and to consider both the official and "unofficial" immovable and intangible cultural heritages in Hong Kong. To further gain societal acceptance towards digitising cultural heritage, projects should strike a balance between the socio-historical values and local significance of the groups or items of cultural heritage to be digitised. Upon deciding on a theme, project deliverables should be transparent in presenting the criteria for selecting the subsets of items to digitise. Strategic selection decisions can also facilitate alignment with project objectives, coordinating resources, and controlling costs.

Future projects should transcend the traditional boundaries in adopting digitisation technologies. Building on traditional digitised contents and materials, there should be more exploration on developing 3D models and immersive experiences for digitising cultural heritage in Hong Kong. Furthermore, for capitalising on emerging technologies, generative artificial intelligence (GenAI) can be leveraged to streamline and improve the digitisation work, such as generating high-fidelity simulations of the appearance of lost or damaged cultural heritage items and checking the correctness and consistency of metadata. For effectively democratising digitised cultural heritage without compromising the quality of output, the best practice of high-tech digitisation for low tech-barrier access should continue to be adopted. It should be the goal that individuals from diverse socio-economic and technical backgrounds can afford accessing the digitised materials.

Regarding the delivery of cultural heritage contents and information, future projects should make every effort to follow international standards on creating metadata of cultural heritage items, to maintain consistency and maximise interoperability. They should ensure multiple modalities of information will be presented

for the diversified users' needs and preferences. Awareness of the international audience should also be demonstrated, through giving essential contextual information, defining terms specific to the local context, and supplying English translation for as much content as possible. The educational potential of the digitised materials can also be highlighted, such that the project deliverables can fulfil their educational values in classrooms and other learning contexts (e.g., community centres).

For improving user engagement, projects can learn from how social media campaigns empower their users in contributing materials and opinions, participating in intellectual exchanges, and also how social media managers interact with these users. These practices might also help boost the publicity of the projects and their deliverables. The publicised materials from the projects should also be added with clear disclaimers about the copyrights of digitised materials and their possible non-commercial usage if applicable. For self-protection, after their efforts in maintaining the quality of presented information, the projects should also remind users about the possible inaccuracy of the information and encourage their own further investigation. Finally, apart from making more efforts in conducting user evaluations, project teams producing online platforms should consider examining their system logs and analytics for a partial picture of the impacts achieved by their projects.

5 Conclusion

This chapter reviewed projects on cultural heritage digitisation in Hong Kong, both from the literature and in practice. A total of 69 projects were identified, where we examined not only the literature but also publicly available materials, such as project websites, news articles, and reports. We summarised the approaches, various important aspects of these projects and the development trends across projects. Upon identifying the facilitating factors and reflecting on limitations and challenges of these projects, we discussed the recommended best practices for future projects in Hong Kong, among which some can be applied to other contexts.

Notes

1 "To protect and safeguard artefacts, monuments, and groups of buildings and sites that have a diversity of values including symbolic, historic, artistic, aesthetic, ethnological or anthropological, scientific and social significance".
2 "By 2030, ensure that all learners acquire the knowledge and skills needed to promote sustainable development, including, among others, through education for sustainable development and sustainable lifestyles, human rights, gender equality, promotion of a culture of peace and non-violence, global citizenship and appreciation of cultural diversity and of culture's contribution to sustainable development".
3 https://www.aab.gov.hk/en/historic-buildings/results-of-the-assessment/index.html#new.
4 https://www.icho.hk/en/web/icho/the_first_intangible_cultural_heritage_inventory_of_hong_kong.html.
5 https://www.europeana.eu/.
6 https://artsandculture.google.com/project/openheritage.
7 https://www.ichandmuseums.eu/.
8 http://www.sucho.org/.

9 https://accademia.stanford.edu/mich/.
10 https://www.louvre.fr/en/what-s-on/life-at-the-museum/the-mona-lisa-in-virtual-reality-in-your-own-home.
11 https://www.digitalmuseums.ca/.
12 http://www.koreanhistory.org/.
13 https://nrch.culture.tw/.
14 https://arquives.ca/.
15 LGBTQ2+ is an acronym for "Lesbian, Gay, Bisexual, Transgender, Queer or Questioning, and Two-Spirit".
16 https://www.saada.org/.
17 https://web.archive.org/web/20230828122002/https://www.singaporememory.sg/.
18 https://sahris.sahra.org.za/.
19 http://historicplacesla.org/.
20 https://en.unesco.org/covid19/cultureresponse/exploring-world-heritage-from-home-with-unesco.
21 https://www.amo.gov.hk/en/about-us/welcome/index.html.
22 https://www.tourism.gov.hk/en/about-us.php.
23 https://hk.art.museum/en/web/ma/tea-ware.html.
24 The Hong Kong Jockey Club is one of the biggest non-governmental organizations in Hong Kong.
25 Toponymy refers to the study of place names.

References

Affleck, J., & Kvan, T. (2008). A virtual community as the context for discursive interpretation: A role in cultural heritage engagement. *International Journal of Heritage Studies, 14*(3), 268–280. https://doi.org/10.1080/13527250801953751

Ahmad, R., Rafiq, M., & Arif, M. (2023). Global trends in digital preservation: Outsourcing versus in-house practices. *Journal of Librarianship and Information Science*. https://doi.org/10.1177/09610006231173461

Allsop, A. (2022). Walking books? The oral history method at The Hong Kong Heritage Project. *Collections, 18*(3), 362–381. https://doi.org/10.1177/15501906221096814

Barton, C., Cox, A., Cruz, S. D., & Hansen, J. (2017). Cultural-heritage inventory implementations: The versatility of the Arches system. *APT Bulletin: The Journal of Preservation Technology, 48*(4), 19–28. https://www.jstor.org/stable/26382563

Bekele, M. K., Pierdicca, R., Frontoni, E., Malinverni, E. S., & Gain, J. (2018). A survey of augmented, virtual, and mixed reality for cultural heritage. *Journal on Computing and Cultural Heritage (JOCCH), 11*(2), 1–36. https://doi.org/10.1145/3145534

Boboc, R. G., Băutu, E., Gîrbacia, F., Popovici, N., & Popovici, D. M. (2022). Augmented reality in cultural heritage: An overview of the last decade of applications. *Applied Sciences, 12*(19), 9859. https://doi.org/10.3390/app12199859

Bond, S., & Worthing, D. (2016). *Managing built heritage: The role of cultural values and significance*. John Wiley & Sons.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

Bradley, J. (2016). No job for techies: Technical contributions to research in the digital humanities. In Mccarty, W. & Deegan, M. (Eds.), *Collaborative research in the digital humanities* (pp. 11–25). Routledge.

Bruno, F., Bruno, S., De Sensi, G., Luchi, M. L., Mancuso, S., & Muzzupappa, M. (2010). From 3D reconstruction to virtual reality: A complete methodology for

digital archaeological exhibition. *Journal of Cultural Heritage, 11*(1), 42–49. https://doi.org/10.1016/j.culher.2009.02.006

Burri, M. (2010). Digital technologies and traditional cultural expressions: A positive look at a difficult relationship. *International Journal of Cultural Property, 17*(1), 33–63. https://doi.org/10.1017/S0940739110000032

Champion, E., & Rahaman, H. (2020). Survey of 3D digital heritage repositories and platforms. *Virtual Archaeology Review, 11*(23), 1–15. https://doi.org/10.4995/var.2020.13226

Chen, W. W. (2020). Body as echoes: Cyber archiving of Buddhist deities in the cave temples of China. *Digital Cultural Heritage*, 119–137. https://doi.org/10.1007/978-3-030-15200-0_9

Ching, S. H. (2018). Turning a service learning experience into a model of student engagement: The Lighthouse Heritage Research Connections (LHRC) Project in Hong Kong. *The Journal of Academic Librarianship, 44*(2), 196–206. https://doi.org/10.1016/j.acalib.2018.02.007

Chow, S. K., & Chan, K. L. (2009). Reconstruction of photorealistic 3D model of ceramic artefacts for interactive virtual exhibition. *Journal of Cultural Heritage, 10*(2), 161–173. https://doi.org/10.1016/j.culher.2008.08.011

Civantos, A. M., Brown, M., Coughlan, T., Ainsworth, S., & Lorenz, K. (2016). Using mobile media creation to structure museum interpretation with professional vision. *Personal and Ubiquitous Computing, 20*(1), 23–36. https://doi.org/10.1007/s00779-015-0895-3

Corallo, A., Esposito, M., Marra, M., & Pascarelli, C. (2019). Transmedia digital storytelling for cultural heritage visiting enhanced experience. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics* (pp. 221–229). Springer. https://doi.org/10.1007/978-3-030-25999-0_19

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology, 13*(1), 3–21. https://doi.org/10.1007/BF00988593

De Bruycker, L., & Girault, Y. (2018). Constraints and stakes in enhancing archaeological landscapes in the digital age. *International Journal of Geoheritage and Parks, 6*(1), 75–94. https://doi.org/10.17149/ijg.j.issn.2210.3382.2018.01.006

Dombrowski, Q. (2022). Saving Ukrainian Cultural Heritage Online (SUCHO): Digital relief work relationships more than technology. *International Leads, 37*(2), 3–12.

Du Cros, H., & Yok-shiu, F. L. (Eds.). (2007). *Cultural heritage management in China: Preserving the cities of the Pearl River Delta*. Routledge.

Dümcke, C., & Gnedovsky, M. (2013). The social and economic value of cultural heritage: Literature review. *European Expert Network on Culture Paper*, 1–114.

Finefter-Rosenbluh, I. (2017). Incorporating perspective taking in reflexivity: A method to enhance insider qualitative research processes. *International Journal of Qualitative Methods, 16*(1), 1–11. https://doi.org/10.1177/1609406917703539

Ginzarly, M. (2021). Social media data for the conservation of historic urban landscapes: Prospects and challenges. In Rauterberg, M. (Ed.), *International Conference on Human-Computer Interaction* (pp. 209–223). Springer International Publishing. https://doi.org/10.1007/978-3-030-77431-8_13

Ginzarly, M., & Srour, F. J. (2022). Cultural heritage through the lens of COVID-19. *Poetics, 92*, 101622. https://doi.org/10.1016/j.poetic.2021.101622

Gireesh, K. T. K., & Raman, N. R. (2022). Conserving knowledge heritage: Opportunities and challenges in conceptualizing cultural heritage information system (CHIS) in the Indian context. *Global Knowledge, Memory and Communication, 71*(6/7), 564–583. https://doi.org/10.1108/GKMC-02-2021-0020

Guidi, G., & Frischer, B. D. (2020). 3D digitisation of cultural heritage. In Liu, Y., Pears, N., Rosin, P. L., & Huber, P. (Eds.), *3D imaging, analysis and applications* (pp. 631–697). Cham: Springer. https://doi.org/10.1007/978-3-030-44070-1_13

Hijazi, A. N., & Baharin, H. (2022). The effectiveness of digital technologies used for the visitor's experience in digital museums. A systematic literature review from the last two decades. *International Journal of Interactive Mobile Technologies, 16*(16), 142–159.

Ho, D., & Hou, H. (2019). Enabling sustainable built heritage revitalisation from a social and technical perspective: A case study. *Facilities, 37*(11/12), 704–722. https://doi.org/10.1108/F-05-2017-0055

Hou, Y., Kenderdine, S., Picca, D., Egloff, M., & Adamou, A. (2022). Digitising intangible cultural heritage embodied: State of the art. *Journal on Computing and Cultural Heritage, 15*(3), 1–20. https://doi.org/10.1145/3494837

Hu, X., Ng, J., & Lee, J. H. (2019). VR creation experience in cultural heritage education: A preliminary exploration. *Proceedings of the Association for Information Science and Technology, 56*(1), 422–426. https://doi.org/10.1002/pra2.42

Hu, X., Ng, J., & Xia, S. (2018). User-centered evaluation of metadata schema for nonmovable cultural heritage: Murals and stone cave temples. *Journal of the Association for Information Science and Technology, 69*(12), 1476–1487. https://doi.org/10.1002/asi.24065

Hu, X., Ng, J. T. D., & Liu, R. (2021). Development and evaluation of a digital museum of a National intangible cultural heritage from China. In K. Toeppe, H. Yan & S. K. W. Chu (Eds.), *Diversity, Divergence, Dialogue. iConference 2021*. Lecture Notes in Computer Science, 12646. Springer. https://doi.org/10.1007/978-3-030-71305-8_42

Hu, X., Yu, B., Alman, S., Renear, A. H., & Carbo, T. (2017). Teaching information science and technology to the world? Practices, challenges and visions. *Proceedings of the Association for Information Science and Technology, 54*(1), 566–569. https://doi.org/10.1002/pra2.2017.14505401074

İnanç, G., & Liew, J. H. (2017). Teaching the heritage of "others" and making it "ours": The power of cultural heritage education. In Walsh, M. J. K. (Ed.), *The Armenian Church of Famagusta and the complexity of Cypriot heritage* (pp. 285–299). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-48502-7_14

Jagielska-Burduk, A., Pszczyński, M., & Stec, P. (2021). Cultural heritage education in UNESCO cultural conventions. *Sustainability, 13*(6), 3548. https://doi.org/10.3390/su13063548

Jung, T. H., Lee, H., Chung, N., & tom Dieck, M. C. (2018). Cross-cultural differences in adopting mobile augmented reality at cultural heritage tourism sites. *International Journal of Contemporary Hospitality Management, 30*(3), 1621–1645. https://doi.org/10.1108/IJCHM-02-2017-0084

Kee, T. (2019). Sustainable adaptive reuse–economic impact of cultural heritage. *Journal of Cultural Heritage Management and Sustainable Development, 9*(2), 165–183. https://doi.org/10.1108/JCHMSD-06-2018-0044

Kenderdine, S., Forte, M., & Camporesi, C. (2012). Rhizome of Western Han: An omnispatial theatre for archaeology. *Revive the Past. Computer Applications and Quantitative Methods in Archaeology (CAA). Proceedings of the 39th International Conference, Beijing, April* (pp. 12–16).

Kim, E. S. (2019). The intersection of digital humanities and artistic augmented/virtual reality research in Hong Kong. In Wong, S. R., Li, H., & Chou, M. (Eds.), *Digital humanities and scholarly research trends in the Asia-Pacific* (pp. 19–41). https://doi.org/10.4018/978-1-5225-7195-7.ch002

Kontiza, K., Antoniou, A., Daif, A., Reboreda-Morillo, S., Bassani, M., González-Soutelo, S., Lykourentzou, I., Jones, C. E., Padfield, J., & López-Nores, M. (2020). On how technology-powered storytelling can contribute to cultural heritage sustainability across multiple venues—Evidence from the CrossCult H2020 Project. *Sustainability, 12*(4), 1666. https://doi.org/10.3390/su12041666

Koukopoulos, Z., & Koukopoulos, D. (2019). Integrating educational theories into a feasible digital environment. *Applied Computing and Informatics, 15*(1), 19–26. https://doi.org/10.1016/j.aci.2017.09.004

Koya, K., & Chowdhury, G. (2020). Cultural heritage information practices and iSchools education for achieving sustainable development. *Journal of the Association for Information Science and Technology, 71*(6), 696–710. https://doi.org/10.1002/asi.24283

Kużelewska, E., & Tomaszuk, M. (2020). European human rights dimension of the online access to cultural heritage in times of the COVID-19 outbreak. *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique, 35*, 1067–1079. https://doi.org/10.1007/s11196-020-09712-x

Lee, H., Chung, N., & Jung, T. (2015). Examining the cultural differences in acceptance of mobile augmented reality: Comparison of South Korea and Ireland. In I. Tussyadiah & A. Inversini (Eds.), *Information and communication technologies in tourism 2015* (pp. 477–491). Springer. https://doi.org/10.1007/978-3-319-14343-9_35

Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., & Fulk, D. (2000). The digital Michelangelo project: 3D scanning of large statues. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 131–144). https://doi.org/10.1145/344779.344849

Li, D., Du, Z., Zhu, Y., & Wang, T. (2010). Digital cultural heritage and its applications: Case studies of Digital Mogao Grottos and Chi Lin Nunnery reconstruction. In *Sixth International Symposium on Digital Earth: Data Processing and Applications* (Vol. 7841, pp. 539–545). SPIE. https://doi.org/10.1117/12.873340

Li, J., Peng, G., & Zhang, N. (2022). Exploring the implementation of web, mobile, and VR technology for cultural heritage presentation. In N. A. Streitz & S. Konomi (Eds.), *Distributed, ambient and pervasive interactions. Smart environments, ecosystems, and cities. HCII 2022*. Lecture Notes in Computer Science, 13325. Springer. https://doi.org/10.1007/978-3-031-05463-1_4

Liang, X., Hua, N., Martin, J., Dellapiana, E., Coscia, C., & Zhang, Y. (2022). Social media as a medium to promote local perception expression in China's world heritage sites. *Land, 11*(6), 841. https://doi.org/10.3390/land11060841

Liang, X., Lu, Y., & Martin, J. (2021). A review of the role of social media for the cultural heritage sustainability. *Sustainability, 13*(3), 1055. https://doi.org/10.3390/su13031055

Liao, H. T., Zhao, M., & Sun, S. P. (2020). A literature review of museum and heritage on digitisation, digitalization, and digital transformation. In *6th International Conference on Humanities and Social Science Research (ICHSSR 2020)* (pp. 473–476). Atlantis Press. https://doi.org/10.2991/assehr.k.200428.101

Liu, Y., & Lin, H. W. (2021). Construction of interpretation and presentation system of cultural heritage site: An analysis of the old city, zuoying. *Heritage, 4*(1), 316–332. https://doi.org/10.3390/heritage4010020

Lo, P., Chan, H. H., Tang, A. W., Chiu, D. K., Cho, A., See-To, E. W., Ho, K. K. W., He, M., Kenderdine, S., & Shaw, J. (2019). Visualising and revitalising traditional Chinese martial arts: Visitors' engagement and learning experience at the 300 years of Hakka Kungfu. *Library Hi Tech, 37*(2), 269–288. https://doi.org/10.1108/LHT-05-2018-0071

Macrì, E., & Cristofaro, C. L. (2021). The digitalisation of cultural heritage for sustainable development: The impact of Europeana. In P. Demartini, L. Marchegiani, M. Marchiori & G. Schiuma (Eds.), *Cultural initiatives for sustainable development. Contributions to management science*. Springer. https://doi.org/10.1007/978-3-030-65687-4_17

Mendoza, M. A. D., De La Hoz Franco, E., & Gómez, J. E. G. (2023). Technologies for the preservation of cultural heritage—A systematic review of the literature. *Sustainability, 15*(2), 1059. https://doi.org/10.3390/su15021059

Meng, Y., Chu, M. Y., & Chiu, D. K. (2023). The impact of COVID-19 on museums in the digital era: Practices and challenges in Hong Kong. *Library Hi Tech, 41*(1), 130–151. https://doi.org/10.1108/LHT-05-2022-0273

Mudička, Š., & Kapica, R. (2022). Digital heritage, the possibilities of information visualisation through extended reality tools. *Heritage, 6*(1), 112–131. https://doi.org/10.3390/heritage6010006

Ng, J., Hu, X., & Que, Y. (2022a). Towards multi-modal evaluation of eye-tracked virtual heritage environment. In *Proceedings of International Learning Analytics & Knowledge Conference 2022*. https://doi.org/10.1145/3506860.3506881

Ng, J. T. D., Sun, Y., & Hu, X. (2022b). An exploratory study on students' digital curation competency and experience in a general education course. In *Proceedings of the 16th International Conference of the Learning Sciences-ICLS 2022* (pp. 1517–1520). International Society of the Learning Sciences. https://dx.doi.org/10.22318/icls2022.1517

Ocón, D. (2021). Digitalising endangered cultural heritage in Southeast Asian cities: Preserving or replacing? *International Journal of Heritage Studies, 27*(10), 975–990. https://doi.org/10.1080/13527258.2021.1883711

Ocón, D. (2022). Digitalisation and cultural heritage tourism in Asia: Stepping in without stepping out. In E. Borin, M. Cerquetti, M. Crispí & J. Urbano (Eds.), *Cultural leadership in transition tourism. Contributions to management science* (pp. 157–181). Springer. https://doi.org/10.1007/978-3-031-14121-8_9

Ocón, D. (2023). Low-cost digital tools to preserve cultural heritage 'blind spots': The case of Kubor Kassim in Singapore. *International Journal of Heritage Studies*, 961–987. https://doi.org/10.1080/13527258.2023.2231903

Pandey, R., & Kumar, V. (2020). Exploring the Impediments to digitisation and digital preservation of cultural heritage resources: A selective review. *Preservation, Digital Technology & Culture, 49*(1), 26–37. https://doi.org/10.1515/pdtc-2020-0006

Pérez, R. J., López, J. M. C., & Listán, D. M. F. (2010). Heritage education: Exploring the conceptions of teachers and administrators from the perspective of experimental and social science teaching. *Teaching and Teacher Education, 26*(6), 1319–1331. https://doi.org/10.1016/j.tate.2010.01.005

Pisoni, G., Díaz-Rodríguez, N., Gijlers, H., & Tonolli, L. (2021). Human-centered artificial intelligence for designing accessible cultural heritage. *Applied Sciences, 11*(2), 870.

Poole, A. H. (2020). The information work of community archives: A systematic literature review. *Journal of Documentation, 76*(3), 657–687. https://doi.org/10.3390/app11020870

Quintero, M. S., Awad, R., & Barazzetti, L. (2020). Harnessing digital workflows for the understanding, promotion and participation in the conservation of heritage sites by meeting both ethical and technical challenges. *Built Heritage, 4*(1), 1–18. https://doi.org/10.1186/s43238-020-00005-7

Russo, M. (2021). AR in the architecture domain: State of the art. *Applied Sciences, 11*(15), 6800. https://doi.org/10.3390/app11156800

Sankofa, N. (2022). Critical method of document analysis. *International Journal of Social Research Methodology*, (6), 1–13. https://doi.org/10.1080/13645579.2022.2113664

Song, G. (2023). Cross-cultural exchange of heritage in Museums: A study of Dunhuang art in the Hong Kong heritage museum from the perspective of cultural translation. *Critical Arts*, (2), 1–15. https://doi.org/10.1080/02560046.2023.2237094

Tasker, G., & Liew, C. L. (2020). 'Sharing my stories': Genealogists and participatory heritage. *Information, Communication & Society, 23*(3), 389–406. https://doi.org/10.1080/1369118X.2018.1510533

Tsenova, V., Wood, G., Dolfini, A., Tindley, A., & Kirk, D. (2020). Un-authorised view: Leveraging volunteer expertise in heritage. In *Proceedings of the 2020 CHI Conference on human factors in computing systems* (pp. 1–14). https://doi.org/10.1145/3313831.3376558

UNESCO Institute for Statistics. (2009). *Cultural heritage – Definition*. Retrieved from https://uis.unesco.org/en/glossary-term/cultural-heritage

Virtual Reality Society. (2017). *What is virtual reality*. Retrieved August 17, 2023, from https://www.vrs.org.uk/virtual-reality/what-is-virtual-reality.html

Wang, Z., Ng, J. T. D., Liu, R., & Hu, X. (2022). Learning analytics enabled virtual reality content creation platform: System design and preliminary evaluation. In *2022 International Conference on Advanced Learning Technologies (ICALT)* (pp. 161–163). IEEE. https://doi.org/10.1109/ICALT55010.2022.00055

Wong, M. (2019). Recontextualizing cultural heritage and identity through three-dimensional spaces: The case of the Hong Kong Maritime Museum. *The International Journal of the Inclusive Museum, 12*(2), 27–46. https://doi.org/10.18848/1835-2014/CGP/v12i02/27-46

Xue, C. Q. (2016). *Hong Kong architecture 1945–2015*. Springer. https://doi.org/10.1007/978-981-10-1004-0

Yang, S., Xu, S., & Huang, W. (2022). 3D point cloud for cultural heritage: A scientometric survey. *Remote Sensing, 14*(21), 5542. https://doi.org/10.3390/rs14215542

Yung, E. H., & Chan, E. H. (2011). Problem issues of public participation in built-heritage conservation: Two controversial cases in Hong Kong. *Habitat International, 35*(3), 457–466. https://doi.org/10.1016/j.habitatint.2010.12.004

Zhang, X., Yang, D., Yow, C. H., Huang, L., Wu, X., Huang, X., Guo, J., Zhou, S., & Cai, Y. (2022). Metaverse for cultural heritages. *Electronics, 11*(22), 3730. https://doi.org/10.3390/electronics11223730

8 Digital memory construction for cultural heritage

Methodology and applications in China

Huiling Feng, Linqing Ma, Tianjiao Qi and Wenhong Zhou

1 Introduction

In 2022, the Central and State Council of China jointly issued 'Opinions on Promoting the Implementation of the National Strategy for Cultural Digitisation' (General Office of the CPC Central Committee, General Office of the State Council, 2022) and thereby launched a holistic project for the construction and development of cultural heritage digitisation at the national level. In June 2023, General Secretary Xi Jinping emphasised at a symposium on the proliferation of culture in Beijing that Chinese civilisation manifests a salient capacity for innovation, particularly exemplified by the untapped potential of Chinese cultural heritage in the realm of digital metamorphosis. Digital memory is a new form of digital storage capable of collecting, organising, storing, and displaying the historical and cultural information of a specific object in a digital manner, and carrying, reproducing, and disseminating such information in cyberspace (Feng, 2020). In contrast to the digitisation of the carrier of existing cultural heritage, digital memory focuses on the emotional interaction between people and cultural heritage, thus emphasising the systematic and complete preservation of the cultural heritage itself and various humanistic elements in the historical and cultural space and time in which it is located. Digital memory construction for cultural heritage, with the digital transformation of cultural heritage as the core, connects all types of tangible and intangible resources that record and reflect the core of cultural memory, forms a resource network that conveys cultural memory in multiple dimensions, and realises the innovative transformation and creative development of cultural memory.

Current research focuses on the construction of digital memory from the perspective of cultural heritage protection, as manifested in the following aspects. First, there is a consensus on the application of digital technology to cultural heritage protection. Discussions lies on how various digital technologies, such as 3D scanning (Sanders, 2014; Wachowiak, 2009), artificial intelligence (AI; Zhou et al., 2023), digital twin technology (Jia & Yan, 2022), digital visualisation (Siddiqui et al., 2022; Zhao, 2022), and virtual reality (VR; Comes et al., 2018; Man & Gao, 2023), can be employed in cultural heritage protection.

Second, innovations in methods and pathways for the digital preservation of cultural heritage have been explored (Mohanty & Swain, 2022; Zhao, 2021).

DOI: 10.4324/9781032707211-10

This includes the comprehensive design and optimisation of specific processes for the collection (Lu, 2007), construction (Wu, 2019), organisation (Zhou et al., 2017), and exploitation of digital cultural heritage resources (Ma, 2022). Topics such as digital processing and digital heritage platforms have been frequently discussed (Li, 2018; Quintero et al., 2017).

Third, strategies have been identified to deepen the practice of digital preservation of cultural heritage (Beel et al., 2017), such as making collaborative recommendations for the digital transformation of heritage preservation institutions like libraries, archives, and museums (Chen & Xiao, 2022; Zhang & Li, 2020).

Fourth, with more practical experience, the construction of digital memory projects is gradually becoming a mainstream approach for cultural heritage preservation (Feng et al., 2022; Qi, 2022). This process subsequently evolves from exploratory practices to higher-level cognition and methodological construction, thereby enriching knowledge on the methodology of digital memory targeted at cultural heritage (Gorbul & Rusakov, 2022).

Different academic perspectives, including the fields of information resource management (Feng, 2020), communication studies (Liu, 2023), psychology (Shi & Zhou, 2022), and digital humanities (Feng, 2021), offer varying interpretations of the concept of 'digital memory'. For instance, communication studies define digital memory based on the digitisation of media memory (Li & Yang, 2022). By contrast, the field of information resource management, particularly archival science, explicitly puts forward a complete concept for digital memory construction, based on which it defines the concept and features of digital memory and clarifies its theoretical connotations, essential methods, and action strategies within the cultural memory framework. It also proposes a China-specific approach by undertaking specific projects in China (Feng, 2003, 2012; Jia et al., 2019; Wang et al., 2021; Wu et al., 2019).

Consequently, projects related to urban, rural, and other types of memory have been implemented on a large scale in China, with participants spread across cultural heritage institutions, government departments, social organisations, and academic institutions. Examples include the China Memory Project initiated by the National Library of China (Ministry of Culture and Tourism of People's Republic of China, 2012), archival projects for traditional villages proposed by the Ministry of Housing and Urban–Rural Development (The Central People's Government of the People's Republic of China, 2023), and various memory initiatives to boost urban development and rural revitalisation. These initiatives have led to memory projects such as 'Beijing Memory' (Capital Library, 2013), 'Sichuan Memory' (Provincial Party Committee Propaganda Department of Sichuan, 2013), an online exhibition hall for overseas Chinese letters (Fujiang Municipal Archives, 2023), the 'World Memory List—Nanjing Massacre Archives' database (National Archives Administration of China, 2022), and the 'Moment in Eternity—Oracle Culture Exhibition' (Shandong Museum, 2023), among others.

The Digital Memory Team of Renmin University of China, which includes us as authors, was the first scientific research team to conduct digital memory construction in China. Since 2013, our team has explored and built a large-scale digital

memory resource platform known as 'Beijing Memory', and successfully accomplished the digital transformation of 22 Beijing cultural heritage topics over the past decade. Our team's chief expert, Dr. Huiling Feng, played a pivotal role in this endeavour. She was instrumental in introducing the concept of 'digital memory' in China, provided theoretical foundations, organised the inaugural international forum on digital memory in 2015, and presented a keynote speech 'Cultural Power of Digital Memory' at the sixth Digital China Construction Summit in 2022. As integral members of this team, we aim to systematically elucidate the theoretical connotations and methods of digital memory construction in this chapter. We will also analyse its application in the 'Beijing Memory' project, presenting it as a case study to offer a reusable framework for the comprehensive mining, in-depth interpretation, digital reproduction, and widespread dissemination of cultural heritage in the digital era.

2 Understanding of digital memory

2.1 The concept of digital memory

The memory literature represents a multifaceted, interdisciplinary domain in which the fields of psychology, sociology, anthropology, and media theory intersect. Although several studies on memory have touched upon its digitisation from the perspective of media change, the term 'memory' used in these contexts is broadly defined. For example, the act of saving digital information is considered 'digital memory'. This situation suggests that the existing research is not yet sufficient to form a specialised academic field dedicated to digital memory. Since the 1990s, digital memory projects have proliferated globally, characterised by a diverse range of constructors and thematic categories. These encompass memory initiatives at national and regional levels, as well as those focused on universities, corporations, rural communities, and various social groups, each contributing to the ever-expanding tapestry of memory projects. However, despite the high level of engagement in digital memory practices, theoretical research remains relatively underdeveloped.

In the 2010s, Chinese scholars and practitioners in the field of information management began interpreting digital memory from an information management perspective and initiated relevant projects contributing to the establishment, expansion, and deepening of digital memory research in China. The Digital Memory International Forum, held in October 2015 by the School of Information Resource Management at Renmin University of China, marked the first systematic discussion on digital memory and had a significant impact on digital memory research in China. Dr. Huiling Feng, who convened the forum, initially elaborated on the concept, background, motivation, features, and goals of digital memory in her keynote address (Feng, 2015). Grounding her discussion on cultural memory theory, she defined digital memory as the digital construction of cultural memory.

The concept of 'cultural memory' was introduced by German scholars Jan and Aleida Assmann in the 1980s and has since been continuously enriched and accepted

by many cultural scholars (Assmann, 2011). Cultural memory is a construct of past experiences that have moved beyond the human brain to reside in 'external storage devices', symbolised by various information symbols such as text, images, sound, and various metaphors. Based on this understanding of cultural memory, and in conjunction with observations of burgeoning digital memory projects on urban and rural memories in recent years in China, this study explicates the concept of digital memory as follows: Digital memory is the digital construction of cultural memory. Simply put, it involves capturing, organising, storing, and displaying the historical and cultural information of specific entities in digital formats to sustain, represent, and disseminate these memories in cyberspace. In a broader sense, all human activities recorded in digital codes, including microblogs, WeChat, webpage information, photographs taken onsite, and videos, can be considered digital memory.

2.2 Relationships between 'digital memory' and 'digital humanities'

With the increasing prominence of digital humanities research in recent years, a growing number of Chinese scholars have become attuned to the intertwined relationship between digital memory and digital humanities and have attempted to integrate digital humanities methodologies into urban memory projects and beyond. For example, Hou employed the concepts and methods of digital humanities to examine the modes of urban memory resource construction and advocated the use of text mining, linked data, geographic information systems (GIS) and VR, and text visualisation technologies to develop urban memory resources. This mode of resource development, through the adoption of digital humanities methodologies, has received widespread scholarly approval (Hou & He, 2018; Zhao et al., 2023; Zhou & Cui, 2021).

In addition to the aforementioned methodological borrowings, we have, including Feng's expertise, conducted an in-depth analysis of the interlaced relationship between digital memory and digital humanities. This collaborative effort highlights significant overlaps between the two, particularly in terms of the humanistic attributes of the subject matter, diversity of digital resources, similarity of digital methods, and complexity of the forms of outcomes. Furthermore, as part of our joint analysis, we elucidate some of the primary distinctions between these fields:

1 The mission of digital humanities primarily revolves around knowledge organisation, discovery, and creation, whereas digital memory emphasises evoking emotional resonance or identity through the construction of memory;
2 Digital humanities predominantly focus on data-driven resources mainly comprising text, whereas digital memory employs a range of both movable and immovable resources, incorporating both data-driven and factual expository approaches; and
3 Digital humanities primarily use distant reading as the main method for resource analysis, whereas digital memory employs both distant and close reading, pursuing a memory as close to reality as possible and thus, necessitating detailed reading and cultural analysis of certain resources.

3 Methodologies for constructing digital memory focused on cultural heritage

3.1 Principles for constructing digital memory focused on cultural heritage

Cultural heritage, having gone through the sieve of history, captures and embodies human cultural memory, making it a crucial source of digital memory. Meanwhile, the idea, technology, and methodologies of digital memory for digitising human resources offer novel approaches for online preservation, long-term storage, and innovative dissemination of cultural heritage. Considering the interrelation between cultural heritage and digital memory, the methodology for constructing a digital memory focused on cultural heritage requires both a profound sense of historical and cultural richness and cutting-edge digital technological innovation. The key to constructing such a digital memory includes the principles outlined below.

3.1.1 Multi-tiered interactions among memory agents

Digital memory is a cultural memory constructed through new ideas, technologies, and methods in the digital age. It not only must respect the traces of time retained in the history of cultural memory but also transcend the limitations of the times and reinterpret and convey the value of cultural memory with contemporary cultural logic. This way of constructing is not unique to digital memory, and cultural memory scholars have criticised and explained it as follows: 'This reconstructed characteristic of memory may make it unreliable' because the process of building cultural memory always relies on the 'external memory' of human beings, and the incompleteness of this storage system is probably to blame. Given a long historical period, the stability and singleness of the main body of cultural memory record, to a certain extent, aggravate the systematic deviation of 'external memory' such that 'unreliable' memory lacks outside evidence for mutual verification. In terms of the integrity, continuity, and systemisation of social memory, cultural memory preserved and transmitted by a single cultural memory institution is bound to be incomplete (Xia, 2019). Therefore, when constructing digital memory, the use of 'multiple-evidence reference systems' must be explored to allow multi-level memory subjects and their stored memory evidence to be cross-referenced and verified, so as to ensure the authenticity and integrity of the constructed social memory. Digital memory does not deny the possibility of constructing memory but rather through open-ended construction, it allows incorporating different sources, such as the subjects' recording of historical documents, cultural memory participants, cultural memory inheritance, and transmission subjects, for constructing cultural memory through written records, historical descriptions, oral interviews, and other ways, thus capitalising on the interaction between different memory subjects. Ensuring that rich cultural memory evidence can be mutually verified compensates, to a certain extent, for the deficiencies of single thematic resources and cognition and enriches the expression dimension of memory items (Feng, 2020).

3.1.2 Multi-source and multi-modal cultural heritage integration

The pluralism of the subject of digital memory construction inevitably leads to problems with multi-source and heterogeneous digital memory resources. Among them, the multi-source aspect is mainly reflected in the different memory subjects when forming memory resources, based on the memory content and form of memory being different. Heterogeneity is mainly reflected in the different types, structures, and contents of cultural memory resources, which may manifest in immovable cultural relics such as buildings, sites, blocks, and landscapes; movable cultural relics such as archives and documents, books, newspapers and periodicals, monographs, and handicrafts, and intangible cultural heritage such as rituals, beliefs, religions, family customs, and village rules and conventions. The construction of digital memory, while allowing the participation of multiple subjects, also creates the possibility of integrating multi-modal cultural heritage elements. Looking at the current typical cases of digital memory, almost all are active attempts at the digital integration of multi-modal cultural heritage elements, which can break through the modal and preservation restrictions of cultural heritage and allow gathering of all kinds of text, images, audio, and video, as well as model resources originally scattered among different custodians and locations into specific digital memory platforms or databases. This way, the new cultural heritage data form can realise the splicing, correlation, and fusion of memory and finally construct a relatively complete original appearance of memory.

3.1.3 Multi-platform digital technology connectivity

To realise the convergence of multi-source and heterogeneous cultural memory resources and effectively integrate multi-modal cultural heritage elements, it is necessary to make full use of existing digital technology and realise memory connectivity based on the digitisation and datafication of cultural memory resources. At present, the combination of semantic web technology, social network analysis, machine learning, AI, and other technologies can effectively transform the original digital memory resources in various analogue states into data, carry out in-depth reading, description, identification, and correlation according to the organisational standards of digital memory resources, and finally form a new data system beyond the original content system and narrative logic. In addition, in the display of resources, digital memory must mobilise a variety of technologies and platforms and adopt flexible methods to enhance the interaction between digital memory and the public. Compared with traditional cultural memory, digital memory can better demonstrate the 'synaesthesia' effect of memory by means of text, pictures, audio, video, physical objects, models, and so on. Nowadays, the development of VR, augmented reality (AR), magnetic resonance (MR), and other technologies has brought more opportunities for multimedia integration of digital memory, such as perception, hearing, and touch.

3.1.4 Multi-dimensional narrative iteration

Digital memory not only emphasises the connection and integration of digital memory resources, but also underscores the multi-dimensional representation of the connotation of cultural memory. The essence of digital memory construction is reflective observation and expression of the past, and this process depends on the inclusion of 'present' elements such as the subject's present consciousness and resources. After basic resource convergence and the correlation and integration of digital memory, digital memory will eventually show different aspects of cultural memory in the form of digital narrative and present fragmented memory resources to different audiences in a complete historical context and cultural framework. In the digital narration process, the narrative logic of the cultural memory is neither single nor fixed. Characters, time, place, events, and a certain cultural heritage element can all be used as narrative starting points and can trigger different faceted memory narrations. In the process of digital memory construction, memory resources are constantly enriched and the understanding and interpretation of memory resources are constantly updated. Therefore, the facets of digital narratives are constantly iterated. By combining the advantages of different digital platforms and technologies, digital memory projects can develop different exhibition works or products focusing on different aspects of memory and presenting them in an open and interactive manner, thus increasingly realising additions, extensions, and corrections to memory elements.

3.2 Methods for digital memory construction oriented towards cultural heritage

3.2.1 Investigation and modelling of cultural memory resources

The purpose of cultural memory resource investigation and modelling is to clarify comprehensively the foundation and needs of digital memory construction, that is, to investigate what digital memory should and can be constructed. Generally speaking, the investigation of cultural memory resources includes the investigation of the current situation of cultural memory resources and the investigation of the gap in cultural memory resources; that is, the discrepancy between the 'reality' and 'should' of cultural memory resources.

In investigating the current situation, the type, carrier, format, content, background, quantity, state, preservation, and preservation location of cultural memory-related resources must be identified. A cultural memory resource vacancy survey is conducted to determine which cultural memories should be recorded but lack the corresponding historical evidence. Investigations of both the current situation and the gap in cultural memory resources must be carried out simultaneously. This is the process that allows understanding the connotation and essence of cultural memory. After understanding the state of cultural memory resources, a cultural memory model based on a precise interpretation of cultural memory must be constructed. A cultural memory model facilitates the further management, organisation, mining,

and presentation of cultural memory resources in the digital environment. As the language used in the formation of most cultural memory resources is spontaneous and unconstrained in expressing the connotations of cultural memory, different memory subjects may use different ways of expressing the same cultural memory, or the same expression forms may explain the different memories of various subjects. Digital memory projects need to associate and integrate multi-modal memory resources; therefore, it is necessary to standardise the concept of cultural memory elements and corresponding resources based on a complete understanding of the current situation of cultural memory resources to provide a standard basis for an accurate description, identification, and association of digital memory resources.

3.2.2 Digital memory resource collection and management

Based on a clear investigation of cultural memory resources and accurate modelling of cultural memory, a digital collection of cultural memory can be carried out more efficiently, and a more complete digital memory resource system can be constructed. The digital memory resource system is designed to collect digitally all cultural memory resources that can explain the core value of cultural memory, including basic memory resources such as all kinds of collected historical archives, physical materials, audio and video, and newly recorded memory resources through field research, oral interviews, on-site shooting, and scanning of records. It also includes derivative resources such as an in-depth interpretation of the core of cultural memory around basic resources and creative cultural products created by various humanistic, artistic, and technological means. Both basic and derived resources of digital memory have distinct multi-modal characteristics; therefore, it is necessary to carry out standardised multi-modal resource aggregation and management approaches based on a digital memory resource database.

Most of the physical forms of memory resources collected, such as various historical archives, local records, newspaper articles, and family materials, need to be digitally scanned and read in accordance with the standardised process, and entered into the digital memory resource database. It is also necessary to select specific topics in a planned manner to supplement the original digital memory resources through modern information recording methods such as shooting pictures, making oral histories, producing and editing videos, electronic hand-drawing, and 3D modelling. After all digital memory resources are stored, it is necessary to carry out an in-depth description and memory identification of digital memory resources according to the cultural memory model, create standardised digital memory metadata and label libraries, and form the association and fusion of multi-modal digital resources based on semantic association.

3.2.3 Exploration and interpretation of cultural heritage value

After connecting and integrating the digital memory resources, the value of cultural heritage will surpass the original single entity value and achieve a higher level of ‘play’ through large-scale and connected digital resources. This cultural heritage

value is realised through an in-depth cultural interpretation and a flexible fusion of narratives. Cultural interpretation is the soul of digital memory construction and is the fundamental reason for digital memory projects to go beyond technology and methods and return to the historical significance and social value of cultural memory. Cultural interpretation refers to the understanding, mining, reference, sorting, and expression of cultural memory resources and the memory objects they record. It embodies the deep integration of the interpreter's knowledge, emotion, historical view, and existing resources and is the basis for selecting the themes, cultural structure, presentation mode, and style design for each digital memory project. In-depth cultural interpretation requires in-depth mining of cultural memory resources, and by correlating the data, more resources can be found for an evidence-based and mutual verification until the spatial and temporal scope, historical background, development context, cultural nature, memory structure, and value characteristics of cultural memory are clarified. Therefore, cultural interpretation is not only the micro interpretation of a single memory resource but also the macro observation of the extensive correlation of large-scale memory resources based on resource association and integration, which would reflect the grand historical pattern of digital memory projects and exquisite micro-historical research.

3.2.4 Digital presentation and interaction of cultural interpretations

After fully exploiting the multilayer value of cultural memory resources, digital memory projects should spread and pass down stories of cultural memory through digital exhibitions and interactions. Digital display and interaction are often the last link in digital memory construction, as well as the core link of direct contact with the public. Therefore, the effect of digital display and interaction will directly affect the public's evaluation of the project. The exhibition and interaction of digital memory items must reflect the characteristics of fun, flexibility, openness, and the iterations of digital memory. Among these elements, fun means that the interpreted cultural memory resources can be transformed from serious content into something that the public can enjoy. Flexibility means that different aspects of cultural interpretation can be converted into several topics for digital presentation and interaction, and work with a specific digital platform or product technology. Openness means that any display and interaction of digital memory is not superior 'memory output' but rather leaves enough space to accept new memory subjects and memory resources that may appear, and can accommodate different ways of memory narration to supplement and modify the present display and interaction. Iteration comprises two aspects: the iteration of memory content and the iteration of digital technology. The iteration of memory content arises from the tolerance of digital memory to new memory resources, which requires the exhibition and interaction of digital memory to adapt constantly to the public's demand for memory cognition and emotion, whereas the iteration of digital technology requires the narrative structure, mode, and content of digital memory to keep pace with the times. Constant adaptation to new technologies and platforms may bring about digital memory inheritance and communication innovation.

4 Chinese digital memory application: The 'Beijing Memory' digital platform

The 'Beijing Memory' digital platform stands as the longest-running and largest-scale project among China's digital memory initiatives. This cultural heritage digital preservation and dissemination project was completed through the aforementioned methodologies for constructing digital memory. This chapter introduces the project in four parts: project overview, overall framework and core technologies, project achievements, and future development prospects.

4.1 Project overview

As China's capital, Beijing is a world-renowned historical and cultural city with over 3,000 years of city-building history, and has been the country's capital for more than 800 years. It has rich tangible and intangible cultural heritage. Beijing memory is not just the memory of Beijing residents, it encapsulates traditional Chinese sentiments of family and nation, as well as spiritual ethos, and thus exerts a centripetal force among Chinese culture enthusiasts both domestically and globally. However, during the process of urban modernisation, Beijing underwent dramatic transformations. Traditional cultural remnants are eroding, visible historical landmarks such as the ancient city wall of Beijing and archways have disappeared, and invisible traditional values are on the verge of marginalisation under the critique of instrumental rationality.

Initiated in 2013, the 'Beijing Memory' Digital Platform aims to aggregate the multi-modal digital resources of Beijing's cultural heritage by offering a multi-dimensional digital panorama of Beijing's culture and disseminating new digital cultural experiences. It is currently the largest urban digital memory platform in China. The project was spearheaded by Dr. Huiling Feng, a national first-class professor accredited by China's Ministry of Education. Dr. Feng also serves as the dean of the Digital Humanities Research Institute at Renmin University of China. The project leverages resources from Renmin University's School of Information Resource Management and the Digital Humanities Research Institute by collaborating with multiple departments such as the School of Chinese Studies, School of History, and School of Arts. The project team consists of experts from over ten disciplines, including information science, literature, history, Chinese studies, arts, journalism, and digital technology; they were trained in historical archives, digital records, information management, data mining, and AI. In addition, the project team currently collaborates with various cultural heritage entities and digital culture regulatory bodies such as the National Federation of Literary and Art Circles, Beijing Municipal Archives, Beijing Municipal Administration of Cultural Heritage, Beihai Park in Beijing, and the China Digital Culture Group. It also serves as a collaborative node for UNESCO's Asia-Pacific Heritage Centre, which is part of UNESCO's South–South cultural cooperation initiatives.

4.2 Overall framework and core technologies

4.2.1 Overall framework

The 'Beijing Memory' digital platform employs an overarching architecture that combines a 'Repository of Thematic Cultural Resources' with 'Scenario-based Visualisation Applications'.

The 'Repository of Thematic Cultural Resources' serves as a digital urban memory repository, essentially creating a thematic database that accommodates multi-modal urban memory resources. This repository acts as a foundational and supportive element of the project structure, offering resources for contextualised applications. It facilitates semantic association, data mining, and visualisation to meet the needs of traditional cultural researchers and the general public. Leveraging its resource advantages, the 'Beijing Memory' project team collaborates closely with local governments, archives, and libraries to acquire various forms of digital urban memory resources, including text, images, videos, audio, and 3D models, all of which are standardised and semantically enriched.

'Scenario-based Visualisation Applications' serve as the dissemination platform for urban memory, employing themes or scenarios as 'memory landscapes'. Depending on the memory subjects and scenarios, various presentation methods are employed such as digital storytelling, digital reconstruction, GIS, visualisation, 3D modelling, data modelling, VR/AR, and animated games. These applications build a comprehensive digital dissemination platform for urban cultural heritage elements. The scenario-based visualisation application presentations on the 'Beijing Memory' digital platform are mainly divided into thematic website clusters and a user-generated content website. The thematic websites display Beijing's historical and cultural narratives thematically, thus offering a comprehensive narrative framework and content system for urban memory. Each theme is led by experts with full student participation. The thematic construction team integrates various resources based on in-depth research and uses advanced presentation technologies to create a high-quality, highly experiential, and comprehensive cultural platform that includes websites, mini-programs, H5 pages, short videos, games, animations, and models. There is also a 'My Beijing Memory' interactive space on the platform, which is built through public participation and resource crowd-funding to facilitate a shared cultural experience.

4.2.2 Core technologies

The core technologies and methods are categorised into the following three types.

1 Multi-modal memory resource integration: This approach allows handling not only text data, but also multiple types of data such as images, videos, audio, and 3D models. By integrating and using various types of urban memory resources, a multi-modal urban memory database is constructed. For instance, historical photos,

videos, audio, and documents can be combined to generate a comprehensive multi-dimensional urban memory model that allows the audience to understand and experience urban memory from multiple angles and dimensions.

2 Integration of management and presentation technologies: This approach employs cutting-edge digital technologies for front-end memory presentations and back-end resource management. It combines the strengths of different technologies with specific management and presentation requirements. Digital scanning, 3D modelling, panoramic photography, and photogrammetry are used to collect and digitise cultural memory resources. Technologies such as text recognition, audio recognition, image restoration, and intelligent completion have been applied to read and restore digital memory. Semantic annotation, smart tagging, ontology, and knowledge graphs are used to organise and correlate the digital memory resources. AI, machine learning, and large data models are used for the data mining and analysis of digital memory resources. Digital exhibitions, publishing, collections, and smart apps are methods for the open use and product transformation of digital memory resources.
3 Multi-dimensional display of urban memory: Using the digital narrative theory, this technology applies various means to display and disseminate urban memory. For instance, it can generate rich multimedia content containing text, images, videos, and audio to showcase historical events, landmark buildings, and cultural figures in cities. It can also produce VR or AR models, enabling audiences to experience the city's history and culture in a more realistic and intuitive manner. Moreover, according to different needs and scenarios, urban memory can be displayed in a customised manner, for example, by creating urban memory content specifically designed for various groups such as children, students, tourists, and researchers.

4.3 Project achievements

4.3.1 'Beijing Memory' digital repository

The 'Beijing Memory' project fully leverages the advantages of information science. It has built a 'Beijing Memory' digital repository and employs intelligent, standardised management processes and methods capable of carrying out the preservation and management of urban memory digital resources throughout their entire life cycle. The construction of this digital repository adopts a three-step, fine-grained urban memory digital resource governance plan, termed 'Store–Analyse–Illuminate'.

First, the project achieves standardised entry and management of urban memory digital resources using flexible metadata schemes paired with systematic cataloguing and encapsulation. Second, it performs deep content calculations on urban memory digital resources through interactive data governance, semantic layer data mining, and vectorised data association. Finally, it realises intelligent discovery of the cultural elements of urban memory resources through multi-dimensional categorisation, automatic entity identification, and cross-modal full resource search.

By 2021, the repository already had relatively complete resource management processes, including cataloguing, review, transfer, and approval, as well as related functional modules such as data management, category management, metadata management, data visualisation, and named entity management. It successfully achieved the first-phase goal of 'storing' data. The repository currently has an approximate resource volume of 8.2 TB. After 2021, the repository initiated its second-phase construction, which focuses on data analytics and the computational capabilities of the memory platform. This enables the structuring, semantic understanding, and intelligent management of memory resources, and thereby provides a data infrastructure for the development and research of digital memory resources for AI, data mining, VR, and geospatial and other digital humanities technologies.

4.3.2 'Beijing Memory' digital narrative website cluster

The 'Beijing Memory' Digital Narrative Website Cluster began construction in 2013 with the aim of meticulously excavating and digitally explicating Beijing's urban cultural heritage. The platform uses the Internet to present a digital panorama of Beijing's culture, disseminating new experiences in a digitised culture. It is currently the largest integrated platform for in-depth research on Beijing's urban memory and historical culture, and features multi-dimensional digital presentations. The platform aggregates cutting-edge techniques and tools from the fields of digital humanities and digital cultural heritage and is used to mine, interpret, and recreate digitally representative examples of Beijing's historical and cultural heritage, such as the Grand Canal, Confucius Temple, Wuying Hall, Imperial Historical Archives, the former Qing Army Office, Shijia Hutong, Chinese Royal Ice Sport Ceremony, Ancient City Wall of Beijing, Cuandixia Village, and Peking Opera facial makeup.

All specialised websites are led by experts and focus on specific cultural heritage elements. To date, 22 specialised websites have been launched, and seven are under construction (see Table 8.1). New specialised topics are continuously being proposed. The completed and ongoing topics are presented in Table 8.1. For instance, the 'Grand Canal of Beijing' uses digital techniques like animated modelling and 3D imagery to depict the 700-year history of the Grand Canal in Beijing. The 'Beijing Confucius Temple of Classical Learning' uses digital art to transform complex and hard-to-understand ancient Confucian scriptures, conventionally experienced through linear reading, into immersive multimedia digital spaces. The 'Former Site of Qing Army Office' employs 3D models to recreate the fusion and collision of traditional Chinese culture and Western thought atop the 'Iron No. 1' bricks and tiles.

Each thematic website of 'Beijing Memory' is akin to a super digital publication, characterised by the features enumerated below.

1 Multiple resources: Resources include official archives, books and publications, audio-visual content, oral histories, private documents, and field collection, among others.

Table 8.1 Thematic websites for 'Beijing Memory': Launched and under construction

Completed thematic websites (22)			
Traditional Beijing Oral and Performing Arts	Beijing Confucius Temple	Beijing City Gates	Beijing Gate Piers
Beijing Nursery Rhymes	Beijing Old Professions	Beijing's War of Resistance (Against Japan)	Beijing–Zhangjiakou Railway
Beijing Food Culture	Chinese Royal Ice Sport Ceremony	Former Site of the Qing Army Office	Royal Wuying Hall
Peking Opera Facial Makeup	Yan Nan Garden	Eight Great Water Courtyards in the Western Hills	Shijia Hutong
Cuandixia Village	Lao She (a famous writer)	Beijing Pagoda Tree	Qian Shi Hutong
The Beauty of Antithetical Couplets	Beijing as a 'Dual-season Olympic Host City'		
Thematic websites under construction **(7)**			
Grand Canal of Beijing	Beijing Tickets and Certificates	Imperial Historical Archives	Nei Lian Sheng (a famous traditional shoemaker)
Shui Yu Village	'New Youth'—The Trumpet of the New Culture Movement		The Long Corridor in the Summer Palace

2 Multi-dimensional narration: Various dimensions, including time, space, spiritual, and material aspects are covered.
3 Application of multimedia technologies: As an example, the 'Former Site of the Qing Army Office' uses panoramic VR and WebGL technologies for architectural scanning and modelling. Topics such as gate piers, nursery rhymes, the Chinese Royal Ice Sport Ceremony, and old professions incorporate digital animation. The topic of Chinese Royal Ice Sports Ceremony even includes game interactivity.
4 Multidisciplinary and multistakeholder: The creation of these websites involves cooperation among experts from multiple disciplines such as information science, history, philosophy, arts, and geography. The collaborating entities include Renmin University Archives, Beijing Normal University, Beijing Municipal Archives, Renmin University Digital Technology Company, the web design firm Monokeros, and animation companies.
5 Multiple Benefits: The 'Beijing Memory' Project Team collaborates with large digital publishing groups and digital technology companies to complete the digital publication of topics such as the 'Chinese Royal Ice Sport Ceremony' and 'Peking Opera Facial Makeup'. There are plans to develop 'Beijing Memory' into a digital cultural intellectual property, aiming to generate greater impact and benefits.

By integrating these elements, the 'Beijing Memory' platform offers a comprehensive, multilayered experience that goes beyond traditional storytelling methods,

and delivers a rich, interactive platform for the exploration and appreciation of Beijing's cultural and historical heritage.

4.3.3 'Four-dimensional Beijing' digital reconstruction and exhibition

'Four-Dimensional Beijing' is China's first large-scale digital cultural product focused on 'urban digital reconstruction', developed based on the city's digital memory resources. Supported by the extensive digital cultural heritage platform of 'Beijing Memory', this project draws on historical and cultural resources and scholarly content. It aims to employ GIS, digital modelling, AI, and VR/AR technologies to recreate the evolving urban layout of Beijing from 1046 BC to the present day, covering more than 3,000 years. The result is a four-dimensional representation of Beijing's development history that integrates a physical three-dimensional space with time.

According to statistics, among current international urban digital reconstruction projects, 'Four-Dimensional Beijing' has the broadest geographical scope and longest time span, covering the most types of historical and cultural elements. This project innovatively embeds historical events and specialised topics within the city's digital models, allowing for a visualisation that superimposes historical landscapes through digital exhibitions.

Construction of 'Four-Dimensional Beijing' began in 2019. Up to now, it has completed the digital modelling of Beijing during the Qianlong period of the Qing Dynasty. Ongoing work includes digital modelling and historical event embedding for Beijing during the May Fourth Movement, reconstructive modelling of ancient Beijing's founding sites associated with 'Zhao Gong Feudal Estate', fine-grained embedding of the historical topic 'Grand Canal', and thematic modelling embedding of 'Beijing Temples' and 'Shuiyu Village'.

For just the Qianlong period, the 'Four-Dimensional Beijing' project team has, based on the 'Comprehensive Map of the Capital during the Qianlong Era' and relevant historical data, completed digital modelling that covers 60 km^2 of urban layout, 36 km of the ancient city walls, 16 watchtowers and arrow towers, eight corner towers, an 8-km^2 area on both sides of the central axis, 2,049 roads, over 40 temples, and more than 5,000 three-dimensional building models (see Figure 8.1).

In 2022, the project team initiated a historical embedding and activation project centred on the 'May Fourth Movement', truly realising a city's 'four-dimensionality'. This effort aims to cultivate a digital surrogate for urban culture, establish a common memory platform, and encourage cultural root seeking.

To date, the team has collected 1,354 types of documents related to the May Fourth Movement. These include over ten precious maps of Beijing from the Republican era, approximately 1,178 related newspaper reports from that period, and over a hundred old photographs. The collection also consists of major domestic research works, memoirs, notes, local histories, literary works, albums, paintings, and original foreign material. The team also organised information on more than 300 individuals, marked over 100 key locations, and recorded more than 30 major events.

Figure 8.1 Aerial view of Beijing during the Qianlong period.
Source: Photograph by the author.

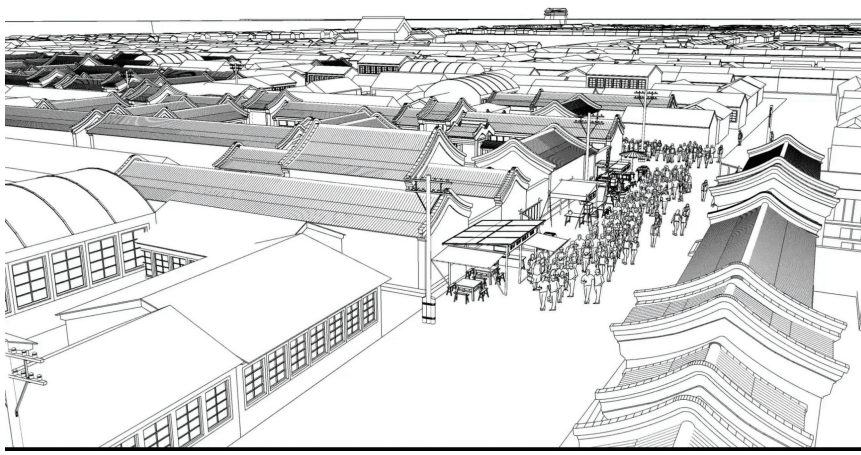

Figure 8.2 National Peking University Normal Department (hypothetical restoration).
Source: Photograph by the author.

Relying on original maps, authentic historical photographs, and meticulously verified historical data, the technical team of the project completed the modelling of relevant areas and buildings in Beijing during that time (see Figure 8.2). They have also 'activated' the historical events of the May Fourth Movement day.

In the future, based on the achievements of the 'Beijing Memory' thematic website cluster and the 'Four-dimensional Beijing' digital reconstruction, the 'Beijing Memory' digital exhibition and experience centre will become a new landmark for the integrated experience of Beijing's historical and cultural heritage. It will also contribute to the sharing of stories about Beijing and China with the world.

4.4 Future development prospects

In future development plans, this project will further optimise and upgrade the application of AI technology to enhance the ability and efficiency of processing and generating city memory content. For example, through more advanced machine-learning algorithms, AI can generate stories and articles with a greater sense of history and local characteristics, thus presenting Beijing's urban memories in a more vivid and attractive manner.

In terms of dissemination and application, the project will employ VR and AR technologies to provide a more realistic and immersive experience of Beijing's history and culture. Audiences can use VR technology to immerse themselves in the everyday lives of Beijing's hutongs and feel the charm of Peking Opera. It can also be used to examine the historical evolution of modern Beijing's landmark buildings and understand the stories behind them.

In terms of objectives, the project aims to become an exemplary digital initiative for urban memory. It seeks to enable more people to understand and feel the history and culture of Beijing while also ensuring better preservation and inheritance of Beijing's urban memories.

5 Discussion and conclusions

5.1 Prospects and challenges of digital memory development

In the digital age, the construction of digital memory has become important for the protection and inheritance of cultural heritage. In particular, in countries such as China, which has a long history and rich cultural heritage, the potential and value of digital memory are self-evident. In recent years, the Chinese government has attached great importance to the application of digital technology in cultural fields. Since 2022, various directives and comprehensive plans such as 'Opinions on Promoting the Implementation of the National Cultural Digitalisation Strategy' and 'Digital China Construction Overall Layout Planning' have been issued by the Central Committee of the Communist Party of China and the State Council. These have provided new opportunities for the construction of digital memory projects with goals like 'comprehensive presentation of Chinese culture' and 'shared digital achievements of Chinese culture for all citizens'.

Additionally, the rapid development of emerging digital technologies such as cloud computing, AI, and blockchain technology has brought about technical opportunities. The digital tools available for constructing digital memory projects are becoming increasingly abundant. In particular, the application of AI technology to text recognition and semantic analysis has enabled large-scale, high-quality resource digitisation.

However, the construction of digital memory projects faces several challenges. For example, legal issues related to intellectual property and privacy protection exist. Project organisers must clarify the ownership of all types of materials and digital products involved in the digitisation process, including whether permission for republishing or reusing is required. This issue not only involves intellectual property law, but may also relate to contract law. Moreover, any digital memory project containing personal or sensitive data must comply with privacy protection laws.

Furthermore, the theoretical foundation of digital memory construction is currently weak. Most builders of digital memory projects are not short of humanistic sentiments and place considerable emphasis on innovative design in practice; however, they pay less attention to abstract refinement and theoretical construction regarding principles, values, functions, thought, and methodology. Strengthening the theoretical construction and methodological research in practice would be beneficial for drawing nourishment from the vast soil of digital culture, thereby enhancing the scientific and standardised nature of digital memory projects.

5.2 Concluding remarks

This chapter, focusing on the construction of digital memory oriented towards cultural heritage, delves deeply into the understanding of digital memory and the methodologies for building digital memory for cultural heritage. By elaborating on the 'Beijing Memory' digital platform as an example, it demonstrates the utility of such methodologies and thereby provides valuable theoretical and practical reference for related fields.

As digital technology continues to advance, digital memory will play an increasingly important role in protecting and passing down a country's cultural heritage. Hence, this topic warrants further exploration. It is also crucial to strengthen cross-disciplinary cooperation and research in areas such as intellectual property, privacy protection, and the theoretical construction of digital memory to promote the continuous development of this field.

Acknowledgments

We would like to express our gratitude for the funding support for this research from the Beijing Social Science Fund's Project 'Research on the Digital Dissemination and Innovation of Beijing's Historical and Cultural Resources' (Project Code: 21ZDA05).

References

Assmann, J. (2011). *Cultural memory and early civilization: Writing, remembrance, and political imagination*. Cambridge University Press.

Beel, D. E., Wallace, C. D., Webster, G., Nguyen, H., Tait, E., Macleod, M., & Mellish, C. (2017). Cultural resilience: The production of rural community heritage, digital archives, and the role of volunteers. *Journal of Rural Studies, 54*, 459–468. https://doi.org/10.1016/j.jrurstud.2015.05.002.

Capital Library. (2013, September 19). *Beijing memory.* https://bjmemory.clcn.net.cn/#/?en=Home&id=1&isDelete=0&status=0&tlevel=1&type=0

Chen, M., & Xiao, P. (2022). Technology adoption and responsible innovation of libraries, archives, and museums (LAM) in the meta-universe era: Centred on NFT. *Library Development, 44*(1), 121–126. https://doi.org/10.19764/j.cnki.tsgjs.20220033.

Comes, R., Tompa, V., Bodi, Ş., Neamtu, C., & Török, K. F. (2018). From theory to practice: Digital reconstruction and virtual reality in archaeology. *Journal of Ancient History and Archaeology. 4*(4). https://doi.org/10.14795/j.v4i4.287.

Feng, H. (1998). Having new memories: Research on electronic document management. *Archives Science Bulletin, 25*(1), 57–59.

Feng, H. (2012). The archival memory view and resources view and the construction of the digital resources 'The Memory of China'. *Archives Science Bulletin, 34*(3), 4–8. https://doi.org/10.16113/j.cnki.daxtx.2012.03.010

Feng, H. (2015). Landscape of memory in the digital age. *China Archives News*, 11–19(3)

Feng, H. (2020). Digital memory: Digital palace for cultural memory. *Journal of Library Science in China, 46*(3), 4–16. https://doi.org/10.13530/j.cnki.jlis.2020019.

Feng, H. (2021). Digital memory from the perspective of digital humanities: Concurrently discussing the method characteristics of digital memory. *Digital Humanities Research, 1*(1), 87–95.

Feng, H., Liang, J., & Ma, L., (2022). Research on the construction of digital memory platforms for ancient villages in Taizhou. *Chinese Archives, 61*(5), 74–75.

Fujiang Municipal Archives. (2023, May 29). *Online Exhibition Hall for Overseas Chinese Letters.* https://www.720yun.com/vr/51322ui5a1s.

General Office of the CPC Central Committee, General Office of the State Council. (2022, May 22). Opinions on Promoting the Implementation of the National Strategy for Cultural Digitisation . https://www.gov.cn/xinwen/2022-05/22/content_5691759.htm

Gorbul, T., & Rusakov, S. (2022). Cultural heritage in the context of digital transformation practices: The experiences of Ukraine and the Baltic States. *Baltic Journal of Economic Studies, 8*(4), 58–69. https://doi.org/10.30525/2256-0742/2022-8-4-58-69.

Hou, Y., & He, S. (2018). The path to integrate and develop urban memory resources from the perspective of digital humanities. *Archives Science Study, 32*(01), 29–34. https://doi.org/10.16065/j.cnki.issn1002-1620.2018.01.005

Jia, H., & Yan, J. (2022). Construction of a digital heritage resource platform based on digital twin technology. *Mathematical Problems in Engineering*, 2022. https://doi.org/10.1155/2022/4361135.

Jia, W., Li, Y., Xu F., & Qu, S. (2019). Review of the construction status and development trends of the Urban Memory Project' in China. *Beijing Archives, 36*(11), 30–31.

Li, C. (2018). Construction of a mobile visual search mechanism for cultural heritage digital libraries based on big data. *Information Studies: Theory and Application, 41*(4), 139–144. https://doi.org/10.16353/j.cnki.1000-7490.2018.04.025

Li, H., & Yang, R. (2022). Bringing the individual back: Digital media, social practice, and imagination in memory research. *News and Writing, 39*(2), 5–15.

Liu, H. (2023). Participation, network, and warehousing: Digital memory construction under the path of memory practice. *Journalism & Communication Review, 76*(4), 60–70. https://doi.org/10.14086/j.cnki.xwycbpl.2023.04.006

Lu, Q. (2007). Collection and management of nonphysical cultural heritage in the local college library. *Library Development, 30*(6), 60–63.

Ma, Y. (2022). Research on the development of Museum Cultural Heritage data resources from the perspective of the digital humanities. *Chinese Museum, 39*(4), 102–106.

Man, S., & Gao, Z. (2023). Digital immersive interactive experience design of museum cultural heritage based on virtual reality technology. *Journal of Electronic Imaging, 32*(1), 011208. https://doi.org/10.1117/1.JEI.32.1.011208.

Ministry of Culture and Tourism, People's Republic of China. (2012, February 24). *The National Library of China plans to launch a 'Chinese memory' project.* https://www.mct.gov.cn/whzx/zsdw/zggjtsg/201202/t20120224_825734.htm

Mohanty, L., & Swain, S. C. (2022). Use of digital technologies by the MSMES to preserve India's cultural heritage and achieve sustainable development goals. *ECS Transactions, 107*(1), 14343. https://doi.org/10.1149/10701.14343ecst.

National Archives Administration of China. (2022, November 07). *'Memory of the World Register - Nanjing Massacre Archives' and 'Rabe's Diary' were launched in Nanjing.* https://www.saac.gov.cn/daj/yaow/201712/fec7cd3eaec348fd8d16bbc205123dcf.shtml

Provincial Party Committee Propaganda Department of Sichuan. (2013, June 15). *Sichuan memory*. http://www.scview.cn/

Qi, T. (2022). From historical archives to historical big data: Exploration based on the 10-year path of Venice Time Machine. *Journal of Library Science in China*, 48(5), 116–129. https://doi.org/10.13530/j.cnki.jlis.2022044

Quintero, M. S., Georgopoulos, A., Stylianidis, E., Lerma, J. L., & Remondino, F. (2017). The CIPA's mission: Digital documentation of cultural heritage. *APT Bulletin: Journal of Preservation Technology, 48*(4), 51–54. https://www.jstor.org/stable/26382567.

Sanders, D. H. (2014). Virtual heritage: Researching and visualising the past in 3D. *Journal of Eastern Mediterranean Archaeology and Heritage Studies, 2*(1), 30–47. https://doi.org/10.5325/jeasmedarcherstu.2.1.0030.

Shandong Museum. (2023, January 22). *Moment in eternity—Oracle culture exhibition.* http://www.sdmuseum.com/art/2022/3/22/art_292605_4702.html?uid=582333&pageNum=2.

Shi, Y., & Zhou, X. (2022). The debate on 'the right to be forgotten' in the era of big data: The ideological dilemma of memory and forgetting. *Editorial Friend, 42*(12), 88–99. https://doi.org/10.13786/j.cnki.cn14-1066/g2.2022.12.012

Siddiqui, M. S., Syed, T. A., Nadeem, A., Nawaz, W., & Alkhodre, A. (2022). Virtual tourism and digital heritage: Analysis of VR/AR technologies and applications. *International Journal of Advanced Computer Science and Applications, 13*(7). https://doi.org/10.14569/IJACSA.2022.0130739.

The Central People's Government of the People's Republic of China. (2023, March 21). *Notice of the Ministry of Housing and Urban-Rural Development and other departments on announcing the list of the sixth batch of villages included in the List of Chinese Traditional Villages.* https://www.gov.cn/zhengce/zhengceku/2023-03/21/content_5747708.htm?eqid=b3d3a40c000019b100000004645891c1

Wachowiak, M. J., & Karas, B. V. (2009). 3D scanning and replication of museum and cultural heritage applications. *Journal of the American Institute for Conservation, 48*(2), 141–158. https://doi.org/10.1179/019713609804516992.

Wang, G., Pei, Y., & Chang, J. (2021). Integration and services of urban memory resources based on digital huminites. *Archives Management, 39*(4), 71–72. https://doi.org/10.15950/j.cnki.1005-9458.2021.04.025

Wu, C., Peng, D., & Meng, Y. (2019) Development strategy and value realisation of cities' sound memory from the perspective of digital humanity: Taking Xi'an as an example. *Library Theory and Practice, 31*(10), 105–109+112. https://doi.org/10.14064/j.cnki.issn1005-8214.2019.10.021

Wu, R. (2019). Integration and protection mechanisms of nonheritage resources from the digital humanities perspective. *Journal of Library Science, 40*(10), 50–54. https://doi.org/10.14037/j.cnki.tsgxk.2018.10.042

Zhang, B., & Li, Z. (2020). Study the digital storytelling-driven model of collection utilisation in libraries, archives, and museums. *Library Tribune, 41*(5), 30–39.

Zhao, N., Li, J., Liu, Q., & Xu, L. (2023). Construction of urban characteristic memory resources from the perspective of digital humanities—Taking the Old Harbin Consulate as an example. *Lantai World, 38*(2), 65–68+72. https://doi.org/10.16565/j.cnki.1006-7744.2023.02.16

Zhao, Q. (2021). Digital unification: New idea and discourse of Chinese International. *Cultural Heritage Protection Library Development, 44*(4), 178–184. https://doi.org/10.19764/j.cnki.tsgjs.20210317

Zhao, X. (2022). Use of interactive data visualisation technology in cultural digital heritage displays using edge computing. *International Journal of Grid and Utility Computing, 13*(2–3), 118–127. https://doi.org/10.1504/IJGUC.2022.124396.

Zhou, L., & Cui Y. (2021). Development of rural archival memory resources for digital humanities: Value, mechanism and path selection. *Beijing Archives, 38*(10), 10–14.

Zhou, W., Wen, L., Zhao, J., &Yang, L. (2023). Preserving artificial intelligence as digital heritage: Prospects for actions in the archival field. *Beijing Archives, 40*(4), 10–14. https://doi.org/10.3969/j.issn.1002-1051.2023.04.005.

Zhou, Y., Zhao, Y., & Sun J. (2017). Research path for intangible cultural heritage information resource organisation and retrieval: Investigation and design based on the ontology method. *Journal of Intelligence, 36*(8), 166–174.

9 China's exploration and experience in digital twinning of tangible cultural heritage

Yi Su, Ziqi Zhou, Liang Zhao and Yujue Wang

1 Introduction

In recent years, with the popularisation and development of the Internet and digital technology, digital techniques have played an increasingly important role in protecting and inheriting tangible cultural heritage. Digital museums provide channels to view exhibitions online, and technologies such as augmented and virtual reality enhance the visitors' experience by breaking through time and space for visitors, turning physical heritage into replicable and editable data, and creating encounters between artefacts and people in virtual spaces. Achieving a higher level of tangible cultural heritage dissemination using emerging digital technologies has an important theoretical and practical value.

The Digital Twin is a concept introduced at the beginning of the 21st century, with a conceptual model, originally proposed by Professor Michael Grieves of the University of Michigan (Grieves, 2002, 2015). The concept was first adopted for tangible industrial products and has since expanded to all manner of products and services (Grieves, 2023). The digital twin technology has been widely used in aerospace, military, industrial manufacturing, and simulations. By recording high-precision, multi-dimensional, all-aspect dynamic simulations of entities in the physical world, digital twin can establish virtual models according to certain mapping rules and realise real-time interaction and association between virtual models and physical entities (Yang et al., 2022). Currently, intelligent examples such as digital twin cities (Deng et al., 2021) and medical digital twin systems (Armeni et al, 2022; Haleem et al., 2023) have emerged. The theoretical frameworks and technological tools for digital twin have matured (Qin & Zhang, 2018). The concepts and terminology of digital twin in this chapter follow International Organization for Standardization ('Digital twin Concepts and terminology', 2023), mainly including digital twin, digital entity, and digital twin system. The digital twin is digital representation of a target entity with data connections that enable convergence between the physical and digital states at an appropriate rate of synchronisation. The digital entity is a computational entity that comprises data elements and procedural elements, which is part of digital twin. The digital twin system provides functionalities for the digital twin composed of inter-operating target entities, digital entities, data connections, and models, data and interfaces involved in

DOI: 10.4324/9781032707211-11

the data connection process. Besides, digital twin technology refers to a range of technologies used to implement digital twin (Liu et al., 2021).

Digital twin technology provides unprecedented opportunities for the inheritance, protection, and utilisation of tangible cultural heritage sites. Tangible cultural heritage has the characteristics of being immovable and non-renewable. By expressing, describing, and modelling the shape, state, and characteristics of these physical entities, digital entity can accurately restore the physical entity of heritage in the virtual space and reproduce the real state of the tangible cultural heritage (Qin & Zhang, 2018), so as to achieve 3D displays, reproduction, and dissemination beyond the limitations of space and time and to facilitate systematic research while enlarging the dissemination audience. Furthermore, the digital entity can simulate various physical attributes of a heritage entity and carry out condition monitoring and restoration tests on the twin to overcome the preventive protection problems brought about by immovable cultural relics and ancient architectural complexes that are immovable (Yang et al., 2021). The twin can also monitor the facilities and passenger flow of scenic spots in real time, simulate emergency scenarios and activity arrangements to achieve intelligent management allocation and operational stability (Qi & Wu, 2022; Yu, 2022; Zhang, 2022), and provide decision-making assistance and optimisation strategies for multiple cultural heritage services. Digital twin in cultural heritage digital 3D recovery, virtual museums (Li et al., 2017), and intelligent scenic spots have great practical value in heritage protection and inheritance, cultural education, and cultural services. In recent years, application framework designs and discussions of various types of cultural heritage digital twin in the sectors of science and technology, culture and museums, have increasingly contributed to the protection and inheritance of cultural heritage, enriched the connotation of cultural heritage digitisation, and expanded the path towards its realisation as well as the space for its development, thus realising the trend of digitisation and intelligentisation (Chen & Li, 2021; Shen & Lv, 2022; Zhan, 2022).

As one of the four ancient civilisations, China has a splendid history and culture. Museums, heritage sites, and scenic spots, abbreviated as MHS, are three typical carriers of tangible cultural heritage protection, management, and services. Large-scale comprehensive museums such as the Palace Museum aim to integrate all three carriers and multiple operation and management roles. However, this goal poses significant challenges to the service and application of the digital twin of tangible cultural heritage. Consequently, this study selects and analyses typical examples of applications of digital twin to tangible cultural heritage sites in China—the Shanghai History Museum, the Longmen Grottoes, and the Yellow Emperor's Native Place—under the MHS perspectives. Notably, the selected examples have developed their own digital twin platforms. The study further employs the Palace Museum as a typical representative that combines its cultural heritage digitisation and intelligent computing application practices on cultural relics protection, visitor services, and operation and maintenance to explore China's pre-exploration of application of digital twin in large-scale comprehensive museums. The basic framework and operation logic for the integration of digital twin and tangible cultural heritage through application case studies is then expounded. Eventually, based on

the need for integrated twinning of the MHS tangible cultural heritage, four aspects were employed in arriving at the integration of digital twin and tangible cultural heritage, namely, application scenarios, data models, relevant technologies, and cultural value, as reference points for the theoretical construction and associated processes of development of tangible cultural heritage digital twin.

2 Development of digital twin research for tangible cultural heritage

This section considers the three carriers of tangible cultural heritage—MHS—as a clue to illustrate the progress of digital twin research on tangible cultural heritage at home and abroad, respectively.

2.1 Digital twin research for museums

Digital twin research for museums is mainly applied to the digital reproduction of cultural relics, construction of digital resource platforms, and restoration and protection of cultural relics. In addition, museums have an obligation to manage relics and serve visitors; therefore, the digital twin research also improves management. From the theoretical framework, Lin (2022) discussed multiple application scenarios of digital twin museums in planning and construction, operation and maintenance, display and exhibition, archaeological excavation, propaganda activities, cultural relic restoration, coordination, and control. Xiao (2022) proposed a new paradigm of creating a digital entity and building a digital resource platform, such as a digital twin museum, for rare Chinese opera and folklore cultural relics in overseas collections; Wang (2018) analysed the necessity and feasibility of constructing a digital entity of underwater cultural relics and heritage to better present the underwater cultural relics. Dunhuang Academy established a globally shared database for massive, multi-source, and heterogeneous digital information resources, including not only unstructured textual and visual resources for caves, murals, coloured sculptures, and expert manuscripts, but also documentary resources for Dunhuang studies. On that basis, Dunhuang Academy realised semantic enhancement with smart data methods thereby enriching the limited resource organisation system framework (Wang et al., 2020). In addition, the Guizhou Dragon fossil (Zheng, 2021), stone statues (Wang, 2021), and other cultural relics have also been integrated into the concept of digital twin to achieve digital reproduction, 3D printing, and virtual restoration. The Shanghai Museum of History built China's first digital twin museum system, which provides an ultrafine 3D digital restoration of the museum and its surroundings (Zhou & Tan, 2022). The system has several scenarios for management and services, including guided tours, monitoring, and security. The Natural History Museum in the UK was among the first museums to adopt digital twin to monitor the environment of the artefacts through sensor data, applying digital twin in areas such as restoration of cultural relics and risk assessment (Richardson, 2020). The Borghese Museum in Rome analyses visitor trajectory data to understand visitor behaviour and visit needs (Centorrino et al., 2021).

2.2 Digital twin research for heritage sites

Existing digital twin research on heritage sites has proposed application scenarios for historical reproduction, digital resource platforms, and virtual reality interactions based on a three-dimensional presentation. Sang et al. (2021) divided the digital twin application scenarios of the Great Wall cultural heritage into three sections: the online platform and mobile terminal of the Great Wall National Cultural Park, the Great Wall digital twin network management platform, and the digital exhibition hall of the Great Wall National Cultural Park venues, elaborating a 'whole life cycle' digital twin structure that integrates the evolution of the historical process and the simulation of the future construction. In his research on the application of digital twin in the protection of ancient villages, Tang (2021) proposed the construction of dynamic village twins based on spatial and temporal information through digital twin to reproduce the history and evolution of villages. At the same time, the interactive experience of real and virtual interaction via time travelling and immersive experience was realised, while Song and Yuan (2021) studied the digitisation of the capital's historical street and alleyway cultural resources, and elaborated the construction of digital twin in the urban area based on modern civilisation, traditional cultural resources, and the cultural relics of the hutongs and alleyways. The Notre Dame de Paris restoration project constructed a 3D building information model based on a pre-fire model, which was applied to the intelligent management planning of the reconstruction works (Veyrieras, 2019). The Monastery of Simon Petras in Greece introduced the extended reality (XR) spatial experience, which combines static architectural model data generated by photogrammetry with dynamic user data collected via sensors and VR headsets to provide visitors with an immersive experience where virtual meets reality (Vlavianos & Nagakura, 2021).

2.3 Digital twin research for scenic spots

Research on the application of digital twin to tangible cultural relics not only considers the twinning of the artefacts themselves but, more importantly, adds the perspective of the manager. Typically, scenic spots are established based on immovable cultural relics, such as ancient sites and buildings, in order to organise tourist visits. Consequently, some scenic spots have applied digital twin to their operations and management. The digital twin platform of the Yellow Emperor's Native Place has, on a replica of 1:1, restored the scenic area and produced an online three-dimensional wisdom park and visualisation platform (Qi & Wu, 2022). The platform can grasp the situation at the spot and the working status of the equipment in real time, in addition to comprehensive analysis, intelligent applications, emergency commands, intelligent property, scenic area marketing, underground pipelines, and other business functions. Figure 9.1 displays the research regarding the application of digital twin to tangible cultural heritage.

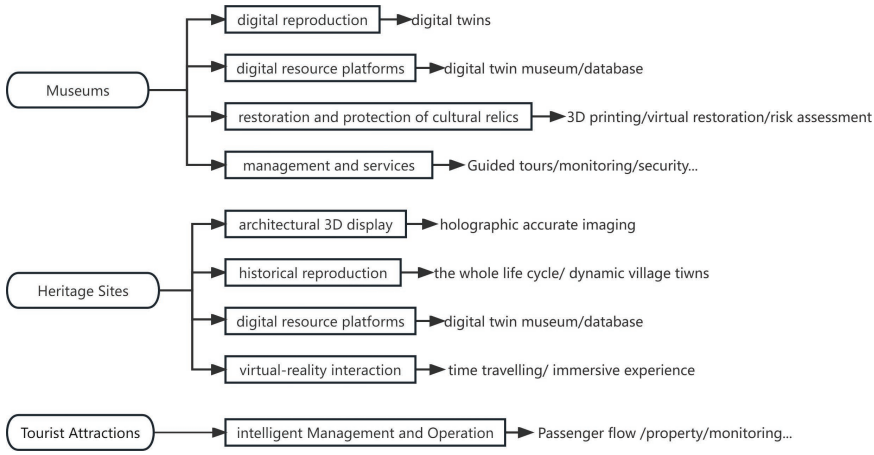

Figure 9.1 Research on the application of digital twin to tangible cultural heritage.

3 Typical cases of digital twin of Chinese cultural heritage in a single format

3.1 Case study of a museum: Shanghai History Museum

The Shanghai History Museum, established in 1984, is a prominent topographical museum that provides a comprehensive and synthetic portrayal of Shanghai's revolutionary history. The building itself is an important historical monument in Shanghai. It has a collection of approximately 110,000 items, mainly transferred from higher authorities, other public institutions, market collections, auctions, and donations. The collection covers 15 categories, namely, paintings and calligraphy, metals, ceramics, handicrafts, certificates and seals, documents, printed matter, textiles, stone carvings, coins, photographs, paper cuttings, stamps, records, and other miscellaneous items.

In September 2021, based on the framework of Shanghai's integrated urban operation system and the concept of urban digital operation, the Shanghai History Museum officially launched the Digital Twin Museum System, marking a significant milestone in the digital transformation of China's museum industry. The museum conducts ultra-high-resolution 3D digital restoration of its internal and surrounding environments. Utilising a powerful digital twin engine, the museum achieved a 1:1 digital replica of its architecture by integrating multi-dimensional real-time dynamic data from Internet of Things (IoT) sensors, government and museum operational data, and environmental meteorological conditions. Meanwhile, by adding regular management mechanisms and process control through the algorithm and model training, the system becomes a 'smart brain' and can promote the museum's more efficient high-speed operation. This comprehensive integration empowers venue management and enhances the visitors' experience.

The Digital Twin System was designed with seven major sections: a digital history museum, digital services, museum perspective, museum sensation, museum voice, urban pulse, and digital cultural preservation. It encompasses 14 digital scenarios, including exhibition guides, dynamic emergency plans, virtual and physical video inspections, message walls, cultural heritage protection, fire safety, security protection, historical building preservation, surrounding traffic guidance, surveys, energy analysis, hotlines and information analysis, and visitor flow monitoring. The application scenarios and services of the system are detailed in Table 9.1.

Table 9.1 The application scenarios and services of the Digital Twin Museum System of the Shanghai History Museum

Module	*Application scenario*	*Services provided*
Digital management and services	Exhibition overview	Introduces the museum combining physical models
	Collection statistics	Introduces information about the museum's collections
	Hot news	Provides the latest news
	Security duty	Daily security arrangements and relevant contact information
	Public service space	Provides availability and usage status of public service areas
Digital cultural preservation	Showcase environment monitoring	Classifies by floor, displays real-time environmental monitoring values for each showcase and humidity fluctuations
	Exhibition hall environment monitoring	Geographically displays environmental monitoring data for exhibition halls and storerooms on each floor
	Cultural Relic protection equipment monitoring	Uses a 3D model map to update the operation and maintenance status of cultural relic protection equipment systems in real time
	Open-air cultural Relic monitoring	Monitors exposed cultural relics through visual surveillance
Visitor service experience	Real-time guide	The intelligent system can call out a route map via voice assistant, guiding visitors to the needed exhibits. Upgraded in version 2.0 to a guide system on the WeChat mini-programme platform, it navigates not only to indoor facilities but also offers remote visiting and on-site viewing modes
	Real-time crowd management	Presents the visitor heat map of the exhibition area using image recognition systems, cloud computing capabilities, and smart models
	Real-time service data	Obtains service data related to exhibitions, exhibits, cultural and creative shops, etc., via mobile terminals
	Comments	The museum accumulates visitor feedback through the Digital Twin system, understanding visitor needs to improve museum services

Focusing on the museum's artefacts and architecture, the Digital Twin Museum System aims at dynamically building capabilities for cultural relic protection. This includes four scenarios: a showcase of environmental monitoring, exhibition hall environment monitoring, monitoring of cultural relic preservation equipment operations, and monitoring of exposed relics, primarily utilising digital technologies like IoT for real-time monitoring and early warning for museum collections and buildings. This system also includes monitoring sensors for high falls, settlement, and tilting of the building. On one hand, it achieves comprehensive monitoring. The powerful computing and compatibility of the digital twin base have unified previously scattered and isolated data, enabling rapid acquisition of monitoring results and 24-hour real-time monitoring of both the macro and micro-environment, small artefacts, and large buildings. On the other hand, it optimises energy consumption. By correlating outdoor temperatures with indoor air conditioning temperatures, and obtaining the museum's layered, time-segmented, and area-specific energy consumption, the system can scientifically plan annual energy-saving schemes. This function was further enhanced in the upgraded 2.0 version in February 2022, providing trend displays for the past 7 and 30 days, monitoring energy consumption.

Regarding visitor service experience, the Digital Twin Museum System's interactive screens can activate artificial intelligence services through voice commands, catering to diverse visitor needs.

In terms of museum management and services, the digital service module improves the museum's management efficiency and resource utilisation through eight sections, including exhibition overviews and collection statistics.

Through the digital twin museum system, the Shanghai Museum of History has achieved a digital transformation that enhances management effectiveness, the protection of cultural relics, and the ability to provide visitor services. This digital transformation not only enhances the operational efficiency and management level of the museum, but also improves the visitor experience, making the museum more modern, intelligent, and sustainable. However, the twin construction of the Shanghai History Museum currently focuses mainly on the digital restoration of the museum building and its surrounding environment, providing corresponding management and services. It emphasises integration with smart city initiatives, reflecting the characteristics of intelligent buildings. There is still considerable room for expansion in the digital preservation and display of the museum's collection of cultural relics.

3.2 Case study of a heritage site: The Longmen Grottoes

The Longmen Grottoes, located in Luoyang City, Henan Province, are a treasure trove of huge stone carving art and many statues; they are rated by UNESCO as the highest peak of stone carving art in China. Longmen Grottoes rank among the major grottoes in China; moreover, they are a World Heritage Site and a key national cultural relics protection unit. Longmen Grottoes were first excavated in the Northern Wei Dynasty during the reign of Emperor Xiaowen, flourished during

the Tang Dynasty, and finally, at the end of the Qing Dynasty. After more than ten dynasties, and more than 1,500 years, there are now 2,345 niches in the caves and more than 110,000 statues, most of which were built for the royal family. The Longmen Grottoes are a record of many historical accounts, and because the grottoes were excavated by important figures from different historical periods and regions, they contain a variety of architectural styles and value for cultural exchange.

In March 2021, to serve tourists and disseminate culture, Longmen Grottoes launched the construction of a Smart Culture and Tourism Digital Twin Platform project, which uses digital twin technology, cloud computing, the Internet of Things (IoT), Big Data, artificial intelligence, virtual reality, and other technologies. The platform integrates multifunctionality and multiple scenarios, such as data integration, panoramic display, territory-wide management, coordinated planning and public services. The platform digitally presents the main Buddha area and its surroundings and is committed to improving the digital level of scenic spot management, operations, and marketing.

The digital reproduction of the Longmen Grottoes was divided into three levels. At the first level, rendering engine, interaction engine, real-time light and shadow, digital restoration, and other technologies are used to carry out medium-precision restoration of the 31.7 km^2 of topography, traffic networks, cultural relics, mountains, fields, forests, rivers, and ecological vegetation around the scenic area. In addition, a 3D model of the surrounding villages was built. At the second level, laser-point cloud technology was used to scan and objectify the grotto area with medium precision, and its appearance and texture were restored. At the third level, Buddha statues, such as the Lushena Buddha of the Fengxian Temple, are restored with a high precision of 1 mm in structural accuracy and 1 mm in textural accuracy to achieve the purpose of information retention and protection and the production of imitations. On the basis of restoration, Longmen Grottoes also carry out digital restoration work. Through scientific restoration and recording of cultural relic information, digital archives can be established such that cultural relics can be a near-equivalent restoration and sustainable presentation, providing digital preservation of cultural heritage.

With regard to management, service, and marketing, the Longmen Grottoes Intelligent Cultural Tourism Digital Twin Platform creates a one-stop solution for the entire process, from digital management to intelligent operation to branded marketing.

First, in the realm of digital management, by amalgamating the data resources of existing information systems within a scenic area, the digital twin cultural and tourism platform achieves real-time precision and digitalised management of the scenic area across networks, platforms, and regions. Through this platform, all resources and apparatus of the scenic area can be instantaneously observed and controlled via a unified interface, akin to a panoramic map. This approach not only facilitates constant temporal and spatial monitoring of the daily operations of the scenic area but also expedites synchronised responses to unforeseen incidents.

Second, regarding intelligent operations, the platform integrates infrared observations, video recognition, and sensor data. It incorporates spatiotemporal data

intelligence analysis based on digital twin technology by leveraging methods such as Big Data analysis and AI algorithms, precise calculations, perceptions, monitoring, and prediction of the visitor traffic, distribution of visitors, behavioural tendencies, and environmental and commercial activities within the scenic area. The platform substantively bolsters scientifically guided operations and decision-making optimisation for scenic areas.

Third, in regard to brand marketing, the digital twin is utilised to digitally restore the cultural heritage of a scenic area, thus crafting a digital archive that metamorphoses tangible cultural heritage into intangible digital assets. By delving deeply into the cultural connotations inherent in digitalised scenes, a cultural heritage intellectual property (IP) ecosystem is nurtured, revolving around creative industries. This initiative serves to fortify economic benefits. Furthermore, through the convergence of cutting-edge technologies such as VR, AR, and AI, a revitalised touring experience is provided to visitors, infusing renewed vitality into cultural heritage. Such convergence orchestrates the exploitation of the expeditious dissemination capabilities, extensive outreach, and elevated user engagement associated with new media, a multifaceted dissemination and marketing of the cultural heritage IP within the scenic area, leading to a strategy for the cultural heritage to transcend into a prominent brand.

Relative to the Shanghai History Museum's digital management of buildings and environments, as a World Cultural Heritage, the Longmen Grottoes focus more on high-precision digital records and virtual restoration of artefacts to achieve digital preservation of cultural heritage.

3.3 Case study of a scenic spot: Yellow Emperor's native place

The Yellow Emperor's ancestral haven, nestled in Henan Province, serves as a celebrated realm in which the lineage of the Xiong family has been etched into historical records. It stands as the site of the birth, pioneering endeavours, and regal establishment of the progenitor of Chinese civilisation, Emperor Xuanyuan, known as the Yellow Emperor. During the Han Dynasty, the Xuanyuan Temple was erected, while the Xuanyuan Bridge was constructed during the Ming era. A monument, bearing the inscription 'Homeland of Xuanyuan Valley', was erected during the Qing Dynasty. Expanded through numerous developments, the area now encompasses a sprawling expanse of 70,000 m^2, its overarching design resonating with the theme of the roots of the Chinese nation. Among the revered sanctums for descendants of the Yellow Emperor, both at home and abroad, it is a famous scenic spot in the area where each year, on the third day of the third lunar month, many gather to honour Emperor Xuanyuan.

The Digital Twin City Information Model (CIM) platform of the Yellow Emperor's Native Palace, born from the ethos of digital twinning, ingeniously fuses the realms of 3D visualisation and City Information Modelling (CIM). It has emerged as a realm of intellectual three-dimensional visualisation, restoring the intricacies of landscapes and internal structures within architectural layers at true 1:1 magnification. The Digital Twin CIM platform divides its functional domains into six

sectors: Comprehensive Analysis, Intelligent Applications, Emergency Command, Smart Property Management, Scenic Area Marketing, and Subterranean Utility Networks. This nucleus, which serves as the nerve centre of the park, diligently monitors and maintains the operational symphony of the entire domain.

Within the realm of Comprehensive Analysis, a treasure trove of features includes prominent spot rankings, analysis of equipment alert frequencies, real-time scrutiny of entrance and exit passages, peak periods of security events, temporal patterns of restroom footfalls, and origin analysis of visitor flows. Overall, the platform can monitor and analyse people and equipment in the landscape.

The sphere of Intelligent Applications orchestrates 3D multi-scenario, multi-perspective utilities, in which the park exercises its managerial prowess through diverse scenes. It encompasses several functions: enveloping the realms of intelligent security, fire management, alerts, environmental monitoring, visitor flow analysis, energy consumption oversight, illuminative sagacity, access control, security patrols, intelligent lavatories, public broadcasting, and person and object tracking.

In the domain of intelligent security, the tapestry of intelligent video surveillance systems paints a spatial tableau of the park's camera layouts and operational statuses. This artistic tapestry enables trajectory tracing, heatmap of footfall intensity, focal region monitoring, anomaly detection, and ingress/egress thermal monitoring. The fabric of intelligent fire management delineates the structural tapestry of the park, interior floor plans, escape routes, and spatial embroidery of firefighting command centres and indoor/outdoor firefighting installations. Should exigencies emerge, swift and precise locations of incidents and the evacuation of surrounding areas are facilitated. Real-time environmental monitoring includes real-time displays of environmental temperature, humidity, PM2.5 concentrations, and ambient noise levels. Visitor flow analysis involves interlocking with ticketing systems and gleaning real-time ingress and egress data. Upon exceeding the designated thresholds, the symphony of security systems resounds in harmony, unleashing the overture of emergency command protocols. Key venues resonate with a symphony of real-time visitor density displays, echoing a refined analysis of visitor demographics across various timespans.

The Emergency Command realm embodies an orchestra of comprehensive coordination, incident assessment, protocol management, emergency drills, and archival dispositions. Within this grand symphony, pre-event risk evaluation and early warning notifications are harmonised, accompanied by auxiliary decision-making and system synchrony during events. Real-time commands, monitoring, and broadcasting resources weave a masterpiece of emergency orchestration.

The Smart Property Management tapestry includes threads of maintenance oversight, inspection management, work-order supervision, and equipment upkeep. This fabric weaves a seamless tapestry of online service request queues and systematic task dispatch, thus ensuring seamless stewardship of the park's operational symphony.

Scenic Area Marketing manifests as a digital renaissance, casting forth three-dimensional imagery of the park's panorama. In parallel, it harmonises with the

Table 9.2 Adaptation of typical cases to theoretical studies on the application

	Museums	*Heritage sites*	*Scenic spots*
	Shanghai History Museum	*The Longmen Grottoes*	*Yellow Emperor's Native Place*
Digital reproduction/3D display	√	√	√
Digital resource platforms		√	
Restoration and protection of cultural relics		√	
Historical reproduction	√		
Virtual-reality interaction	√		
Management and service/intelligent operation	√	√	√

park's network and the expansive Internet, synchronously broadcasting panoramic magnificence while inviting immersive virtual reality exploration.

The Subterranean Utility Networks tableau is choreographed to the cadence of underground pipeline CAD schematics, the birth of three-dimensional dioramas, and their intricate narrative unspooling through the platform's panoramic canvas. Additional supporting functions include a symphony of the subterranean pipeline three-dimensional simulacra, elemental attribute enquiries, real-time data vigilance, and repair status scrutiny.

Evidently, the digital twin CIM platform of the Yellow Emperor's hometown scenic area focuses on the management perspective, and the services provided to the tourists are limited to simple scenic guides, with lower accuracy of digital restoration of cultural relics. Table 9.2 summarizes the aadaptation of the three typical cases and the applied theoretical research.

4 Pre-practice explorations of digital twin in MHS-comprehensive museums

Comprehensive museums, such as the Palace Museum, encompass a vast array of cultural heritage, including extensive ancient architectural complexes and collections of cultural artefacts, consolidating MHS identities into one. Comprehensive museums face the urgent challenge of safeguarding a large volume of cultural heritage and encounter the complex challenges of multi-role operational management. They must consider the new eco-audience service from the perspective of museums, the preventive conservation of cultural relics from the perspective of heritage sites and the intelligent collaborative management from the perspective of tourist attractions. Therefore, the digital twin of the MHS-comprehensive museums is not just a simple collocation of digital twin services in the form of museums, heritage sites, and scenic spots. The systematic digital twin service for all scenarios, multiple roles, and multiple demands requires more rigorous and detailed design and architecture and is still in the early exploration stage. To the best of our knowledge, there is no mature digital twin application of the MHS comprehensive museum for

all scenarios, roles, and needs. The Palace Museum, one of the globally renowned MHS-comprehensive museums, holds a pioneering position in exploring digital twinning. It has accumulated several mature and practical experiences in heritage restoration and preservation, digitalised audience services, and digital operational management, all of which can directly serve the digital twin system, culminating in the preliminary exploration of comprehensive digital twin services.

The Palace Museum is a large-scale comprehensive museum that integrates ancient architectural complexes, palace collections, and culture and art through the ages, established based on the imperial palaces of the Ming and Qing dynasties and their collections. Covering an area of more than 1 million square metres and preserving more than 8,700 ancient buildings and over 1.86 million pieces/sets of collections, the Palace Museum is the largest and best-preserved complex of ancient wood-framed palace complexes in the world. The Palace Museum has many valuable, movable, and immovable cultural relics. Therefore, it is an important representation of China's tangible cultural heritage, as inscribed on the UNESCO World Heritage List in 1987. It's worth mentioning that the Palace Museum received 19.34 million visitors in 2019, making it the most visited museum in the world.

The rich collection resources and unique historical and cultural values of the Palace Museum have brought about high foot traffic. However, there is always a tension between preservation and accessibility because of the limited carrying capacity of cultural heritage, so it has been a long-standing consensus to utilise digital means to archive and openly share cultural heritage resources and provide digital services to the public. The digitisation construction history of the Palace Museum can be traced back to the establishment of the Resource and Information Centre in 1998. To date, it has undergone 25 years of development. Digitisation of the Palace Museum has laid a solid practical foundation for the application of digital twin for heritage conservation and audience service levels. Therefore, this section analyses the pre-practice exploration of the Palace Museum from three dimensions: heritage digital archiving, preventive monitoring and preservation, and digital audience services.

4.1 Heritage digital archiving

The conservation of buildings and cultural relics is the primary task of the Palace Museum and forms the basis for various steps. In 2002, the Palace Museum launched the 18-year 'Overall Restoration and Conservation Project of the Palace Museum', which is the largest restoration activity of the Palace in a century. The lifespan of ancient buildings was maximised without altering the original state of the heritage (Yu, 2023). Digitally archiving and further organising the tangible cultural heritage based on the restoration and conservation of cultural objects are also important initiatives for salvage conservation.

Since 2000, through the cooperation with the Toppan Printing Company of Japan to carry out the 'Cooperative Study on the Digital Application of the Cultural Assets of the Palace Museum', the Palace Museum has continued to carry out high-precision three-dimensional digitised collection of the central axis area, the Hall of Mental Cultivation, the Studio of Weary Diligence, the Garden in the Palace

of Tranquil Longevity, Palace of Lasting Fortune, and the archaeological site on the east side of the Garden in the Palace of Benevolent peace, among other areas. As an example, in 2015, the Research Conservation Project on the Hall of Mental Cultivation was established, wherein the Palace Museum upholds the following three major principles: (1) maximum retention of the historical information of the ancient buildings, (2) do not change the ancient architecture of the original heritage element, and (3) inherit the traditional repair techniques of ancient buildings in the process of restoration. The project explored the historical and cultural values of built heritage, assessment of the current state of heritage, conservation and display of heritage, data management and virtual display of image collection, and future management and monitoring. Since 2015, the Palace Museum also continued to carry out a 360-degree panoramic image of the data collection work, and has completed all the open areas of high-precision panoramic image recording work.

4.2 Preventive monitoring and preservation

In March 2005, the Palace Museum established the ancient architecture of the science and technology protection working group, responsible for ancient buildings and their restoration projects in science and technology protection projects, the ancient architecture of environmental monitoring, and the testing and use of new materials and technologies. The Palace Museum conducts 'non-movable cultural relics body deterioration risk monitoring and analysis of technology and equipment research and development' projects. The project begins with the need for preventive conservation of cultural relics and an assessment of their current status. By producing a series of equipment to monitor the deterioration of immovable cultural heritage in different types of environments, the project establishes a cultural relics risk prediction and traceability model as well as the formation of a systematic monitoring and early warning system solution for the ancient architectural complex of the Palace Museum.

The 'Research and Demonstration Project on Scientific Cognition of Building Techniques of Ming and Qing Dynasty Official Buildings and Key Technologies for Ontological Protection' is an important study of construction techniques, traditional materials, disease mechanisms, conservation and restoration, and management and display. The project brings about conceptual and technological advances in architectural heritage conservation, enables intelligent sensing of potential risks in hidden parts of the building, and constructs an information management technology system based on records of the entire process of protection of official Ming and Qing buildings (Wang, 2023).

In response to the building's safety risk potential, the Palace Museum has implemented a safety risk early warning key technology research project based on multi-dimensional characteristics and historical risk data, such as the scale, density, structure, material, overall space and landscape of the buildings, and the number and preservation of movable cultural relics. This project studies the index system for the safety risk assessment of ancient buildings and compiles the safety risk determination and early warning system as well as the comprehensive prevention

and control guidelines. It also constructs an early warning algorithmic model for the risk analysis of abnormal personnel behaviour to enable dynamic and accurate counting of high-density crowds. Moreover, combined with the ancient building's own attributes and historical risk data, the ability of dynamic fire risk perception and the analysis of early warning methods under the influence of different seasons, time periods, and complex weather are studied. Eventually, the Palace Museum constructed a security risk perception and early warning platform for ancient buildings based on artificial intelligence and Big Data technologies.

4.3 Digital audience services

To better meet the audience's needs for cultural consumption, since the official website was launched in 2001, the Palace Museum has provided a diverse range of digital services and created a series of both interesting and practical cultural products so that buildings and cultural relics are able to 'walk' out of the Palace in the virtual space to activate the new value of heritage. Since 2013, the Palace Museum has released more than ten 'Palace Museum Production' series of apps, including 'Daily Palace Museum', 'Forbidden City 365', 'Palace Museum Exhibitions', 'Palace Museum Ceramics Museum', and others (The Palace Museum, 2022). These apps allow audiences to view cultural relics and learn about them anytime and anywhere from home. They utilise various digital interactive technologies, such as in the 'Han Xizai Evening Banquet' app, to deeply interpret the value and cultural connotations of the relics. Additionally, through large-scale digital exhibitions and virtual reality (VR) works such as 'The Palace Museum as a Museum', 'Discovering the Hall of Mental Cultivation: A Digital Experience', 'Pattern as the Carrier of Meaning: Interactive Immersive Digital Exhibition', 'Forbidden City: The Emperor's Palace', the museum employs high-immersion projection screens, virtual reality headsets, motion capture devices, touch screens, and other apparatus. Using artificial intelligence (AI), virtual reality (VR), voice and image recognition, and other technologies, they create high-precision interactive digital models to lead visitors in an immersive experience of the Palace Museum's architecture and collections.

In 2020, the Palace Museum released the 'Digital Palace Museum' mini-programme, integrating various digital resources of cultural relics collected and processed over more than 20 years and a wealth of digital products developed and produced, all on one platform. This programme has become a museum transcending time and space. It includes a 'Digital Relics Library' with 100,000 high-definition cultural relic images, a 'Palace Museum Famous Paintings Record' with over 1 billion pixels for appreciation and research of paintings and calligraphy, a 'Digital Curio Pavilion' for appreciating 3D relics at 4K resolution, virtual roaming projects like 'Panoramic Palace Museum' and 'V Palace', and youth-oriented projects such as 'Pocket Craftsman' and 'Youth Exclusive Website'. The map navigation section in the 'Palace Museum' mini-programme uses location-based services (LBS) for precise navigation, gauges visitor comfort based on crowd density for guided tours, and also offers AR real-world navigation functionality (Zheng, 2019).

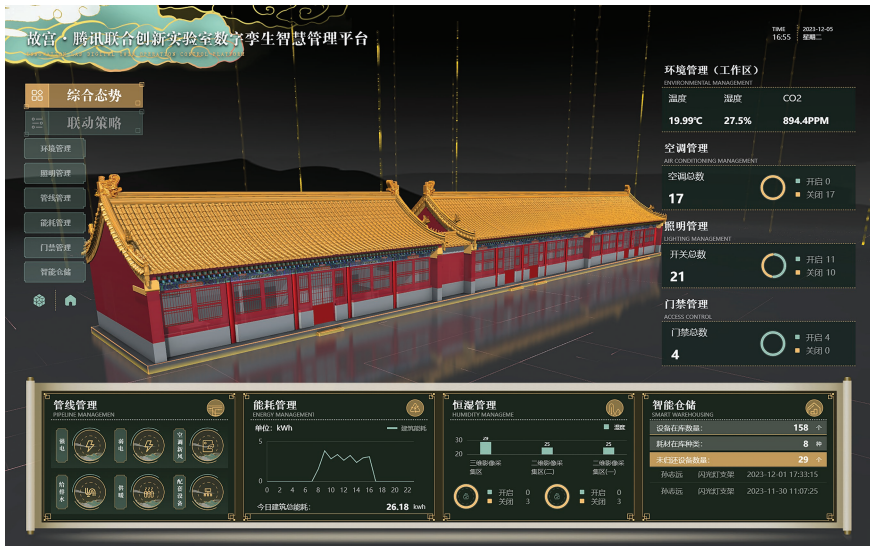

Figure 9.2 Digital twin technology enables integrated control of environmental equipment.

On the basis of the above aspects of construction exploration, in May 2023, the 'Palace – Tencent Joint Innovation Laboratory' was inaugurated, as the Palace's first 'digital twin' application landing project, further integrating and extending the functionality of the Digital Palace. The total area of the laboratory is about 500 m^2, and the digital twin team of Tencent jointly developed the 'Digital Twin Intelligent Management Platform', making the laboratory a smart building that integrates reality and interconnectivity, as shown in Figure 9.2. The laboratory includes areas for the collection and processing of artefactual data. In the process of cultural relics data collection, by real-time monitoring of indoor temperature and humidity environmental data from sensors, the digital twin intelligent platform can adjust environmental parameters according to preset standards in advance, providing appropriate environmental parameters for the safety of cultural relics. In short, the platform is able to provide a high-quality demonstration of the application of new technologies for the digital acquisition and processing of cultural relics.

At present, the Palace Museum is still constantly researching and introducing the world's advanced digital technology, for the protection of cultural heritage and audience service needs to carry out various applications. However, in general, despite the considerable data foundation and practical experience, as a result of the set of museums, heritage sites, and scenic spots in one of the three identities, more complex problems need to be solved by integrating and synergising digital technology and platforms, and higher requirements for compositing protection and service needs to be met for the Palace Museum. Therefore, in the next stage of the construction of digital twin in the Palace Museum, it is necessary to summarise the experience and clarify the direction of the existing digital twin application

practice of tangible cultural heritage discussed, covering a comprehensive range of digital twin application scenarios to complete the MHS-comprehensive digital construction.

5 Comprehensive framework and operational logic for digital twin of tangible cultural heritage

From the above representative case study, irrespective of the MHS perspective under a single form of tangible cultural heritage or the MHS-comprehensive digital twin museums, only by comprehensively and systematically sorting out the application scenarios and analysing the requirements can we build a highly available and adaptable digital twin integrated framework and operational logic. Based on the analysis and generalisation of existing cases, we explain the comprehensive digital twin framework for tangible cultural heritage from the three levels of protection, management, and service, corresponding to the three facets of artefacts, environments, and users to cover the multi-role operation and maintenance management and the full-scenario demand for the twin service of tangible cultural heritage.

5.1 Preservation: From archiving to restoration

For cultural heritage, conservation is always the primary reason to use digital twin. The digital archive will then be reproduced in real time in the virtual world for conservation purposes.

First, **digital archiving** accomplishes a 1:1 restoration of heritage in the virtual space, collects physical data, and digitally restores the scenic area. The scope of the replica is extended to the whole area of the scenic spot while restoring different precision levels for core artefacts, overall heritage, and surrounding environments. This ensures wide visibility and clear artefact restoration, establishing a symbiotic relationship between virtual and physical realms and furnishing the foundation for further functionalities. Accordingly, scientifically detailed recording of heritage information is conducted to create digital archives, achieving digital preservation of artefacts to support subsequent restoration work.

Second, **virtual restoration** simulates cultural relics restoration based on digital twin for high-precision data acquisition, restoration scheme simulation, and physical positioning recovery. This process mitigates the risk of secondary damage, enhancing the efficiency and accuracy of restoration.

Third, **preventive conservation** deploys sensors for real-time collection of dynamic data, including real-time environmental, equipment, and crowd data. Dynamic data primarily applies to scenarios, such as display case environment monitoring, exhibition hall environment monitoring, the operational status of relic preservation equipment, exposed artefact surveillance, and crowd distribution and density. Utilising data mining and AI computing, it enables automatic warnings and simulated evaluations. For instance, real-time monitoring of temperature and humidity within the venue using sensors could ensure they remain within suitable ranges, thus preventing secondary damage to cultural artefacts.

Especially for the virtual restoration and preventive conservation of cultural relics, the requirements for the accuracy of their models are high, the computational complexity of the models and the scale of the required training data are large, and the cost of constructing high-quality and large-scale expert knowledge-annotated data is high. This applies especially to MHS-comprehensive museums, such as the Palace Museum, which has a wide range of movable and immovable cultural relics and aims to realise twin-based restoration and preventive conservation of a huge volume of cultural relics, requiring large-scale, high-quality, and high-efficiency data collection and acquisition. Coupled with the specialisation and specificity of the data of the cultural heritage itself. One of the challenges lies in the fact that (1) large-scale sensor-based data collection on cultural artefacts should avoid secondary damage to the artefacts. Another major challenge is (2) how to develop and design intelligent human-computer collaborative data acquisition solutions to rapidly acquire heritage data with high-quality expert knowledge annotations for different types of heritage and reduce labour costs and improve efficiency.

5.2 Management: Intelligent operations

Unlike traditional static virtual simulation techniques, a significant feature of digital twin is the achievement of real-time monitoring and operation in the MHS management processes. Therefore, management is one of the features of applying digital twin technology, leveraging technical advantages to conduct dynamic data analysis across physical entities, surrounding environments, and visitors. It improves the management efficiency of the whole museum or scenic spot, realises intelligent operation, and supports scientific decision-making.

First, **visitor management** employs the digital twin system using technologies such as image recognition and cloud computing to calculate in real time the density of foot traffic, distribution status, and visitor behavioural habits of each exhibition hall by collecting infrared observation, video recognition, and sensor data. For instance, it can identify peak visiting hours in a day, months with high visitor influx throughout the year, primary visitor routes, and the most frequented spots. Based on crowd data, the staff can dynamically guide visitors during peak traffic and manage surrounding parking in real time. The system intelligently plans visitor routes, facilitating timely coordination among visitors and preventing accidents such as congestion and trampling. Additionally, long-term accumulation of visitor flow data can analyse preferences and habits, providing decision-making support for exhibition curation.

Second, with **continuous space-time monitoring**, comprehensive museums face numerous challenges given high visitor flow, complex exhibition hall structures, dispersed visitor distribution, and numerous artefacts, posing complexities in managing people and objects. To address this, through 24-hour real-time monitoring across the entire scenic area, the digital twin system encompasses resource monitoring and equipment control and covers external environments, overall internal environments, and localised micro-environments. This continuous space-time monitoring includes every small artefact and large structure, facilitating the timely

mobilisation of all human and material resources within the scenic area. Regarding emergencies, the digital twin platform can swiftly coordinate responses, offering a one-stop solution, and promoting intelligent operations.

Third, regarding **refined energy consumption** regulation, energy consumption such as water and electricity requires specialised oversight in traditional museum management. Moreover, it is challenging to concurrently monitor all energy consumption data and equipment operations. However, with electricity, for instance, the digital twin system can correlate indoor and outdoor temperatures, corresponding to energy consumption in different periods and areas within the venue such that it can scientifically devise an annual energy-saving plan for the museum. This plan can help fine-tune room temperatures across different floors, areas, and times and display recent energy consumption trends.

In the case of MHS-comprehensive museums, such as the Palace Museum, with numerous and varied cultural relics, a large site area, and a high flow of visitors, the daily operation of the site, daily management of cultural relics, environmental monitoring, visitor management, and other relevant matters may be in the tens of millions of records. Management and monitoring based on various sensors undoubtedly face the challenge of (3) how to reduce energy costs and improve the efficiency of effective management without causing secondary damage to the heritage. Meanwhile, (4) designing an effective data collection and analysis schema for different things is another major challenge given the various transaction types.

5.3 Service: Guiding and marketing

User-centred spatially accurate guided tours and in-depth interactive cultural experiences are another innovative feature of the cultural heritage digital twin.

First, under **intelligent guidance**, by collecting cultural artefact data, physical data, and geographical spatial data of the scenic area, visitors can explore the entire area anytime and anywhere while supporting VR browsing experiences. Additionally, it provides detailed artefact information and enables retrieval and positioning, furnishing visitors with detailed routes for artefact viewing, thus enabling offline exhibition guidance services for users. Further, the digital twin system can offer more convenient mobile terminals, allowing online access to accurate guidance services.

Second, **crowd distribution** visualises crowd data for users, including real-time congestion levels, popular attraction rankings, and intelligent route recommendations. Additionally, museums can acquire user demand information through the digital twin system. After viewing exhibitions, visitors can leave feedback or ideas through the system. Accumulating service data allows for improving museum services and optimising measures based on specific needs.

Third, with **cultural services**, the initial step is to convert digital archives into digital assets, delve into their valuable content, create IP, and launch cultural and creative products. Visitor interaction experiences must then undergo innovation via devices such as VR and AR. Finally, new media must be utilised to disseminate cultural heritage and IP products, facilitating diverse cultural communication.

Clearly, at the service level, the core challenge lies in (5) leveraging the expertise of cultural and museum domain specialists, user psychology, principles of communication, and AI technological means to fully unearth the cultural value and genetic code of artefacts. Unlike smaller museums, large comprehensive museums, such as the Palace Museum possess an immensely rich collection of heritage, offering a vast potential for cultural content exploration. For museums, comprehensively exploring, integrating, and summarising the commonalities and characteristics of cultural genetic codes from various historical artefacts, holds significant importance for the inheritance and dissemination of Chinese culture, emphasising a 'comprehensive' perspective. For ordinary visitors, identifying and showcasing the most distinctive and appealing cultural symbols further develops successful cultural peripheral products and brands, providing more popular and effective cultural services, focusing on 'precision'.

6 Development path of comprehensive tangible cultural heritage digital twin

Based on the existing research and practice, digital twin has emerged as a new direction in tangible cultural heritage preservation and management. The practical results demonstrate the numerous positive impacts of tangible cultural heritage digital twin. However, it is important to acknowledge that while digital twin is well developed in industrial sectors such as smart manufacturing (Park, 2020), it remains in the exploratory stage in the cultural domain. Moreover, completed systems are not yet widespread. In particularly, the MHS-comprehensive museums, represented by the Palace Museum, express the complex needs of large-volume heritage conservation, multi-role operation and maintenance management. A comprehensive framework for the construction and realisation of the above-mentioned digital twin would better suit the multiple needs of cultural heritage regarding preservation, management, and services. Based on the previous practice, combined with the three-level framework, the digital twin top-level architecture planning of the Forbidden City is shown in Figure 9.3. The digital twin visualisation platform carries out different precision reproduction from point, line, and surface levels to comprehensively realise heritage protection, museum management, and visitor services.

However, as analysed in the previous section, the process of realising the framework faces several digital construction challenges. Consequently, this study proposes solutions via four aspects—(1) application scenarios, (2) data models, (3) technology integration, and (4) cultural value uncovering, aiming at providing assistance and guidance for the further development of digital twin for tangible cultural heritage.

6.1 Integration of application scenarios

In theoretical research, the current applications of digital twin in tangible cultural heritage lack a comprehensive and systematic integration of application scenarios. The digitisation of the Palace Museum primarily involved application scenarios for

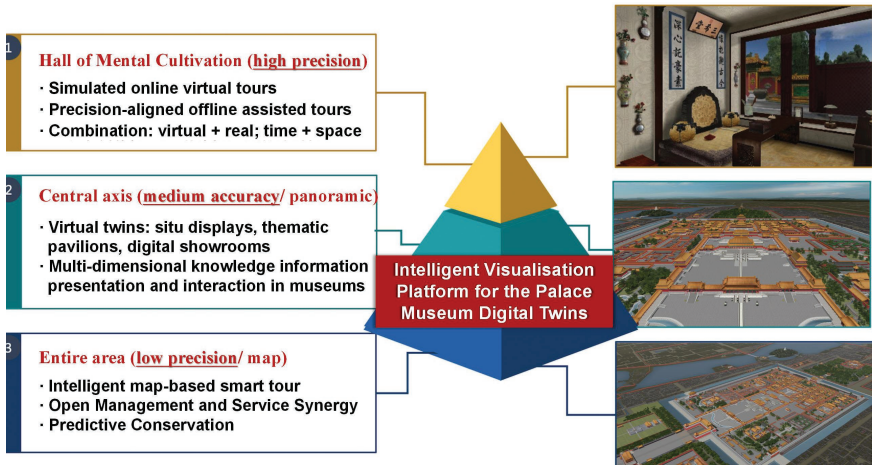

Figure 9.3 Framework of the intelligent visualisation platform for the Palace Museum Digital Twin.

conservation and services. A comprehensive and systematic identification of application scenarios from the demand side is essential for providing clear guidance for digital twin construction.

Current research on digital twin in tangible cultural heritage mainly resides in the stage of theoretical conceptualisation and high-level design and suffers from insufficient depth and excessive specificity in application research. Although scholars have identified the potential of digital twin to enhance tangible cultural heritage, they have primarily focused on smaller-scale heritage sites. Application scenarios often lack diversity and seldom cover multi-role, multi-type, and multi-scenario public service models. Research has predominantly delved into the significance of digital twin, technical implementation, and functional modules while conceptualising specific application scenarios for practical implementation remains underdeveloped.

Meanwhile, innovative applications of digital twin theories have been explored across various domains such as museums, historical architecture complexes, and tourist attractions. Examples include the creation of smart scenic areas, the digitalisation of cultural heritage, and digital twin museums. However, these applications often remain isolated within each domain, far from achieving holistic cross-domain integration for full-scene synergy and coverage across the entire value chain of cultural artefacts.

In practical applications, the existing case frameworks require further enrichment beyond the Protection-Management-Service framework discussed earlier. Expansion of application scenarios is necessary, as demonstrated below:

Protection: Beyond current digital archiving applications, there is a demand for virtual restoration and preventive protection. Virtual restoration should be realised through digital means to closely approximate the original appearance of artefacts,

thereby enhancing existing archives. Moreover, driven by technology and empowered by science, protection efforts should transition from being reactive to being proactive. The digital twin system, with its bi-directional interaction and virtual-physical interplay, can analyse sensor data, compare corresponding indicators, and provide alerts when needed. In addition, it can simulate the developmental dynamics of physical entities and minimise interference with artefacts. This approach aims at achieving the goal of monitoring cultural heritage changes, predicting risks, controlling hazards, and enabling early protection measures.

Management: Management applications for scenic areas and museums have advanced with real-time mapping, Big Data analysis, and intelligent operations and have reached a mature state.

Service: Although virtual reproduction has been achieved, user demands extend beyond this realm. Unlike manufacturing-focused digital twin, historical data and public experience constitute the essence of digital twin of cultural heritage. The reproduction of cultural heritage should not only present the past but also depict the past and present of artefacts through digital twin models. This encompasses historical evolutionary processes, significant temporal and spatial changes, and pivotal historical events that contributed to the propagation and inheritance of deep-rooted cultural heritage values. Immersive experiences are essential to engage visitors. Given the continuous advancement of cultural consumption concepts and the growing consumer demand, experiential value has become a significant pursuit. It is particularly important to incorporate interactive technologies that merge virtual and real data in the reproduction process. For example, integrating immersive technologies, such as VR/AR glasses, holographic projection, auto-stereoscopy, and extended reality (XR), creates interactive experiences that bridge the gap between reality and virtuality, resulting in a lifelike encounter. Additionally, 3D printing and derivative digital artefacts have emerged as consumer trends.

6.2 Optimising data models

Data serve as the core driving force behind digital twin, and practical applications rely heavily on the organisation and design of twinning metadata frameworks. Consequently, a pivotal challenge in implementing digital twin in the domain of tangible cultural heritage is the construction of twin data models.

In the realm of culture, a series of well-established generic data models and metadata standards have emerged that encompass metadata standards, ontologies, and subject vocabularies to support the data-driven protection and utilisation modelling of cultural heritage. At the artefact collection level, commonly used data models and standards primarily stem from an artistic perspective, including DC metadata, FOAF, GeoNames, Categories for the Description of Works of Art (CDWA) in the art domain, VRA Core metadata, International Image Interoperability Framework (IIIF), The Canadian Heritage Information Network (CHIN)'s humanistic data dictionary and Object ID metadata. Regarding the architectural aspect of tangible cultural heritage, widely used data models and standards include

the CIDOC-CRM conceptual reference model and a profile of it, Linked Art Data Model which uses the Getty Vocabularies including the Art and Architecture Thesaurus (AAT) as core sources of identity for domain-specific terminology, MIDAS Heritage—The UK Historic Environment Data Standard, the Augmented Representation of Cultural Objects (ARCO) Metadata Element Set (AMS). The aforementioned data models cover a broad spectrum of resource types at tangible cultural heritage sites. They are mainly applied to resource attribute descriptions for visual, textual, and artefact-based information, and extend to encompass digital resources related to the artefacts' environment, history, literature, and copyrights. CDWA describes the physical appearance and content characteristics of artefacts and architectural features, along with preservation, management, and cataloguing information. Furthermore, VRA Core, IIIF, CHIN, and Object ID describe the visual resources, digital images, humanistic collections, and movable artefacts, respectively. The ARCO AMS is utilised for three-dimensional virtual presentations of collections, and MIDAS Heritage extends to temporal and spatial environments and historical information on artefacts and historical sites. CIDOC-CRM covers descriptions of historical facts, humanistic arts, archaeological sites, time and location, biographical information, and copyright statements associated with cultural heritage. FOAF and GeoNames describe the personal information and geographic locations of the structures.

However, based on the aforementioned analysis, the data dimensions of existing metadata standards are limited primarily to the artefacts themselves. This encompasses the artefact features, content attributes, artefact management, artefact history, and artefact environment. The description framework is confined to three-dimensional visualisation, digital resource platforms, and certain management scenarios. Moreover, the current data models lack a universally adaptable system for constructing digital twin specifically for tangible cultural heritage. It is challenging to integrate the diverse existing demands and complex application scenarios into a cohesive framework, which hampers their maximum efficacy. For example, most current data models lack real-time monitoring and predictive data for people and objects, as well as immersive experience-related data. Furthermore, as the foundation of the entire twinning system, the data were closely integrated with the service scenarios required for different fields and needs in the twinning construction process. Therefore, the design of the twinning metadata models must cover comprehensive application scenarios. However, existing tangible cultural heritage digital twin require multiple scenarios and dynamic data for application. The application scenarios extend beyond the artefacts themselves to include management institutions and visitors. This resulted in multiple sources of heterogeneous data originating from various perspectives and scenarios. Organising and standardising such data remain unresolved challenges. On the other hand, the digital twin of tangible cultural heritage requires the collection of real-time data, such as temperature and humidity, from display cases. The acquisition and storage of such dynamic data pose higher requirements for data models and databases.

6.3 Integrating relevant technologies

Digital twin technology is not a single technology but an integration of various technologies. The process of human-machine-environment intelligent fusion and collaboration involves a series of technologies ranging from perception and access to services and transactions. However, when applying these technologies to tangible cultural heritage, it is essential to consider whether the foundation of the digital twin technology should incorporate relevant technologies from the cultural heritage domain. The answer is in the affirmative.

Virtual restoration and preventive protection are considered examples. Damaged artefacts and heritage sites can be virtually restored using digital technology to infer their original state. Techniques such as 3D laser scanning, close-range photography, and 3D printing for virtual restoration can be integrated into a digital twinning framework. Using digital twin technology, lightweight and high-precision data acquisition, simulation-based restoration scenarios, and precise physical location restorations can be achieved. This effectively avoids the risk of secondary damage and reduces human and material-resource consumption. In the process of preventive protection, technologies like AI computing, Big Data, cloud computing, and visualisation must be applied to data mining and analysis. A large database is necessary to provide a decision-making basis for the assessment and prediction of artefacts.

6.4 Uncovering cultural value

The construction of databases for organising and integrating relevant digital resources and data has become a consensus in the age of informatisation. This holds true for museum and cultural heritage sectors. Most museums digitise their collections and offer open retrieval, browsing and knowledge acquisition services to the public. However, existing data often suffer from weak correlations, lack of in-depth semantic descriptions, and do not meet the requirements of discoverability, findability, accessibility, interoperability, and reusability (FAIR), thereby hindering higher-level knowledge application services and satisfactory discovery of deeper core values.

Smart data involves operations such as the fusion, correlation, and analysis of multi-source data to create machine-understandable, versatile, and efficient data methods. Currently, theories and methods related to Big Data, deep learning, linked data, and knowledge graphs have been extended to fields such as archives and museums, providing new ideas for the construction of digital twin data at cultural heritage sites. For example, the Digital Dunhuang Project constructs a thematic vocabulary, publishes linked data, builds domain ontologies, annotates mural images, and creates systems to manage digital assets and interactive narratives. This smart approach transforms Dunhuang's data into intelligent resources, facilitating digital humanities research and making murals more accessible to audiences.

7 Conclusion

The digitalisation journey of China's cultural heritage can be traced back to the late 1990s, when the heritage industry began to experiment with the digital registration and recording of ancient buildings and collections (Shi, 2017). During this period, the focus was primarily on creating digital images of artefacts. In the early 21st century, significant progress was made in the digitisation of cultural heritage with the advancement of new digital technologies and international standards. Technologies such as 3D scanning and the use of metadata standards such as the Dublin Core have enabled cultural institutions to create more detailed and accurate digital replicas of artefacts and provide comprehensive metadata, thereby enhancing the discoverability and accessibility of cultural heritage data resources. With the rising demand for cultural experiences in recent years, there has been a need to provide visitors with intelligent digital resources related to cultural heritage. This calls for the integration of new technologies such as digital twin to make cultural heritage more accessible.

In this study, four typical digital twin applications—the Shanghai History Museum, Longmen Grottoes Scenic Area, Yellow Emperor's Hometown Place, and the Palace Museum—were used as case studies, particularly in response to the complex twinning needs of comprehensive museums, such as the Palace Museum, that combines MHS. Based on China's exploration and experience, this study systematically summarises and constructs a comprehensive framework for digital twin of tangible cultural heritage, adapted to meet the needs of all-round and multi-role twinning services. Moreover, building on existing research and case analyses, this study presents four development suggestions for the implementation of digital twin in comprehensive tangible cultural heritage sites: integrating application scenarios, optimising data models, integrating relevant technologies, and uncovering cultural value.

The construction of digital twin for tangible cultural heritage not only enables the deep exploration of cultural heritage value and promotes the development of digital humanities but also lowers the threshold for people to access culture. This process drives the digital transformation and upgrade of the cultural industry and the development of cultural tourism. It is believed that with the further development of digital twin technology, more and more tangible cultural heritage, especially MHS-comprehensive museums represented by the Palace Museum, can achieve a win-win coordination of management and service, protection and openness, thereby making a major leap forward in the cultural industry.

Acknowledgments

We would like to express our gratitude for the funding support for this research from the National Key Research and Development Program of China "Key Technology Research and Service Demonstration of Digital Twin for Large Comprehensive Museums" (Project Code: 2022YFF0904300).

References

Armeni, P., Polat, I., De Rossi, L. M., Diaferia, L., Meregalli, S., & Gatti, A. (2022). Digital twins in healthcare: Is it the beginning of a new era of evidence-based medicine? A critical review. *Journal of Personalized Medicine, 12*(8), 1255. https://doi.org/10.3390/jpm12081255

Centorrino, P., Corbetta, A., Cristiani, E., & Onofri, E. (2021). Managing crowded museums: Visitors flow measurement, analysis, modeling, and optimization. *Journal of Computational Science*. https://doi.org/10.1016/j.jocs.2021.101357

Chen, W., & Li, Y. (2021). A strategic study of the communication of NRM in the context of the meta-universe. *Voice & Screen World*, (S1), 6–9.

Deng, T., Zhang, K., & Shen, Z-J. (2021). A systematic review of a digital twin city: A new pattern of urban governance toward smart cities. *Journal of Management Science and Engineering, 6*(3), 125–134. https://doi.org/10.1016/j.jmse.2021.03.003

Grieves, M. (2002, October). Completing the cycle: Using PLM information in the sales and service functions [slides]. In *SME Management Forum*.

Grieves, M. W. (2023). Digital twins: Past, present, and future. In N. Crespi, A. T. Drobot & R. Minerva (Eds.), *The Digital Twin* (pp. 97–121). Springer. https://doi.org/10.1007/978-3-031-21343-4_4

Haleem, A., Javaid, M., Singh, R. P., & Suman, R. (2023). Exploring the revolution in healthcare systems through the applications of digital twin technology. *Biomedical Technology, 4*, 28–38. https://doi.org/10.1016/j.bmt.2023.02.001

International Organization for Standardization. (2023). *Digital twin concepts and terminology (ISO Standard No. 30173:2023)*. https://www.iso.org/standard/81442.html

Li, C., Zhang, X., & Qin, X. (2017). The future technology application and development of the library from the ten major strategic technological trends. *Library & Information*, *38*(6), 37–47.

Li, T. (2021, September 27). *Shanghai's first digital twin museum system launched, Shangri-La Museum has a Smart 'Digital Brain'*. wenhui.whb.cn.https://wenhui.whb.cn/third/baidu/202109/27/425974.html

Lin, C. (2022). Research on digital twin technology and its application in museums. *Public Communication of Science & Technology*, *14*(1), 106–108. https://doi.org/10.16607/j.cnki.1674-6708.2022.01.034

Liu, M., Fang, S., Dong, H., & Xu, C. (2021). Review of digital twin about concepts, technologies, and industrial applications. *Journal of Manufacturing Systems*, 346–361. https://doi.org/10.1016/j.jmsy.2020.06.017

National Cultural Heritage Administration. (2017). *Law of the People's Republic of China on the protection of cultural relics*. Retrieved from http://www.ncha.gov.cn/art/2017/11/28/art_2301_42898.html

Park, S. S. (2020). A cloud-based digital twin manufacturing system based on an interoperable data schema for smart manufacturing. *International Journal of Computer Integrated Manufacturing, 33*(10–12). https://doi.org/10.1080/0951192x.2020.1815850

Qi, D., & Wu, F. (2022). Research on the application of the digital twin CIM platform in Yellow Emperor's Native Place scenic spot. *Railway Construction Technology*, *60*(10), 202–206.

Qin, X., & Zhang, X. (2018). Application of digital twin technology in the digital construction of material cultural heritage. *Information and Documentation Services*, *39*(2), 103–111.

Richardson, J. (2020). What digital twin technology means for museums. MuseumNext. Retrieved from https://www.museumnext.com/article/what-digital-twin-technology-means-for-museums/

Sang, Y., Zhao, L., Sun, K., Wang, X., & Meng, X. (2021). Research on the framework and application of digital twin technology for the cultural heritage of the Great Wall. *Technology Innovation and Application*, (35), 19–23+27. https://doi.org/10.19981/j.CN23-1581/G3.2021.35.005

Shen, R., & Lv, L. (2022). Research on the innovative integration and application of Yuan Cosmos in the digitisation of Beijing carved lacquer. *Packaging Engineering,* (S1), 283–288. https://doi.org/10.19554/j.cnki.1001-3563.2022.S1.047

Shi, X. (2017). Digital management of collections at the palace museum. *China Cultural Heritage Scientific Research*, *12*(4), 38–43.

Song, K., & Yuan, H. (2021). Digitisation of cultural resources in the core area of the capital's functions. *Urban and Rural Development*, *66*(20), 62–63.

Tan, J. (2021). Research on the application of digital twin technology in the conservation of ancient villages. *Informatization of China Construction*, *27*(15), 65–67.

The Palace Museum. (2022). *The new arrival of graceful swallows—Digital Palace Museum press conference held at the Palace Museum*. Retrieved from https://www.dpm.org.cn/classify_detail/257844.html

United Nations Educational, Scientific and Cultural Organization. (1972). *Convention concerning the protection of the world cultural and natural heritage*. Retrieved from https://www.un.org/zh/node/182113

Veyrieras, J. (2019). A digital twin for Notre-Dame. *CNRS News*. Retrieved from https://news.cnrs.fr/articles/a-digital-twin-for-notre-dame

Vlavianos, N., & Nagakura, T. (2021). An architectural metaverse that combines dynamic and static 3D data in XR. In CHNT – ICOMOS Editorial board. *Proceedings of the 26th International Conference on Cultural Heritage and New Technologies*. Propylaeum, Heidelberg.

Wang, J. (2018). The necessity and feasibility of building a digital twin system for underwater cultural heritage in China. *Yangtze River Series*, *40*(9), 3.

Wang, l. (2021). Digitalized restoration technique applied in protection of the stone statues. *Northern Cultural Relics*, (4), 64–68. https://doi.org/10.16422/j.cnki.1001-0483.2021.04.007

Wang, X. (2023). Practice and exploration of the protection of the World cultural heritage of the Palace Musuem on the new development concepts. *Palace Museum Journal*, *66*(1), 19.

Wang, X., Tan, X., & Xia, S. (2020). Research and practice of Dunhuang smart data. *Digital Humanities*, *1*(4), 11–23.

Xiao, Y. (2022). Digital twin of Chinese opera and folklore relics in overseas collections and the construction of digital resource platform. *Cultural Heritage*, *16*(1), 89–96.

Yang, G., Li, M., Cui, C., & Guo, Y. (2021). Digital twin: Connotation, challenge and application. *Software Guide*, *20*(1), 6–10.

Yang, Y., Zou, J., Shi, M., Li, Y., Yang, B., Wang, H., Shi, C., Jin, L., & Lu, X. (2022). Analysis of the research status of digital twin technology. *Journal of Technology*, *22*(2), 176–184+188.

Yu, F. (2022). Digital twin technology unlocks holographic intelligent scenes for cultural tourism. *China Culture Daily*, 7.

Yu, F. (2023). The Palace Museum provides 'China's Program' for World Heritage Protection. *China Culture Daily*, 001.

Zhan, Q. (2022). The non-legacy meta-universe: Empowerment, integration, immersion. *Today's Mass Media*, *31*(9), 101–103.

Zhang, Y. (2022). Longmen Grottoes Intelligent Cultural Tourism Digital Twin Platform creates new scenes and applications for inheriting and promoting traditional culture. *China Culture Daily*, 4. https://doi.org/10.28144/n.cnki.ncwhb.2022.003618

Zheng, S.W. (2019). *Approach the 'digital Palace' Palace Museum launched seven digital products*. Retrieved from http://culture.people.com.cn/n1/2019/0716/c1013-31237991.html

Zheng, Y. (2021). *Digital reconstruction and 3D printing of Keichousaurus fossils*. (Doctoral dissertation, Guizhou Normal University).

Zhou, Q., & Tan, S. (2022). Digital twin system construction of Shanghai History Museum and the digital transformation of museums. *Chinese Museum*, *39*(2), 30–35.

10 Digital deduction theatre

An experimental methodological framework for the digital intelligence revitalisation of cultural heritage

Xiaoguang Wang, Ke Zhao, Qi Zhang and Chengyong Liu

1 Introduction

Cultural heritage is the resort of human history and culture, which includes buildings, artefacts, folklore, traditional knowledge, language, customs, and so on (Sullivan, 2015). Cultural heritage research has undergone unprecedented changes in the present digitally driven world. This revolution is not only reflected in the implementation of technology but also profoundly affects the way in which science and technology are merged with culture, thus propelling cultural heritage research into a new era of digital intelligence. Digital intelligence refers to the ability to interact effectively with digital technologies and the environment (Adams, 2004). It goes beyond technical skills such as digitisation, Big Data, artificial intelligence, and digital media technologies, encompassing the ability to utilise digital information creatively and efficiently in various contexts. Digital humanities (DH) increase ease for researchers to access and analyse large-scale data and discover new questions (Schich, 2014); augmented reality, virtual reality, and extended reality technologies can enhance the attractiveness and interactivity of cultural heritage (De Paolis et al., 2023; Fu et al., 2020); and artificial intelligence-generated content (AIGC) driven by large language models enhances humanities data processing efficiency, facilitating the organisation of information and knowledge services (Liu et al., 2023). For example, a DH laboratory is a multifunctional space that combines digital cultural heritage resources, tools, and interdisciplinary research methods (Pawlicka-Deger, 2020). It encourages collaboration, innovation, and public engagement, and promotes digital preservation, interdisciplinary research, and the wide dissemination of cultural heritage. However, DH lacks design theory and practice to some extent, with an emphasis on graphical presentation and publishing interfaces (Liu, 2013), and has not yet reached the point of being used as a method for knowledge discovery and generation (Burdick et al., 2012, p. 14). There is an urgent need to seek new interactive and participatory design methods for the digital intelligence revitalisation of cultural heritage. The laboratory's theatre prototype is a renewal concept that allows the fusion of cultural heritage with digital intelligence environments, providing a new experimental space for humanities research. Such a theatre can be seen as a situated practice that involves the physicality, virtuality, and design of space; the application of tools and resources; research collaboration; and the spatial representation of practice (Oiva, 2020).

DOI: 10.4324/9781032707211-12

This chapter constructs an experimental methodological framework for digital intelligence revitalisation of cultural heritage by exploring the digital deduction theatre (DDT) driven by digital intelligence and DH. In the first part, we briefly compare the differences between scientific and humanities research and highlight the changes in humanities research in the digital intelligence environment, especially the emergence of DH research. Next, we discuss the challenges digital intelligence poses to humanities research. Based on this analysis, we traced the theatre prototype of the laboratory, identified the innovative development of the humanities laboratory based on the concept of the scientific laboratory, and specifically addressed the theatre properties of the DH laboratory. Thus, we aim to capture the changes in humanities research powered by digital intelligence, understand the unique features of DH laboratories, and reflect on their development and characteristics. Finally, DDT was proposed as an experimental methodological framework for cultural heritage digital intelligence revitalisation with core concepts and forms, construction paths, and technology infrastructure.

2 Humanities research in the digital intelligence environment

Science and Humanities (or in British English "Arts") are essential elements of human civilisation and innovation. Science is about the "how" and the humanities look into the "why" (Strosberg, 2013, p. 13). Science, which encompasses a wide range of disciplines, including mathematics, physics, chemistry, biology, astronomy, geology, engineering, and medicine, is an evidence-based study of phenomena, rules, and processes in the natural and material worlds (Scott, 2008, pp. 3–15). Science research has discovered the laws, theories, and principles that explain and predict phenomena through empirical methods such as observation and experimentation (Kuhn, 2012, p. 10). Its application is in problem solving and technical innovation. Humanities research primarily use qualitative research methodologies for analysis, interpretation, and creation to understand "human experience, agency, identity and expression" (Bakhshi et al., 2008, pp. 9–10). It focuses on the disciplines of literature, history, classics, art, and music (Burdick et al., 2012, p. 4). Scientific research is a product of rationality, objectivity, precision, repeatability, and universality (Strosberg, 2013, pp. 11–13). Humanities research is an imaginative activity characterised by subjectivity, ambiguity, and complexity (Burdick et al., 2012, pp. 92–93).

Nonetheless, the advancement of digital technologies has progressively provided humanities research with the capability to respond to controversies and questions about research materials and understanding using rational methods (e.g. experimentation, logical operations, universal theories) (Spiro, 2012, pp. 28–29). Humanities research can now focus on vast amounts of digital content, such as texts, images, sounds, videos, animations, 3D models, maps, and other media files (Drucker, 2020, p. 1), which are rich in variety and scope, as opposed to the limited research objects (types and quantities) of traditional methods. However, the development of databases, platforms, and digital libraries in the digital intelligence environment has made using humanities materials more convenient and efficient (Gavin & Smith, 2012, p. 65). Additionally, both the accumulation of original

materials and ever-generating born-digital resources require humanities academics to research, preserve, expand, and disseminate them using new digital technologies and tools (Berry & Fagerjord, 2017). Overall, the expanded functionalities and usefulness of digital technologies for humanities materials have allowed the paradigm of humanities studies to expand and strengthen the evidence-based character of humanities research.

DH can be seen as a systemic paradigm shift in humanities research in a digital intelligence environment that utilises data and technologies to drive development and innovation in the humanities (Svensson, 2010). DH, originating from humanities computing, is an interdisciplinary field of study that utilises digital technologies for humanities research at the intersection of computational methods and humanities materials (Berry & Fagerjord, 2017; Drucker, 2021, p. 1). The first generation of humanities computing (Humanities 1.0) represented data-driven research, with digital scholarly work characterised by digital incunabula and machine actionability (Sahle, 2016). Since the late 1990s, vast collections of cultural heritage from galleries, libraries, archives, and museums (GLAM) have been digitised into multimodal representational resources, such as text, images, video, audio, and 3D models (Windhager et al., 2018). A digital resource with deeper and richer semantic information was created by computers processing the data and metadata of digitised items (Arnold & Tilton, 2022, p. 183). Thus, the data and computational components from empirical research are included in humanities research, thereby expanding the potential applications of digital methodologies. Primarily, DH have transformed our research methodology, expanded the range of research questions we may pose, and enhanced the comprehensiveness, intricacy, and precision of the responses we can offer (Davidson, 2008).

The DH Manifesto 2.0 (Presner et al., 2008) emphasises "qualitative, interpretive, experimental, emotional, and generative". Instead of simply applying computing technology to humanities research, DH focuses on the multilayer interpretation of humanities research through interdisciplinary collaboration. As DH is characterised by digital, its core is still humanities (Burdick et al., 2012, p. 4). Under the paradigm of data-driven research, DH must investigate issues of value, cultural significance, and deeper meaning. Moreover, scientific methodologies are required to transform data into theory, given that theory, whether declared or implicitly, transforms data into interpretation (Davidson, 2008). After 15 years of DH 2.0, it remains uncertain whether Humanities 3.0 will centre on the effects of AI technology, particularly the AIGC explosion. However, it is evident that AI technologies influence humanities research, covering linguistics, poetry, film, classical art, history, and more, through textual, graphic, audio, and video elements (Liu et al., 2023). Thus, new instruments and approaches to humanities research and enquiry are being developed using Big Data and AI technology. DH has led to the development of interdisciplinary studies that have deepened our understanding of human culture, history, and society. This trend heralds the intertwined development of humanistic science and scientific humanities. In the context of digital intelligence, the research questions, methods, and ideologies of humanities research will continue to evolve, and may even give rise to new research paradigms.

3 Challenges for humanities research

The digital age presents a paradigm shift whereas Hockey (2004) points out, "two cultures" exist; humanities research not only maintains critical analysis and insight, but also converges with the experimentation and rigorous argumentation of scientific research. Cultural heritage studies, which are an important part of DH, have also been motivated by innovations in digital technology (Ulutas Aydogan et al., 2019). Large-language models enable intelligent parsing and knowledge extraction from historical texts and improve information processing efficiency (Chen et al., 2021). Argument, virtual, and extended reality technologies can enhance the quality of knowledge dissemination (Bekele et al., 2018). However, opportunities and challenges often coexist. Humanities and cultural heritage research face similar challenges in the organisation of digital resources, application of digital methods, and transformation of research paradigms.

Information overload poses challenges for resource management and digital literacy. The digital transformation of humanities research can be traced back to the use of information technology to digitise medieval Latin by the father of the DH, Roberto Busa (1999). It started with literature and linguistics and then spread rapidly to various fields such as history, museology, archaeology, music, and architecture (Davidson, 2008). Currently, the humanities have amassed a vast array of digitised resources transitioning from print publications, as well as numerous resources in born-digital formats, notably a series of digitisation efforts of cultural heritage collections made by GLAM (Macrì & Cristofaro, 2021). Digitisation has transformed humanities scholars from reading one paper book to being able to browsing many digital texts (Jänicke, 2015). However, these massive digital resources have exacerbated the growing information overload since the invention of the printing press, causing an inevitable academic crisis for humanities scholars. When we are surrounded by vast amounts of disordered information, it is easy to blindly hoard all kinds of digital resources (Sweeten et al., 2018). However, due to insufficient digital literacy and a lack of effective digital competence, it is difficult to scientifically manage and use these digital resources (Eshet, 2004). Therefore, a new model is required to manage digital resources and enhance digital acceptability.

The mechanisation of digital tools poses a challenge to the value of using digital methods in academia. In response to the growing abundance of digital resources, scholars employ various research software programmes and tools for academic research. Lightweight coding languages like Python and its simplified toolkit enhance usability, but at the expense of increased and obscured complexity (Tenen, 2016, p. 86). This results in the analytical processes of these tools becoming a black box, as humanities scholars struggle to effectively analyse results without understanding their algorithmic principles and operational mechanisms. As Dennis Tenen (2016, p. 85) emphasises, it is crucial not just to use these tools or learn coding, but to understand the results and delve into the underlying logic of their production. Additionally, digital tools may contain concepts and techniques conflicting with traditional humanities (van Es et al., 2021). Drucker (2012, p. 88) notes that

humanities scholars often engage in digital humanities projects in a neutral role, practicing as positivists. For instance, content recognition in text using markup is seen as interpretive, but it requires explicit computational paraphrasing to limit ambiguity. Despite the active application of digital methods in generating humanities knowledge, there is a need for critical reflection on integrating humanities theory, humanism, and hermeneutics into mechanical computational practices (Drucker, 2022, p. 15). The focus should be on making digital tools more humane and humanistic to enhance the cultural value and significance of applying digital methods.

The novelty of academic topics raises the challenge of merging research paradigms across different disciplines. As the humanities develops, the use of digital and computational approaches in humanities projects is no longer new and has led to a multitude of emerging topics. Examples include generative agents and social simulation (Park et al., 2023), human-agent collaborations (Cila, 2022), user-centred cultural experience (Doukianou et al., 2020), virtual reconstruction of cultural heritage (Pietroni & Ferdani, 2021), and so on. These topics have significant interdisciplinary attributes that require the exploration of new methodologies. Similarly, resources in novel forms such as video games place greater demand on researchers' innovative thinking (Coltrain & Ramsay, 2019; Holloway-Attaway, 2021), which also calls for interdisciplinary collaboration. Additionally, the integration of technology, arts, and humanities can create more interactive and immersive experiences that significantly enhance public engagement, which requires interdisciplinary work (Kenderdine, 2021, p. 29).

These challenges spur the creation of a new methodological framework focused on efficiently managing digital resources and enhancing their acceptability, increasing user engagement through transformed digital technology applications, and integrating interdisciplinary methods and theories to advance new academic fields and break traditional disciplinary boundaries. Consequently, this approach leads to exploring new hybrid experimental spaces in which the laboratory and theatre thrive in the digital era.

4 Laboratory and theatre: Reinventing experimental research

The origins of laboratories include medieval monasteries, anatomical theatres (see Figure 10.1), apothecary workplaces, and domestic kitchens. These laboratory prototypes primarily comprise a series of practical operations performed in real physical spaces. In 1594, there was an anatomical theatre in Padua, Italy, where professional knowledge of human anatomy was taught to the public through dissections in architectural spaces (Wershler et al., 2022, p. 49). Compared to other prototypes, theatre redefines the boundaries between the public and privateness of the space, aiming to give the audience a clearer understanding of knowledge through pedagogical performances.

In the early modern period, the laboratory, often used as a site to produce knowledge (represented by the alchemy laboratory), also played the role of a theatre of the witness. In the famous Hobbes-Boyle controversy, Boyle achieved theatrical witnessing of experiments using three technical solutions (Shapin & Schaffer,

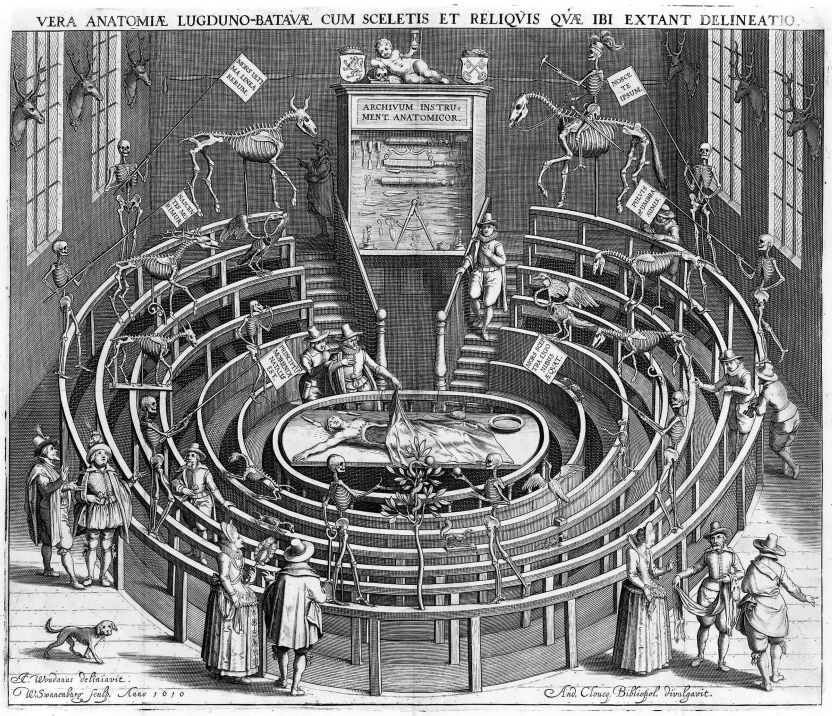

Figure 10.1 Anatomical theatre at Leiden University in the early 17th century. Willem Isaacsz Swanenburg-Museum Boerhaave, Public Domain, https://commons.wikimedia.org/w/index.php?curid=194044.

2011, p. 25). Specifically, Boyle established a public laboratory space with experimental apparatus and equipment, inviting nobles and scholars to observe the experimental operations as witnesses and leave a record of their visits. Thus, people can witness and testify to the production of scientific facts with their own eyes. Boyle then used text and illustrations to bring together the experimental scenes, processes, operation details, and results into a book for publication. These comprehensive, faithful, detailed, and vivid experimental records provide readers with imaginative witnesses. Moreover, Boyle organised an experimental community in which performers of scientific experiments and philosophers rationally discussed and resolved disputes, which set the stage for public witness. Following the 19th century, Edison's laboratory constructed the role of objects as witnesses. Various experimental objects on the table, such as precision instruments, telescopes, mirrors, flasks, and drafts, were imagined as witnesses (Parikka, 2016); thus, things and people worked together to complete the experimental narrative. Overall, tracing the historical development of the laboratory reveals the importance of theatre as a physical space for connecting things and people. The theatre form of the early laboratory aimed to demonstrate the knowledge-producing practices of scientific

experiments to the public, improve the authenticity and trustworthiness of the experiments, and facilitate knowledge dissemination.

In the 1970s, space broke free of time and became productive (Lefebvre, 2012, p. 37). The term theatre extends from the original meaning of physical space to the connotations of spatial energy. In 1991, Brenda Laurel used Aristotle's Poetics as a framework for her analogy between human-computer interaction and theatre to propose a poetics of interactive art (Laurel, 2013, p. 41), which makes the computer a digital space for theatre. Later, Janet Murray developed the term cyberdrama (Murray, 2017, pp. 232–233), which emphasises the enactment of stories in the special fictional space of the computer. With the development of new media technologies, theatre-related terms, such as virtuality, simulation, and interactivity, continue to emerge in discussions on digital culture and the Internet (Balme, 2008, p. 202), further contributing to the theatricalisation of space. Additionally, Georgina Guy (2016, p. 3) proposed theatricalised exhibition, which is an effective presentation of non-linear narratives through the interaction of scenes, media, and audience. By creating a virtual world with a unique connection to reality, theatre involves the expression and presentation of different dimensions, such as material, cultural, and spiritual spaces. The experimental nature of theatre also permits it to host a series of logical, scientific, and creative activities that achieve emotional experiences, including spiritual resonance and connections to history, art, and faith. It conveys the abstract cultural meanings of values, traditions, customs, and social systems carried by a cultural heritage, as well as enable users to perceive, contemplate, and reflect on them in an immersive manner.

However, the growth and diversification of laboratories in the humanities presents a challenge to the traditional interpretation of science laboratories. Science laboratories are usually based in physical locations, with hands-on practices by people with specialised knowledge and skills, using materials, instruments, and equipment (Hannaway, 1986). Humanities laboratories draw on scientific concepts to develop distinctive models of laboratory (Pawlicka-Deger, 2020). Examples include Makerspaces (School of Advanced Study, University of London), virtual laboratories (Alfalab, a project of the Royal Netherlands Academy of Fine Arts), and experimental theatres (CAMLab, Harvard University). The concept of the laboratory has shifted from a space focused on knowledge science to a space dedicated to practice science (Pickering, 1992, pp. 8–9), and from the creation of physical forms of laboratories to concept-based laboratories (Pawlicka-Deger, 2020).

The DH laboratory, centred around innovation, interdisciplinarity and collaboration, is a driving force in creating a laboratory space that demonstrates the new character of the humanities laboratory with a new theatre design in terms of the knowledge practice and the research dissemination. First, the DH experiment adopted a data-driven humanities computing method (e.g. Humanities + Design, a research lab at Stanford University) to deepen the analysis and promote the exploration of traditional humanities research towards quantification and modelling. Second, DH uses intelligent technology and digital media to recreate historical and cultural scenarios, thus providing new perspectives for macro- and micro-cultural analyses. For instance, the CAMLab at Harvard University uses AI-powered

immersive installations of dance forms from Buddhist caves, which enable audiences to access knowledge in unique ways. Third, DH facilitates an open and collaborative research culture (El Khatib et al., 2020), which helps break down academic barriers, accelerate the flow of knowledge, and promote research innovation on a global scale.

The DH laboratory advanced the development of theatre-based research laboratories in the 21st century. It is oriented towards authenticity, verifiability, and reproducibility and promotes the innovation of experimentation and research paradigms of humanities characteristics in the context of Big Data and AI. However, digital technologies fall short in to realise their full potential to transform and enhance humanities studies if they make users passive recipients, rather than active participants in knowledge production (Moreshead & Salter, 2022, p. 95). Within this new perspective, the integration of theatre with digital intelligence continues to advance, contributing to the creation of theatre-based experimental spaces that integrate people, objects, and technology. An experimental theatre is characterised by imagination, seriousness, technicality, and exhibitionism. First, imagination is reflected in the theatre's ability to create imaginary virtual worlds that have a special relationship with reality and evoke the imagination and creativity of the audience through immersive experiences. Next, although theatre is often seen as an entertaining activity, its seriousness lies in the fact that the visual, represented by deduction, can stimulate the user's thinking, promote reflection and insight, and even explore deeper philosophical, ethical, and other issues. Moreover, theatre is suitable for human-computer interaction activities and thus has the potential to apply digital technologies that can support logical, scientific, and intelligent deductive activities. Finally, the goal of theatre is to present imaginary worlds that expand, enhance, and enrich people's abilities to think, feel, and act (Laurel, 2014, p. 39). With this in mind, we propose the DDT of cultural heritage as a composite art form that is creative, professional, intelligent, and expressive. The emergence of this theatre form promoted the use of science, intelligence, and thinking to actively engage in experiments in the digital intelligence revitalisation of cultural heritage.

5 Cultural heritage digital deduction theatre

Cultural heritage data have emerged as pivotal informational assets in the digital era. The synergy between digital technology and cultural heritage has become a crucial driver of productivity, with intelligent curation of cultural heritage proving to be an invaluable skill. As ancient cultural heritage converged with modern theatre, a new type of experimental theatre has emerged. Driven by digital intelligence and DH concepts, the DDT of cultural heritage is a cutting-edge interdisciplinary experimental space that explores a new paradigm of experimentation in the humanities. Based on tools and equipment like 3D immersive projection, interactive screens, XR storytelling, and advanced modelling, it combines techniques and methods from a variety of fields, including history, information management, humanities, literature, art, and AI. The DDT establishes a complete innovation chain of Digital Incubation – Virtual-Real Fusion – collaborative deduction-immersive

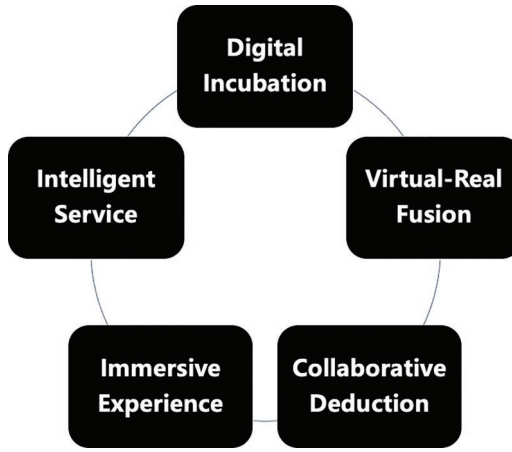

Figure 10.2 The innovative chain for Cultural Heritage Digital Deduction Theatre. Drawing by the author.

experience – Intelligent Service (see Figure 10.2). This chain supports the research and practice of cultural heritage to revitalise the historical, cultural, scientific, and artistic knowledge embodied within it.

5.1 Core concept and forms

- *Digital Incubation.* DDT uses digital devices and high-performance computing to record and archive cultural heritage sites. The purpose is to build open-access knowledge graphs, text databases, image databases, 3D model libraries, cultural gene libraries, GIS platforms, and other cultural heritage smart data. Digital resources provide powerful data support for the preservation, inheritance, and innovation of cultural heritage sites.
- *Virtual-Real Fusion*. By employing digital twins, GIS technology, and XR devices, DDT aims to establish precise connections, mappings, and overlays between cultural heritage artefacts and their interpretive information and knowledge. This enables seamless integration of the physical world with the digital space.
- *Collaborative Deduction.* DDT uses methods such as human-intelligence collaboration, open crowdsourcing, data mining, simulated reasoning, and artistic imagination to perform deductive processing of cultural heritage towards the restoration and reproduce historical and cultural scenes. Through collaborative deduction, experts and participants from different fields can collaborate to conduct interdisciplinary research and teaching, thereby offering a comprehensive understanding and representation of cultural heritage from multiple perspectives.
- *Immersive Experience.* Using XR devices and interactive systems, DDT aims to offer users a simulated reality an interactive three-dimensional visual narrative environment. Users can enter virtual scenes of cultural heritage as digital avatars and interact with characters and their surroundings to experience a sense of presence.

- *Intelligent Service.* With AIGC technology and large language models such as ChatGPT, DDT aims to provide content generation, intelligent conversation, information retrieval, publication, and dissemination services related to cultural heritage. Through these intelligent services, DDT can support the revitalisation, regeneration, and sustainable inheritance of outstanding Chinese traditional culture, meet users' knowledge demands, and offer personalised, interactive, and engagement experiences.

5.2 Construction path

Knowledge has been the subject of extensive research in various disciplines (McInerney, 2002). The knowledge life cycle indicates that knowledge presents different values at different stages (McElroy, 1999). Understanding the characteristics and value of knowledge at each stage and adopting appropriate management strategies and applications are important for stimulating the vitality of knowledge, realising its value, and meeting users' knowledge demands. Additionally, the knowledge life cycle is the process of interacting with people. Therefore, it is necessary to pay attention to the interaction between knowledge and people and promote knowledge sharing and inheritance. In the DDT of cultural heritage, the specific meaning of the knowledge life cycle can be understood as the process in which cultural heritage knowledge starts from the most basic data, gradually undergoes analysis, understanding, and interpretation, and finally creates value through practice and dissemination. The use of the knowledge life cycle can optimise the knowledge management of cultural heritage and maximise its role and impact on continuous innovation. It also enables more people to participate in, experience, and understand the knowledge and value of cultural heritage. Therefore, we propose a construction path for knowledge production, practice, and experience for the digital intelligence revitalisation of cultural heritage sites.

5.2.1 Knowledge production: Scientific-deductive thinking for knowledge demands

Knowledge production is a type of digital cultural heritage activity based on scientific deductive thinking that generates and innovates knowledge for specific knowledge demands. Scientific deductive thinking emphasises reasoning and analysis based on evidence to interpret cultural heritage data and reveal their meanings, values, and connotations. Thus, it provides an accurate and valuable support for cultural heritage practices.

- *Knowledge Demands.* Cultural heritage knowledge demands involve both content and form aspects in different scenarios, aimed at solving problems in the process of the digital revitalisation of cultural heritage or filling the gaps in related knowledge. At the content level, people expect to understand the historical backgrounds and meanings of cultural heritage sites. Regarding form, people look forward to a novel and intuitive digital experience brought about by the

application of new technologies (Valtolina, 2016) to interact and engage with cultural heritage and pursue personalised knowledge acquisition (Pujol et al., 2012).

- *Data Reconstruction.* Cultural heritage data reconstruction benefits from the DH approach. It aims to construct a more complete and comprehensive knowledge system by integrating and linking heterogeneous cultural heritage data from multiple sources to reconstruct the veins and relationships of cultural heritage knowledge. This process involves the transformation of data, information, and knowledge and requires attention to the quality and relevance of the information extracted from the data, as well as ensuring that the entire process is scientific and reliable.
- *Knowledge Innovation.* When existing cultural heritage knowledge resources are insufficient to meet the needs of digital intelligence revitalisation, it becomes imperative to embark on new research and exploration. This was initially intended to bridge the gap in existing knowledge and provide novel solutions for the digital intelligence revitalisation of cultural heritage sites. Interdisciplinary collaboration and communication are instrumental in fostering deep integration and innovation of knowledge.

5.2.2 Knowledge practise: Promoting diversified knowledge applications with smart data

In cultural heritage knowledge practice, smart data provide the foundation and support for knowledge sharing and application. Knowledge sharing promotes the diversification and innovation of knowledge applications. The ultimate objective and useful outcome of smart data and knowledge sharing is knowledge applications.

- *Smart Data.* The combination of digital technology and cultural heritage provides opportunities for smart data practice. By constructing smart data, it is possible to identify the connections, developmental trends, and potential value of cultural heritage sites. It can transform knowledge into productivity and provide support for digital deductions.
- *Knowledge Sharing.* Knowledge sharing plays an important role in cultural heritage. This serves as a bridge connecting smart data closely with knowledge applications for wide dissemination and social engagement. Based on the concept of open data and the FAIR principle, knowledge sharing enables more people to access and use smart cultural heritage data to participate in cultural heritage knowledge practices (Basaraba, 2018).
- *Knowledge Application*. Through knowledge contextualisation, narrative design, and iteration, cultural heritage knowledge can be better applied to practical scenarios to meet user needs. Knowledge contextualisation combines knowledge with specific contexts and user groups to improve knowledge applicability and usability. Furthermore, narrative design presents cultural heritage knowledge through storytelling to enhance user engagement and emotional empathy (Ioannidis et al., 2013). Moreover, iteration and optimisation can continuously improve the quality and effectiveness of cultural heritage knowledge applications.

5.2.3 Knowledge experience: Innovating user intellectual experience with theatre

Knowledge experience refers to innovation in the DDT of cultural heritage sites. Connecting the digital world with the real-world cultural heritage facilitates the flow of cultural heritage knowledge and creates new value.

- *Phygital Continuum.* The phygital continuum emphasises the interaction and collaboration between the digital and physical, enabling users to experience and explore cultural heritage in the virtual world (Turco & Giovannini, 2020). Through the aesthetic presentation of digital technology, the audience connects with cultural heritage knowledge and engages in the digital deduction process using various in-depth interaction approaches (Nofal et al., 2017), acting as creators of knowledge value.
- *Knowledge Flow.* Knowledge flow, i.e., the process of wisdom generation, begins with adequate access to data, processing it into information, and then internalising the information to form cognition and implement it into action (Kakabadse et al., 2003). In the knowledge experience of DDT, users first gain a vague understanding of cultural heritage knowledge, then analyse it, obtain a precise elaboration during the experience, and finally organise and implement it. Thus, cultural heritage knowledge has the maximum value.
- *Value Realisation.* Value realisation of the cultural heritage knowledge experience is reflected in three aspects: promoting learning, inspiring thinking, and expanding cognition. Through virtual exploration in DDT, people can gain a deeper understanding of history, art, and culture and promote holistic and in-depth learning. Such immersive experiences contribute to exploring the historical stories and cultural significance of cultural heritage, and stimulate creativity and imagination. In this context, knowledge experience is no longer merely passive acceptance, but a process that inspires thinking and helps enhance cognitive abilities.

5.3 Technology infrastructure

The DDT of cultural heritage takes data as the basis, models as the engine, and technology as the support to analyse, express, and represent cultural heritage in an all-round and multi-dimensional way. A modular technology architecture with adaptability and flexibility was built to guarantee the digital revitalisation of cultural heritage.

- *Data Analysis and Management.* Using statistical analysis, social network analysis, machine learning, and other professional knowledge, cultural heritage data can be processed and analysed to build various databases that provide data resources for the digital revitalisation of cultural heritage.
- *Model Development and Application.* Based on the results of the data analysis and management, various algorithms and techniques have been applied to

data mining, model construction, and optimisation, including knowledge graph models, digital-real symbiosis models, interpreter collaboration models, digital storytelling models, and large language models, to support the interpretation and presentation of cultural heritage.

- *Hardware and Software Support.* Enhances the adaptability and flexibility of the theatre, including intelligent voice dialogue platforms, multiscreen systems, project management systems, and visualisation software.
- *Situation Simulation and Narrative Development*. Allows users to explore and understand different types of cultural heritage deduction forms through role-playing or narrative involvement, deepening their understanding of cultural heritage and empowering them to become more effective knowledge creators.
- *Interactive Exhibition and Experience.* Provides users with diversified interactive displays and experience methods, including AR, VR, XR, and interactive installations, to better present the history, culture, and value of cultural heritage sites.
- *Research and Design.* Designing interactions based on user needs in a user-friendly manner enables professionals and non-specialists to engage in research, data exploration, computational modelling, and immersive experiences that promote cultural heritage research and education.
- *Sustainable Development Management.* Establish a solid management system, including personnel training, technical support, and quality control, to safeguard the sustainable development of DDT of cultural heritage.

6 Conclusion

The development of digital technologies and their widespread application in the humanities have fostered DH research. DH laboratories based on Big Data and AI are becoming a new frontier for research on the integration of the humanities with science and technology. Although we are still in the digital incunabula age, there is an urgent need to integrate humanities values into the machine-readable systems led by digital intelligence. This requires not only deep insights and critical thinking, but also the enhancement of scientific social ethics and values so that digital technologies and the humanities can be mutually reinforcing and empowering. This chapter introduces DDT as a novel methodological framework for the digital intelligence revitalisation of cultural heritage through a new theatre design. In the DDT, deduction is assigned a richer meaning. It integrates rational thinking and artistic expression, not only using scientific logic and reasoning ability to deeply analyse and interpret the connotation and value of cultural heritage, but also vividly representing cultural heritage through digital curation and visual display, forming a comprehensive methodological framework. Under its guidance, the Intellectual Computing Laboratory for Cultural Heritage at the Wuhan University is vigorously promoting the construction of a DTT to provide a feasible path. At present, theatres have launched a number of plays, such as the Aggregating Linked Data of Cultural Artefacts Images – China in Overseas Museums, high-precision three-dimensional reconstruction and immersive experience of the Arrow Nock Great

Wall and Dunhuang Caves, interactive digital storytelling of Mount Wutai Murals in Dunhuang, and virtual reality storytelling of the nine-coloured deer murals.

DDT is a transformative space that combines rationality and sensibility, science and humanities, materiality, and civilisation. It encompasses the entire process of user engagement, research, and practice, which not only organises and represents the full cycle of cultural heritage knowledge in a multi-dimensional and multi-perspective manner but also facilitates the interpretation and use of cultural heritage. The DDT fosters the development of future-oriented frontier research fields, redefines the paradigm of practical knowledge production grounded in digital logic, and shapes a new mode of knowledge innovation driven by material forces. By crossing disciplinary boundaries and challenging the cognitive and research models of traditional disciplines, DDT contributes to the production of forward-looking and innovative outcomes in practice, and promotes the research, preservation, and dissemination of cultural heritage. It also provides new insights for research in the Digital Humanities Laboratories. However, this experimental methodological framework has certain limitations. DDT relies on advanced digital technology, which may lead to high costs and a dependence on specialised technical personnel, limiting its widespread application. Additionally, differences in users' cultural backgrounds and technical proficiency can affect their experience and understanding of cultural content. Overall, as an emerging methodological framework, DDT holds great potential for advancing the protection and digital intelligence revitalisation of cultural heritage. Yet, it also faces challenges in technology, cost, and ethics. This methodological framework is expected to be further developed and refined in the future through continual technological innovation and interdisciplinary collaboration.

Acknowledgements

We thank for the funding of this research by the major project "Resources Development and Service of Cultural Heritage Smart Data" (Grant No. 21&ZD334) supported by the National Social Science Foundation of the People's Republic of China and the national key research and development program "Research and Development of Key Technologies and Platforms for Digital Services of National Archives of Publications and Culture Collection Resources (Typical Application Demonstration)" (Grant No. 2022YFF0901902) supported by the Ministry of Science and Technology of the People's Republic of China.

References

Adams, N. B. (2004). Digital intelligence fostered by technology. *Journal of Technology Studies*, *30*(2), 93–97. https://eric.ed.gov/?id=EJ905130

Arnold, T. B., & Tilton, L. (2022). *Analyzing audio/visual data in the digital humanities*. Bloomsbury Publishing. https://scholarship.richmond.edu/cgi/viewcontent.cgi?article=1272&context=mathcs-faculty-publications

Bakhshi, H., Schneider, P., & Walker, C. (2008). *Arts and humanities research and innovation*. Nesta. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b96b44bd446c3f9d17a3780f44497d5125594a97

Balme, C. B. (2008). *The Cambridge introduction to theatre studies*. Cambridge University Press.

Basaraba, N. (2018). Communication model for nonfiction interactive digital narratives: A study of cultural heritage websites. *Frontiers of Narrative Studies, 4*(s1), s48–s75. https://doi.org/10.1515/fns-2018-0032

Bekele, M. K., Pierdicca, R., Frontoni, E., Malinverni, E. S., & Gain, J. (2018). Survey of augmented, virtual, and mixed reality for cultural heritage. *Journal on Computing and Cultural Heritage (JOCCH), 11*(2), 1–36. https://doi.org/10.1145/3145534

Berry, D. M., & Fagerjord, A. (2017). *Digital humanities: Knowledge and critique in a digital age*. John Wiley & Sons.

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnapp, J. (2012). *Digital_Humanities*. MIT Press.

Busa, R. (1999). Picture a man... Busa Award Lecture, Debrecen, Hungary, 6 July 1998. *Literary and Linguistic Computing, 14*(1), 5–9. https://doi.org/10.1093/llc/14.1.5

Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. D. O., Kaplan, J., & Zaremba, W. (2021). Evaluation of large language models trained on codes. arXiv preprint arXiv:2107.03374. https://doi.org/10.48550/arXiv.2107.03374

Cila, N. (April 2022). Designing human-agent collaborations: Commitment, responsiveness, and support. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–18). https://doi.org/10.1145/3491102.3517500

Coltrain, J., & Ramsay, S. (2019). Can video games be used in humanities scholarship? In M. K. Gold & L. F. Klein (Eds.), *Debates in the digital humanities 2019* (pp. 36–45). University of Minnesota Press.

Davidson, C. N. (2008). Humanities 2.0: Promises, perils, and predictions. *PMLA, 123*(3), 707–717.

De Paolis, L. T., Gatto, C., Corchia, L., & De Luca, V. (2023). Usability, user experience, and mental workload in a mobile Augmented Reality application for digital storytelling in cultural heritage. *Virtual Reality, 27*(2), 1117–1143. https://doi.org/10.1007/s10055-022-00712-9

Doukianou, S., Daylamani-Zad, D., & Paraskevopoulos, I. (2020). Beyond virtual museums: Adopting serious games and extended reality (XR) for user-centred cultural experiences. *Visual Computing for Cultural Heritage*, 283–299. https://doi.org/10.1007/978-3-030-37191-3_15

Drucker, J. (2012). Humanistic theory and digital scholarship. In M. K. Gold (Ed.), *Debates in the digital humanities* (pp. 85–95). University of Minnesota Press.

Drucker, J. (2021). *The digital humanities coursebook: An introduction to digital methods for research and scholarship*. Routledge.

Drucker, J. (2022). Normative digital humanities. In J. O'Sullivan (Ed.), *The Bloomsbury handbook to the digital humanities* (pp. 7–10). Bloomsbury Publishing.

El Khatib, R., Arbuckle, A., Siemens, L., Siemens, R., & Winter, C. (2020). *An "Open Lab?" Electronic textual cultures laboratory in the evolving digital humanities landscape*. http://www.digitalhumanities.org/dhq/vol/14/3/000480/000480.html

Eshet, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia, 13*(1), 93–106. https://www.learntechlib.org/primary/p/4793/

Fu, X., Zhu, Y., Xiao, Z., Xu, Y., & Ma, X. (2020, April). RestoreVR: Generating embodied knowledge and situated experiences of Dunhuang mural conservation via interactive virtual reality. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). https://doi.org/10.1145/3313831.3376673

Gavin, M., & Smith, K. M. (2012). Interview with Brett Bobley. In M. K. Gold (Ed.), *Debates in the digital humanities* (pp. 61–66). University of Minnesota Press.

Guy, G. (2016). *Theatre, exhibition, and curation: Displayed & performed*. Routledge.

Hannaway, O. (1986). Laboratory design and the aim of science: Andreas Libavius versus Tycho Brahe. *Isis, 77*(4), 585–610. https://doi.org/10.1086/354267

Hockey, S. (2004). History of computing humanities. In S. Schreibman, R. Siemens & J. Unsworth (Eds.), *A companion to digital humanities* (pp. 1–19). Blackwell Publishing.

Holloway-Attaway, L. (2021). Interviews with Stella Wisdom and the digital curator at British Library. *gamevironments*. https://doi.org/10.26092/elib/925

Ioannidis, Y., El Raheb, K., Toli, E., Katifori, A., Boile, M., & Mazura, M. (2013, October). One objective of many stories: Introducing ICT in museums and collections through digital storytelling. In *2013 Digital Heritage International Congress (DigitalHeritage)* (Vol. 1, pp. 421–424). IEEE. https://doi.org/10.1109/DigitalHeritage.2013.6743772

Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G. (2015). Close and distant reading in digital humanities: Surveys and future challenges. *EuroVis (STARs), 2015*, 83–103. http://dx.doi.org/10.2312/eurovisstar.20151113

Kakabadse, N. K., Kakabadse, A., & Kouzmin, A. (2003). Review of knowledge management literature: Towards taxonomy. *Journal of Knowledge Management, 7*(4), 75–91. https://doi.org/10.1108/13673270310492967

Kenderdine, S. S. (2021). Experimental museology: Immersive visualisation and cultural (big) data. In M. Achiam, M. Haldrup & K. Drotner (Eds.), *Experimental museology: Institutions, representations, users* (pp. 15–34). Routledge.

Kuhn, T. S. (2012). *The structure of scientific revolutions*. University of Chicago Press.

Laurel. (2013). *Computers as theatre*. Addison-Wesley.

Lefebvre, H. (2012). From the production of space. In J. Collins & A. Nisbet (Eds.), *Theatre and performance design* (pp. 81–84). Routledge.

Liu, A. (2013). Meaning of digital humanities. *PMLA, 128*(2), 409–423. https://doi.org/10.1632/pmla.2013.128.2.409

Liu, C., Wang, D., Zhao, Z., Hu, D., Wu, M., Lin, L., Shen, S., Li, B., Liu, J., Zhang, H., & Zhao, L. (2023). *SikuGPT: A generative pretrained model for intelligent information processing of ancient texts from the perspective of digital humanities*. arXiv preprint arXiv:2304.07778. https://doi.org/10.48550/arXiv.2304.07778

Liu, Z., Li, Y., Cao, Q., Chen, J., Yang, T., Wu, Z., Hale, J., Gibbs, J., Rasheed, K., Liu, N., Mai, G., Liu, T., & Liu, T. (2023). *Transformation vs. tradition: Artificial General Intelligence (AGI) for the arts and humanities*. arXiv preprint arXiv:2310.19626. https://doi.org/10.48550/arXiv.2310.19626

Macrì, E., & Cristofaro, C. L. (2021). The digitalisation of cultural heritage for sustainable development: The impact of Europeana. *Cultural Initiatives for Sustainable Development: Management, Participation and Entrepreneurship in the Cultural and Creative Sector*, 373–400. https://doi.org/10.1007/978-3-030-65687-4_17

McDonald, C., Schmalz, M., Monheim, A., Keating, S., Lewin, K., Cifaldi, F., & Lee, J. H. (2021). Describing, organising, and maintaining video game development artefacts. *Journal of the Association for Information Science and Technology, 72*(5), 540–553. https://doi.org/10.1002/asi.24432

McElroy, M. (1999, April). Knowledge life cycle. In *ICM Conference on KM*, Miami, FL. https://www.academia.edu/download/35876565/KnlgLifeCycle.pdf

McInerney, C. (2002). Knowledge management and the dynamic nature of knowledge. *Journal of the American Society for Information Science and Technology, 53*(12), 1009–1018. https://doi.org/10.1002/asi.10109

Moreshead, A., & Salter, A. (2022). Digital humanities and media studies. J. O'Sullivan (Ed.), *The Bloomsbury handbook to the digital humanities* (pp. 93–100). Bloomsbury Publishing.

Murray, J. (2017). *Hamlet on the Holodeck, updated edition: The future of narrative in Cyberspace*. MIT Press.

Nofal, E., Reffat, M., & Vande Moere, A. (2017). Phygital heritage: An approach for heritage communication. *Immersive learning research network conference* (pp. 220–229). Verlag der Technischen Universität Graz; Graz, Austria. https://doi.org/10.3217/978-3-85125-530-0-36

Oiva, M., & Pawlicka-Deger, U. (2020). Lab and Slack. Introduction of situated research practices in the digital humanities to the DHQ special issue. *DHQ: Digital Humanities Quarterly, 14*(3). http://digitalhumanities.org:8081/dhq/vol/14/3/000485/000485.html

Parikka, J. (2016). Lab imaginary: Speculative practices in situ. Across & beyond. A transmediale reader on post-digital practices, concepts, and institutions, 78–91. https://archive.transmediale.de/content/the-lab-imaginary-speculative-practices-in-situ

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). *Generative agents: Interactive simulation of human behaviour*. arXiv preprint arXiv:2304.03442. https://doi.org/10.48550/arXiv.2304.03442

Pawlicka-Deger, U. (2020). Laboratory turn: Exploring discourses, landscapes, and models of the humanities laboratory. *DHQ: Digital Humanities Quarterly, 14*(3). http://www.digitalhumanities.org/dhq/vol/14/3/000466/000466.html

Pickering, A. (1992). From science as knowledge to science as practice. In A. Pickering (Ed.), *Science as practice and culture* (pp. 1–26). University of Chicago Press.

Pietroni, E., & Ferdani, D. (2021). Virtual restoration and virtual reconstruction in cultural heritage: Terminology, methodologies, visual representation techniques, and cognitive models. *Information, 12*(4), 167. https://doi.org/10.3390/info12040167

Presner, T., Schnapp, J., Lunenfeld, P., & Crowdsourced Contributions. (2008). *The digital humanities Manifesto 2.0*. http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf.

Pujol, L., Roussou, M., Balet, O., Vayanou, M., & Ioannidis, Y. (2012, March). Personalising interactive digital storytelling in archaeological museums: The CHESS Project. *The 40th annual conference of computer applications and quantitative methods in archaeology* (pp. 93–100). Amsterdam University Press. https://www.madgik.di.uoa.gr/sites/default/files/2018-06/caa2012_paper_final.pdf

Sahle, P. (2016). What is a digital scholarly edition? *Digital Scholarly Editing: Theories and Practices, 1*, 19–39. http://dx.doi.org/10.11647/OBP.0095.02

Schich, M., Song, C., Ahn, Y. Y., Mirsky, A., Martino, M., Barabási, A. L., & Helbing, D. (2014). The network framework of cultural history. *Science, 345*(6196), 558–562. https://doi.org/10.1126/science.124006

Scott, E. C. (2008). *Evolution vs. creationism: An introduction*. Bloomsburg Publishing.

Shapin, S., & Schaffer (2011). *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton University Press. https://doi.org/10.1515/9781400838493

Spiro. (2012). This is why we fight: Defining the values of the digital humanities. In M. K. Gold (Ed.), *Debates in the digital humanities* (pp. 16–34). University of Minnesota Press.

Strosberg, E. (2013). *Art and science*. WW Norton.

Sullivan, A. (2015). Cultural heritage and new media: A future for the past. *The John Marshall Review of Intellectual Property Law, 15*, 604. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/johnmars15&section=28

Svensson, P. (2010). Digital humanities landscape. *Digital Humanities Quarterly, 4*(1). http://www.digitalhumanities.org/dhq/vol/4/1/000080/000080.html

Sweeten, G., Sillence, E., & Neave, N. (2018). Digital hoarding behaviours: Underlying motivations and potential negative consequences. *Computers in Human Behavior, 85*, 54–60. https://doi.org/10.1016/j.chb.2018.03.031

Tenen, D. (2016). Blunt instrumentalism: Tools and methods. In L. F. Klein & M. K. Gold (Eds.), *Debates in the digital humanities 2016* (pp. 83–91). University of Minnesota Press.

Turco, M. L., & Giovannini, E. C. (2020). Towards a vital heritage approach to museum collections. *Journal of Archaeological Science: Reports, 34*, 102639. https://doi.org/10.1016/j.jasrep.2020.102639

Ulutas Aydogan, S., Münster, S., Girardi, D., Palmirani, M., & Vitali, F. (2019, October). A framework to support digital humanities and cultural heritage studies research. In *Workshop on research and education in urban history in the age of digital libraries* (pp. 237–267). Springer International Publishing. https://doi.org/10.1007/978-3-030-93186-5_11

Valtolina, S. (2016). A storytelling-driven framework for cultural heritage dissemination. *Data Science and Engineering, 1*(2), 114–123. https://doi.org/10.1007/s41019-016-0007-z

van Es, K., Schäfer, M. T., & Wieringa, M. (2021). Tool criticism and the computational turn. A "methodological moment" in media and communication studies. *M&K Medien & Kommunikationswissenschaft, 69*(1), 46–64. https://www.nomos-elibrary.de/10.5771/1615-634X-2021-1-46.pdf

Wershler, D., Emerson, L., & Parikka, J. (2022). *The lab book: Situated practices in media studies*. University of Minnesota Press.

Windhager, F., Federico, P., Schreder, G., Glinka, K., Dörk, M., Miksch, S., & Mayr, E. (2018). Visualization of cultural heritage collection data: State of the art and future challenges. *IEEE Transactions on Visualization and Computer Graphics, 25*(6), 2311–2330. https://doi.org/10.1109/TVCG.2018.2830759

11 Combining ontology and nanopublication models to reconstruct digital commentaries on ancient Chinese books

Mengjuan Weng, Xiaoguang Wang, Jueying Lei and Xilong Hou

1 Introduction

Chinese Classics are significant because they reflect the cultural values and orientations of the Chinese people. Commentaries are derivative texts formed through commentators' interpretations of classic texts. These commentaries not only reflect the commentators' understanding and values in their era but also play an irreplaceable role in contemporary people's comprehension of classic texts. In the Chinese context, we refer to Chinese Classics as "original documents" (原典), and the books where commentaries reside as "Zhushu documents" (注疏文献). Figure 11.1 presents an example of original and Zhushu documents, revealing that the original and commentary texts appear in different styles.

Zhushu documents serve as the primary carriers of commentaries. There are two common organisational methods for Zhushu documents. The first one is the commentator-centred organisation method, where the commentaries of one or more classic texts by the same commentator are compiled into one book. This results in reviews of the same classic text by different commentators being scattered across different types of literature in the form of printed books or their digital versions, making it challenging to obtain, read, analyse, and review them. The other method is the classic-centred organisation method, which involves combining different people's commentaries on the same classic text into one book. Nevertheless, the inclusion or exclusion of commentaries entirely depends on the compilers' intentions, leading to the possibility of some excluded commentaries being ignored and lost in the course of history. Furthermore, compilers generally express their opinions based on the collection of numerous commentaries. However, currently, access, reading, and research on commentaries are primarily based on printed or digitalised documents, lacking correlation and integration with external related documents. This inconvenience hampers reading comprehension and academic research on commentaries.

Today, with the vigorous development of digital humanities, preserving ancient books digitally alone is no longer sufficient to meet people's needs. Based on digital texts of ancient books, building a data model and using the model for semantic annotation of ancient books to create a knowledge base such as a knowledge graph or linked data to support digital reading and digital humanities research is the new

DOI: 10.4324/9781032707211-13

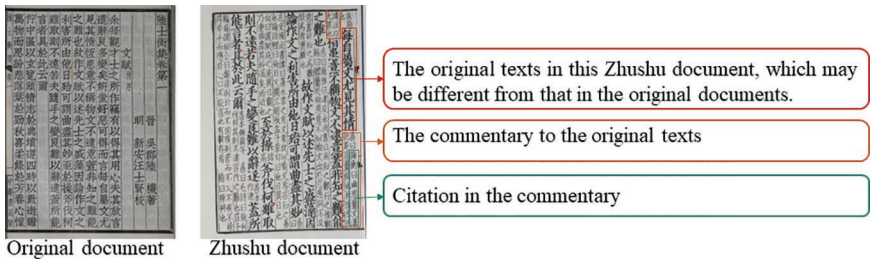

Figure 11.1 Example of original and Zhushu documents.

trend in the digitalisation of ancient books. It is also becoming an integral part of the collation of ancient books in the digital environment.

As interpretive texts of classic texts, commentaries play an almost equally important role as classic texts themselves. Using semantic annotation to present and correlate various commentaries or changes in the contents of different versions can provide additional references and research perspectives for studies based on ancient books. Therefore, it is necessary to digitally reconstruct commentaries to support digital reading and digital humanities research.

Guided by the concept of digitally reconstructing ancient books, this chapter proposes a path for reconstructing commentaries primarily based on ontology and nanopublications. This path aims to re-represent and reorganise commentaries based on existing Zhushu documents, transforming existing resources into smart data. This approach provides a foundation for innovative compilation, publication, reading, and analysis of commentaries.

2 Basic concepts

2.1 Digital reconstruction of ancient books

The digital reconstruction of ancient books is a systematic engineering endeavour, and its theoretical foundation is continuously evolving. It is closely intertwined with fields such as Big Data, data science, digital humanities, and others. Its theoretical underpinnings primarily encompass knowledge organisation, cultural heritage, digital preservation, and smart data utilisation (Lei et al., 2022).

To be more specific, the digital reconstruction of ancient books entails the re-exploration, reorganisation, and re-expression of the content contained within digital representations of ancient books, which serve as the tangible carriers of this historical knowledge. It represents an extension of the practice of digitising ancient books in the age of data intelligence, with the overarching goal of breathing new life into these ancient texts. The core essence of this process lies in employing data models, semantic web technologies, and other tools to semantically enhance and interconnect digitised ancient resources. This is achieved through excavation, reorganisation, design, expression, display, and dissemination of the historical knowledge and cultural significance encapsulated within ancient books. The aim is to transform digital resources of ancient books into smart data resources, thereby

achieving the digital reconstruction and activation of their content, presentation formats, and application scenarios. This, in turn, provides robust support and assurance for academic research in the humanities and cultural dissemination.

The digital reconstruction of ancient books encompasses three key aspects:

1 Content Carrier Aspect: In this aspect, the focus is on converting material carriers (such as rubbings, manuscripts, and rare books) into digital media formats. Digital media refers to data presented in various machine-readable formats, including digital text, images, videos, and online games.
2 Content Organisation Aspect: Concerning content organisation, the objective is to transform isolated text into a comprehensive knowledge network. Ancient books house an extensive repository of cultural knowledge from bygone eras. To externalise this complete knowledge system, it necessitates an accurate depiction of the semantic relationships between various knowledge elements and groups. This involves the collection and processing of multi-version, multimodal, and multimedia ancient book resources, along with semantic-level annotation, aggregation, association, and reorganisation. Furthermore, through knowledge unit deconstruction and semantic analysis, correlation mining, and other methodologies, fine-grained identification and representation of conceptual entities and knowledge units within ancient books can be achieved. This results in the creation of fine-grain data units, facts, and knowledge, forming collections like character knowledge bases and cultural gene repositories. Finally, with the aid of intelligent computing, human interpretation, and media conversion, it becomes possible to reorganise knowledge and present it across various themes and documents.
3 Presentation Form Aspect: In terms of expression, the goal is to shift from linear reading to hypertext reading. Full integration of ancient book resources not only assists relevant disciplines in uncovering new research questions but also enriches the ways in which ancient book culture is experienced by the public. The digital reconstruction of ancient books improves large-scale, cross-temporal and spatial, fine-grained content mining by systematically considering ancient book resources. It enhances the presentation and expression of ancient book content through the reorganisation and reuse of multidimensional and multimodal knowledge. By understanding the dynamics of digital activities and the unique characteristics of ancient books, this process fully unlocks the knowledge and resource value of ancient books, enriching the methods for their preservation and inheritance.

2.2 Commentaries and their characteristics

Commentaries are texts formed through the act of commenting and re-commenting on classic texts written by predecessors. In other words, commentaries represent the accumulation of knowledge by predecessors in the collation of ancient Chinese books. The interpretation of classic texts by scholars from different eras not only held positive significance for the society of their respective times but also reflected the values of those eras. For contemporary readers, commentaries serve as an essential foundation for understanding, inheriting, and disseminating the ideological connotations present in these classics.

Chinese scholars began annotating and interpreting ancient books as early as the Warring States Period (475/403 BC–221 BC; Fan, 2022). Prior to this, commentaries were written as separate works. However, later, for the sake of convenience in reading, commentaries gradually became attached to the original texts. Most of the existing commentaries are structured in the form of "original text – commentaries – re-commentaries" and various variations of this structure. This also constitutes the primary method for organising the content of the Zhushu documents.

Through time, the style of commentaries has become increasingly diverse, encompassing *zhuan* (传), *zhu* (注), *jian* (笺), *shu* (疏), *xun* (训), *gu* (诂), *kao* (考), *zheng* (证), *yin-yi* (音义), *zhang-ju* (章句), *jie-gu* (解诂), *jiao-zhu* (校注), *yi-shu* (义疏), *shu-zheng* (疏证), *shuo* (说), *lun* (论), *yan-yi* (衍义), *ping-zhu* (评注), *yi-zhu* (译注), *ping-yi* (评议), and more (Dong, 1997). Simultaneously, the scope of commentaries expanded from classic texts to include literary works.

Commentaries can be broadly categorised into four types: self-commentaries, commentaries by others, supplementary commentaries, and collective commentaries (Rao, 1996). The level of granularity for annotation varies, encompassing words, phrases, sentences, or sentence groups, as well as entire articles. Commentaries serve different purposes in understanding the original text, including phonetic notation, interpretation, grammatical explanations, research into and introductions about the author's life, thoughts, creative intentions, and the context of the book's writing. They also involve the analysis, evaluation, and exploration of the ideological significance of the work, textual research, explanations, and the supplementation of historical facts and allusions to famous figures, as well as the appreciation and evaluation of literary and artistic works, and the collection and analysis of supplementary materials. In general, the functions of commentaries can be categorised into three main areas: one primarily focuses on explaining words with an emphasis on knowledge, another emphasises the expression of principles with an emphasis on meaning, and the third combines exegesis and the expression of principles while also encompassing dimensions of knowledge and meaning.

Commentaries often contain numerous references to other texts. From a content perspective, these citations may include references to other people's commentaries, original classics such as *The Analects of Confucius* (论语), and reference books like dictionaries. Although there are no strict norms and standards for citing other texts in commentaries, from a structural perspective, there are two main forms of citation: the association between the author of the quoted sentence and the quoted sentence itself, and the connection between the ancient books containing the quoted sentence and the quoted sentence itself.

2.3 Ontology

Ontology originates from the field of philosophy and explains the characteristics and interrelationships of things. It is a systematic description of objective entities in the world. In the field of artificial intelligence, Gruber (1993) proposed a widely recognised definition, which asserts that "An ontology is an explicit specification of a conceptualization" characterised by clarity and formalisation (p. 199). In other

words, ontology defines the fundamental terms and their relationships that compose the vocabulary of a specific subject area, along with the rules governing the expansion of this vocabulary (Gruber, 2009; Tambassi, 2021).

In terms of its structural composition, ontology primarily comprises classes, attributes, and instances. Attributes encompass both object properties and data properties. Among these, a class provides a standardised description of domain concepts, object properties represent the relationships between different classes, and data properties of a class denote its attributes.

In recent years, ontology engineering has progressively emerged as the core of knowledge organisation and management. The concepts and methodologies for constructing ontologies have matured significantly, including approaches such as IDEF-5, the skeleton method (Uschold & Gruninger, 1996), ontology design patterns (Gangemi & Presutti, 2009), and NeOn (Suárez-Figuero et al., 2011).

2.4 Nanopublications

Nanopublication is a novel form of publication based on the nanopublication model, and its implementation aligns with the Semantic Web, affording nanopublications the advantage of generating machine-readable linked data. The nanopublication model is a contemporary knowledge representation and semantic framework introduced in the early 21st century. It has found widespread applications in various fields, including life sciences, computer science, humanities, and social sciences (Amith & Tao, 2018; Beck et al., 2012; Dimitrova et al., 2021; Ehrhart et al., 2021; Fu et al., 2015; Kuhn et al. 2015; Kuhn et al., 2018; Li et al., 2021).

The term "nano" in nanopublication signifies a formalised and machine-readable way of communicating the smallest possible units of publishable information (Robin & Sven, 2020). After more than a decade of development, the nanopublication model has achieved a stable structure (Groth et al., 2010; Mons et al., 2009). The core concept of this model revolves around associating information units with their contextual information. Through the formal representation of these information units and their context, it supports the integration, querying, and reasoning of information units. Specifically, the nanopublication model comprises three modules: assertion, provenance, and the publication information module. Each module constitutes an RDF graph, with the RDF graph consisting of "subject-predicate-object" triples. Among these modules, the assertion module refers to the nanopublication statement; the provenance module aims to elucidate how the assertion was generated, who generated it, when it was generated, and where it was extracted, and the publication information module provides provenance information about the nanopublication itself, including, at a minimum, the creator and timestamp.

According to Gray et al. (2023), well-formatted nanopublications must adhere to the following eight standards:

1 A nanopublication comprises a set of RDF quadruples, which include subject-predicate-object+context.
2 The context of each triplet (such as an RDF graph) must be specified as a valid URI.

3 There exists one and only one quadruple in the form of "[N] rdf:type np:Nanopublication [H]," where [N] represents the nanopublication URI, and [H] is the head URI.
4 Given [N] and [H], there exists one and only one quadruple in the form of "[N] np:hasAssertion [A] [H]," where [A] denotes the assertion graph URI.
5 Given [N] and [H], there exists one and only one quadruple in the form of "[N] np:hasProvenance [P] [H]," where [P] signifies the provenance graph URI.
6 Given [N] and [H], there exists one and only one quadruple in the form of "[N] np:hasPublicationInfo [I] [H]," where [I] represents the publication information URI.
7 [N], [H], [A], [P], and [I] URIs must all be distinct.
8 All triples must be placed within [H], [A], [P], or [I].
9 The triple in [P] must have at least one reference to [A].
10 [I] must have at least one reference to [N].

Additionally, to facilitate checking whether the nanopublication has been altered, ensuring the immutability of the nanopublication is mandatory, and the use of trusted URIs is recommended (Kuhn & Dumontier, 2014).

3 Reconstruction path of ancient book commentaries

This section discusses the process of reconstructing ancient book commentaries. To better explain this process, we will first summarise the five data states and four data conversion paths for the digital reconstruction of ancient books, based on the general requirements for such reconstructions (Figure 11.2). The proposed semantic representation path primarily focuses on transforming text data into semantic data.

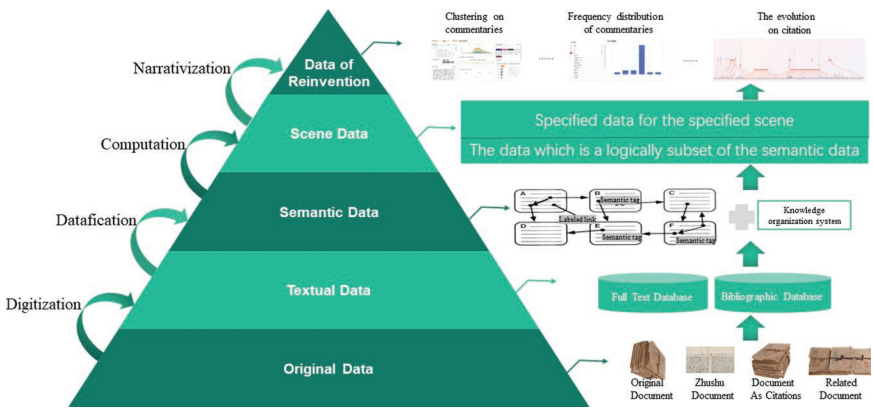

Figure 11.2 The reconstruction path of ancient book commentaries.

3.1 Digitisation: Transformation of raw data into text data

The original data consist of ancient books that exist in various material forms, including printed copies, inscriptions, unearthed documents, and bamboo slips. These original data are scattered across numerous public and private collection institutions, individuals, and ancient book auctions and business markets both domestically and abroad. In the reconstruction of ancient book commentaries, the relevant original ancient books must include at least the original documents, Zhushu documents, the documents citing the commentaries, the documents cited in the commentaries, and other related documents.

Text data refers to text stored in an encoded format that can be processed by a computer. In this context, we are referring to text encoded using text encoding standards such as the Unicode Standard, GB2312, and the ancient Chinese book character-encoding scheme.

The conversion of original data into text data primarily involves two steps: scanning or copying and textualisation.

1 Scanning involves using devices such as scanners or cameras to capture text images from ancient books. The data generated at this stage consists of digital image resources and their associated metadata information.
2 Textualisation refers to the process of converting scanned or copied ancient books into text format, also known as transcription. During this stage, optical character recognition (OCR) technology and crowdsourcing are typically employed (Sturgeon, 2018). It should be noted that the textualisation process of Zhushu documents is considerably more complex than that of original ancient books and other texts. This complexity arises because in Zhushu documents, the original text and the annotated text often appear in different formats, with the annotated text typically having a smaller font size and less fixed positioning.

3.2 Datafication: Transformation of text data into semantic data

Semantic data refers to text data that has been enhanced semantically. The process of semantic enhancement includes three main components: enhancement based on punctuation, on metadata, and on external data linking. Punctuation-based enhancement involves segmenting ancient Chinese texts because Chinese books, including Zhushu documents, were originally written and printed without punctuation before the May 4th Movement. This segmentation can be achieved either manually or automatically (Wang et al., 2016; Qian et al., 2021).

This section introduces a semantic representation scheme based on nanopublications and ontologies to facilitate enhancement through metadata and external data linking. Through this semantic representation method, we transition from an isolated commentary text to a commentary knowledge network.

Nanopublications offer advantages in minimising the independent publication of knowledge units, generating linked data, ensuring traceability, and maintaining credibility. Ontologies are valuable for semantic modelling. By combining these

two approaches and using them for the semantic representation of commentary knowledge, we not only reduce the need for separate publications of commentary knowledge while revealing the internal semantic relationships within it but also ensure the traceability of responsible entities related to republishing activities and maintain the credibility of the published content. Therefore, the term "semantic data" in this chapter refers to the entirety of data composed of ontology, nanopublications, and related instances, primarily consisting of commentaries and their attributes.

Upon analysing the characteristics of commentaries, we identify four types of minimum knowledge units within commentaries: interpretation, citation, provenance, and alignment. The first three knowledge unit types pertain to metadata-based enhancements and necessitate the design of metadata or data models. The data model utilised in this study is based on Wang's (2023) contribution. The fourth type of knowledge unit is suitable for enhancement based on external data linking.

1 Interpretation knowledge unit: Essentially, commentaries result from the interpretation of texts. Once an author completes the text creation, the understanding of the text's meaning becomes dynamic and subject to change, and the author's intention becomes uncertain. Owing to differences in knowledge structures and interpretation methods, different interpreters will produce varying commentary texts for the same text. These commentary texts, when combined with the original text, constitute an interpretation knowledge unit. In the HERU ontology, the classes *heru:InterpretationObject* and *heru:InterpretationResult* are defined as subclasses of the first-level class *heru:TextSegment*. Instances of both *heru:InterpretationObject* and *heru:InterpretationResult* can be of varying lengths to accommodate different types of interpretation unit representations, such as titles, words, sentences, paragraphs, articles, and more. The relationship between *heru:InterpretationObject* and *heru:InterpretationResult* is represented by *heru:hasInterpretation*.
2 Citation knowledge unit: Over time, commentators' original intentions behind their comments may no longer be accessible. However, commentators often quote scriptures when providing commentary (Shu et al., 2019), offering an alternative means to understand their commentary. The commentary text and the quoted text together form a citation knowledge unit. In the HERU ontology, the classes *heru:InterpretationResult* and *heru:Citation*, respectively, represent the citing unit and the cited unit. They are also subclasses of *heru:Text Segment*. The relationship between the *heru:Interpretation Result* and *heru:Citation* classes is expressed as *cito:cites*.
3 Provenance knowledge unit: Provenance units encompass two categories: physical provenance and original source. Physical provenance refers to the physical document containing the text, whereas the original source is the document in which the text was initially created. The significance of provenance units lies in describing the source of the text and ensuring the accuracy and verifiability of the text's content. In the HERU ontology, three provenance attributes are designed: *dcterms:creator*, *prov:hadPrimarySource*, and *hico:isExtractedFrom*.

The *dcterms:creator* attribute describes the author of the text, while the latter two source attributes help differentiate identical or nearly identical content objects from different sources.

4 Alignment knowledge units: Alignment units are divided into two categories: alignment of the interpreted text with the original text and alignment of the cited text in the commentary with the text in the cited documents. The former type of alignment unit enables the automatic collection of related commentaries within the same classic, while the latter type establishes a connection between citations in the commentaries and the cited documents. This reduces the spatial distance between the citations in the commentaries and the cited documents in printed media, making the connection between citations and commentary texts more evident, accessible, and calculable. In the interpretation ontology, alignment takes five forms: (a) alignment between the interpretation object and the original text, (b) alignment between the citation and the cited text, (c) alignment between the citation and the interpretation result, (d) alignment between the citation and the text segments, and (e) alignment between text segments. Among these, the alignment relationship between the interpretation object and the original text exists because many existing Zhushu documents are integrated into the classic text, which may differ from the original documents where the classic texts are originally found. The alignment relationship between the citation and cited texts exists because some commentaries cite other ancient books, such as dictionaries, or other classic documents, such as Analects of Confucius. This alignment relationship facilitates the connection between the cited documents and Zhushu documents. An alignment relationship between citations and interpretation results exists, because some commentators cited the commentaries of their predecessors while commenting. This relationship helps establish the connection between Zhushu documents and aids in organising the inheritance of Zhushu documents.

The interpretation, citation, and alignment units serve as the assertion module of the nanopublication, while the provenance unit serves as the provenance module of the nanopublication (Figure 11.3). Additionally, publication information content should be supplemented based on the publishing information of the nanopublication.

3.3 Computation: Transformation of semantic data into scene data

Scenario data refers to data oriented towards user needs. Data for new scenarios can be obtained through semantic queries, calculations, and other methods. These data may involve reorganising the original data or generating new data.

There are two methods for converting semantic data into scene data. The first is semantic queries, in which users use natural language to make queries. To obtain the data needed for a specific scene, the natural expression of scene data must be converted into clear query statements. The second is text vectorisation calculations. Some specific scenarios require data that cannot be obtained solely through

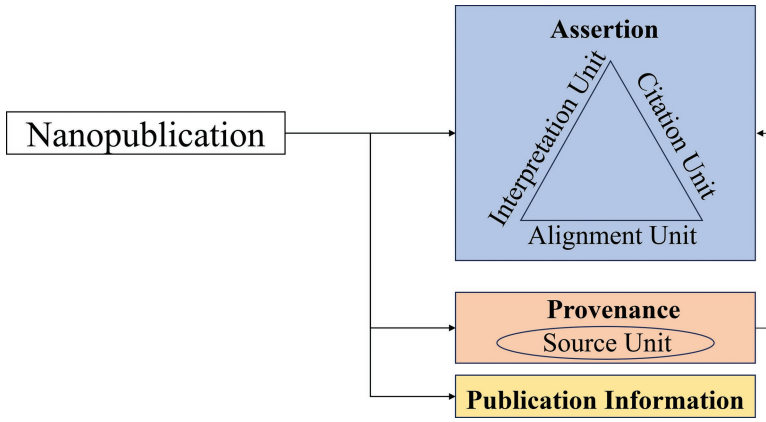

Figure 11.3 Commentary semantics based on ontology and nanopublication.

queries. In such cases, it is necessary to perform vectorisation calculations on text using technologies like natural language processing, representation learning, and reasoning based on knowledge graphs.

3.4 Narration: Transformation of scene data into data of reinvention

Reconstruction refers to the creation of new media products based on the narrative expression of scene data.

Reconstruction can take two basic forms: one involves the reorganisation and representation of scene data, while the other involves the comprehension and re-expression of scene data. The first form aims to preserve the original content of ancient books, albeit fragmented and reconnected, and present it in a new media format. In other words, by systematically exploring ancient book resources across time and space, one can achieve detailed content mining. Through the reorganisation and reuse of multidimensional and multimodal knowledge, the presentation and expression of ancient book content can be enriched. The other form involves changing the way people engage with ancient books by using media platforms such as variety shows, documentaries, virtual reality, and interactive games. In this approach, the narrator must possess a profound understanding of the ancient book's content and recreate it in a medium different from the original book. By creating immersive reading experiences, VR/AR reading, interactive reading, and other innovative formats for ancient books, readers can engage with these texts in entirely new ways.

Both forms require strong narrative skills, which can be categorised into two modes: creator- and audience-driven narration (Segel & Heer, 2010).

4 Digital reconstruction of *Wen Fu*'s commentaries

We examine Wen Fu due to its short length, which allows for a quick implementation of prototypes. *Wen Fu (文赋)* is a literary theory work by Lu Ji (陆机) of the Western Jin dynasty (266–316). The complete text is found in the

Wen Xuan (Selections of Refined Literature, 文选) compiled by Xiao Tong, who was formally known as the Crown Prince. *Wen Fu* is a masterpiece in the history of Chinese literary theory and criticism, serving as the first comprehensive and systematic study of the fundamental theories of literary creation. We have gathered various versions of *Wen Fu* and their corresponding commentaries from libraries, converting them into textual data. Subsequently, we employ ontology and a nanopublication template to represent commentary knowledge contained within the commentaries. In other words, we facilitate the transformation from textual data to semantic data through content extraction, type identification, and template completion steps. Finally, utilising the semantic data, we demonstrate how to acquire scene data and subsequently redesign it to present a reconstructed state in a new media format.

4.1 From original ancient books to codable text

Using Zhang Shaokang's *Wen Fu Ji Shi (文赋集释)* as a reference, we identified a total of 24 works of *Wen Fu* and their commentaries (Table 11.1).

We initially utilised scanning equipment to obtain digital images of *Wen Fu* and its commentaries. The content of *Wen Fu* is depicted in Figure 11.1 (left), while the content of *Wen Fu*'s commentaries (using *Wen Xuan* as an example) is presented in Figure 11.1 (right). Subsequently, we employed OCR recognition and simplified and traditional Chinese conversion to transform the text into unicode-encoded text. As the OCR recognition accuracy did not reach 100%, we engaged in crowdsourcing to break down the text proofreading tasks and successfully completed the proofreading for the aforementioned books.

4.2 From encoded text to semantic data

As previously mentioned, we utilise the HERU ontology as a data model for classical texts and their commentaries. Following the nanopublication representation framework, we can derive semantic data. This process comprises the following steps:

Table 11.1 Books on *Wen Fu* and its commentaries

Book title	*Author*	*Publication year*
Wen Fu (The Ten Volumes Collection of Lu Shiheng 陆士衡文集十卷)	Write by Lu Ji; Proofread by Wang Shixian	1936
Wen Fu (Wenxuan Noted by Six Ministers六臣注文选)	Compiled by Prince Zhaoming of Liang	1929
Wen Fu (Proof of works of Lu Ji 陆机集校笺)	Write by Lu Ji; Proofread by Yang Ming	2016
Wen Fu (Commentary on Selected Works of Zhaoming 评注昭明文选)	Compiled by Prince Zhaoming of Liang; Compiled by Yu Guanghua of Qing dynasty	1919
Wen Fu (Selected Works Sixty Volumes, 文选)	Compiled by Prince Zhaoming of Liang; Compiled by Hu Kejia of Qing dynasty	1912–1949

1 Extraction of resources and attributes involves identifying text fragments suitable as resource objects and attributes, treating attributes as the smallest knowledge units, and assigning URIs to the resource objects.
2 Semantic annotation, based on classes and relationships within the ontology, reveals the semantic categories of resource objects and establishes semantic relationships between resource objects, as well as between resource objects and attributes.
3 Template filling encompasses module and attribute filling. We insert the structured knowledge units into their respective modules, as illustrated in Figure 11.3, and subsequently supplement the nanopublication with the relevant publication information. At this point, a nanopublication is created. Although the nanopublication includes a URI, it does not indicate whether the content within it has changed, rendering the URI untrustworthy.
4 To generate a trusted URI, we employ signature encryption technology, such as a hash algorithm, to create a summary of the information. This summary is then appended to the original URI, resulting in a trusted URI.

Considering the relatively small volume of literature, we have completed the aforementioned steps manually.

4.3 User demand identification and scene data acquisition

Based on the results of the literature survey, we have identified that the needs of users can be categorised into two main groups: first, the need for commentary on specific sentences within the text and their chronological distribution, and second, the need for information regarding the documents cited in *Wen Fu*'s commentaries. To acquire this type of data, we have employed the method of creating SPARQL query templates (Figure 11.4).

Additionally, taking into account the needs of humanities scholars for remote reading within the digital humanities paradigm, we have also designed cluster analysis and word frequency analysis of commentary texts to assist users in quickly acquainting themselves with the text's content. Clustering was performed using the IDC algorithm, as proposed by Sherkat et al. in 2018. The scene data for cluster analysis includes information such as the document's cluster assignment, the similarity between the document and the cluster, the cluster keywords, and the correlation between the keywords and the cluster. The method for acquiring this type of scene data involves utilising NLP technology to quantitatively analyse the text based on SPARQ query results.

```
construct {?ot heru:hasInterpretation ?ir. ?ir prov:startedTime ?year. ?ir
cito:cites ?q} {
    ?io owl:sameAs ?ot.
    ?ot a heru:OriginalText.
    ?ot rdfs:label ?ot_name.
    ?io heru:hasInterpretation ?ir.
    ?ir a heru:InterpretationResult.
    ?io a heru:InterpretationObject.

    ?io rdfs:label ?io_name.
    ?ir rdfs:label ?ir_name.
    optional{?ir prov:startedTime ?year.}
    optional{?ir cito:cites ?q1.?q1 owl:sameAs ?q}
}
```

```
select ?bf1_name ?p11 ?irr ?bf2_name ?p22 ?c1 where {
    ?ir cito:cites ?c. ?ir a heru:InterpretationResult. ?c a heru:Citation.
    ?c rdfs:label ?c1. ?ir rdfs:label ?irr.
    optional {
        ?ir prov:hadPrimarySource ?bf1. {?bf1 a bf:Item} union {?bf1 a bf:Instance}
        ?bf1 rdfs:label ?bf1_name.}
    optional {
        {?c10 owl:sameAs ?c. ?c11 owl:sameAs ?c10. } union {?c11 owl:sameAs ?c.}
        ?c11 rdfs:label ?cc11.
        ?c11 prov:hadPrimarySource ?bf2. {?bf2 a bf:Item} union {?bf2 a bf:Instance}
        ?bf2 rdfs:label ?bf2_name.}
    optional {?p1 a foaf:Person. ?ir dcterms:creator ?p1. ?p1 rdfs:label ?p11.}
    optional {?p2 a foaf:Person. ?p2 rdfs:label ?p22. ?c dcterms:creator ?p2. }
}
```

Figure 11.4 SPARQL query templates.

4.4 Re-expression of scene data

Designing a reconstructed state based on scene data necessitates narrative skills, and it is crucial to ensure that reconstructions are presented in a user-friendly manner. We have designed various reconstruction states based on scene data, as depicted in Figure 11.5. Unfortunately, we have not yet assessed to what extent the user needs have been satisfied. The data are available online (http://47.104.72.126:40102/#/).

Specifically, we have achieved four reconstruction states based on four sets of scenario data. These four reconstruction states all belong to the first category, as described in Section 3.4. In other words, it involves the breaking up and reorganisation of the text.

1 Commentary on specific sentences in *Wen Fu*: As shown in Figure 11.5(a), we present all commentaries on specific sentences in *Wen Fu* in a readable format. While reading *Wen Fu*, users can click on a specific sentence to view its commentaries and the Zhushu documents where they are located.
2 Distribution of commentaries: As shown in Figure 11.5(b), we present the number of commentaries from different eras in a bar chart, while allowing users to explore commentaries from a specific era further. Specifically, when a user clicks on the bar chart representing a particular era, the data on the left will respond dynamically in real time, enabling the user to delve into further details of the Zhushu literature from that era.
3 Documents cited in commentaries: In Figure 11.5(c), we present the cited documents in the commentaries in the form of a dashboard. These cited literature pieces serve as the foundation for understanding *Wen Fu*. The dashboard comprises three sub-windows: a Sankey diagram, a list of cited documents and their respective locations, and a bar chart illustrating document citation frequencies.

Figure 11.5 Reconstruction states of *Wen Fu*'s commentaries.

The Sankey diagram displays the cited documents from all Zhushu documents. Users can click on Zhushu documents to view their citations. The thickness of the lines connecting Zhushu documents to cited documents indicates the frequency of citation. The thicker the line, the higher the citation frequency. The length of a Zhushu document represents the number of documents it cites. The longer the length, the more documents it cites. Users can further explore the relevant details of specific Zhushu documents and cited documents. When a user clicks on a cited document, the list below it will respond dynamically in real-time, displaying the Zhushu documents that cite the document, the frequency of citations, and the location of each citation. The bar chart and location list respond dynamically in a manner similar to the distribution of commentaries.

4 Cluster analysis: Figure 11.5(d) presents the cluster analysis in the form of a dashboard. Cluster analysis comprises multiple views. From top to bottom and left to right, they are as follows: (a) Title panel: The user can set the confidence level, click a button to restore the previous clustering result, and click a button to initiate re-clustering. (b) Cluster tree view: Display clusters and documents in a folder tree format. (c) Cluster word cloud view: Present a word cloud diagram for a specific cluster. (d) Document cluster view: A parallel coordinator diagram displays the correlation between documents, clusters, and the text content of the document. (e) Clusters of Zhushu documents in different time periods. (f) Documents belonging to a specific cluster in the selected period and the similarity within the cluster. (g) Content of documents in a specific cluster during the selected period. (h) Keyword view of the cluster: By default, each rectangle contains the first five keywords of each cluster. (i) Term-cluster view: A parallel coordinator diagram shows the relevance of one or more terms to the cluster. When a user clicks on a cluster in (a), the word cloud diagram for that cluster is updated, as shown in (b); when a user clicks on a document in (a), the similarity between the document and each cluster is updated, as shown in (c). Additionally, by clicking the "Switch" button, users can view the document's text content. Similarly, when clicking on a specific cluster in (h), (i) is updated in real time with the top five keywords for the cluster and the correlation between each keyword and the cluster. The cluster analysis interface also allows users to customise clustering results. Users can set their confidence level, adjust document or keyword clustering, add new categories, etc., and then click the "Re-execute Clustering" button to obtain personalised clustering results.

5 Discussion

5.1 Platform tool requirements for digital reconstruction of commentaries

The Zhushu documents are not only the result of compiling ancient books but also a concentrated expression of the academic innovation of ancient scholars. This innovation is based on the systematic sorting and integration of documents.

Building upon the semantic representation path of commentary knowledge proposed in Section 3.2, ancient book organisers can transform their personal

understanding of the connotation of commentary knowledge into explicit semantic relationships. This transformation provides a more structured and intelligent data foundation for researchers, offering an essential corpus for text analysis and comprehension using artificial intelligence technology.

However, ancient book organisers often have a limited familiarity with semantic technology. The commentary semantic publishing platform can simplify complex technical processes into user-friendly interfaces, reducing the learning curve for ancient book organisers and enhancing the efficiency of ancient book collation. Currently, support for semantic publishing platforms is lacking in China.

The core distinction between the nanopublication publishing platform and existing digital text annotation platforms or knowledge graph construction platforms lies in its support for nanopublication templates, the generation of trusted URIs, and version control. Nevertheless, the use of named graphs in nanopublications renders it incompatible with existing text annotation platforms for publishing. Future attention should be given to adding relevant functions to existing text annotation platforms to create a user-friendly nanopublication publishing platform.

5.2 Automatic construction of commentary knowledge units

This chapter presents the initial introduction of the commentary knowledge, along with its application through several examples. In the future, there will continue to be numerous challenges in structuring, correlating, and semantically representing extensive commentaries. Some studies have made progress in aligning interpretation objects and classic original texts using technologies like machine learning. These efforts have laid the groundwork for constructing knowledge networks for large-scale Zhushu documents. However, in the process of automatically building commentary knowledge units, we need to focus not only on automatically aligning interpretation objects with the classic original text but also on automatically aligning citations and quoted sentences in the interpretation results and automatically identifying interpretation objects and interpretation results.

The Chinese Text Project (cText) currently only enables hypertext reading of citations and cited documents (please note that this may not be entirely accurate), and it has not yet reached the level of refining quoted sentences. One of the primary reasons for the difficulty in advancing this work is that different commentaries on the same classic have varying degrees of digitisation.

Currently, the majority of Zhushu documents related to the same original documents are still in paper format, which makes it challenging to automate the identification of interpretation objects and interpretation results in Zhushu documents, let alone the identification of citations in the interpretation results. Moreover, the cited documents have not been digitised yet, resulting in a lack of data for automatically aligning citations and quoted sentences in the interpretation results. Additionally, aligning references presents a challenging problem as it is impossible to determine which version of the content the references in the commentaries correspond to. This determination needs to be made by analysing various factors such as the era of the commentator and the acceptable version.

5.3 Commentary scene requirement identification and reconstruction design

Understanding and clarifying the usage habits and needs of different target users regarding commentaries will help further improve and promote the digital reconstruction of commentaries. Faced with the transformation of humanities academic research driven by digital humanities, the digital reconstruction of commentaries requires a deeper understanding of user experience demands in the digital reading of ancient books, academic exchanges, fine-grained content analysis, and knowledge mining. It also aims to improve the efficiency and accuracy of resource acquisition for scientific researchers.

In response to the public's demand for preserving excellent traditional culture, the digital reconstruction of commentaries should be provided through diversified and three-dimensional mass communication methods. This will enable the creation of richer and better digital products, such as audio readings, knowledge explanations, VR/AR books, 3D ancient books, interactive games, and historical restoration. Alternatively, with the help of short videos and other internet platforms, we can explain and disseminate the knowledge within commentaries in a format that resonates with the public. By innovatively interpreting the cultural connotations of commentaries, we can serve the public in cultivating their understanding of excellent traditional cultural literacy and basic education. Commentaries originally served the purpose of understanding pre-Qin texts, and today, they can still play an essential role in helping contemporary people understand ancient thoughts.

6 Summary

Based on the overall requirements of digital reconstruction, this chapter proposes a commentary reconstruction path, especially focusing on refining the representation of commentary semantic data. Finally, this chapter uses *Wen Fu* and its commentaries as an example to illustrate the practice of digitally reconstructing commentaries.

Acknowledgments

We would like to express our gratitude for the funding support from the major project "Resources Development and Service of Cultural Heritage Smart Data" (No. 21&ZD334) supported by National Social Science Foundation of China. We also thank Wuhan University Library for providing the collection data related to *Wen Fu*, and Yajing Lv, Xuanyi Zhao, Shuya Pan and others for their contributions in data proofreading and annotation.

References

Amith, M., & Tao, C. (2018). Representing vaccine misinformation using ontologies. *Journal of Biomedical Semantics, 9*, 22. https://doi.org/10.1186/s13326-018-0190-0

Beck, T., Free, R. C., Thorisson, G. A., & Brookes, A. J. (2012). Semantically enabling a genome-wide association study database. *Journal of Biomedical Semantics, 3*, 9. https://doi.org/10.1186/2041-1480-3-9

Dimitrova, M., Senderov, V. E., Georgiev, T., Zhelezov, G., & Penev, L. (2021). Infrastructure and population of the OpenBiodiv biodiversity knowledge graph. *Biodiversity Data Journal, 9*, e67671. https://doi.org/10.3897/BDJ.9.e67671

Dong, H. L. (1997). *Interpretation of ancient books*. Liaoning Education Press.

Ehrhart, F., Willighagen, E. L., Kutmon, M., van Hoften, M., Curfs, L. M. G., & Evelo, C. T. (2021). A resource to explore the discovery of rare diseases and their causative genes. *Scientific Data, 8*(1), 124. https://doi.org/10.1038/s41597-021-00905-y

Fu, G., Batchelor, C., Dumontier, M., Hastings, J., Willighagen, E., & Bolton, E. (2015). PubChemRDF: Towards the semantic annotation of PubChem compound and substance databases. *Journal of Cheminformatics, 7*, 34. https://doi.org/10.1186/s13321-015-0084-4

Gangemi, A., & Presutti, V. (2009). Ontology design patterns. In S. Staab & R. Studer (Eds.), *Handbook on ontologies* (pp. 221–243). Springer Berlin Heidelberg.

Gray, A., Burger, K., Kotoulas, S., Loizou, A., Tkachenko, V., Waagmeester, A., … & Wilkinson, M. (2023, November 17). *Nanopublication guidelines*. Nanopublications. https://nanopub.net/guidelines/working_draft/

Groth, P., Gibson, A., & Velterop, J. (2010). The anatomy of a nanopublication. *Information Services & Use, 30*(1–2), 51–56. https://doi.org/10.3233/ISU-2010-0613

Gruber, T. (2009). Ontology. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of database systems*. Springer. https://doi.org/10.1007/978-0-387-39940-9_1318

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199–220. https://doi.org/10.1006/knac.1993.1008

Kuhn, T., Chichester, C., Krauthammer, M., & Dumontier, M. (2015). Publishing without publishers: A decentralized approach to dissemination, retrieval, and archiving of data. In *Proceedings of the 14th International Semantic Web Conference (ISWC)*. http://arxiv.org/abs/1411.2749

Kuhn, T., & Dumontier, M. (2014). Trusty URIs: Verifiable, immutable, and permanent digital artifacts for linked data. In *Proceedings of the Semantic Web: Trends and Challenges: 11th International Conference* (pp. 395–410). Springer International Publishing. https://arxiv.org/abs/1401.5775

Kuhn, T., Meroño-Peñuela, A., Malic, A., Poelen, J. H., Hurlbert, A. H., Ortiz, E. C., … & Dumontier, M. (2018). Nanopublications: A growing resource of provenance-centric scientific linked data. In *Proceedings of the 14th IEEE International Conference on E-Science (E-Science)* (pp. 83–92). IEEE. https://doi.org/10.1109/eScience.2018.00024

Lei, J., Jueying, H. X., Xilong, W., & Xiaoguang, X. (2022). The logic and approach of digital reconstruction of ancient books in the data intelligence. *Digital Humanities Research, 2*(2), 46.

Li, X. Y., Peng, S. Y., & Du, J. (2021). Towards medical knowmetrics: Representing and computing medical knowledge using semantic predications as the knowledge unit and the uncertainty as the knowledge context. *Scientometrics, 126*(7), 6225–6251. https://doi.org/10.1007/s11192-021-03880-8

Mons, B., & Velterop, J. (2009, October). Nano-publication in the e-science era. In T. Clark, J. S. Luciano, M. S. Marshall, E. Prud'hommeaux, & S. Stephens (Eds.), *Workshop on semantic web applications in scientific discourse (SWASD 2009)*. CEUR-WS.org.

Qian, W., Dongbo, W., Bin, L., & Chao, X. (2021). Deep learning based automatic sentence segmentation and punctuation model for massive classical Chinese literature. *Data Analysis and Knowledge Discovery, 5*(3), 25–34. https://manu44.magtech.com.cn/Jwk_infotech_wk3/article/2021/2096-3467/2096-3467-5-3-25.shtml

Rao, S. K. (1996). *Semantic hermeneutics of ancient books*. Xinjiang Renming Publishing House.

Robin, R., & Sven, B. (2020). What are nanopublications? *Nanopub*. https://nanopub.readthedocs.io/en/latest/getting-started/what-are-nanopubs.html

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics, 16*(6), 1139–1148. https://doi.org/10.1109/TVCG.2010.179

Sherkat, E., Nourashrafeddin, S., Milios, E. E., & Minghim, R. (2018). Interactive document clustering revisited: A visual analytics approach. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces* (pp. 281–292). https://doi.org/10.1145/3172944.3172964

Shu, F., Feng, L., Qiu, J., Shu, S., & Lariviere, B. (2019). Exploring the function of citations in ancient Chinese literature. *Journal of Information Science, 40*(12), 1338–1346. https://doi.org/10.3772/j.issn.1000-0135.2021.12.011

Sturgeon, D. (2018). Large-scale optical character recognition of pre-modern Chinese texts. *International Journal of Buddhist Thought and Culture, 28*(2), 11–44. https://doi.org/10.16893/ijbtc.2018.12.28.2.11

Suárez-Figueroa, M. C., Gómez-Pérez, A., & Fernández-López, M. (2011). The NeOn methodology for ontology engineering. In M. C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, & A. Gangemi (Eds.), *Ontology engineering in a networked world* (pp. 9–34). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24794-1_2

Tambassi, T. (2021). *The philosophy of geo-ontologies*. Springer. https://doi.org/10.1007/978-3-030-78145-3_2

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The Knowledge Engineering Review, 11*(2), 93–136. https://doi.org/10.1017/S0269888900007797

Wang, B., Shi, X., Tan, Z., Chen, Y., & Wang, W. (2016). A sentence segmentation method for ancient Chinese texts based on NNLM. In *Proceedings of Chinese Lexical Semantics: 17th Workshop (CLSW 2016)* (pp. 387–396). Springer International Publishing. https://doi.org/10.1007/978-3-319-49508-8_36

Wang, X., Weng, M., Hou, X., & Lei, J. (2023). The knowledge representation and semantic modeling of ancient books commentaries. *Journal of Library Science in China, 49*(3), 75–91.

Zong, F. (2022, October 8). Collation of ancient books in Chinese history. *Guang Ming Daily*. https://news.gmw.cn/2022-10/08/content_36069440.htm

Index

Note: **Bold** page numbers refer to tables and *italic* page numbers refer to figures.