



UNIVERSITY
OF LONDON
PRESS

A Matter of Trust

Building Integrity into Data, Statistics and Records
to Support the Sustainable Development Goals

edited by Anne Thurston



INSTITUTE OF COMMONWEALTH STUDIES

A Matter of Trust

Building Integrity into Data, Statistics and
Records to Support Sustainable Development



**HUMAN
RIGHTS
CONSORTIUM**

SCHOOL OF
ADVANCED STUDY
UNIVERSITY
OF LONDON

I see a crisis before us. An evidence crisis. I want to convince you that evidence – which is different from data, information or facts – is critical to accountability, identity and memory, and ultimately to democracy. If we are going to survive these perilous times for the world – and they are perilous – we need evidence.¹

1 L. Millar, *A Matter of Facts: The Value of Evidence in an Information Age* (Chicago: American Library Association, 2019), preface.

Published by the University of London Press

Institute of Commonwealth Studies,
School of Advanced Study, University of London, 2020
<https://www.sas.ac.uk/institutes/institute-commonwealth-studies>

Text © contributors, 2020

Images © contributors and copyright holders named in captions, 2020

ISBN 978-1-912250-34-9

University of London Press
School of Advanced Study
University of London
Senate House
Malet Street
London WC1E 7HU

While every attempt has been made to provide correct URLs, the World Bank, WHO and other institutions are constantly updating statistics and replacing online information with new URLs so this has not always been possible.



**HUMAN
RIGHTS
CONSORTIUM**

**SCHOOL OF
ADVANCED STUDY
UNIVERSITY
OF LONDON**

Contents

Acknowledgements	vii
About the authors	ix
Introduction	1
1. Records as evidence for measuring sustainable development in Africa <i>Anne Thurston</i>	7
2. The state of data and statistics in sub-Saharan Africa in the context of the Sustainable Development Goals <i>Paul Komba and Ngianga-Bakwin Kandala</i>	19
3. Data, information and records: exploring definitions and relationships <i>Geoffrey Yeo interviewed by James Lowry</i>	49
4. The potential – constructive and destructive – of information technology for records management: case studies from India <i>James Manor</i>	67
5. Statistical accuracy and reliable records: a case study of mortality statistics in The Gambia <i>Andrew Griffin</i>	83
6. Mainstreaming records and data management in sustainable development: lessons from the public and private sectors in Kenya <i>Justus Wamukoya and Cleophas Ambira</i>	97
7. Open data and records management – activating public engagement to improve information: case studies from Sierra Leone and Cambodia <i>Katherine Townsend, Tamba Lamin, Amadu Massally and Pyou Chung</i>	109
8. Assuring authenticity in public sector data: a case study of the Kenya Open Data Initiative <i>James Lowry</i>	121

9.	Preserving the digital evidence base for measuring the Sustainable Development Goals <i>Adrian Brown</i>	135
10.	Preserving and using digitally encoded information as a foundation for achieving the Sustainable Development Goals <i>David Giarretta</i>	153
11.	Transparency in the 21st century: the role of records in achieving public access to information, protecting fundamental freedoms and monitoring sustainable development <i>Victoria Lemieux</i>	173
12.	Information management for international development: roles, responsibilities and competencies <i>Elizabeth Shepherd and Julie McLeod</i>	189
13.	The quality of data, statistics and records used to measure progress towards achieving the SDGs: a fictional situation analysis <i>John McDonald</i>	211
	Index	243

Acknowledgements

In 2016, the Institute of Commonwealth Studies at the University of London was awarded an Arts and Humanities Research Council grant (AHRC Networking Grant, AH/P006205/1) to enable me to explore, with a team of multidisciplinary experts, issues affecting how development is measured and the consequences. This involved examining the requirements for quality controls across the closely related disciplines of data, statistics and records, not only for reliably measuring the Sustainable Development Goals but for strengthening accountable governance. I greatly appreciated the Institute's generous support throughout this process.

John McDonald's help has also been invaluable. Not only has he been a point of reference for innumerable professional issues, but his patience and kindness have been remarkable. James Lowry, too, has been consistently generous and supportive, and the members of the multidisciplinary team have been outstanding collaborators.

I am pleased to dedicate this volume to the many colleagues around the world with whom I have explored the issues covered here. In particular, I want to express gratitude to Pino Akotia, whose openness, integrity and love of his country have always inspired me.

Anne Thurston, Editor

Notes on contributors

Cleophas Ambira has long been committed to developing professional capacity in archives and records management in Kenya, particularly in the digital environment. He is the training manager at the Kenya Commercial Bank Group and chair of the Kenya Association of Records Managers and Archivists (KARMA), which he founded in 2011. He has carried out records management consultancies for a number of public and private sector institutions in Kenya that have involved restructuring recordkeeping systems, reviewing records management policies and developing appropriate records management training curricula. Dr Ambira has authored works on records management and computer studies and on the relationship between records management and risk management in the banking sector in relation to international good practice. His particular focus has been how the effective management of digital records in Kenya can facilitate e-government and the good practice framework and skill requirements needed to support it, particularly in the light of the rapidly changing technologies that require increasingly robust and efficient systems to support transparency and accountability.

Adrian Brown began his career in the mid-1990s as an archivist at English Heritage, where he wrote and published a digital archiving strategy and built a digital repository, pioneering features now standard in digital repositories internationally. Moving to the UK National Archives in 2002, he developed practical solutions for digital preservation. He led key elements of the Seamless Flow and Digital Continuity Programme, which made it possible to preserve UK government digital records through time and technology change, and he oversaw the National Digital Archive of Datasets. Joining the Parliamentary Archives in 2009, he led the development of a new digital repository and delivered a full digital preservation capability to the UK Parliament. In addition to giving numerous presentations on digital preservation, digitisation and web archiving at conferences, seminars and workshops internationally, Adrian has served on a wide range of bodies, including the Digital Preservation Coalition, the International Records Management Trust and the UK Web Archiving Consortium. An expert adviser to the European Commission and UNESCO, he has received a number of awards for his work, including the Emmett Leahy Award for Outstanding Contributions to the Records and Information Profession in 2016.

Pyrou Chung directs the East West Management Institute's programmes on natural resource, land and data initiatives in southeast Asia under the Open Development Initiative. As an ecologist and sustainable development expert focused predominantly on the non-profit and non-government sector, her work has been focused on the nexus between environmental governance, conservation and human rights, with an emphasis on implementing integrated biodiversity, natural resource management and land rights initiatives with indigenous, forest-dependent or rural communities. More recently this work has intersected with the open data movement to utilise GIS data and information in an open data platform, Open Development Mekong, which promotes evidence-based decision-making, digital gender inclusion and indigenous data sovereignty. The platform is a one of a kind open data portal in southeast Asia focused upon the Lower Mekong Countries – Cambodia, Laos, Myanmar, Thailand and Vietnam.

David Giaretta has worked on digital preservation since 1990 and has led many of the important international developments in this area. He chaired the panel that produced the Open Archival Information System (OAIS) Reference Model (ISO 14721), which has become the de facto standard for protecting and preserving digital archives. He also led the group that produced the ISO standards for audit and certification of trustworthy digital repositories (ISO 16363) and the requirements for bodies conducting audit and certification of digital repositories (ISO 16919). In addition, he led a number of large digital preservation projects, which represented a major investment by the European Union, working with more than 50 partner organisations. These projects built on his experience working in and leading large data digital repositories and software systems. Dr Giaretta has been involved with the Alliance for Permanent Access (APA) from its establishment and became its director in July 2010. His extensive publications include his book *Advanced Digital Preservation* (Springer, 2011). Among many awards, David received the Emmett Leahy Award for Outstanding Contributions to the Records and Information Profession in 2012.

A senior records and information manager with 40 years of UK and international experience, **Andrew Griffin** has had a wide range of international experience, working for the India Office Library and Records, the University of Papua New Guinea, the British Library, the Cayman Islands National Archive and St Bartholomew's Hospital in London. Joining the International Records Management Trust in 1997, he managed and contributed to a range of multi-year projects in The Gambia, Ghana, Sierra Leone, Tanzania, Belize and the East African Community, as well as many short-term projects. In particular, he was actively involved in designing and implementing practical solutions

to records and information management challenges, notably in relation to financial, human resource, hospital and legal records. He also worked for the International Monetary Fund in Washington, DC, where he led a major improvement programme to introduce standards and procedures for managing records. Most recently, he has worked with the National Records Service of the Government of The Gambia supporting the implementation of an electronic records management system.

Prior to becoming a professor of biostatistics at Northumbria University, **Nganga-Bakwin Kandala** worked as head of the Health Economics and Evidence Synthesis Research Unit at the Luxembourg Institute of Health. He was an associate professor in health technology assessment (joint appointment Oxford University and the University of Warwick). He also worked at the University of Southampton and King's College London, as a Mellon Foundation Fellow at the University of Montréal, as associate professor at the University of Botswana and as associate lecturer at the University of Lagos. He is affiliated with the University of Witwatersrand in South Africa as a professor of biostatistics. Dr Kandala has published widely in the field of statistics and health in diverse populations. His books include *Advance Techniques in Modelling Maternal and Child Health in Africa* (Springer Science, 2014) and *Female Genital Mutilation Around the World: Analysis of Medical Aspects, Laws and Practice* (Springer Science, 2018), which he co-wrote with Paul Komba. His current research interests include health economics, health technology assessment, statistical methods applied to epidemiology, survival analysis, longitudinal data analysis, meta-analysis and Bayesian modelling, spatial analysis and disease mapping.

Paul Komba is an academic and a barrister based at Northumbria University. His background is in human rights law, global health rights and international development. He has been director of the African Commercial Law Foundation in London and a legal adviser to the Cambridge Refugee Support Group. His research interests include African commercial courts' handling of records in the digital environment, statistical data for addressing the right to health for victims of conflict-gender violence and the use of ICT in the judiciary. Dr Komba has held a wide range of university teaching and postdoctoral research posts, for instance at the Faculty of Law at the University of Johannesburg, the Kinshasa School of Public Health, the University of Warwick and the University of Cambridge Centre of African Studies. He joined the Department of Engineering and Environment at Northumbria University in 2018. He is co-author with Professor Nganga-Bakwin Kandala of *Female Genital Mutilation Around the World: Medical Aspects, Law and Practice* (Springer, 2018).

Tamba Lamin is a business entrepreneur and technology and data architect with a background in technology management, information management telecommunications, network security and computer networking. He is committed to helping design, build and run digital applications that make people's lives more meaningful and productive. His experience working with global multicultural teams across Asia, Africa, North America, South America and Europe has equipped him to advocate, promote and sponsor civic technology initiatives that focus on reducing corruption and improving accountability in Africa, particularly Sierra Leone. One of the companies that he founded designed and built an open-source election data platform (<http://electiondata.io>) to facilitate free, fair, safe, secure and transparent elections, with functions including data collection and data-sharing of election results. Tamba is now a senior technology architect for Accenture as well as a part-time doctoral student at Pace University in New York, where he is researching how software framework architecture can affect end-user performance of applications built using such a framework.

Victoria Lemieux is an associate professor of archival science at the School of Information at the University of British Columbia (UBC), where she leads a multidisciplinary blockchain research cluster. Her current research is focused on risk to the availability of trustworthy records, in particular in blockchain and distributed ledger recordkeeping systems, and how these risks impact transparency, financial stability, public accountability and human rights. Dr Lemieux's many awards for her professional work and research include the 2015 Emmett Leahy Award for Outstanding Contributions to the Field of Records Management, a 2015 World Bank Big Data Innovation Award and a 2016 Emerald Literati Award for her research on blockchain technology. She is a faculty associate at multiple units within UBC, including the Peter Wall Institute for Advanced Studies, Sauder School of Business and the Institute for Computers, Information and Cognitive Systems. Previously, she was a senior public sector specialist at the World Bank, and she has held positions as a professional archivist, records manager and risk manager within the public and private sectors. In addition, she has consulted for the United Nations, the Commonwealth Secretariat and the World Bank. She has been a Certified Information Systems Security Professional since 2005.

James Lowry joined the University of Liverpool in 2015 and was co-director of the Centre for Archive Studies until taking up the post of assistant professor at Queens College, City University of New York in 2020. Previously, as deputy director of the International Records Management Trust (2009 to 2015), he developed expertise in policy and system design projects for public sector records and archives management programmes across Europe, Africa and the Caribbean. This experience is reflected in his research focus on official

records, data and power, particularly in postcolonial contexts, as well as on transparent government through the intersection of open data and access to records. Focusing in particular on how records and archives management practices can improve the quality of open government data, Dr Lowry led the development of the UK government's commitment on records management in its Open Government Partnership National Action Plan (2013 to 2015). His publications include *Integrity in Government through Records Management*, which he edited with Justus Wamukoya (Ashgate, 2014). James has served as chair of the Association of Commonwealth Archivists and Records Managers, secretary to the International Council on Archives' Africa Programme and trustee of the International Records Management Trust.

During a career of more than 25 years with the National Archives of Canada, **John McDonald** held a number of positions that facilitated the management of records and information across the Canadian federal government. He was the first winner of the Emmett Leahy Award for Outstanding Contributions to the Records and Information Profession to be recognised for his pioneering leadership and accomplishments in the field of digital records management (in 1999), which had an 'extraordinary impact on the profession, not only in Canada but around the world'. In his subsequent consulting career, he developed strategic and operational plans for enhancing information management in public sector organisations in Canada and overseas. He has also designed and delivered graduate-level courses at the School of Library, Archival and Information Studies at the University of British Columbia and the Faculty of Information at the University of Toronto. He also contributed to several projects managed by the International Records Management Trust, including developing training material. He is a past chair of the International Council on Archives' Committee on Electronic Records and Founder and past chair of the Canadian Federal Government's Information Management Forum.

Julie McLeod, Professor of Records Management in the iSchool at Northumbria University in the Faculty of Computing and Information Sciences, joined the university in 1994 after a career managing information for research and development scientists. Her work has been characterised by a commitment to linking good practice principles to practical challenges. Her research interests focus on digital records management and research data management, and she has created high-quality education programmes that integrate records and information management with other disciplines, for instance law and computing. She has edited the highly respected *Records Management Journal* since 1995 and continues to be its consulting editor. Its coverage spans all continents, with articles submitted from dozens of countries, covering research and contemporary practice in areas including records management, information

governance and risk, often from different disciplinary perspectives. Professor McLeod regularly speaks at conferences, publishes extensively and has co-edited several books. She received the Emmett Leahy Award for Outstanding Contributions to the Records and Information Profession in 2014.

James Manor is a leading world expert on Indian politics. He is the Emeka Anyaoku Professor of Commonwealth Studies at the School of Advanced Study, University of London, and the former VKRV Rao Professor at the Institute for Social and Economic Change, Bangalore, India. He was professor of government at Harvard University (1985 to 1987), and was appointed as professorial fellow of political science at the Institute of Development Studies at the University of Sussex in 1987. He was director of the Institute of Commonwealth Studies and part-time professor of Commonwealth politics at the University of London (1993–7). Professor Manor has done consultancies for agencies including the World Bank, UNDP and the OECD, as well as for national governments, including the Swedish, Dutch, Norwegian and Colombian governments. His research has mainly been focused in South Asia but also includes sub-Saharan Africa. He has recently done studies of democratic decentralisation in Asia and Africa, elections and politicians' survival strategies. In addition to his work on state–society relations, his research interests include contemporary history, globalisation and development, human rights, local government, modern history and political institutions.

Amadu Massally is passionate about bringing together Africans in the diaspora. In 2009, he won the National Organization for Sierra Leoneans in North America Diamond Award, which is given to a Sierra Leonean who demonstrates exceptional service to his country and people. He has been involved in many programmes and activities that promote Sierra Leone, including pioneering the Millennium Challenge Corporation's establishment of the Open Government Partnership (OGP) Diaspora Mobilization. He was instrumental in establishing an OGP unit in Sierra Leone, working in the Office of the President. Still actively involved in national programmes in Sierra Leone, he is currently co-coordinating a massive online open education programme, working with TpISENT, a leading ICT professional services company in Sierra Leone, for which he serves as Chief Financial Officer. He attended Morehouse College in Atlanta and the University of Massachusetts, where he earned a degree in Management and Accounting. He is a Certified Public Accountant and currently holds certificates as an Information Systems Auditor and a Professional Scrum Master.

Elizabeth Shepherd joined the Department of Information Studies at University College London in 1992 to update and revise the MA in Archives

and Records Management, becoming programme director in 2002, professor of archives and records management in 2011 and head of department in 2018. Previously, she worked in archives and records management in local government. Her research interests include the role of records management in effective information policy compliance and information governance. Professor Shepherd has published numerous articles and (with Geoffrey Yeo) *Managing Records: A Handbook of Principles and Practice* (Facet Publishing, 2003). Professor Shepherd was instrumental in establishing the International Centre for Archives and Records Management Research and User Studies (ICARUS) to help meet the rapid technological and intellectual changes facing the archives and records management discipline. She presents her research regularly at national and international conferences through invited papers/keynotes and peer reviewed papers. She was a member of the Lord Chancellor's Advisory Council on National Records and Archives (2000 to 2006).

Anne Thurston has worked on defining international solutions for managing public sector records in lower-resourced countries for nearly three decades. As an academic at University College London and as Director of the International Trust, she has had extensive experience of working with many governments, particularly in Africa, to find practical solutions for protecting records as evidence of citizens' rights, accountability and transparency, and national development. Between 1970 and 1980 she conducted research in Kenya and was employed by the National Archives. Joining the staff of the School of Library, Archive and Information Studies at University College London in 1980, she developed an international records studies programme. The findings of her groundbreaking survey of recordkeeping systems across the Commonwealth provided the basis for shaping the programme as well as the inspiration for the work of the Trust. Dr Thurston was a member of the UK Lord Chancellor's Advisory Council on Public Records between 1994 and 2000, was granted an OBE by the Queen in 2000 for services to public administration in Africa, received a lifetime achievement award from the UK Records Management Society in 2006 and was awarded the Emmett Leahy Award for Outstanding Contributions to the Records and Information Profession internationally in 2007. In recent years, she has worked on a range of projects, including several at the World Bank that were aimed at defining the relationship between effective records management and international development.

Katherine Townsend is a passionate advocate of open data and open government and an expert in policy development, public-private partnerships and civic engagement. She supports civic tech movements across the globe to improve in-person and online connections for social entrepreneurship and local growth and transparency. She worked for the innovation and transparency

initiative at the Department of State and the US Agency for International Development (USAID), where she helped launch its Open Data Initiative with the aim of making information transparent, useable and accessible as the norm for international assistance. She also worked with the Code for Africa's Open Government Fellows, SDG Global Data Partnership and Data Collaboratives for Local Impact in East Africa and co-developed the Global Open Data for Agriculture and Nutrition and the Africa Open Data Conference series to demonstrate the benefit of transparent and participatory government. Most recently, she has worked to develop a platform (data.org) for partnerships seeking to increase data science in the social sector, drive data science for social impact and convene and catalyse NGOs to integrate data science into their strategies and decision-making.

For over three decades, **Justus Wamukoya** has been a leading figure in the crusade to change the culture of recordkeeping in Africa, particularly eastern and southern Africa, with the aim of seeing records and other forms of documentary evidence contribute positively towards democratic governance and accountability. As a staff member of the Kenya National Archives, a senior lecturer in the Department of Library and Information Studies at the University of Botswana, and as professor and later dean of the School of Information Sciences at Kenya's Moi University, he has gained wide experience of records management issues. A scholar and a practitioner, he has carried out an extensive range of research and consultancy projects and published widely. Professor Wamukoya has played an active part in the development of the East and Southern African Regional Branch of the International Council on Archives (ICA) and a leading role in the ICA's Africa Programme, which aims to strengthen the profession across the continent. In addition, Professor Wamukoya has been closely involved in numerous research and educational development projects, contributing substantially to records systems restructuring and other governance and records management improvements, as well as to educational development projects.

Geoffrey Yeo is an academic researcher in the field of records studies. He has published widely on archives and records management topics, including the challenges and opportunities of making, keeping and using records and archives in digital environments. With his latest book *Records, Information and Data* (Facet Publishing, 2018), he has taken a cutting-edge approach to exploring ideas about convergence among information-related disciplines and attitudes to information and its governance. He is a frequent speaker at international academic and professional conferences and is also co-author (with Elizabeth Shepherd) of *Managing Records: A Handbook of Principles and Practice* (Facet Publishing, 2003). Yeo worked as an archivist in London for the Corporation

of the City of London, St Bartholomew's Hospital and the Royal College of Physicians before joining the staff of University College London's School of Library, Archive and Information Studies (now the Department of Information Studies) in 1995, where he taught until he became an honorary researcher in 2014. He has also worked with the International Records Management Trust on projects to restructure hospital records in The Gambia and Ghana and to design training materials in Botswana.

Introduction

Background

The Sustainable Development Goals (SDGs) initiative has the potential to set the direction for a future world that works for everyone. The SDGs were approved by 193 United Nations member countries in September 2016 to help guide global and national development policies in the period to 2030. The 17 goals build on the successes of the Millennium Development Goals, while also including new priority areas such as climate change, economic inequality, innovation, sustainable consumption, peace and justice. Each of the 17 SDGs is to be assessed against agreed targets and indicators.¹ One hundred and sixty nine targets set out quantitative and qualitative objectives, with 232 indicators. Individual countries, supported by international organisations, notably the UN Statistics Division, are responsible for collecting and processing the data and generating the statistics required to measure the indicators.

Each goal presents a considerable challenge in terms of collecting and analysing relevant data and producing the statistics needed to measure progress. Measuring the indicators is intended to guide policy development, strategy design and, in general terms, the future direction of individual countries. Taken across countries, the measurements are widely expected to foster greater intergovernmental cooperation and the development of regional and even global strategies. However, as Morton Jerven² has pointed out, most governments in lower resourced countries (his research focused on Africa) have yet to introduce the control systems needed to produce high-quality, reliable data and statistics; those responsible for data collection and the production of statistics tend to be too few in number and to lack the expertise needed to introduce the necessary policies, standards, procedures and accountability structures. Jerven questions

1 The global indicator framework developed by the Inter Agency and Expert Group on SDG Indicators (IAEG-SDGs) was agreed by the UN Statistical Commission in June 2017. It is supported by the SDG database dissemination platform, maintained by the UN Statistics Division, which provides a metadata repository containing the latest information available about the indicators.

2 M. Jerven, *Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It* (Ithaca, NY: Cornell University Press, 2013).

how anyone can rely on the data and statistics generated under these conditions to make decisions and set direction.

The implications are significant, not only for measuring the Sustainable Development Goals but for the broader ability to plan and achieve development. If progress cannot be measured accurately because of inadequate, inaccurate or flawed statistics, the results can be misguided decisions and doubts about the achievement of the goals. Failure to ‘get the statistics right’ can result in wrong decisions being made, wrong strategies being adopted, and wrong laws, policies and standards being established. It can also lead to a needless waste of resources.

Getting the statistics ‘right’ depends upon the quality and integrity of the data used to produce the statistics. These, in turn, depend upon the quality of the processes that support the collection, manipulation and analysis of the data and the production of the statistics. Ultimately, the quality of these data management and statistical processes depends on the availability, completeness and integrity of the records that document them. Without a documentary record to provide evidence of how the data were gathered and analysed or how statistics were produced and disseminated, it is not possible to confirm that the statistics used to measure the SDG indicators are complete, accurate, relevant and meaningful.

Moreover, records are important sources of information in their own right. They contain information about how, when and where the processes supporting the measurement of the SDG indicators were undertaken as well as information about the data and statistics themselves. This information, when well-managed, can be manipulated with other information contained in other records to support a wide range of purposes. For instance, it can be used to identify and act upon opportunities for merging data from related sources, to analyse trends in the quality of the processes, data and statistics, and to produce management statistics that support the administration of the processes that generate the data and statistics.

The significance of the quality and integrity of data and statistics for measuring the SDG indicators reliably has received considerable attention from a variety of global organisations, including the Sustainable Development Solutions Network and the Global Partnership for Sustainable Development Data.³ However, as yet, relatively little attention has been given to the role of

3 With the approval of the global indicator framework, the IAEG-SDGs has formed three working groups to address specific areas relevant to SDG indicator implementation: Statistical Data and Metadata Exchange, Geo-Spatial Information and Interlinkages. In November 2017, the IAEG produced a consultation draft outlining guidelines and some best practices on data flows and global data reporting SDGs. The guidelines highlight many of the challenges that UN member countries face in producing the high-quality statistics required to measure the SDG indicators.

records in providing evidence to demonstrate that the data and statistics are trustworthy and can be used reliably. Processing data to produce statistics is one thing but processing authentic and reliable data using auditable processes in line with international standards, such that the statistics can be trusted, is quite another.

In order to explore this issue and its implications, a UK Arts and Humanities Research Council project was set up towards the end of 2016 at the University of London's Institute of Commonwealth Studies. Under the banner 'Digital records as evidence to underpin Global Development Goals', two workshops were delivered, one in 2017 and one in 2018, to explore the relationship between data, statistics and records as primary types of information for measuring the goals and to initiate an interdisciplinary dialogue among humanities scholars, development experts and information professionals, including data experts, statisticians and records management professionals.

The members of this team, each a specialist in one of these areas, saw the need to reach beyond the worlds of data and statistics to address the role of records in enabling countries to prove the integrity of not only the data and statistics but also of the processes used to collect and analyse them. Approaches to managing data, statistics and records are different, but viewing them as parts of a whole helps to ensure the quality and integrity of each and to identify errors and weaknesses. The workshop participants recognised that quality, completeness and integrity are difficult if not impossible to achieve without effective policies, procedures, standards and systems and without records management expertise. They considered, for instance, the challenges of achieving reliable data, statistics and records when it is not clear where the information has come from, why it was compiled and how it is to be protected for future use. They recognised that data, statistics and records are being lost regularly on a large scale, particularly in digital formats and particularly in lower resourced countries, where structures often are not in place to protect and preserve them.

They decided to explore these issues further from their own perspectives and to produce chapters that, together, would present an interdisciplinary perspective. The chapters explore a range of interrelated development issues which have not previously been articulated, but which affect the quality, veracity and trustworthiness of the data, statistics and records that are fundamental to measuring and achieving the SDGs. They focus particularly on Africa, which illustrates the substantial challenges for managing information. However, the issues identified are generic and will resonate with any country that is grappling with the challenges of managing the quality and integrity of the data, statistics and records they generate and use to measure the SDG and indicators.

The first three chapters explore the historical context for the challenges of managing data, statistics and records and the relationships between them. Anne Thurston provides an overview of background of the recordkeeping challenges and realities that African countries tend to face as they measure and implement the SDGs. Paul Komba and Ngianga-Bakwin Kandala offer a similar perspective from the world of statistics, tracing the developments and challenges for measuring development in Africa statistically. Geoffrey Yeo, in an interview with James Lowry, looks at the different meanings that have been attached to the terms data, statistics and records and the different ways in which their relationships have been interpreted and understood.

The second set of four case studies offers a glimpse of the realities 'on the ground' based upon country experiences. James Manor uses the Mahatma Gandhi National Rural Employment Guarantee Act and the Aadhaar initiative in India to explore how well-managed digital records can contribute to constructive development programmes but how, when unmanaged, they can undermine programme objectives, waste resources and lead to misguided decisions and actions. Andrew Griffin examines the relationship between data, statistics and records in the context of mortality statistics in The Gambia to illustrate the complexities of acquiring reliable information to measure the achievement of the SDGs, especially in a low resource environment. Justus Wamukoya and Cleophas Ambira examine the status of records in Kenya and draw on examples from mobile banking in Kenya to highlight the significant new risks that society faces in conducting financial transactions online through the use of smart phones. They also suggest that the sensitive nature of the transactions is focusing attention on the integrity and trustworthiness of the data, statistics and records that these transactions generate. Katherine Townsend, Tamba Lamin, Amadu Massally and Pyrou Chung present case studies from Sierra Leone and Cambodia that highlight the power of open data to promote democratic principles, increase transparency and empower citizens to contribute to policy making and corruption control. They also explore how records management could strengthen the quality, integrity and longevity of the data.

A third group of chapters focuses on the technical challenges of managing and preserving the data, statistics and records that support SDG initiatives. Information recorded in digital form is especially susceptible to loss and corruption because of poor storage conditions, dependence on changing technology and the lack of metadata to facilitate retrieval of the records. These chapters demonstrate that maintaining the integrity and accessibility of records requires careful attention to the formats in which they are stored, the standards for their classification and description, the conditions under which they are protected from alteration and unauthorised access, and the procedures for maintaining their integrity and accessibility through time in spite of changes in technology. James Lowry argues that the principles and techniques developed

over centuries in the field of recordkeeping for the purpose of assuring authenticity can also be used to improve data quality, so that the information needed to implement and monitor the SDGs is not only available but authentic. Adrian Brown considers the practical implications of developing the digital preservation capabilities needed to ensure that data collected to measure the SDGs can be compared through time and that decision-makers can be held accountable for how it was gathered. David Giaretta then explores the complexities of collecting, using and preserving digitally encoded information, in particular scientific data, so that conclusions and actions arising from them are based on authentic and accurate information. He highlights the technical challenges of managing data in relation to the SDGs, the importance of international standards and the key issues that need to be resolved if the goals are to be achieved.

The final chapters identify strategies for managing the digital information needed to measure the SDGs. Victoria Lemieux presents the findings of a World Bank research programme on transparency and information management. She describes a tool developed for use in high-level assessments of systems of record to predict whether the records created and held in these systems will be available and trustworthy through time to support development goals. Elizabeth Shepherd and Julie McLeod use a maturity model to identify the competencies required to ensure that strategies are comprehensive, relevant and effective. They relate each level of competency to international standards and address the roles, responsibilities and competencies needed to manage information for development, particularly for measuring the SDGs reliably. John McDonald concludes the study by using a fictional scenario to illustrate both the issues that lower resourced countries face and the comprehensive strategies that they can introduce to enhance their capacity to manage the data, statistics and records needed to support the SDG initiative.

Taken together, these chapters open a window to an evidence-based approach to development and to the practical actions needed to address the information management issues the SDGs raise. To illustrate, even as SDG 3 seeks to ensure healthy lives and promote well-being for people of all ages, a worldwide global health crisis, the Covid-19 pandemic, is spreading human suffering, destabilising the global economy and upending the lives of billions of people. Bringing the virus under control requires a global solution supported by high quality data, statistics and records based on internationally accepted standards and protocols. The absence of such standards and protocols is undermining efforts to address the pandemic. More broadly, and by its example, Covid-19 is highlighting and bringing into stark focus the serious challenges countries are facing as they struggle

to ensure the quality of the data, statistics and records required to support achievement of the SDGs by 2030.

Much can be learned from the data, statistics and records issues associated with Covid-19 that could help to reinforce the credibility, relevance and effectiveness of the data, statistics and records used to support the SDGs. Covid-19 has aggravated an already serious development crisis, and urgent action to accelerate progress toward addressing infections, hospitalisations and deaths is required worldwide. How can global leaders make difficult decisions about bringing the virus under control and dealing with its after-effects without reliable, verifiable and complete information that they can trust? Similarly, how can the results of the SDGs be trusted if the information used to support their achievement can't be trusted?

Hopefully, this book will contribute a new perspective to the SDG initiative by highlighting the value of creating, managing and using high-quality data, statistics and records to achieve meaningful and realistic global and national development policies, now and in the critical period to 2030 and beyond.

1. Records as evidence for measuring sustainable development in Africa

Anne Thurston

The expectation that the United Nations Sustainable Development Goals (SDGs) will provide a basis for addressing global economic, social and environmental crises assumes the availability of trustworthy, accessible evidence of measurable progress. The lack of policies and systems for managing this information reliably inevitably means that there are gaps in the information and that often its reliability cannot be demonstrated, in Africa or elsewhere. It is important to understand how this situation came about and the impact it has on development.

The SDG approach to measuring progress does not yet take account of the challenges for managing records. Addressing the challenges would make a substantial difference to governments' ability to measure the goals accurately and protect and preserve development information for future use. Records should document processes, decisions, actions, activities and communications, protect rights and entitlements, inform policy and hold officials accountable for their actions. Any set of information, regardless of its structure or form, can be managed as a record. This includes information in the form of a document, a collection of data or other types of digital or analogue information that are created, captured and managed in the course of business.¹

Poorly managed records can easily be lost, altered, fragmented, corrupted or destroyed. With each of these losses, transparency and accountability are diminished and the ability to measure compliance, extract meaningful data and use the information as a reliable measure of development is compromised. This chapter explores the loss of control of official government records in Africa in the decades following independence as background for understanding the consequences for the ability to trust and use it.²

1 ISO 15479-1:2016 Information and Documentation – Records Management – Part 1: Concepts and Principles.

2 The chapter draws on an earlier article: A. Thurston, 'Records management in Africa: old problems, dynamic new solutions', *Records Management Journal*, 6 (1996): 187–99. It also reflects extensive field experience in Africa over a period of 30 years as director of the International Records Management Trust.

Today, thanks largely to the determination and leadership of teaching staff in the field of records and archives in universities across Africa, the records profession in Africa is emerging from what have often seemed like insurmountable challenges. The role of records still tends to be largely underrecognised and misunderstood, but there is now greater potential for records to make a more meaningful contribution to society than has been possible in the past.

Breakdown of records systems in Africa

At the time of independence, many of the outgoing colonial governments in Africa set up national archives to preserve the historical record of what they left behind. In some cases, these administrations, recognising the power of the information that the records documented, destroyed or hid the more sensitive records rather than handing them over to the new governments.³ Overall, however, the colonial governments left behind basic systems, policies and procedures for managing records. Initially, some of the independent African governments invested resources in their national archives as the agencies responsible for implementing national records policies. However, with immediate development needs and political realities to address, the national archives had to compete with other government agencies for funds, and ironically, as the volume of government records grew, the archives' budgets declined. National archives in Africa lacked the adequately trained staff and resources needed to develop the legal frameworks and professional systems to support this growth.

The International Council on Archives (ICA) was established in Paris in 1948 to promote the use of records, preserve their integrity, advance the documentation of human experience and make the information in the records available to promote international cooperation. However, while UNESCO did, in principle, support Africa's new national archives, the wider international community did not see this as a priority and did not invest in addressing the growing issues affecting records in African countries. The small ICA office in Paris lacked the resources to support significant development, and the African national archives remained focused primarily on historical records. The common legal closure period of 30 to 50 years from when records ceased to be in active use often made the archives irrelevant as immediately useful development information for independent governments.

Through the 1970s and 1980s, there was very little investment in the national archives in Africa, and many deteriorated and stagnated as institutions. For

3 Notably, just before independence, records documenting political activities in Kenya were airlifted to England, where they remained hidden for decades until they surfaced during a British High Court case in which the British government was charged with brutality during the Mau Mau emergency.

instance, equipment needed to protect the records often ceased to function and was too costly to replace. Staff establishments did not grow, and only a small number of professionals were trained abroad, often in skills that did not fit the needs of their institutions. On their return from training, their remuneration and status were low, despite their enhanced skills and the immense challenges they faced. The national archives often found it impossible to accept transfers of closed records from government departments, not only because they lacked the staff to do so but because the repositories constructed at the time of independence now tended to be full to overflowing.

Registry systems for managing active records in ministries and departments also broke down, and often the only procedures available to guide staff had been developed in the colonial period. In many ministries, files were simply opened in a running sequence, with no classification or indexing system and no means of locating or tracking the movement of individual files. It was not unusual for cleaners to be tasked with caring for records. The value of the records as information sources was gradually compromised; file titles did not match file contents, information relating to the same issue was scattered through files with similar titles and policy papers were mixed with papers of ephemeral value. The result was that government policy often was developed and implemented on an ad hoc basis, and officials suffered daily embarrassments as they were unable to access the information they needed to make decisions or to take action.

The breakdown of recordkeeping systems had a direct and growing impact on the ability to govern and on citizens' lives. When police or court records could not be found, citizens' rights were denied, judicial processes were manipulated and citizens could be incarcerated without due process of law. Case precedent broke down in the courts, and prisoners detained on remand had little hope of a court hearing if their case records could not be found. When a patient's medical history could not be located or did not exist, tests had to be repeated unnecessarily or patients were given inappropriate, even risky, treatment. When a civil servant's personnel file was missing, it was often impossible to claim pension rights, and it was not unusual for civil servants with low qualifications to manipulate the payroll to be paid higher salaries than they were qualified to receive. When land records could not be traced, it was not possible to establish ownership. Legitimate landowners were not able to borrow against title deeds. Financial transactions were difficult to track, and theft of financial assets and corruption became increasingly common. The impact on citizens was increasingly severe, but often the cause went unrecognised.

Records professionals in Africa worked against unequal odds to maintain the integrity of the profession and its contribution to national stability. With inadequate recognition, inadequate resources and inadequate training to address the problems they faced, it was very difficult to see how to reverse the situation.

Records management, structural adjustment, public sector reform and computerisation

The loss of control of public sector records in Africa coincided with several significant trends in international development, which included donor and lender pressure on governments to reduce budget deficits, efforts to reform the structure of the public service and recognition that computerisation was fundamental to controlling government resources and improving efficiency. From the early 1980s, the World Bank, the International Monetary Fund (IMF) and other donors and lenders began promoting structural adjustment and public sector reform programmes. The emphasis was on good governance (efficiency, accountability and transparency) and on achieving savings by reducing overall civil service size, often as part of conditionality for loans from Bretton Woods organisations.⁴ Computerisation was viewed as fundamental to achieving these objectives.

When the international community began to focus on developing laws, procedures, organisational structures and management approaches to support these new emphases, it became increasingly apparent that the information needed to underpin accountable, transparent and efficient government was not available. The failure to modernise the records systems needed to support the growth of government made it ever harder to find and use essential information. As Zambia's deputy minister for home affairs observed:

Most countries in this part of Africa are undertaking structural adjustment and public service reform programmes aimed at good governance. This is being done by introducing changes in the management of public affairs and the protection of human rights. These objectives cannot be achieved in the absence of reliable and accurate information, which has become a vital resource for governments in the management of public affairs.⁵

The solution to the breakdown of government records systems seemed, for many, to lie in computerisation, which would enable countries to leapfrog

4 In 1944, a new international monetary system was agreed by delegates from 43 nations in Bretton Woods, New Hampshire; the IMF and what became the World Bank Group were established. The IMF and the World Bank developed structural adjustment policies due to a series of global economic disasters during the late 1970s, such as the oil crisis, the debt crisis and multiple economic depressions, in the belief that deeper intervention was necessary to improve a country's overall wellbeing. During the 1980s, the IMF and the World Bank created loan packages for the majority of countries in sub-Saharan Africa as they experienced economic crises. Although reducing the budgetary deficit was a key policy measure, ultimately, economists could point to few, if any, examples of substantial economic growth in lower-income countries under structural adjustment programmes. See R. Lensink, *Structural Adjustment in Sub-Saharan Africa* (London: Longman, 1996).

5 Meeting of the East and Southern Africa Regional Branch of the International Council on Archives, July 1996.

developments that had taken place elsewhere and provide a modern efficient information base for development. By the mid-1990s, computerisation was a feature of virtually every major donor assistance programme across the public sector in sub-Saharan Africa, particularly in relation to the control of key resources, notably finance and personnel. Paper-based records, which were widely viewed as being disordered, incomplete and difficult to share, were, in many cases, dismissed in favour of digital records and information systems. These newer information systems were viewed as the responsibility of information technology staff, who rarely had training in, or knowledge of, records standards and requirements. The national archives, which lacked information technology experience and tended to be very poorly funded, were largely dismissed as irrelevant to the modern state.

In many sub-Saharan African countries, the national archives, which could have made invaluable contributions to information management reform, were isolated from mainstream policy-making and administration. In the circumstances, they tended to focus on protecting historical records. As computer analysts and system designers were granted increasing respect and status, donors and officials remained unaware of the potential contribution of records professionals and the significance of the international records standards and practices they were developing.

Peter Mazikana, a records specialist from Zimbabwe, was one of the first African records professionals to articulate the significance of records and to advocate involvement of the national archives. At the pan-African Conference on Archival Policies and Programmes in Africa, held in Abuja, Nigeria, in 1994, he noted:

It is not possible to account for expenditure unless there are records to show what revenue was received and how it was expended. It would be impossible to determine an organisation's viability and profitability unless there were a way of monitoring the inputs going into the production process and the quality of products resulting. Materials management is a non-starter unless there are records of items received into stock and those issued. There would be no human resource management unless there were a record of who is employed, to do what and what remuneration. In these respects, therefore, it becomes clear that records management is an underpinning function in all business activities. But the question still remains as to why it is hardly considered as important and relevant and is taken so much for granted?⁶

Some administrators did recognise the significance of the records issue. For instance, in Ghana, where the head of the civil service encouraged the production of a documentary video film on the relationship between records

6 P. Mazikana, 'The role of records management in business during market reform programmes', *Janus*, 1 (1996): 43.

management and citizens' rights, the secretary (minister) for education spoke on camera of his concerns:

The impact on citizens is even greater than you would believe. Up to now, I don't think we fully realise that record keeping is important. Whenever there's a problem, instead of finding out what has happened before, what we did in the past and why it didn't succeed or what successes we had, we simply start afresh, we go on inventing the wheel all the time.⁷

As the move towards computerisation accelerated, it was clear that there were more challenges than originally anticipated. In many African countries, power supplies were erratic, hardware and storage media were difficult to obtain, technicians and repair services were not always available. Many countries found themselves locked into foreign-supported systems but were unable to finance maintenance and upgrades. Moreover, the evidentiary value of the digital records created through computerisation depended on the ability to maintain and protect their authenticity, but resources were not available to develop digital records management systems for preserving and protecting the reliability and integrity of the records through time.

When digital records did become more common across Africa, few people in the records profession, in government or in international agencies realised how quickly digital records would become the predominant medium of government communications or how easily they could be lost or distorted. As governments and citizens rapidly came to rely on digital records (created on desktop computers, in databases, in email, on mobile devices, via websites and on social media platforms), there was little understanding of the skills and structures needed to manage them or even of which government agency should be responsible. Whereas previously, government records had been kept in registries/records units and in national archives, now they were often fragmented across multiple systems. Sometimes responsibility for digital records was split between several government agencies, for instance the one responsible for ICT development, the one responsible for access to information and the one responsible for culture; often it was unclear which should lead on policy. Ministries and departments often pursued their own computerisation projects without government-wide coordination.

These issues were by no means limited to Africa. The Canadian information commissioner, for instance, noted in a speech on information management in the public sector in July 2004 that:

Earlier audit reports have dealt with other examples of poor recordkeeping: the files related to Goods and Services taxation fraud,

7 K.B. Asante, Secretary of State for Education, Government of Ghana, *Protecting the People: Records Management and Citizens' Rights in Ghana*, International Records Management Trust film, produced in 1996 and distributed for educational purposes.

improper tendering of government contracts, the inability to locate costly commissioned reports, the lack of security for sensitive information and other examples. The Auditor General has said that some programs are so poorly documented that an audit cannot even be completed. The records are simply not there, are incomplete or are unreliable. Neither a paper trail nor its digital equivalent is in place ... The implications of poor recordkeeping are a serious matter.⁸

While the challenges were generally the same throughout the world, they were particularly difficult to address in lower resource environments, where awareness, professional capacity and financial resources were more limited.

Consequences for Africa of losing control of records

The consequences of this situation had a growing impact. In 2004, the World Bank manager for the E-Government Applications Group Informatics Program noted, 'Without effective and efficient records management in place, the desired impact of financial and governance reforms is often minimal at best'.⁹ The same year, the head of Sierra Leone's personnel management office noted:

Over the years, important records have deteriorated considerably, been tampered with or even disappeared. The lack of accurate and accessible information hinders efficient personnel administration as well as long-term staff development for capacity building. It also hampers effective planning and implementation of development programmes and leads to mismanagement of finances and the inability of government to maintain accountability ... Reform in this area will lay the basis for other public sector reform programmes, the introduction of computerisation and the restructuring of manual information systems.¹⁰

In 2012, an article by staff from the University of Botswana noted, 'The chaotic state of records and collapsing recordkeeping systems in most African countries makes it impossible to determine responsibility for official actions and to hold individuals accountable for their actions'.¹¹ Corruption investigators,

8 *Information Management in the Government of Canada*, Notes for an address by The Hon. John Reid, P.C., Information Commissioner of Canada, for the Information Management and Government Conference, July 28, 2004 (Office of the Information Commissioner, Ottawa, July 30, 2004).

9 'IFMS implementation: aspects for consideration', Deepak Bhatia, PowerPoint Presentation at the World Bank, September 2004.

10 Interview with Osho Coker, head of the Personnel Management Office, International Records Management Trust/World Bank Consultations on Evidence-Based Governance in the Electronic Age, March 2003.

11 Dithapelo Lefoko Keorapetse, Political and Administrative Studies, and Segomotso Masegonyana Keakopa, Library and Information Studies, University of Botswana, 'Records management as a means to fight corruption and enhancing accountability in Botswana', *Journal of the Eastern and Southern Africa Regional Branch of the International Council on Archives*, 31 (2012): 24–35.

prosecutors, regulators, auditors and lawyers all recognised the importance of being able to access reliable evidence of economic crimes, whether they involved tax fraud, payroll fraud, illicit financial flows, money-laundering, bribes, stolen assets or unauthorised allocations.

In 2013, the high costs of the records management gap in Africa were dramatically illustrated by two significant press reports. The *New York Times* noted that Sierra Leone's 29 top health officials had been indicted by the government anti-corruption agency on charges of misappropriating funds from a global vaccine provider:

At hospitals in the interior, and at the central medical store in Freetown, they have found no records to support the dispensing of drugs worth thousands of dollars; and they could not find records for 23 of the Health Ministry's 55 bank accounts. Record keeping has been abysmal, an anticorruption investigator wrote in a report.¹²

In Malawi, the BBC reported on 'Cashgate', the biggest financial scandal in the country's history:

At the centre of the scandal is a computer-based financial information storage system. Some government officials have allegedly been exploiting a loophole in the system to divert millions from government coffers. It is estimated that up to \$250m (£150m) may have been lost through allegedly fraudulent payments to businessmen for services that were not rendered. According to a report in the local media, an audit by managers of the financial system has established that records of some transactions carried out between July and September 2013 were deleted.¹³

Regular warnings of the consequences of poor management of official records continued to appear in audit reports, anti-corruption investigations, expenditure tracking surveys, research reports and press reports. Still, donors and development planners generally felt that computerised systems offered the best basis for planning, monitoring and measuring national and international development goals and tended to believe that computer-generated information was different from records, even when it was the primary evidence of actions and transactions. The same debate went on in many parts of the world.¹⁴ The lack of awareness by key stakeholders, including senior managers, programme

12 'Sierra Leone's health care system becomes a cautionary tale for donors', *New York Times*, 13 April 2013, <http://www.nytimes.com/2013/04/14/world/africa/sierra-leone-graft-charges-imperil-care-and-aid.html?hpw>.

13 "'Cashgate': Malawi's murky tale of shooting and corruption', BBC, 27 January 2014, <http://www.bbc.co.uk/news/world-africa-25912652>.

14 For instance, John McDonald of the Automated Information Systems Division, Government Records Branch, National Archives of Canada, noted in a presentation to the American Society of Archivists in 1988 that government information system managers in Ottawa 'assumed that the records manager only looked after paper records. And anyway, the electronic information in computers wasn't a record – so it didn't count'. John McDonald,

planners, IT staff, development planners and sometimes even records professionals, of the need to protect the integrity, authenticity and long-term accessibility of digital records, put governments at significant risk.

Digital records could be altered, fragmented, corrupted or deleted, either through malicious interference or through inadequate management. Computerised payrolls, for example, contained increasing amounts of inaccurate and incomplete data as payroll changes often were made on the basis of inaccurate or incomplete authorising evidence (paper or electronic), for instance, letters of appointment, promotion or transfer. When metadata, which should have described the context, content, structure and management of the records was not captured, was imprecise or became separated from the records as technology changed, the audit trail of changes to the payroll could not be verified. Far from solving the problem of ‘ghost workers’, computerised systems often added to them. Pino Akotia, at the University of Ghana, noted in 2013 that:

Payroll fraud has resulted from the prevailing weak records system and practices and the technical vulnerability of the Integrated Pay and Personnel system. Available information on payroll fraud illustrates the implications for the national economy ... Indeed there is no single location in the public service where data on all employees paid under the consolidated fund is complete and available. Sections of public servants have no personnel files. One of the effects is that personnel have wrong job titles: a ‘driver’ with designation as ‘cook’ and a cleaner as ‘Certificate “A” teacher’ distorting the actual number of teachers on the payroll, with funds wrongly expended. Personnel who have left government service continue to be paid.¹⁵

Governments and international organisations saw digitisation as a quick way to make records accessible and end dependence on paper records. By the 1980s, digitisation initiatives were widespread across Africa. Many development planners did not understand that management frameworks were needed to protect the digitised records and their integrity through time, just as they were for born-digital records. Many digitisation projects failed to incorporate requirements for legal admissibility, reliability and usability, such as metadata capture, image resolution, standardised indexes, classification structures, and retention and disposition schedules. As a result, digitised records were often difficult to retrieve, use and rely upon as legal evidence, leaving the creating agency and civil society at risk. For example, where the scan was poor or where the digital copies deteriorated through time, the legal value of the record was

¹⁵ ‘Records management and data management: closing the gap’, *Records Management Journal*, 20 (2010): 53–60.

15 P. Akotia, *Audit and Accountability in the Government of Ghana, Integrity in Government through Records Management* (Farnham: Ashgate, 2014), p. 132.

questionable. In some cases, when digitisation projects were introduced, the records were found to be in such disarray that records management teams had to be brought in to organise them before they could be digitised.

The rapid obsolescence of software and computer systems added to the risks. There is no doubt that the new technologies did help to streamline many processes. For instance, new tools and detection methods supported compliance monitoring and ongoing analysis of corruption risk through statistical, text mining and visual analysis. This did not, however, change the requirement to capture and preserve accurate evidence; just as in a paper environment, well-structured systems and training were essential to protecting the quality of the evidence. The World Bank's 2016 World Development Report noted correctly that it is 'fair to say that long-term preservation of digital records and information in most countries in the world is at serious risk'.¹⁶

Throughout this period, however, records professionals across the world continued to work steadily towards building international records and metadata standards, requirements and management tools for digital as well as paper records. Although this work was little known within the global development community, it was widely shared through the international records community, for instance through the International Council on Archives, the InterPARES Project¹⁷ and the International Records Management Trust, and gradually it had an impact on teaching programmes across Africa and elsewhere.

Open data and records management

From the 1990s, when public sector transparency, accountability and openness emerged as predominant international development themes, opening data to civil society was seen by many as a powerful way forward in facilitating sustainable development, making it possible to move beyond official secrets acts and lengthy closure rules.¹⁸ Open data can be used and reused immediately and freely, so long as it is attributed, does not refer to identifiable individuals and does not violate security restrictions. The benefits to using open data include improving economic performance, supporting human rights and making it

16 World Development Report 2016: 'Digital dividends: One step forward, two steps backward: Does e-government make governments in developing countries more transparent and accountable?' Victoria Lemieux, World Bank Development Report Background Papers, Open Knowledge Repository, World Bank, 2016.

17 The International Research on Permanent Authentic Records in Electronic Systems (InterPARES) aims at developing the knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form and providing the basis for standards, policies, strategies and plans of action capable of ensuring the longevity of such material and the ability of its users to trust its authenticity. See <http://www.interpares.org>.

18 The normal legal closure period had been reduced to 20 from 30 years, which still did not serve development needs.

possible for citizens to participate more fully in decision-making. Citizens can, for instance, use government data to track public expenditure against budgets, reuse it to support business development or track incidents of abuse.

However, inaccurate or incomplete data or otherwise flawed data can skew development findings, undermine confidence in government or endanger citizens' rights. Incorporating records standards in open data schemes would make a significant contribution to strengthening data quality, accessibility and usability.

Conclusion

While data and statistics provide the essential basis for measuring the SDGs, records' contribution in terms of documenting processes, protecting integrity and enabling preservation is also essential. High-quality records provide evidence of how data was created, when and why. They can verify where the data came from, how it was compiled, how it was used and how it was mapped together with other datasets to arrive at composite statistical findings. Records standards make it possible to extract, disaggregate, protect and preserve data, statistics and records documenting SDG measurements to 2030 and beyond. This audit trail is an essential aspect of the ability to trust and use data and statistics and to use them effectively.

As Africa grows increasingly dependent on digital information, it will be essential to ensure that the information created remains authentic, trustworthy and legally reliable for as long as it is needed. Building international standards into system design will go a long way towards ensuring that systems are capable of capturing and preserving quality records and data through time.¹⁹ Thirty years ago, records professionals in Africa faced apparently insurmountable challenges that they have worked hard to overcome. Today's challenges are of a different nature. The continually evolving dynamics of using and managing digital information are immensely complex and cannot be solved by the records profession alone.

Today, the words data and records are often used interchangeably. What were traditionally called records now are often referred to as data. Hospital patients' records are often referred to as disease data; records of births and deaths created through an official registration process are referred to as birth and death data; and records created to document the day-to-day activities of the state are referred to as administrative data. There are even references to

19 See, e.g., ISO 14721 the Open Archival Information System (OAIS) Reference Model (2012); the European Framework for Audit and Certification of Digital Repositories (2014); A. Brown, *Practical Digital Preservation: A How-To Guide for Organizations of Any Size* (London: Facet Publishing, 2013). J. Lowry's chapter in this volume ([Chapter 8](#)) explores this issue in great depth.

open data records. Records now can be created in databases, outside the control of recordkeeping systems, and often no one is certain who is responsible for protecting and preserving them. What is their provenance? What is their context as part of an audit trail? Who decides? Who is responsible for ensuring that this is done? How long does the information need to survive? How is it to be stored? What happens when the technology with which it was created is upgraded or changed and what is the cost?

These challenges offer an opportunity to clarify the unique role that records play in defining sustainable development and in tackling global development issues that affect all people, everywhere. They point to the need to re-examine the relationship between data, statistics and records; to explore complementary standards, policies, practices, systems, structures, capabilities, technologies and tools for managing digital information. They underscore the need to articulate the unique and significant role that the records profession plays in making it possible to capture, document and protect evidence for accountability.

The data, statistics and records communities offer different but complementary approaches to creating and using information, making it vital that information professionals cooperate across the boundaries of their professions. New organisational alliances, for instance between national statistical offices and national archives, will have enormous benefits for measuring and achieving the Sustainable Development Goals. In the future, when most, if not virtually all, information is created and maintained in digital form, the quality, reliability, accessibility and longevity of data, statistics and records will be fundamentally important for meeting the challenges of sustainable development. Harmonising their contributions, without losing sight of their unique roles, will offer far greater opportunities for success than addressing them separately.

2. The state of data and statistics in sub-Saharan Africa in the context of the Sustainable Development Goals

*Paul Komba and Ngianga-Bakwin Kandala**

The Sustainable Development Goals (SDGs) were set up to support sustainable health, tackle poverty and enhance peace and prosperity for present and future generations, at all levels, locally to globally.¹ The implication is that the SDGs can be achieved through governmental and non-governmental interventions, supported by data and statistics to ensure that they are on course to deliver those goals and targets. The rise of an evidence-based policy paradigm and the idea of managing by results has led aid agencies and international policy-makers to place statistical measurement at the heart of monitoring and evaluation of official development assistance.²

Statistical development in Africa has attracted the interest of international policy-makers as well as regional and national bodies across the continent. There is a growing sense that statistics should be the backbone of sound policy decisions.³ There are now increasingly persistent calls for African policies to be driven by evidence-led research, turning away from gut feelings or ideological-driven agendas as nations embark upon the process of achieving the SDGs. Poor statistics hurt African governments' ability to make good policy decisions; reversing this requires the collection of sound data and its effective use in addressing the issues of transparency and accountability.

Reliable statistics provide the evidence needed to assess solutions to socio-economic problems facing Africa. For example, no government can build schools without prior knowledge of the numbers of children likely to be

* The authors would like to thank Anne Thurston and Christopher Nnanatu for their comments on the initial draft of this chapter.

1 S. Morton, D. Pencheon and N. Squires, 'Sustainable Development Goals (SDGs) and their implementation: a national global framework for health, development and equity needs a systems approach at every level', *British Medical Bulletin*, 124 (2017): 81–90.

2 This interest in statistical evidence has grown ever stronger since the UN's adoption, in September 2015, of the 17 SDGs.

3 A. Awiti, 'Poor data no excuse for our bad policies', *The Star*, 18 July 2017, <http://data.adaialogueseries.org/spatial-inequalities/poor-data-no-excuse-for-our-bad-policies/>.

enrolled. Similarly, no government can claim to have reduced crime rates unless it can compare statistics on current crime rates to those of previous years. A country needs to know what crops it grows well, and where, if it is to prevent famine and malnutrition in children. Donors can only know whether their assistance is changing lives if they have access to quality data, stored securely and readily accessible for decision-making purposes. Development programmes should produce measurable results, and developmental decisions should be informed by the analysis and interpretation of data by government and/or educational agencies. In short, statistics constitute the barometer for measuring whether governments are making progress in addressing the concerns of their populations.

In this context, in 2014, the United Nations launched an appeal for a ‘data revolution’ prior to launching the SDGs.⁴ The concept of a data revolution highlighted the need for reliable statistics to address the widening gap between developed and developing countries in terms of access to and use of information. Statistics enable the state to address crucial issues affecting the lives of its citizens. They are a means by which citizens potentially can hold governments and their policies accountable.

In recent years, the state of statistics in Africa has been subjected to intense criticism. The major grounds have been, first, that development data produced by African regimes tends to be fabricated in order to reflect well on the regime and that collecting verifiable data inside closely guarded societies is virtually impossible.⁵ Second, it is argued that in any case, statistics gathered in Africa are often flawed and do not present the true situation on the ground.⁶ These criticisms tend to focus largely on export–import data and the economic sector of Africa,⁷ while pointing to the collapse of African statistical agencies. The concerns, which have also have been expressed by the World Bank, are strong indications of African statistical agencies’ inability to generate reliable and comparable data needed to evaluate the continent’s progress. Indeed, very few of the statistics produced in Africa, and especially sub-Saharan Africa, are sufficiently reliable to use, and critics contend that virtually all are guesstimates.⁸

4 See M. Guerero, ‘Betting on a data revolution to help manage UN development goals’, 2015, <http://www.passblue.com/2015/04/27/betting-on-a-data-revolution-to-help-manage-un-development-goals/>.

5 A. Gladstein, ‘Why dictators love development statistics: they’re an easily faked way to score international points’, *New Republic*, 26 April 2018, <https://newrepublic.com/article/148133/dictators-love-development-statistics>.

6 M. Jerven, ‘On the accuracy of trade and GDP statistics in Africa: errors of commission and omission’, *Journal of African Trade*, 1 (2014): 45–52.

7 A.J. Yeats, ‘On the accuracy of economic observations: do sub-Saharan trade statistics mean anything?’, *World Bank Economic Review*, 4 (1990): 135–56.

8 G. Alexander and J. Endres, *The Trouble with Statistics in Africa* (Johannesburg: Africa Check, 2014).

This situation has persisted for so long that critics suggest that securing reliable data in and for Africa is an unachievable goal.

This chapter, however, takes the view that rather than engage in sheer scepticism, it is more helpful to appreciate the progress made by African states and to focus on the way forward in achieving and monitoring governance in relation to the SDGs.⁹ We present a situation analysis and review the state of statistics across Africa in relation to the SDGs. The main issue here is that Africa offers a contrasting picture. On the one hand, some valuable statistical data exists, even though they are rarely used to plan and implement policies. On the other, many of the statistics that are gathered and published may not be of much help in addressing issues that matter to international and national development agencies. Moreover, there is a lack of technical capacity to analyse these data and make it available to the public as a basis for determining basic needs at the sub-regional or sub-county levels.

We argue that this gap must be addressed if African countries are to tackle real issues facing their populations with a view to achieving critical SDG targets. In this respect, statistics in Africa cannot be understood in isolation from the social conditions in which they are produced, processed and managed. Part of the reason that the available statistical data is underused has to do with the conditions for accessing and storing it in a continent that is traditionally more reliant on paper than on electronic media.¹⁰ Thus, in considering the state of statistics in Africa and in analysing the challenges that statisticians and data collectors face, attention needs to be given to the socio-economic and political conditions in which the information is collected, processed, stored, managed and used.¹¹

We explore these issues in four sections. The first defines ‘data’ and ‘statistics’, terms that are often mistaken for one another but that have distinct meanings. The second offers an overview of statistical censuses as carried out in Africa. To produce this, we used a Strengths, Weaknesses, Opportunities and Threats (SWOT) model to understand the issues associated with producing good quality statistical data in Africa. The third focuses on these key challenges in relation to gathering reliable statistics in Africa. The last section suggests some

9 See Objective 5.

10 It must be noted that once statistical data have been collected and processed, they need to be preserved for as long as they will be needed. On the state of statistical records management in Africa, see A. Thurston, ‘Records management in Africa: old problems, dynamic new solutions’, *Records Management Journal*, 6 (1996): 187–99; B.E. Asogwa, ‘The challenge of managing electronic records in developing countries: implications for records managers in sub-Saharan Africa’, *Records Management Journal*, 22 (2012): 198–211.

11 The same diagnosis is articulated by M. Jerven in his book *Poor Numbers* (Ithaca, NY: Cornell University Press, 2013). See also M. Jerven, ‘Random growth in Africa’, *Journal of Development Studies*, 46 (2010): 274–94.

of the ways that these challenges can be overcome, especially in relation to addressing the schism between demand for and the supply of data.

Defining the terms statistics and data

The starting point for discussing statistics and data in the African context is to consider how these terms should be defined. At its most basic level, data is information about a subject of interest (for example heights in a population), which can come in different forms.¹² Data can be quantitative (numerical) or qualitative (descriptive), for example, the answers to interview questions. The important point about data is that if they are to be of value to policy-makers, they need to reflect what they need to know. For example, data on the number of children of a specific age attending school in a region of the Democratic Republic of Congo will be a good indicator of the number of schools to be built for potential class intakes.

When a census is conducted of children of school-going age in a given region, it is possible, based on the data, to make statements directly about the population, for example, its average age. Generally, we can only estimate a particular characteristic or variable in the entire population; it is practically impossible to collect information on many issues at once. Rather, we can take a representative sample of the population we want to know about. We also want the sample to reflect the diversity of that population (for example, boys and girls, ethnic origin, disability cases). There are techniques for ensuring this as well as for random sampling.¹³

Descriptive statistics are a mathematical tool for analysing and organising data about a given state of affairs in a summary form.¹⁴ They illustrate different characteristics of a particular sample or population, making it possible to present the data in a meaningful way; statistics obtained from a sample of the population can be used to make inferences about the characteristics of the population. For instance, it can be helpful to an international developer or national policy-maker to know the mean, or average, of a particular variable. This is calculated by adding up the value of all the numbers reflecting a particular variable and then dividing that sum by the total of all the numbers.

Descriptive statistics are also about the spread or variability of a dataset, i.e. how much the data clusters around the mean, or whether the values are widely dispersed. The standard deviation is a measure used to quantify the amount of variation or dispersion of a set of data values. The way a standard deviation is

12 T.C. Urdan, *Statistics in Plain English*, 2nd edn (London: Routledge, 2005), p. 89.

13 W.G. Cochran, *Sampling Techniques*, 3rd edn (India: Wiley, 2007), p. 452.

14 J. Bather, 'A conversation with Herman Chernoff', *Statistical Science*, 11 (1996): 335–50. See also T. Porter, *The Rise of Statistical Thinking, 1820–1900* (Princeton, NJ: Princeton University Press, 1986).

calculated for data from a population is different from the way it is calculated for a sample. Data that has been statistically processed using measures including standard deviation, frequency, mode, range and interquartile range, are often referred to as statistical data.¹⁵ It is the raw information from which statistics are created. Statistics cannot exist without data, but it is possible to have data without statistics.

For data and statistics to remain relevant to policy-makers in Africa and beyond, the results should be interpreted in such a way that is easy for the decision-makers to understand. This is critical, not least because they need to have confidence in what statisticians recommend. Effective communication between statisticians and decision-makers is fundamental to using statistical data to decide whether and how to plan an intervention.

This discussion about data and statistics raises two fundamental questions. First, what sort of statistics are available about Africa? Second, to what extent do the existing statistical data provide a reliable foundation upon which to base policy decisions? To address these questions, we first turn to the nature and issues facing one of the most crucial aspects of data in Africa, namely census data.

Census data

The availability of a significant amount of cross-sectional census data in Africa makes it possible, theoretically, to monitor and explore the state of a country's current development across many aspects of social life (e.g. health, economics, education and science). They provide a means by which public policies can be continuously evaluated. This has led some scholars to make an association between good data and good governance.¹⁶ However, while census data have often been collected across sub-Saharan Africa, they tend not to have been collected regularly. We cannot compare statistics gathered at irregular intervals, often of several years, given the existence of many gaps and inconsistencies in data collection. Policy-makers find it difficult to use the data effectively to assist in implementing policies.¹⁷

Investigations using census data are likely to yield meaningful results because census data are a complete enumeration of all individuals in a country at a given time, allowing a meaningful understanding of progress, including monitoring the SDGs, which is crucial for implementing interventions. A population

15 D. Tanner, *Using Statistics to Make Educational Decisions* (London: Sage Publications, 2012), p. 15.

16 W. Baldwin and J. Diers, 'Demographic data for development in sub-Saharan Africa', *Poverty, Gender and Youth Working Paper No. 13* (New York: Population Council, 2009), p. 3.

17 Baldwin and Diers, 'Demographic data'.

census is the result of the process of collecting, compiling, evaluating, analysing and publishing or otherwise disseminating demographic, economic and social data pertaining to all people in a country, or in a well-defined part of a country, at a specified time.¹⁸ Censuses play a vital role in developing the official statistics needed to assist state agencies, businesses, other organisations or the public in planning, decision-making, monitoring or assessing policies.

Census data tend to be collected in such a way that the identity of the respondents is protected and that the data are relevant, accurate, reliable, timely, objective and comprehensive. Generally, such data are compiled, reported and documented in a scientific and transparent manner and disseminated impartially. Moreover, they tend to be collected in accordance with national and international standards and classifications that are appropriate for distribution by gender, disability, region and similar socio-economic features.¹⁹ A census is a complex and costly enterprise, especially in terms of careful planning and mobilisation of people and resources,²⁰ particularly so because all inhabited areas must be visited to provide a fair coverage of the entire population.

The majority of countries in sub-Saharan Africa conducted their first population censuses in the 1970s as a result of the African Census Programme (ACP), which was established by the United Nations Economic Commission for Africa. The ACP provided significant technical and financial assistance, received through the United Nations Population Fund, which enabled many countries to conduct censuses, especially in the 1980s and 1990s. By the 1990s, sub-Saharan Africa had assembled an impressive volume of population data, and national statistical institutes had developed expertise in collecting, processing and analysing these data.

In recent years, however, without ACP support, serious financial difficulties have prevented the organisation of population censuses. This has resulted in increased intervals between censuses or in a lack of censuses, as illustrated in [Table 2.1](#). This, in turn, has reduced the quality and volume of statistics available to governments for planning and formulating policy and for efforts to monitor the MDGs and, later, the SDGs.

Apart from the lack of funding for censuses initiatives, the intervals in census data collection in African countries have also been due to wars, political

18 S. Randall, E. Coast and P. Antoine, 'UN census "households" and local interpretations in Africa since independence', *Sage Open*, 5 (2015): 1–18, <https://doi.org/10.1177/2158244015589353>.

19 K.H. Hill, 'Trends in childhood mortality in sub-Saharan Africa', in K.A. Foote, K.H. Hill and L.G. Martin (eds), *Demographic Change in Sub-Saharan Africa* (Washington, DC: The National Academies Press, 1993), pp. 153–217.

20 L. Diop, 'Organization and financing of population censuses in sub-Saharan Africa: problems and prospects', paper presented at the Symposium on Global Review of 2000 Round of Population and Housing Censuses, New York, 2001, 7–10 August.

Table 2.1. Census frequency in French- and English-speaking sub-Saharan African countries

<i>Countries</i>	<i>Years censuses were conducted</i>
East Africa	
Kenya	1948, 1962, 1969, 1979, 1989, 1999, 2009
Tanzania	1967, 1978, 1988, 2002, 2012
Uganda	1911, 1921, 1948, 1959, 1969, 1980, 1991, 2002
Ethiopia	1984, 1994, 2007
Central Africa	
Angola	1970, 2014
Central African Republic	1988, 2003
Cameroun	1976, 1987, 2005
Gabon	2003
West Africa	
Benin	1978, 1992, 2002, 2013
Nigeria	1866, 1871, 1896, 1901, 1911, 1921, 1952, 1962, 1963, 1973, 1991, 2006
Ghana	1971, 1984, 2000
Guinea	1983, 1996
Ivory Coast	1998, 2014
Senegal	1976, 1988, 2002
Togo	1960, 1970, 1981, 2010
Southern Africa	
South Africa	1911, 1921, 1936, 1951, 1960, 1970, 1980, 1985, 1991, 1996, 2001, 2011
Lesotho	1986, 1996, 2006
Botswana	1904, 1911, 1921, 1936, 1946, 1956, 1964, 1971, 1981, 1991, 2001, 2011
Swaziland	1950, then every 10 years to the present
Malawi	1977, 1987, 1998, 2008
Namibia	1991, 2001, 2011
Zambia	1980, 1990, 2000
Post-conflict countries	
Democratic Republic of Congo	1984

Table 2.1. (continued)

<i>Countries</i>	<i>Years censuses were conducted</i>
Rwanda	1991, 2002
Mozambique	1987, 1997, 2007
Sudan	1973, 2007
Liberia	1843, 1974

Source: the Sub-Saharan Economic and Statistical Observatory, 1996, <https://www.afdb.org/en/documents/document/the-african-statistical-yearbook-2019-109564>.

instability, economic crises, and inadequate policies and leadership.²¹ Unless satisfactory solutions are found, many sub-Saharan African countries will find it impossible to use reliable statistics to monitor SDGs. Strategies are needed to address this issue. Effective awareness-raising, good organisation, rigorous planning of operations and optimum use of new technologies will make censuses less difficult to finance and more likely to produce reliable statistics for monitoring and achieving the SDGs.²²

In countries where population and housing censuses are carried out, the data represent a significant source of information on health matters (for example, immunisation and family planning) regardless of how inadequate and incomplete vital registration programmes may be. This is why it is crucial for governments in sub-Saharan Africa to ensure that full censuses are carried out if they are to meet the growing demand for statistical information at the national and subnational levels and to support the SDGs.

Post-apartheid South Africa is among the few African countries that have made progress in this area. Conducting its first population census in 1996, South Africa subsequently carried out censuses in 2001 and 2011, largely to compensate for the unreliable, uneven statistics produced during the apartheid regime, as illustrated by the overall figures produced under the Native Laws Amendment Act, the Areas Amendment Bill and the Group Areas Act 1950, all of which grossly underestimated the overall figures of urban and city residents.²³

Kenya has also worked to correct past distortions. In 2006, the government passed Law No. 4 of the Statistics Act, making it mandatory for the state to carry out regular censuses for every ten years on the basis of a printed

21 Record of censuses in sub-Saharan Africa. Adapted from the African Census Analysis Project (ACAP). Available at <http://www.acap.upenn.edu>.

22 Diop, 'Organization and financing of population censuses'.

23 Diop, 'Organization and financing of population censuses'.

questionnaire.²⁴ This is part of the broader context for the Kenyan population and housing census that aimed to deliver on the country's vision: 'Counting our People for Implementation of Vision 2030'. By generating information at all administrative levels, the Kenyan government has sought to provide a sound basis for assessing policies relating to its population. Despite this ambition, however, Kenya has not collected any fresh data in keeping with its current population growth.²⁵

The Democratic Republic of Congo illustrates an extreme situation. Relevant statistics do exist, but they date back to collection efforts by the National Institute of Statistics in 1984. Since independence in 1960, the DRC has had a turbulent history, and the earlier statistics have become obsolete; any projections that might be drawn from them will no longer be helpful in planning interventions. The idea of a second census was mooted and planned for July 2011 but did not materialise, even though a decree on its organisation was signed in August 2009.²⁶ It would be particularly valuable to have a census in this post-conflict country, which is heavily in debt, with a very poor population, despite its huge reservoir of mineral resources. The results of an up-to-date, well-conducted census would allow economic and social planning based on reliable statistical data that could contribute to the reconstruction of the country and help build capacity in the National Institute of Statistics. It could provide the nation, as well as international organisations, with reliable data for monitoring the SDGs.²⁷

Having considered issues relating to census data, we now need to consider other statistical activities at the regional and sub-continental levels.

Statistical activities in Africa

The Mo Ibrahim Foundation has examined the main activities undertaken by statisticians in Africa and considered the crucial significance of data for policy-making and service delivery. A report released in 2016 by the Mo Ibrahim Foundation on Africa's data revolution noted that there has been progress in the quantity of data being collected over the past ten years, especially in household surveys and population censuses. It noted, for instance, that:

24 See A.J. Christopher, 'The Union of South Africa censuses 1911–1960: an incomplete record', *Historia*, 56 (2011): 1–18.

25 See Laws of Kenya, Statistics Act No. 4 of 2006, National Council for Law Reporting, <http://www.kenyalaw.org>.

26 K. O-Kongo, 'Geographic information system and implementation of Kenyan vision, 2030', MBA thesis (2016), http://erepository.uonbi.ac.ke/bitstream/handle/11295/98690/Okong%27o_Geographic%20Information%20System%20and%20the%20Implementation%20of%20Kenyan%20Vision%202030%20State%20Department%20of%20Lands.pdf?sequence=1&isAllowed=y.

27 DRC's Decree law of 15 January 2009, Journal Officiel de la Republique Democratique du Congo, Premiere Partie, 2009.

- a third of all Africans lived in a country which had conducted a population census after 2010
- Kenya's revision of its economy meant that the country was recategorised from low-income to lower-middle-income
- Nigeria's rebasing revealed that its economy had surpassed South Africa's and is the largest in Africa.

However, the report also noted that there continued to be challenges in the frequency and the quality of data produced. For instance:

- four out of five known births in Africa occurred in a country without a complete birth registration system
- almost half of Africans lived in a country that had not conducted an agricultural census in the last ten years²⁸
- nine out of ten Africans lived in countries that had conducted a population census in the past ten years, and most Africans lived in countries that had conducted a household survey in the past decade. However, only half lived in countries that had carried out more than two comparable surveys. Their governments therefore could not access timely and comparable data on the changes in poverty levels.²⁹

The most readily available statistical data in Africa have been collected by western-based institutions. These data are important for sectors responsible for budgeting and planning where no other reliable data exists.³⁰ Local-level data are sparse in sub-Saharan statistical systems,³¹ and the available local data often do not provide the information needed for realistic planning. Decision-makers at international, national and local levels need data that are disaggregated down to the lowest level of administration. The ability to disaggregate data (breaking them down into sub-population, district, locality, and so on) enables policy-makers to plan appropriate programmes, determining which evidence-based interventions are most appropriate and deciding where they are most needed.

For instance, small sample surveys do not provide enough information to allow a health service in a given African country to determine precise locations where resources need to be allocated. 'Services are delivered through local authorities who need intelligence on their local communities to know how best to serve the people. Counting people to make people count is what the Data

28 2014 Annual Report on FAO's projects and activities in support of producer organizations and cooperatives, Rome, 2015.

29 Mo Ibrahim Foundation, 'Strength in numbers: Africa's data revolution', introduction, 2015.

30 Mo Ibrahim Foundation, 'Strength in numbers', p. 2.

31 B. Anderson, 'Quantifying the challenges facing data revolution in Africa: a first attempt', blog post for the Africa Open Data Conference, Dar es Salaam, Tanzania, 2015.

Revolution is about.³² Disaggregating data can show where aggregate data are masking discrepancies. For example, by looking at disaggregated data for smaller sub-populations, a national policy-maker or an international developer can recognise whether outcomes vary by sub-population and whether strong results by some sub-populations are masking poorer results by others.

Data collected at irregular intervals tend to be of uneven quality, as we illustrate in Table 2.2, which is based on information compiled from USAID-sponsored Demographic and Health Surveys³³ (DHS) and the UNICEF Multiple Indicators Clusters Surveys³⁴ (MICS).

Over the past 20 years, official statistics in Africa have suffered stagnation and obsolescence due to a progressive lack of sufficient human and financial resources allocated by governments. Beginning in the 1980s, when policy-making in the continent began to be dominated by structural adjustment programmes,³⁵ the main effect has been the steady reduction or curtailment of budgets for data collection and statistical analysis over decades.³⁶

The number of experts involved in data collection has grown over this period, but there has been an absence of basic controls for the reliability and quality of data.³⁷ Big data is an example of information produced outside official controls that should describe the process of drawing together disparate datasets to offer new insights into a population.³⁸ One of the challenges of this new trend is that there is no guarantee that disparate datasets are unbiased or will remain relevant to the sectors that are of interest to international and national policy-makers. Moreover, it is unclear how the rapidly growing pools of data generated through data digitisation and algorithms match publicly held databases, for instance those related to control of diseases in any given country. Another concern is whether big data is to be seen as an outright challenge to the credibility of African statistical institutes or even as a gradual takeover of the growing market for statistical information.

32 Baldwin and Diers, 'Demographic data', p. 8.

33 <https://dhsprogram.com/>.

34 <http://www.mics.unicef.org>.

35 Anderson, 'Quantifying the Challenges'.

36 A. Adepoju (ed.), *The Impact of Structural Adjustment on the Population of Africa: The Implications for Education, Health and Employment* (London: James Currey, 1993). See also S. Devarajan, 'Africa's statistical tragedy', *Review of Income and Wealth*, 59 (2013): S9–S15. Also see S. Chen, F. Fonteneau, J. Jütting and S. Klasen, 'Towards a post-2015 framework that counts: developing national statistical capacity', Paris21, Discussion Paper No. 1 (Paris, 2013).

37 See M. Jerven and D. Johnston (eds), *Statistics Tragedy in Africa? Evaluating the Database for African Economic Development* (London: Routledge, 2016), p. 3.

38 V. Bonnacase, 'Généalogie d'une Evidence Statistique: de la "Réussite Economique" du Colonialisme Tardif à la Faillite des États Africains, (v1930-v1980)', *Revue d'histoire Moderne et Contemporaine*, 62 (2015): 33–63.

Table 2.2. Showing the uneven nature of Demographic and Health Surveys (DHS) and Multiple Indicators Clusters Surveys (MICS), 2000–15.

Red cells: areas where recent data are available.

Green cells: areas which have relatively sufficient data.

Yellow cells: areas which have fairly sufficient data.

Grey cells: no available information.

Country	Latest planned population census	Latest planned household survey	Number of poverty surveys in last ten years (WB 2014)	Latest planned agricultural census	Vital statistics from civil registration (births)	Vital statistics from civil registration (CR) (deaths)	Status of health management information system	Status of education management information system	World Bank statistical capacity indicator (overall average 2014)
Algeria	2008	MICS 2006	Less than 2	2015	97%	Over 75%	Partial	No information	52
Angola	2014	DHS (2015)	Less than 2	2015	5 million children without birth certificate	No CR data used for vital statistics	No information	Partial	49
Benin	2013	DHS 2011 (2016)	More than 2	1992	60%	No CR data used for vital statistics	Pilot	No information	66
Botswana	2011	HIES 2002; DHS 1988	2	2015	72%	No CR data used for vital statistics	No information	Functioning	51
Burkina Faso	2006	DHS 2010	More than 2	2010 (planned for 2016)	No CR data used for vital statistics	No CR data used for vital statistics	Functioning	No information	71

Burundi	2008	DHS 2010	Less than 2	No information	Weak system, not functioning	No CR data used for vital statistics	Partial	Functioning	54
Cabo Verde	2010	DHS 2005	Less than 2	2014	Over 90%	Developing 'model of good practice'	No information	No information	69
Cameroon	2005	DHS 2011	Less than 2	2013	60–69%	No CR data used for vital statistics	No information	Partial	57
Central African Republic	2003	MICS 2010	More than 2	No information	50–59%	No CR data used for vital statistics	No information	Partial	59
Chad	2009	DHS 2014	2	2011 (planned for 2014–15)	No CR data used for vital statistics	No CR data used for vital statistics	No information	Functioning	63
Comoros	2003	DHS 2012	Less than 2	2004	Coverage expanding	No CR data used for vital statistics	No information	Functioning	40
Congo	2007	DHS 2011	2	2014–15	No CR data used for vital statistics	No CR data used for vital statistics	Pilot	No information	48
Côte d'Ivoire	2014	DHS 2011	Less than 2	2014	No CR data used for vital statistics	No CR data used for vital statistics	Pilot	No information	47

<i>Country</i>	<i>Latest planned population census</i>	<i>Latest planned household survey</i>	<i>Number of poverty surveys in last ten years (WB 2014)</i>	<i>Latest planned agricultural census</i>	<i>Vital statistics from civil registration (births)</i>	<i>Vital statistics from civil registration (CR) (deaths)</i>	<i>Status of health management information system</i>	<i>Status of education management information system</i>	<i>World Bank statistical capacity indicator (overall average 2014)</i>
Democratic Republic of Congo	1981	DHS 2013	2	1990	No CR data used for vital statistics	No CR data used for vital statistics	Partial	Partial	57
Djibouti	2009	MICS 2006	Less than 2	1995	No CR data used for vital statistics	No CR data used for vital statistics	No information	Functioning	46
Egypt	2006	DHS 2014	More than 2	2010	70–90%	70–90%	No information	Functioning	90
Equatorial Guinea	2002	DHS 2011	Less than 2	No information	70–90%	No CR data used for vital statistics	No information	Functioning	34
Eritrea	None	DHS 2002	Less than 2	No recent census	No CR data used for vital statistics	No CR data used for vital statistics	No information	Functioning	31

Ethiopia	2007	DHS 2011 (2016)	2	2002	No CR data used for vital statistics	No CR data used for vital statistics	No information	Functioning	61
Gabon	2013	DHS 2012	Less than 2	1982	No CR data used for vital statistics	No CR data used for vital statistics	No information	No information	42
Gambia	2013	DHS 2013	2	2012	No CR data used for vital statistics	No CR data used for vital statistics	Functioning	Functioning	67
Ghana	2010	DHS 2014	Less than 2	2014	60–69%	Over 25%	Functioning	Functioning	62
Guinea	2014	DHS 2012	2	2001	57%	No CR data used for vital statistics	No information	No information	52
Guinea-Bissau	2009	MICS 2010	Less than 2	1988	No CR data used for vital statistics	No CR data used for vital statistics	Pilot	No information	43
Kenya	2009	DHS 2014	2	1979	Expanding system of CR offices	Over 25%	Functioning	Functioning	54
Lesotho	2006	DHS 2014	2	2010	No CR data used for vital statistics	No CR data used for vital statistics	No information	Functioning	72

<i>Country</i>	<i>Latest planned population census</i>	<i>Latest planned household survey</i>	<i>Number of poverty surveys in last ten years (WB 2014)</i>	<i>Latest planned agricultural census</i>	<i>Vital statistics from civil registration (births)</i>	<i>Vital statistics from civil registration (CR) (deaths)</i>	<i>Status of health management information system</i>	<i>Status of education management information system</i>	<i>World Bank statistical capacity indicator (overall average 2014)</i>
Liberia	2008	DHS 2013	2	No recent census	Weak but expansion programme in progress	No CR data used for vital statistics	Functioning	Functioning	47
Libya	2006	NS (2011)	Less than 2	2013–14	No CR data used for vital statistics	No CR data used for vital statistics	No information	Pilot	29
Madagascar	1993	DHS 2008 (2016)	More than 2	2004	No CR data used for vital statistics	No CR data used for vital statistics	No information	Functioning	62
Malawi	2008	DHS 2010 (2015)	More than 2	2007	No CR data used for vital statistics	No CR data used for vital statistics	Partial	Functioning	76

Mali	2009	DHS 2012	More than 2	2015	Coverage in expansion – no specific number found	No CR data used for vital statistics	No information	Functioning	67
Mauritania	2013	MICS (2011)	2	1985	System reliable, but coverage varies from rural to urban. Breakdown available by region	Over 75%	No information	Partial	59
Mauritius	2011	Continuous Multipurpose Household Survey 2013	More than 2	2014	Over 90%	Over 90%	No information	No information	86
Morocco	2014	Household Panel Survey 2011	2	2015	System going informatic, high coverage (no precise number)	60–69%	No information	No information	79

<i>Country</i>	<i>Latest planned population census</i>	<i>Latest planned household survey</i>	<i>Number of poverty surveys in last ten years (WB 2014)</i>	<i>Latest planned agricultural census</i>	<i>Vital statistics from civil registration (births)</i>	<i>Vital statistics from civil registration (CR) (deaths)</i>	<i>Status of health management information system</i>	<i>Status of education management information system</i>	<i>World Bank statistical capacity indicator (overall average 2014)</i>
Mozambique	2007	DHS 2011	2	2010	Existing system, coverage in expansion	No CR data used for vital statistics	Functioning	Functioning	74
Namibia	2011	DHS 2013	2	2014	Existing system, needs to be overhauled	No CR data used for vital statistics	Pilot	Functioning	49
Niger	2012	DHS 2012	More than 2	2008	Below 50%	No CR data used for vital statistics	Pilot	No information	68
Nigeria	2006	DHS 2013	More than 2	2013	35% in 2007 (NPC 2010)	No CR data used for vital statistics	Functioning	Functioning	72
Rwanda	2012	DHS 2014	More than 2	2013	Coverage expanding, census	No CR data used for vital statistics	Functioning	New system launched June 2015	79
Sao Tomé and Príncipe	2012	MICS 2014	More than 2	2012	Complete registration	No CR data used for vital statistics	No information	Functioning	69

Senegal	2013	DHS 2014	More than 2	2013	No CR data used for vital statistics	No CR data used for vital statistics	Pilot	No information	73
Seychelles	2010	Household Budget Survey 2013	Less than 2	2011	Over 90%	Over 75%	No information	Functioning	62
Sierra Leone	2004	DHS 2013	More than 2	1985	Over 70%	No CR data used for vital statistics	Functioning	Functioning	59
Somalia	1987	MICS 2006	Less than 2	No recent census	No CR data used for vital statistics	No CR data used for vital statistics	No information	Partial	20
South Africa	2011	General Household Survey 2013	More than 2	2013	Over 75%	Over 80%	Partial	Functioning	74
South Sudan	2008	National Baseline Household Survey 2009	Less than 2	No recent census	No CR data used for vital statistics	No CR data used for vital statistics	Pilot	Functioning	29
Sudan	2008	National Baseline Household Survey 2009	Less than 2	2015	40–60% (system and coverage expanding)	No CR data used for vital statistics	Pilot	Functioning	43

<i>Country</i>	<i>Latest planned population census</i>	<i>Latest planned household survey</i>	<i>Number of poverty surveys in last ten years (WB 2014)</i>	<i>Latest planned agricultural census</i>	<i>Vital statistics from civil registration (births)</i>	<i>Vital statistics from civil registration (CR) (deaths)</i>	<i>Status of health management information system</i>	<i>Status of education management information system</i>	<i>World Bank statistical capacity indicator (overall average 2014)</i>
Swaziland	2007	MICS 2010	Less than 2	2003	Existing system, strengthen in progress	No CR data used for vital statistics	No information	Functioning	60
Tanzania	2012	DHS 2015	More than 2	2008	Unreliable system	No CR data used for vital statistics	Functioning	Functioning	72
Togo	2010	DHS 2013	2	2012	90–70%	No CR data used for vital statistics	Pilot	No information	64
Tunisia	2014	MICS 2011	2	2014–15	Over 70%	'Satisfactory'	No information	Functioning	72
Uganda	2014	National Household Survey 2012	More than 2	2008–9	Coverage expanding	No CR data used for vital statistics	Functioning	Functioning	64

Western Sahara	None	None	No information	No recent census	No CR data used for vital statistics	No CR data used for vital statistics	No information	No information	No information
Zambia	2010	DHS 2013	More than 2	2000	Weak and unreliable system	No CR data used for vital statistics	Functioning	Functioning	60
Zimbabwe	2012	DHS 2010 (2015)	More than 2	No recent census	Reform still in progress	Over 25%	Functioning	Functioning	58

At the same time, there is evidence that some African governments are beginning to use data in new ways. Moving beyond doubts expressed about technical expertise in national data collection and statistics,³⁹ and beyond long-standing patterns of corruption and misfeasance, countries such as South Africa, Mozambique and Namibia have begun leveraging big data to improve efficiency and effectiveness of government in policy areas such as citizen security, taxation and smart cities. Other countries in sub-Saharan Africa have been slower in appreciating the role of data and statistics as an aid to government in discharging its obligations towards its citizens⁴⁰ and embarking on the road to prosperity.⁴¹

A detailed and richer picture of the state of statistical data in Africa is captured through use of the SWOT technique as described below.

SWOT analysis

A situation analysis aimed at understanding the emerging internal strengths, internal weaknesses, external opportunities and external threats for managing statistics across the continent should help explain the factors that must be addressed if realistic solutions are to be developed.

As Table 2.3 shows, Africa's strengths in the field of statistics include emerging new frameworks, regulations and action plans geared towards enhancing statistical development. After a significant decline in the quality of work by national statistics offices in Africa from the 1970s, the Addis Ababa Plan of Action for Statistical Development in Africa was officially adopted in May 1990 by the United Nations Economic Commission for Africa,

39 The internet has been the main driver behind the notion of big data in Africa, which depends on improving internet connectivity. Most high-intensity data projects make use of a cloud-based component, and this naturally requires a connection to the cloud provider. At the same time, connectivity in Africa is still by no means up to the global standards seen in more developed markets, and this acts as a constraint to fully realising big data's benefits. Another important constraint is a lack of skills. A worthwhile data project requires both the technical skills to manage and analyse the data and the strategic skills to draw meaningful conclusions from the analysis. Finally, it can be quite difficult to decide where to begin building and deploying analytic models. There are so many areas that can benefit from analytics, that service providers can be at a loss as to where they can benefit most from analytics in the short and long term. See A. Shankar, 'Africa's entry into big data and analytics', IntelligentCio.Com, 2017, <http://www.intelligentcio.com/africa/2017/08/27/africas-entry-into-big-data-and-analytics/>.

40 See for example, Politique Africaine, 2014, and P. Lehohla, 'Statistical development in Africa in the context of the global statistical system', background document prepared for the Statistical Commission 39th Session, 26–29 February 2008, <https://unstats.un.org/unsd/statcom/39th-session/documents/bg-africastatdev-E.pdf>.

41 See J. Stiglitz, 'Redefining the role of the state: what should it do? How should it do it?' and how should these decisions be made?', paper presented on the 10th Anniversary of MITI Research Institute (Tokyo, Japan), 1998, p. 3, https://www0.gsb.columbia.edu/mygsb/faculty/research/pubfiles/1494/Stiglitz_RedefiningRole.pdf.

Table 2.3. SWOT analysis for sub-Saharan Africa

<i>Strengths</i>	<i>Weaknesses</i>
1 National regulations governing the management of statistics exist in many African countries.	1 National statistical systems in most African countries are vulnerable and fragile.
2 An infrastructure generally exists for large-scale data collection, including censuses and sample surveys.	2 Statistical capacity tends to be low in ministries, departments and agencies (MDAs) as well as in some pan-African statistical organisations.
3 Training institutions exist for professional and semi-professional statistical staff.	3 There is an absence of registration of actual civil status and of vital statistics.
4 Regional organisations are capable of providing technical and financial support to countries.	4 Low data quality contributes to low use of statistical data by policy-makers.
5 There is regular exchange of knowledge, experience and good practices, including meetings of national statistical institute officials at regional and national levels, statistical newsletters and the African Statistical Yearbook.	5 There is inadequate statistical information on key development indicators, such as environmental/ climate change, gender, governance, HIV/AIDS control.
6 Key stakeholders at all levels in national statistical services are willing to collaborate to achieve synergy and cost-effectiveness in statistical production.	6 There is a lack of incentives and of sufficient capabilities and skills to handle and make use of the available data.
	7 There is insufficient administrative autonomy and insufficient professional independence in African statistical systems.
	8 There is a lack of predictable and sustainable funding for harmonising statistics in Africa.
	9 There is a lack of publicly available disaggregated data along socio-demographic lines that could make a difference in devising policies and targeting interventions at the grassroots levels.
	10 There is political interference in statistical work especially at the national level.

<i>Opportunities</i>	<i>Threats</i>
1 There is a growing demand for statistics and an international consensus that statistics are an indispensable part of the enabling environment for improving the results of development efforts and decision-making at all levels.	1 There is a lack of coordination among international partners that have introduced multiple initiatives for statistical systems.
2 Governments in the region recognise weaknesses in their statistical systems and the need to strengthen them.	2 There are inadequate microdata for sub-counties and municipalities, despite the demand for those data as a means for establishing accountability and assessing how governments are discharging their duties towards citizens.
3 Development partners have been willing to support capacity-building initiatives in Africa both financially and technically.	3 There is reduced investment in statistics by governments and international donors, particularly where richer countries are committed to budget austerities and reducing aid to poorer countries of Africa.
4 International frameworks, standards, guidelines and successful practices exist to support statistical harmonisation.	4 It is difficult to attract and retain statistical staff as governments experience chronic underfunding and ministries compete for financial resources to carry out their mandates.
5 Regional, continental and international partnerships exist for statistical development.	5 There is a lack of commitment to coordination among key stakeholders.
6 Technological advances have made computers cheaper, more powerful and more accessible.	6 There are insufficient legal measures in place to support improved statistical data.

Conference of Ministers for Economic Development and Planning. The Plan, which was at the cutting-edge of statistical advocacy, promoted evidence-based development. This trend has been enhanced by the fact that most African countries have joined the global effort to reduce poverty by supporting the achievement of the SDGs, including sustained and equitable economic growth, which is in line with the objectives of previous pan-African initiatives. The regular exchange of knowledge and experience, as well as agreed good practices among national statistical experts in Africa, has also helped.

These strengths, if anything, are overshadowed by weaknesses and threats for statistical reliability. As the SWOT model suggests, despite positive trends across Africa, the picture is mixed. We recognise the significant efforts by international agencies to ensure that Africa has reliable data. For instance, recent improvements in Nigeria's national accounts compilation methodology have effectively doubled the estimated Nigerian GDP, making Nigeria the leading economic power in sub-Saharan Africa, even before South Africa.

However, this reinforces the argument: it shows real improvement in Nigerian statistics, but nevertheless, the actual data on the country's macroeconomics remains questionable.

In recent years there have been improvements in statistics in Africa in relation to the continuous development of both the Demographic and Health Surveys (DHS) and the Multiple Indicators Clusters Surveys (MICS). However, even here significant challenges remain. As has been discussed and shown in [Tables 2.1](#) and [2.2](#), data are gathered irregularly, and some sources have not been updated in recent years. In some cases, key information is lacking about a particular issue for sub-regions or for counties. Where census data have not been collected consistently, their quality and timeliness can be compromised, especially when there are delays in collecting and releasing data. Timeliness is a serious problem. Where there is a long interval between data collection and reporting, by the time the data are released, the circumstances that led to a demand for them may have changed, making them redundant, irrelevant and difficult to compare. This encourages a culture where policy-makers and the media tend to rely on data estimated on the basis of gut feelings. Many of the DHS and MICS depict situations at the national level that mask the reality at the subnational or regional levels. These gaps, and their impact on the quality and timeliness of data, must be tackled if DHS and MICS sources are genuinely to become major pillars for national and international development programmes in Africa.

Overall, the challenges for statistical development in Africa remain enormous, and the scarcity of capacity to tackle them is pronounced. The shortcomings are largely due to the lack of significant investment in obtaining data needed for planning and implementing policies. National statistical institutes in most sub-Saharan countries have limited human and financial resources, and often there are insufficient skills among staff responsible for data collection and management. For instance, a statistical department may have only two or three competent statisticians and demographers, while the majority of staff lack appropriate training.

There are also technical issues, especially when data are in paper form and liable to destruction or where there is no digital preservation strategy in place. Even where appropriate technology exists, there are the additional challenges of power cuts, poor equipment and low bandwidth, which compound the difficulties of accessing and sharing data. All of this significantly impacts the data that are collected. For instance, data on older women, youth and agricultural activities remain a challenge.⁴² Despite the persuasive UN language ('nobody

42 E.g. NEPAD (New Partnership for Africa's Development). This initiative, a blueprint for Africa's development in the 21st century, was adopted at the 37th Summit of the Organisation of African Unity (now the African Union) in 2001.

left behind'), major investment is needed if all areas of statistical concerns are to be adequately addressed.

The result is that Africa has inadequate statistical information on key development indicators, such as environmental/climate change, gender, governance, HIV/AIDS control. Often, laws relating to statistics are not enforced, and plans for improvement are not implemented, leaving national governments without reliable information about the populations they should be serving. As has been noted, data from various sources (census, surveys, civil registration) tend not to be disaggregated to the community level, which is where interventions are needed.⁴³ Census population data are readily available, and over 80 per cent of African countries have conducted a census in the last ten years. However, there are two significant challenges: first, how can the coverage and quality of the census be reliably assessed, and second, how can we ensure that data are disaggregated to the lowest level of administration where they are needed for planning purposes.

Surveys in Africa have tended to be driven by international organisations, and there is an assumption that well-structured data exist to inform policy planning and implementation. It is true that many African countries have conducted at least one or two household surveys over the last three to five years. But, as World Bank staff have noted, because only 33 per cent of African countries have conducted more than two poverty-related surveys in the past ten years, comparative data are not readily available, and civil registration, which represents the only credible way to count people on an ongoing basis and thus to produce useable vital statistics, has not been adequately resourced.⁴⁴

Another major statistical challenge for sub-Saharan Africa is that much of the data that exist are buried in files across government ministries and agencies. As one scholar has indicated, 'sometimes, sourcing of this administrative data is made very difficult due to administrative bottlenecks'.⁴⁵ End users are often faced with secrecy and confidentiality issues. Significant data may be buried in an obscure special report, in published documents held only in a few libraries or on administrative files. Sometimes staff employed in government-run National Statistics Institutes are simply not aware of the detailed statistics contained in these documents. Weaknesses in records systems often make it difficult to find relevant information on government files. Even if the relevant source is identified, the data may not be arranged according to time series, which

43 S Randall, Ernestina Coast and Philippe Antoine (2015), 'UN census "households" and local interpretations in Africa since independence', SAGE Open (April–June 2015): 1–18, <https://doi.org/10.1177/2158244015589353>.

44 There are exceptions where disaggregated data exist, which include Kenya and South Africa. See Awiti, 'Poor data no excuse'.

45 K. Beegle, L. Christiaensen, A. Dabalen and I. Gaddis, *Poverty in a Rising Africa* (Washington, DC: World Bank Group, 2016).

makes the users' task very daunting. Annual reports and other documents often contain figures for the most recent years only.

Other constraints arise from the social conditions under which the data are collected and recorded. These can include misconceptions about why the statistics are being collected, uncooperative attitudes of participants from whom data are being collected, statistical illiteracy, ineffective statistical legislation and lack of dedication by enumerators.⁴⁶ These challenges are often the result of political influence over how statistics are collected and presented.

Political influence is a challenge for statistics in Africa.⁴⁷ Too often, policies are driven by political views rather than by empirical analysis.⁴⁸ Even where data are available, they may be ignored by African policy-makers, especially where they threaten the personal views or political needs of leaders. Often, data captured through national and international efforts and their potential value are not known to the public. The prevailing culture of secrecy, the fear of misuse of the information and the perceived need for confidentiality add to the difficulties of accessing and using the data.

By way of summary, the issues presenting a challenge in Africa can be categorised as those of a technical nature related to the data and statistics themselves and those associated with the overall management infrastructure (for example, laws and policies, human and financial resources).

Technical data-related issues include:

- data are lacking, gathered irregularly and/or not updated
- data are buried in poorly organised files across ministries and agencies
- data are not timely or are even redundant because of delays in processing and release
- the coverage of certain segments of society is poor (for example, older women, youth)
- data are not disaggregated to enable analysis at a local level
- tools for measuring coverage and disaggregation are lacking
- there is a lack of comparative data through time
- a preservation strategy is lacking
- technical infrastructure is poor (power cuts, poor equipment, low bandwidth).

46 C.C. Nweze, 'Environmental constraints in data sourcing in Nigeria, 1914', unpublished paper, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.560.4352&rep=rep1&type=pdf>.

47 Nweze, 'Environmental constraints'.

48 S. Ellis, 'The current state of international science statistics for Africa', *The African Statistical Journal*, 6 (2008): 177–89.

Infrastructural management-related issues include:

- lack of financial and human capacity
- lack of investment in infrastructure resulting in inadequate infrastructure for managing quality and integrity of data and statistics from creation/ collection, to use, to retention and preservation
- laws are non-existent or, if in place, they are ineffective and/or not enforced
- access to data may be constrained by ‘secrecy and confidentiality issues’
- statisticians and others may not be aware of the existence of data
- social conditions undermine the quality, integrity and availability of data and statistics
- there are misconceptions about why statistics are being collected
- participants are uncooperative
- there is statistical illiteracy in the proposed user population
- enumerators lack dedication
- political influence and views can lead to statistics being ignored and suppressed.

These weaknesses have contributed to the argument that African statistics are a tragedy. The statistical narratives emerging from the continent ‘tell us less than we would like to think (they do) about income, poverty and growth in the region’.⁴⁹ This concern has been echoed by the World Bank’s chief economist for Africa, Shantayanan Devarajan, who has argued that because of the state of statistics in Africa: ‘We cannot be sure whether there is growth or that poverty is declining’.⁵⁰

Overcoming the challenges

We have discussed some of the challenges of gathering reliable data and statistics in Africa. Let us now consider some of the ways that the challenges can be overcome. There is a need for a pan-African approach to transforming the current situation. By 2009, a number of African institutions, including the Economic Commission for Africa, the African Union and the African Development Bank, had developed a Strategy for the Harmonisation of Statistics. The strategy, which was adopted that year by African heads of state, remains a valuable way forward. It is a continent-wide effort to produce and disseminate harmonised statistics to inform African development at the local and regional levels.⁵¹

49 Baldwin and Diers, ‘Demographic data’, p. 5.

50 S. Devarajan, ‘Africa’s statistical tragedy’, *Review of Income and Wealth* (2013), <https://doi.org/10.1111/roiw.12013>.

51 See <https://au.int/en/ea/statistics/statafric>.

The strategy provides a framework for defining policies and good practices for developing, producing and using statistics and is intended to guide the process of harmonising concepts and definitions, adapting international or peer-agreed good practices, such as quality frameworks, and using common methodologies for producing and disseminating statistics to facilitate comparisons of statistics of African Union member states across time through coordination and collaboration of national, regional and international stakeholders.

Advocates of the strategy believe that statistical services in regional institutions⁵² need to be strengthened in order to generate timely, reliable and harmonised statistical information to support political, economic, social and cultural integration across the continent. Working with individual African governments, the strategy offers the potential to tackle the challenges of producing, analysing and disseminating the quality information needed to inform decision-making and evaluate the results of policies. However, to achieve this, it will be necessary to develop statistical capabilities within a reasonable timeframe. Establishing such capabilities will require human, material and financial resources; mobilising these resources will require significant and constant support from all development partners.

The strategy proposes that national statistical systems should be structured to:

- 1 Raise awareness among governments and the public of the strategic importance of statistics for promoting citizens' wellbeing and economic and social development in general. Funding for statistical activities will continue to depend to a large extent on this awareness.
- 2 Redefine national priorities for statistical development to ensure that the basic macroeconomic data needed to strengthen national accounting systems and monitor improved household living conditions are available and are disseminated effectively.
- 3 Ensure that basic statistics and reliable indicators are in place to address issues relating to poverty and sustainable development (for instance, the fight against HIV/AIDS, protecting the environment, developing the private sector, increasing economic performance, promoting equal gender), while also monitoring new requirements for data.
- 4 Improve the quality and timeliness of data production, for instance by acquiring adequate IT equipment and ensuring that quality controls for collecting, processing and disseminating data are in place.
- 5 Establish/strengthen partnerships at the national, sub-regional, regional and international levels to mobilise extrabudgetary resources needed to support good statistical practice for producing and disseminating statistics to meet international standards.

52 M. Jerven, 'Lies, damn lies and GDP', *The Guardian*, 20 November 2012.

- 6 Harmonise and implement the rules by which statistics are collected and disseminated across Africa as part of the reform and modernisation process. This could result in a substantial, but inevitably necessary, increase in statistical requirements for all countries.

With sufficient resource allocation in the coming years and a significant dose of political goodwill, Africa can address the overall poor quality and inconsistent nature of basic statistics in countries throughout the continent and be in a position to monitor the implementation of the SDGs by 2030. However, the goals can be achieved only if the situation becomes a priority for governments and development agencies in coming years.

Conclusion

This chapter has aimed to provide a frank assessment of the current state of statistics in and for Africa. We have noted that statistical development has a chequered past, owing largely to the social and political conditions under which statistics are produced. We have explored the current state of statistics in selected countries, investigated key challenges and considered ways in which those challenges can be overcome. The focus has been on both internationally generated statistics, such as DHS and MICS, and national efforts to collect and manage statistics.

By exploring the strengths, weaknesses, opportunities and threats facing statistical development in Africa, we have considered the major challenges that need to be overcome if African countries are to achieve the SDGs. We have noted, in particular, the importance of ensuring that national statistical institutions are properly resourced and have the autonomy needed to resist pressure to give undue credit to a particular leader. We also noted the importance of making statistics a national priority and building the capacity of existing institutions so that Africa can generate accurate, timely, relevant, accessible and unbiased data. Implementing these changes will require that governments act in partnership with civil society groups and international agencies to support statistical priorities.

3. Data, information and records: exploring definitions and relationships

Geoffrey Yeo interviewed by James Lowry¹

The achievement and measurement of the Sustainable Development Goals (SDGs) depend on the availability of trustworthy data from a variety of sources. Records, especially records created by government agencies, are often identified as one of the most important sources from which such data can be derived. Over centuries, the records and archives management profession has developed approaches to maintaining, controlling and contextualising records, which can help users assess the trustworthiness of the records and perhaps also the quality of the information that can be gained from them. With so much vested in the SDGs, it has become increasingly important to interrogate these terms – records, information and data – to achieve a better understanding of how they are interrelated. In this interview, James Lowry asks Geoffrey Yeo, author of *Records, Information and Data*,² to analyse the distinctions and relationships among these concepts.

JL: I'd like to ask you about some of the themes you discuss in your book and their relevance to the challenges of measuring global progress towards the SDGs. In the book, you explore a number of topics that may perhaps help us to understand what is meant when people speak about 'information', 'data' and 'records', and how they affect the indicators used to measure progress toward the SDGs.

GY: I certainly hope that my book will make a useful contribution. It doesn't specifically address the SDGs, but it looks more generally at records, information and data, the different meanings that people have attached to these terms, and the different ways in which their relationships have been interpreted and understood.

I had a number of aims in writing the book, but one key aim was to examine the growing tendency among records professionals (records

1 For biographies of Geoffrey Yeo and James Lowry, see the list of contributors at the beginning of this volume. See also [Chapter 8](#) in this volume.

2 G. Yeo, *Records, Information and Data: Exploring the Role of Record-Keeping in an Information Culture* (London: Facet Publishing, 2018).

managers and archivists) to try to explain *records* in terms of *information*. Philosopher John Searle has described information as ‘one of the most confused and ill-defined notions in ... intellectual life’.³ Why, then, does it have such a high profile in contemporary discourse? Why have records professionals attributed such importance to it in recent years? These questions turned out to have many ramifications, and investigating the connections – real or supposed – between records and information turned out to have many more.

The approach I took is rather different from the approach followed by most writers about records management and information management. The book looks at many varied ideas about the meaning or meanings of ‘information’ and explores many aspects of what it calls ‘the role of recordkeeping in an information culture’. It argues that concepts of information, although currently fashionable, don’t provide an adequate foundation for understanding how records are made or how they operate. Information is not what records *are*, nor is it what they *contain*, but it is, perhaps, what we may hope to gain from using them intelligently. Information should be associated with the use of records rather than with their creation.

JL: It is generally accepted that data play important roles in assessing and achieving the SDGs, and they have often been considered to have close relationships to records and also to information. Where do data fit in to your analysis?

GY: Originally, I hadn’t intended to say much about data. But, as my work on the book progressed, I found that I couldn’t adequately explore the concepts of records and information, and the associations between them, without considering data and the burgeoning worlds of data science and data management. Relations between records and data became an important theme of the book. As you know, it’s a topic on which there are widely varying opinions.

I think a lot of difficulty arises because of uncertainty about what we mean when we speak of ‘data’. Does the term refer to *anything* that is, or can be, stored on a computer? Or only to digital materials that are in some way meaningful? Or only to materials that exist in structured formats (as, for example, in databases)? Or is ‘data’ a wider term, embracing a range of non-digital as well as digital resources? Each of these views has its advocates, and those who adhere to any one view often take it for granted, not recognising that other people may understand the term ‘data’ very differently. When people talk about the importance of data

3 J. R. Searle, *Making the Social World: The Structure of Human Civilization* (Oxford: Oxford University Press, 2010), p. 71.

for achieving – or measuring progress towards – the SDGs, they don't always take the trouble to explain what they mean by 'data'.

- JL: Other contributors to this book will have much more to say about connections between data, records and the SDGs; I'd like to focus our conversation on records and their relationship (or relationships) to data. You have characterised one view of this relationship as sequential, where records are made prior to the creation of data.⁴ You see an association between this view and environments where digital systems employed structured data, and records were made and kept in paper form.
- GY: This is not my own view, but it's a view that often surfaces in the literature. It sees records as 'source documents' from which data are extracted or derived. For those who support this view, records arise from the conduct of organisational business; data entry clerks then examine the records, identify appropriate content, format or code it, and input it as data into 'structured' database systems. The data in these systems are then used for a range of purposes including administrative and financial control, strategic planning and decision-making; purposes beyond those that led to the original creation of the records. For example, employment data can be derived from records of staff appointments, agricultural data extracted from land surveys, or environmental data aggregated from records of impact assessments.

Sometimes, a chain of processes is involved. Coders or data entry clerks identify relevant details in what are supposedly 'unstructured' records and enter them into a 'structured' database. Or this task may be automated, perhaps using some kind of recognition technology. Either way, this initial 'data entry' is a preliminary to further processing. When the initial entry is complete, a computer program takes the structured textual data and uses them to create processed or 'computational' data, or statistics of various kinds. More complex routes involving multiple stages of further processing are also possible.

From this point of view, the quality of the data depends on the quality of the original records from which they are drawn. The reliability, accuracy and trustworthiness of the records determine the reliability, accuracy and trustworthiness of the data and statistics derived from them. Poorly kept records, it is argued, result in inaccurate, incomplete or unverifiable data, which can lead to organisations wasting resources attempting to process or analyse data that are of poor quality. Worse, governments and donor agencies can be misled into making ill-informed decisions with potentially damaging consequences. Skewed findings, misguided

4 Yeo, *Records, Information and Data*, pp. 111–12.

policy initiatives and misplaced funding can all have devastating effects on people's lives. Open data projects may also rely on data derived from poorly kept records; and so citizens may unwittingly be provided with data that are untrustworthy. Data can be collected from interviews, experiments, surveys, measurements or calculations as well as from records, but when data are extracted from records, the ability to trace the data back to the records from which they are derived is an important issue. Records can also serve to document the procedures by which data are collected and the processing methods that are applied to data, and further serious difficulties can arise if records serving these purposes are not made and kept to appropriate standards.

These views have been expressed in many reports issued by the International Records Management Trust (IRMT) over the past 20 years. They were first articulated by Piers Cain and Anne Thurston in the late 1990s, at a time when the first automated systems were being acquired in low-income countries. Donor agencies were actively encouraging governments in these countries to adopt automated systems, particularly (at that time) for personnel and payroll data. Automation was often seen as the solution to the inadequacies of existing paper systems where records had been poorly maintained and the information that could be obtained from them was frequently incomplete or unreliable. At the same time, however, the existing paper records were seen as 'the primary sources of the data needed for input into the automated system'.⁵ Automation didn't solve the problem of reliability; it simply transferred the problem from a paper-based to a digital environment. The IRMT affirmed that the answer lay in effective records management controls, which would support and ensure not only the systematic creation and survival of the records that were needed, but also their orderliness, trustworthiness and continuing accessibility.

I don't want to suggest that these issues have become outdated or unimportant, but they are characteristic of an era when records chiefly took the form of paper files and data were associated only – or very largely – with structured automated systems; databases often had to be populated from paper sources. Today, in wealthier countries – and increasingly also in many less wealthy ones – paper files are becoming obsolete, records are being maintained in digital rather than paper form, and the world of recordkeeping looks very different.

5 P. Cain, 'Automating personnel records for improved management of human resources: the experience of three African governments', in R. Heeks (ed.), *Reinventing Government in the Information Age* (London: Routledge, 1999), pp. 135–55, at p. 146.

Recently, we've heard much about 'datafication', which may also lead us to rethink our approach to these issues. The term 'datafication' became popular after it was used in a 2013 book by Viktor Mayer-Schönberger and Kenneth Cukier.⁶ To 'datafy a phenomenon', in the words of these writers, is to put it in a format that allows it to be tabulated and analysed. Elsewhere, Cukier and Mayer-Schönberger tell us that 'datafication is ... taking all aspects of life and turning them into data'.⁷ More specifically, it seems, it is about transforming resources so that they can be analysed in depth using new computational and analytical techniques from the realms of big data and artificial intelligence. It has often been noted that, by using these techniques to detect and analyse themes, patterns and relationships in digital materials, we will be able to open up innovative modes of discovery and investigation.

Other commentators have picked up these ideas, and I think that datafication can now be understood in at least two senses: it can be interpreted as a practical imperative to create resources in, or convert them into, datafied forms; more conceptually, it implies an intellectual reframing of all digital objects as data amenable to computation. These changes are having a major impact, not only on governments and businesses, but also on the world of scholarship. Academic writings now often refer to the 'datafication of the humanities' or the arrival of 'computational social science'.⁸

These developments certainly have relevance for the records discipline. Some records professionals, and some computing experts, have begun to look at digital records – and digitised versions of analogue records – from the specific perspective of data science, and see them as candidates for participation in computationally-based data analysis projects. Advocates of datafication argue that reconceptualising records as data – or perhaps transforming records into data – is moving us to 'a world in which ... the whole record can be mined and analysed'.⁹ They have

- 6 V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (London: John Murray, 2013).
- 7 K.N. Cukier and V. Mayer-Schönberger, 'The rise of big data: how it's changing the way we think about the world', *Foreign Affairs*, 92 (2013), <http://www.foreignaffairs.com/articles/2013-04-03/rise-big-data>.
- 8 T. Blanke and A. Prescott, 'Dealing with big data', in G. Griffin and M. Hayler (eds), *Research Methods for Reading Digital Data in the Digital Humanities* (Edinburgh: Edinburgh University Press, 2016), p. 190; R. Kitchin, 'Big data, new epistemologies and paradigm shifts', *Big Data & Society*, 1 (2014): 1–12, at p. 1.
- 9 S. Ranade, 'Traces through time: A probabilistic approach to connected archival data' (IEEE International Conference on Big Data (Big Data), Washington DC, 2016), <https://doi.ieeecomputersociety.org/10.1109/BigData.2016.7840983>, pp. 3260–3265.

also argued that, if this transition is to succeed, digitisation processes for paper and other analogue records will have to be restructured to generate computer-processable data rather than mere digital images. At the same time, we are told, creators of born-digital records should be persuaded to use analytics-friendly formats.

As yet, these changes have had little impact in most low-income countries, where the pace of technological development has been less rapid. At present, many low-income countries are still heavily dependent on paper records; even as digital applications are being introduced to support the current work of government agencies, these countries often have little capacity to manage records in digital form. Indeed, in many of these countries, there is no recognition that digital records are records or that recordkeeping principles should be applied to them. Issues such as these remain of primary concern today. In these circumstances, the so-called 'datafication' of records may appear to be a topic of interest only to wealthier countries and may seem to have little immediate relevance in poorer areas of the world.

Nevertheless, the changes that are under way in wealthier countries suggest that the need to translate data from records into structured databases is becoming outmoded in new environments where records are largely digital and analytical tools can be applied directly to them. Although at present the older models of using paper records as a source for digital data entry and of converting 'unstructured' records to 'structured' data still have validity in many low-income countries, in the future we can expect them to be superseded everywhere as new skills are developed and newer technologies become more widely available around the world. As notions of datafication become more widespread, it will become more apparent that we need no longer see records and data as two distinct kinds of entity; instead, datafication suggests that records themselves can be interpreted as data that we can mine, analyse, reuse and repurpose.

- JL: In my work with the international development community I've noticed that the concept of 'records' is often seen as quaint or irrelevant. Given that advocacy for recordkeeping requires us to speak the language of stakeholders, users, budget holders, etc., should we abandon the language of records in favour of the language of data?
- GY: Despite changes in technology, the challenges that Cain and Thurston identified 20 years ago haven't disappeared. It remains the case that, when one set of resources (let's call it 'A') is analysed or processed to create another resource ('B'), the utility of B always depends on the qualities of A, as well as on the processing methods employed. Regardless

of whether we want to label A and B as ‘data’ or ‘records’, the old adage ‘garbage in, garbage out’ still applies. And irrespective of whether we choose to speak of ‘data’ or ‘records’, the issues of ill-informed decision-making, misguided policies, misplaced funding and failed attempts at open government still arise, with all their consequences for the lives of citizens in lower-income countries. Similarly, in the context of the SDGs, we will find it impossible to measure whether the SDGs have been achieved if the resources for assessing their achievement are unavailable, inadequate or unreliable. In addressing these challenges, we undoubtedly need to find a language or languages to help us articulate our understandings and communicate our concerns and proposed solutions to other stakeholders. There seems to be a case for abandoning the kinds of distinction between records and data we have made in the past and seeing whether we can achieve more practical success if we frame our approaches in a different way.

Where communication with others is concerned, it may sometimes be appropriate, or more effective, to talk about data; talking about records does not always seem to have the same resonance. As I wrote in my book:

the language that now carries weight ... in the corridors of power is the language of data and information, and many records professionals ... feel a political imperative to adopt this language when they seek to convince resource allocators or government policy-makers that they can contribute to the 21st-century digital landscape.¹⁰

But this approach has its own difficulties and drawbacks. One difficulty is that some of the people with whom we speak are likely to assume that data are always created and maintained digitally, and that analogue records have altogether fallen out of the picture, which is certainly not the case. Another difficulty with speaking of records as ‘data’ is the widespread notion that data are simply ‘raw facts’ or ‘sources of truth’ and are wholly or largely independent of social and contextual influences. As I noted in my book, when we look at a database or dataset:

no one seems to be making statements; no one is affirming that they can vouch for the data; the apparent absence of signs of authorship gives the impression that the data are uncontroversial and objective.¹¹

But, of course, data aren’t autonomous, independently valid or context-free. They are always conditioned by the practices used to generate

10 Yeo, *Records, Information and Data*, p. 198.

11 Yeo, *Records, Information and Data*, p. 142.

them and the circumstances that led to their production. Data are rarely as uncomplicated as they seem. The ‘facts’ they present are propositions about the world or about actions or events: propositions stated by humans, or by computing devices programmed by humans, in particular contexts. And this means they can never be exempt from social constraints or from possibilities of error, ambiguity or bias.

So when I said that there seems to be a case for abandoning the kinds of distinction between records and data that we’ve made in the past, I wasn’t trying to suggest that records managers and archivists should forsake the concept of ‘records’ in favour of speaking and thinking only about ‘data’. On the contrary: I wanted to suggest that, rather than simply reconceptualising records as data, records managers and archivists – at least in their own professional discourse – might usefully be encouraged to understand data as records. Viewing records as data opens the way to employing powerful analytic tools, which will enable new modes of future investigation and research; but viewing data as records reminds us that data are shaped by their cultural contexts and that effective use and comprehension of them will only be possible if knowledge of their contexts is safeguarded.

Placing emphasis on records rather than data also reminds us that records do much more than communicate facts, or supposed facts, about the world. They also allow us to express ideas, opinions, emotions and predictions; to pose questions, issue orders, make promises or establish rights and responsibilities. In coordinating human behaviour and social relations, they are part of the way we conduct business and live our lives. In the digital world as much as in the analogue, records are more than ‘data’; they are instruments through which social actions are achieved.

Alan Bell¹² has written about the dangers that can arise when archivists and records managers choose to speak about *information* rather than records – particularly the dangers that they may be led to forget, disregard or even deny the importance of what David Bearman¹³ called ‘recordness’ – and it seems to me that the same dangers will arise if we are over-enthusiastic in adopting the vocabulary of data. Perhaps we can or should use this vocabulary when we think it is politically necessary, while remaining aware that it doesn’t offer us a solid base for reflective professional thinking about records and their keeping. But our professional leaders and professional

12 A.R. Bell, ‘Participation vs principle: does technological change marginalize recordkeeping theory?’, in C. Brown (ed.), *Archives and Recordkeeping: Theory into Practice* (London: Facet Publishing, 2014).

13 D. Bearman, *Electronic Evidence: Strategies for Managing Records in Contemporary Organizations* (Pittsburgh: Archives and Museum Informatics, 1994), p. 133.

associations also need to promote a campaign to overcome the idea that records are a 'quaint' legacy of an archaic paper-based world, and to reaffirm the continuing importance of records in the 21st century, both as instruments of current social action and as bulwarks that support our knowledge and understanding of past events.

JL: Could we agree with the many data scientists, analysts, journalists and others who see paper records as data?

GY: I think you are right when you say that analysts and commentators from many different backgrounds see paper records as data, but I'm not sure that they would all approach this question in the same way. Once again, much depends on what people mean when they speak of 'data'. Those who perceive data as essentially *structured* materials – the kinds of materials we typically find in relational database systems – should have little difficulty in recognising that, before the advent of digital technology, similar materials were created using paper records: ledgers, registers, card indexes and the like. It seems to me that, if we accept this premise, it requires no great conceptual leap to understand these paper artefacts as data, or at least to understand that they *hold* data.

Individuals who use records for purposes of academic scholarship might approach your question in another way. Many scholars – particularly in the field of history, but also in other disciplines – have long had a perception of records as data that can support their research. They use the word 'data' to refer to the materials they can employ to unravel a problem and reach conclusions to their investigations. Historians and other scholars who use the word 'data' in this sense do not seek to limit its scope to digital resources. Nor, I think, do they see data as resources that are necessarily or primarily in structured form. Some historians may see records as the only 'data' they need; others may say that records – whether digital or analogue – are simply one of many different kinds of data they employ in their research.

Despite the appropriation of the term 'data' by the computing industry during the past half-century, my own view is that it can still be applied to paper as much as to digital materials. But I also think that this is an area where we need to proceed cautiously. As I said, one of the risks that records professionals run in adopting the language of data is that other stakeholders in the workplace may assume that data are always digital. The broader scholarly view that 'data' can embrace many different media is not always acknowledged outside academic circles. Any expectation that everyone in government, in business or in the international development sphere will recognise analogue records as data is likely to give rise to misunderstandings and failures in communication.

JL: I think that data are the building blocks of records, whether the records are paper or digital. Does this differ from the ‘sequential’ view of records and data, or do you think they are much the same?

GY: The ‘building blocks’ view of records and data is not the same as the ‘sequential’ view, although both views have had distinguished advocates. Whether you accept the ‘building blocks’ view will partly depend on what you think we mean by data, and on the levels of granularity at which you believe data exist.

If you agree with the computer scientists who tell us that a single bit or byte is ‘the smallest unit of data a computer can handle’,¹⁴ then yes, you can reasonably claim that low-level data such as bits or bytes are not themselves records, but are ‘building blocks’ from which a digital record can be constructed. Similarly, in the paper world, a single pen-stroke, a single letter of the alphabet or a single word might perhaps be construed as ‘data’, but I think they cannot so easily be construed as records; pen-strokes, alphabetical and numerical characters and words are not records but building blocks of records.

But if you think that ‘data’ must refer to something less granular than a single bit, byte, character or word – if you think that data must be capable of conveying a greater degree of meaning – matters become a little more complicated. One widely held view of the term ‘data’ is that it generally refers to structured statements such as ‘President: Joe Bloggs’ or ‘number of widgets in stock: 39’. From a records perspective, statements of this kind can be seen in different ways. If we want, we can certainly see them as building blocks of records, but each of them can also be seen as a complete record in itself. If we are seeking a record and find an entry of this kind in a database, we are not obliged to look for further components; we have found a record of an assertion that Bloggs is the president, or an assertion that 39 widgets are in stock.

From a computer science perspective, too, entries like these can be seen in different ways. ‘Number of widgets: 39’ can be seen as data at one level, but it can also be seen as a building block of a larger set of data at a higher level. I think we should accept that data and records may both exist at different levels of granularity. Lower-level data can be (and often are) used to construct larger aggregations of data; lower-level records can be (and often are) used to construct larger aggregations of records. And perhaps this is little more than two ways of saying the same thing.

JL: I have argued that machines such as autonomous cars and autonomous weapons are record-making devices if they receive, transmit or store data

14 K.C. Laudon and J.P. Laudon, *Management Information Systems: Managing the Digital Firm*, 15th edn (Harlow: Pearson, 2018), p. 242.

with even basic contextual metadata, but this is rooted in the notion that records are data with metadata and structure. Some of the SDGs will depend on sensor data, both to be achieved and to be measured. How should we recognise records in devices and systems?

GY: I'm sure you are right that the data captured by so-called 'smart' or autonomous devices are records. They are records of the functioning of the device and of its sensing of the environment in which it operates. And, given the potential for these devices to act in ways that could have major consequences for human lives, it seems vital that such data should be recognised as records that may need to be retrieved and interpreted in future, and should be preserved and managed accordingly. As with so many initiatives in computer technology, there is a serious risk that the recordkeeping requirements will not be recognised by the developers of these devices or will only be recognised at a stage in their development when it is too late to implement them satisfactorily.

The only point where I might disagree with you concerns the need for separate contextual metadata. Distinctions between data and metadata aren't always clear-cut in data-centric environments; what person X thinks of as metadata may be perceived by person Y simply as further data that the device has captured. I'd argue that the data captured by these devices are records even if their metadata aren't separately identified. And that metadata, whether separately identified or not, are also records.

JL: You have written that 'record-making is always ... bound to contexts of social action'.¹⁵ It could be argued that although data collection takes place in social contexts, it is not necessarily bound to those contexts in the same way as it is for records, since bonds of this kind would require a persistent relationship to contextual metadata. You also wrote that the creation and transmission of records are 'not a matter of information, but a matter of social action'.¹⁶ Can this also be said about data? Could we summarise this line of thinking by saying that a record is data with metadata?

GY: Rather than arguing that records are data, I prefer to explore the idea that data are records. Some of our colleagues have claimed that only some data are records,¹⁷ but I'm increasingly inclined to the view that *all* data, if they persist in a stable form beyond their moment of creation, have record characteristics. In my book, I proposed a number of arguments

¹⁵ Yeo, *Records, Information and Data*, p. 129.

¹⁶ Yeo, *Records, Information and Data*, p. 152.

¹⁷ See, e.g., K. Anderson, 'The footprint and the stepping foot: archival records, evidence, and time', *Archival Science*, 13 (2013): 349–71, at p. 363; D. Hofman, L. Duranti and E. How, 'Trust in the balance: data protection laws as tools for privacy and security in the cloud', *Algorithms*, 10 (2017): 1–11, at p. 3.

in favour of seeing persistent data as records. Data are not context-free but arise from particular acts of statement-making or recording that take place at particular moments. Over time, they are also likely to be subject to interventions from their custodians or users, interventions that add to the richness of their contexts. Even if our knowledge of those contexts is imperfect or has been lost, the data are still shaped by the contexts in which they have been created and stored; the bond doesn't disappear simply because we have little or no knowledge of it.

What about the need for metadata? Well, some people say that 'if there are no metadata, it's not a record'; I think this may be what you are suggesting? Of course, contextual metadata are beneficial, because they help reduce the risk of total loss of contextual knowledge. And let's not forget that there are many other kinds of metadata that serve other equally useful purposes. However, I don't think it's as simple as that. Replying to your previous question, I said that distinctions between data and metadata aren't always clear-cut. In data-centric environments, it's not always necessary to identify metadata as a separate category; data in which assertions are made about context can be very useful even if they don't sit in a little box labelled 'metadata'. We can still have records even when their metadata aren't separately identified; we don't need to find the little box labelled 'contextual metadata' in order to know whether we are looking at a record.

Now I'd like to go further and suggest that we can encounter records even when assertions about their context seem wholly absent. A good example might be the 11th-century survey that we know as *Domesday Book*. Today, of course, it is surrounded by vast quantities of metadata; the book and its contexts have been described on countless occasions. But when it was compiled in 1086 it must have stood alone in glorious isolation, with no metadata and no written contextualisation of any kind. Its contexts were well-known to its users, and no one felt it necessary to inscribe them; arguments about the need for contextual metadata didn't arise in 11th-century England. Advocates of the mantra 'if there are no metadata, it's not a record' presumably have to believe that *Domesday Book* wasn't a record until somebody catalogued it, many years later. But I'm sure that, like me, you will find this absurd. The status of *Domesday Book* as a record – as one of the most valued records that survives from the Middle Ages – has nothing to do with its metadata. Of course, present-day records systems require metadata if they are to function effectively; metadata are far more necessary in a 21st-century era of record abundance than in an 11th-century era of record scarcity. Users of records can be seriously handicapped when metadata are missing or inadequate for their needs. Nevertheless, while the presence of metadata

is always a very good thing, it doth not a record make. Records are records even when the metadata we seek are lacking.

- JL: If data can be combined into records, and if configurations of these data can constitute evidence, do we need to revisit legal theory as a foundation for defining records? It seems that courts often consider many forms of data, information and records to be evidential, whether or not they meet archival standards of trustworthiness.
- GY: As I've said, I'm not sure that it's entirely helpful to talk about data being combined into records. There are other, and, I think, more fruitful, ways of looking at the relationships between data and records. But the question of whether legal theory provides a foundation for defining records is a separate issue; it doesn't depend on our understandings of data and their combination.

In the past, certainly, there has been a long tradition of seeing records in legal terms. When the Public Record Office was set up in London in 1838, its remit was limited to the records of courts of law; the writings of the administrative departments of government weren't formally deemed to be 'records', and the Office's responsibilities weren't extended to administrative writings until the 1850s. In England, the idea that records emanate only from law courts dates back to the early Middle Ages, when 'record' was a formal oral recollection of court proceedings. When oral methods of recalling judicial business were superseded by writing, the word 'record' was applied to their written successors, and definitions of 'record' that confined the record to legal settings persisted down to the 19th century. As early as the 17th century, however, the word 'record' was being used more widely outside the legal world; over time, it became increasingly common for people to speak of the 'records' of any institutional body, and – more recently – of the 'records' of families and individuals. I'm sure that, today, almost no one in England would be likely to restrict the word to the records of the law courts.

In continental European countries with systems of civil law, legal traditions are very different. In these countries, the word 'record' is largely unknown; lawyers, diplomatists and archivists in civil-law countries have traditionally spoken of 'documents', and the evidential function of documents has been analysed in jurisprudence and embedded in law over many centuries. In recent years, some records professionals have chosen to equate the civil-law 'archival document' with the English word 'record',¹⁸ although I think it's open to question whether this equation is fully correct. Of course, in English-speaking countries, the common

18 L. Duranti, *Diplomatics: New Uses for an Old Science* (Lanham: Scarecrow Press, 1998), p. 6.

law also recognises that records (whether emanating from law courts or from other places) have evidential aspects. Indeed, the word ‘evidence’ belongs to the English common-law tradition; civil lawyers have generally preferred to use words such as ‘proof’ and ‘authentication’. But the common-law view of the evidential aspects of records is perhaps less rigorous and systematic than the view you find in the traditions of continental Europe.

You ask how far I think legal theory might still provide a foundation for defining records. The first point I’ve tried to make is that we aren’t dealing with a single legal theory here. Civil-law ideas about ‘documents’ are different from common-law ideas about ‘records’; and there may be other legal traditions, such as sharia law, that could or should be taken into account. My second point would be to sound a note of caution about the idea of ‘defining records’: we can offer definitions that help us examine a range of different perceptions and understandings of records, but I don’t believe that we will ever be able to construct an incontrovertible statement of ‘what a record is’.

Having said this, I think it is important to acknowledge that legal theories have been a major force shaping people’s understandings of records in the past; and the ways in which we understand records today can’t be wholly independent of the understandings we have inherited from earlier generations. And, of course, legal aspects of record-making and recordkeeping still influence our work today. We can see this, for example, in the work of national standards bodies on records’ legal admissibility and evidential weight. But today we recognise, or should recognise, that the role of records is not limited to the provision of evidence. We also recognise that the evidential role of records isn’t confined to legal circles: auditors, journalists, historians and many other users may see records wholly or partly in evidential terms. The legal aspect of making and keeping records is certainly a part of the mix, but it is not the only part – and not even the most significant part, in my view.

- JL: You and others have argued that the evidential paradigm should not dominate recordkeeping theory. You have said that, in addition to evidence, records can offer other benefits including memory and senses of individual and communal identity. What are the important qualities of records if we start from a position where records are testimony of the personal or cultural?
- GY: I didn’t mean to suggest that evidence is unimportant. If they had no means of uncovering evidence, institutions that seek to promote justice and accountability would be unable to function or would

find their functioning severely impaired; individuals would often be unable to assert their rights against powerful vested interests. On some occasions, human witnesses can supply evidence when it is needed; on other occasions, especially when human witnesses are unavailable or untrustworthy, institutions or individuals rely on records to obtain the evidence they require.

But *all* the benefits – I call them ‘affordances’¹⁹ – that records offer can be important to those who rely on them. Consider, for example, the role of records in supporting memory. Human memory, we know, is fallible, and many people depend on records to redress its failings. Some people may claim that they live only for the present or the future and that memories of the past are unimportant to them, but others affirm that their lives would be empty and meaningless without such memories. Information, too, is an affordance of records, and different people and different cultures around the world will assign different values to affordances such as evidence, information and memory. Some will find affordances that others don’t recognise.

I was intrigued by your use of the word ‘testimony’ in the last part of this question. It’s a word that records professionals don’t use as often as one might expect. I like it because it carries resonances of people who say ‘I can tell you about it because I was there. I saw what happened with my own eyes’. Creation of records implies direct participation in, or first-hand knowledge of, an action or event. I concede that it’s possible to find examples of records created by people who don’t have such immediate knowledge: the official record of a birth, for instance, is made by a registrar who was not present when the baby was born, but who relies on statements made by others with first-hand knowledge of the birth. However, I’d argue that, for most people, a key aspect of what we think of as records is that their creators participated in the actions or events they represent or were able to call on first-hand knowledge of them.

An account of events written by someone without first-hand knowledge (such as a school textbook on medieval history) can be valuable in its own right, but we don’t usually think of it as a ‘record’ (or ‘testimony’) of the events that the author has written about. We could perhaps say that one of the ‘important qualities’ of records – I’d prefer to say one of the qualities that people tend to look for in records – is

19 For the concept of ‘affordance’, see O. Volkoff and D.M. Strong, ‘Affordance theory and how to use it in IS research’, in R.D. Galliers and M.-K. Stein (eds), *The Routledge Companion to Management Information Systems* (Abingdon: Routledge, 2018), pp. 232–45.

that they were created by someone closely connected to the matters they represent. Or by a mechanical device with a similarly close connection; the sensing devices you asked me about earlier offer a useful reminder that records in today's world don't have to be created by humans.

Nevertheless, I have a couple of reasons for being cautious about describing records as 'testimony'. First, because the word 'testimony' is closely associated with ideas about 'witnessing', many people will naturally associate it with the role of witnesses in a court of law. Although this isn't the only sense in which we can speak of 'testimony', I feel that the word can't be wholly disengaged from ideas about evidence, and particularly legal evidence presented in court. Yet evidence, as I've said, is only one among many affordances that records offer. When I speak of the connection between records and actions or events, I choose not to describe records as 'testimony'; I prefer to say that they are *representations* of actions or events, created by people who participate in, or have close knowledge of, the actions or events concerned. This terminology, I think, gives no primacy to evidence. Of course, no choice of terminology can be wholly neutral, but speaking of 'representation' seems less weighted in this regard than speaking of 'testimony'.

Second, I think the word 'testimony' always seems to bear connotations of looking back to some action or event that took place in the past: an action or event that is separate or distinct from the 'testimony' that now tells us about it. From the perspective of users consulting records made at an earlier date, this is indeed what records do: they tell us about things that occurred in some other time or place. But at the moment of their creation, records don't merely provide a retrospect on previous actions; the issuance of a record performs an action in itself. To create a representation is to perform an action, and we can also perform many other kinds of action – we can make statements, ask questions, give orders or enter into commitments – by creating and communicating representations of them. The creation of records always has a role in social action; it is always performative; and it is the performativity of records that gives them much of their authoritativeness and their power. 'Testimony' is a valuable concept, but I'd be reluctant to say that 'records are testimony ...' is 'a position we start from'.

- JL: Given that the SDGs are targets for action in many countries with widely differing circumstances, do definitions of records need to be specific to the contexts of the records' creation or use?
- GY: I'm not sure that they *need* to be, but I think that in practice they very often are specific to those contexts. Whenever we construct definitions

of records, we think of records in particular ways, and those ways of thinking are always conditioned by our own circumstances.

Consider, for example, how records are defined in the international standard for records management, ISO 15489: they are said to be ‘information created, received and maintained as evidence and as an asset by an organization or person, in pursuance of legal obligations or in the transaction of business’.²⁰ Ostensibly, this is an all-purpose definition that embraces personal as well as organisational records; although the standard is primarily for organisational use, the authors of the definition took care to state that records could be created, received and maintained by individual ‘persons’ as well as organisations. Nevertheless, their reference to the role of records ‘in the transaction of business’ might be thought to betray an organisational bias; their reference to keeping records ‘as an asset’ was undoubtedly influenced by contemporary ideas about the management of corporate ‘information assets’. The definition almost certainly would not have employed this terminology if it had been written by a keeper or user of personal records, or if it had been written at any time before the late 20th or 21st century. The definition is not universal, but was moulded by its authors’ circumstances, which led them to think of records in a particular way.

- JL: Looking ahead, I think that recordkeeping – or archival science – will become a specialisation within data science, or within computer science more generally. Do you agree?
- GY: No. Undoubtedly, the great majority of records in the foreseeable future will be created and maintained in digital form, and the practical tools we will use to maintain them will be designed using the techniques of computer science. The sheer volume of digital records will make it – is already making it – impossible for records managers and archivists to scale up their traditional manual methods of working, which will have to be replaced by automated processes. The use of computational techniques and artificial intelligence in areas such as description, preservation and access will become an essential part of working life for every records professional. But I don’t believe that archival science as a discipline will be subsumed into computer science. Archival science has concerns for the distinctive societal roles of records and archives, concerns that data science and computer science do not share.

Archival science also embraces – and must continue to embrace – the legacy of many centuries of records created using paper and other analogue media. Human needs for records antedate the invention of

20 ISO 15489-1: 2016, Records Management. Part 1: Concepts and Principles, clause 3.14.

writing, and have endured for about 10,000 years across many shifts in technology; the interests of archival science are not confined to digital records, which are a product only of the last half-century. While the challenges and opportunities of digital technologies increasingly occupy the centre of the stage, I am confident that archival science will remain a distinct discipline concerned with understanding, evaluating and managing the records created in the past by non-digital means, as well as the records created digitally in the present and future.

JL: Although you clearly want to differentiate recordkeeping from data science and computer science, I sense that you are very reluctant to provide conclusive or universal definitions of terms such as ‘records’ and ‘data’.

GY: You’re right; this is not a task I would want to attempt. ‘Records’ and ‘data’ are words that can bear a wide variety of meanings and interpretations, both within and across disciplines, and I believe it would be inappropriate to try to impose a single definition of either term. In my book, I argued in favour of a way of looking at records as persistent representations of actions and events: this is a way of looking that I personally have found very helpful. Although I’d prefer not to label this view of records as a ‘definition’, many commentators – perhaps inevitably – have chosen to refer to it as ‘Yeo’s definition of records’. Regardless of how it is labelled, others will be welcome to use, or adapt, it if they find it beneficial to their own thinking, research or practice. But I certainly wouldn’t want to suggest that my way of looking at records is the only possible or only acceptable way; definitions of records remain a moving target.

In countries where the SDGs are objectives for strategic action, individuals and communities will undoubtedly have varied assumptions, ideas and beliefs about the scope of data and records, their interrelationships and their roles in sustainable development. Individual contributors to this volume come from many different disciplines and will also have different conceptual understandings of records and data. Yet I’m sure you’ll agree that collaborative working will be essential if we are to move forward on the issues and concerns expressed in their contributions. If the chapters of this book help different stakeholders to recognise and understand the diverse viewpoints of others with whom they seek to collaborate, they will play a very valuable part in cross-disciplinary communication and cooperation.

4. The potential – constructive and destructive – of information technology for records management: case studies from India

James Manor

Twenty years ago, it was sometimes difficult to explain the importance of improving records management. Sceptics wondered whether it was worthwhile sorting through floor-to-ceiling stacks of old government files in musty Asian or African offices. They had to be persuaded that orderly sets of records could restore the memories of governments. However, effective records management facilitates government driven by precedent and not by the whims of politicians and bureaucrats. It promotes the rule of law. It prepares the ground for greater government transparency and accountability and for efforts to curb corruption. It enables governments – and citizens – to see which earlier and current policies and programmes worked, and which have misfired.

Moreover, times have changed. The old arguments about the virtues of managing paper-based records management are still valid, and it is still worth wading through heaps of old files, but remarkable advances in information technology (IT) have inspired euphoria about the immense promise of digital records, and many political leaders have embraced IT as a potent weapon in their efforts to promote development programmes and prevent corrupt bureaucrats and middlemen from plundering them. But so have politicians who seek tight top-down control of governments and populations.

Amid the excitement about IT, a reality check is in order. It is possible because solid evidence abounds demonstrating that well-managed digital records can contribute mightily to constructive programmes but that, if mismanaged, digital records can be the source of grave damage.

This chapter considers both phenomena by examining two gargantuan government initiatives in India. First, it briefly summarises key findings from a major study of the immense, positive contribution of well-managed digital records in the world's largest poverty programme, the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA). Then, at greater length, it presents evidence from a diversity of reports on Aadhaar, a universal identification programme in which a government's objectionable aims, haste

and bullying have caused hardships, hunger and even deaths among many extremely poor people, compounded by poor digital records management.

The Mahatma Gandhi National Rural Employment Guarantee Act

Early in its decade in power (2004–14), the multiparty United Progressive Alliance led by the Congress Party passed the National Rural Employment Guarantee Act.¹ It gave every household in rural areas – where two-thirds of Indians live – the right to demand and receive up to 100 days employment per year, doing manual labour on public works sites at a reasonable wage. It was intended as a hedge against destitution, and the response was massive. At this writing in February 2018, 24.85 billion ‘person days’ have been worked under this programme. On 31 January 2018, for example, labour was performed on 947,824 sites.² It is thus the largest poverty reduction programme in history. Since this initiative has been exhaustively analysed elsewhere,³ what follows is a concise summary of the main relevant points.

The wages earned have made an immense material difference to poor families. One day’s earnings enable a family to purchase enough subsidised food to provide two decent meals per day for a month. Wages have also been spent on health care, education, livestock, small plots of land and microenterprises – all of which have eroded bonds of dependency on richer villagers and the resultant subordination of poorer people. Since many key decisions about the programme are made by elected local councils – which receive at least half of its substantial funds – poor people have become more active in the public sphere. That has enhanced their political awareness, confidence, skills and connections – which add up to greater ‘political capacity’ to defend their interests. Many women – who have performed most of the work – have been drawn into the public sphere for the first time. And since they receive wages directly to their bank accounts, their influence within and beyond their households has been bolstered.

The MGNREGA was shrewdly and effectively formulated by people with a realistic, sophisticated understanding of local power dynamics and of how pilfering has occurred in other government programmes. A formidable IT system

- 1 Mahatma Gandhi’s name was added after that government was reelected in 2009, in an effort to stress the Act’s non-partisan character.
- 2 These and further details are available on the programme’s website: <http://mnregaweb4.nic.in>.
- 3 R. Jenkins and J. Manor, *Politics and the Right to Work: India’s Mahatma Gandhi National Rural Employment Guarantee Act* (New Delhi/London/New York: Orient BlackSwan/Hurst/Oxford University Press, 2017), which provides detailed explanations of the points in the text that follow.

records every event in each worker's engagement with the programme: the date each demanded employment, the date on which it was first provided, compensation due to each worker for delayed provision of work, each day's work performed, the wages due to each worker, and so on. Each worker is also given a 'job card', a small booklet on which the same information is recorded manually. It can be checked against the electronic record to detect efforts to cheat workers. The IT system is thus a potent transparency mechanism, and it is supplemented by other devices to promote transparency.

Enlightened, dynamic civil servants in New Delhi took steps to overcome logistical impediments such as poor internet connectivity in villages – problems that have greatly damaged the second programme analysed below, Aadhaar. For example, they overcame the problem of poor internet connectivity and power supplies in many villages by mandating the transfer of handwritten local records to IT centres for insertion into the electronic record. They also interrogated field researchers in order to discover flaws in the programme, and when these became apparent, they took corrective action.⁴ As a consequence, it became more difficult to steal funds from the MGNREGA than from almost any other Indian government programme.

Thanks in substantial measure to the well-managed use of IT in this enormous exercise in records management, this demand-driven programme has been immensely constructive in reducing poverty, promoting rights and village-level democracy, and undermining invidious local power dynamics across India.

Aadhaar

During the national election campaign in 2014, opposition leader Narendra Modi denounced Aadhaar as a 'political gimmick', but when he became prime minister in May of that year, he soon warmed to it. He is an enthusiast for technocratic approaches to government. He is also genuinely interested in reducing corruption.⁵ He regarded Aadhaar as a device to curb spending on poverty programmes that had soared under the previous government, by reducing the illicit diversion resources from them – in order to redirect funds to investments in infrastructure and other spheres that might accelerate economic growth. And crucially, as he radically centralised power in the prime minister's office, he also saw Aadhaar as a means of achieving greater top-down control.

He wanted to show that he – unlike his staid predecessor – was a bold man of action. So, he pursued this initiative (and many others) particularly aggressively.

4 This writer has direct experience of these practices.

5 His efforts to tackle it have produced ambiguous results. See J. Manor, 'Modi stuck between two promises', *Nikkei Asian Review*, 27 July 2015.

To maximise the impact of Aadhaar, he sought to make it mandatory for all manner of things: access to subsidised food and fuels; pensions for vulnerable groups; work opportunities under the aforementioned MGNREGA; school enrolments and free midday meals for school children; mobile telephones; bank accounts; income tax payments, and much more.

The breadth of its coverage, the government's refusal to make it voluntary, and the hasty, forceful pursuit of universal enrolment have raised serious concerns. We need to consider anxieties about possible leaks and the illicit penetration of the Aadhaar system; the government's coercive approach; malfunctions and denials of services which have caused acute hardships; and the opportunities that it has created for new kinds of corruption.

Leaks and the system's vulnerability to penetration

Worries about the security of Aadhaar data have long been expressed, and in 2017 they intensified when it emerged that 'a simple Google search' could give access to 'thousands of databases, that contain Aadhaar numbers with other sundry personal data'. Another report listed nine government departments and educational institutions that had leaked data.⁶

In March 2017, India's IT minister (echoing the prime minister) assured parliament that 'there is no leak' and that Aadhaar data are secure, but his statement was wildly inaccurate. In the three preceding months, at least 13 leaks had occurred. They included cases that affected (respectively) 12,000, 30,000, 500,000–600,000, and one million people. Twelve of those leaks were the fault of central or state government agencies or organisations associated with them.

The authority that oversees Aadhaar has exclusive powers to take legal action when such breaches occur. The Aadhaar Act⁷ explicitly denies citizens the right to sue for damages as a result of leaks. If a fraudster uses an individual's details to steal government benefits, the authorities are not obligated to inform the injured party – and often fail to do so. If citizens learn of this, no procedure exists for them to obtain new numbers. Most attempts to use India's strong Right to Information Act to uncover breaches in the Aadhaar system and the defrauding of individuals have been rejected – on the basis of a section of that Act which refers to 'national security'.⁸

The Aadhaar Authority has initiated criminal proceedings against numerous private parties for leaks, but very few against government officials or agencies.⁹ This is a serious omission. Chandigarh's Department of Food and Civil Supplies

6 TECH2, 24 March 2017 and Trak.in, 24 March 2017.

7 Its official title is the 'Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits and Services) Act, 2016'.

8 Scroll.in, 5 March 2017. The section in question is 33.2.

9 Medianama, 24 April 2017.

reportedly publicised the numbers of 490,000 Public Distribution System beneficiaries, quite vulnerable people. In Jharkhand state, personal details of 150,000 pensioners appeared on the Directorate of Social Security's website. Another government website leaked data on 500,000 minors. A programme officer at the Aadhaar Authority said that several banks (some of them state-owned) were displaying customers' data, which might enable someone to drain their accounts. Most spectacularly, he also found that India's Ministry of Rural Development website revealed the Aadhaar details of 100 million MGNREGA workers – again, extremely poor people. Eventually, 210 government websites were found to have displayed 'the list of beneficiaries along with their name, address, and other details and Aadhaar numbers'. When this was discovered, they were taken down, but they had been up for some time.¹⁰

There are good reasons for the authority to worry about abuses by private sector companies and banks, many of which saw opportunities for exploitation. Even before the Modi government's Aadhaar law came into effect, a private company was advertising its capacity to use it 'to verify your maid, driver, electrician, tutor ... and everyone else instantly'. As the Bill was being discussed in parliament, the government revealed that 'just about any person or company can draw on the Aadhaar system for its purposes. There are no qualifications on who may use it and why'.¹¹ This soon led the authority to investigate three firms – including a major bank, Axis – 'for attempting unauthorised authentication and impersonation' using stored Aadhaar biometrics.¹² A technology lawyer described parallel databases constructed by private firms like Jio, a formidable player, as 'a goldmine for identity theft and fraud'.¹³

As late as December 2017, the authority was claiming that their database had never been hacked. But then it emerged that Airtel, a major telephone company, had compromised the system by misusing data. It took advantage of a government requirement that all SIM cards had to be linked to Aadhaar by 31 March 2018 – part of its excessively aggressive effort to force mass enrolment (see below). Airtel was opening bank accounts for users without their consent. That posed a danger that funds due to them under government programmes would go to these accounts so that genuine beneficiaries would never receive them.¹⁴

The system is clearly vulnerable to criminals. A gang of ten in Kanpur was arrested after the authority discovered that they had cloned Aadhaar client applications which supposedly can only be filed at an authorised centre. They obtained the centre's employees' fingerprints (needed to log in) by using 'butter

10 Scroll.in, 25 April 2017 and 19 November 2017.

11 Scroll.in, 16 March 2016. See also, Scroll.in, 22 December 2016.

12 *Times of India*, 24 February 2017.

13 Scroll.in, 21 July 2017.

14 Scroll.in, 19 December 2017.

paper' and sold replica application forms for Rs 5,000 (US\$78.25) each. They also obtained the system's source code and tampered with it to enable them to bypass other biometric protections like iris recognition. Some employees of Aadhaar centres were also fraudsters. The Authority 'blacklisted 49,000 operators for corrupt practices'.¹⁵

It has also sought to hush up leaks and misdeeds – vowing to 'take action against an individual for reporting a security vulnerability in Aadhaar'.¹⁶ As one commentator complained, it 'should be rewarding those who find breaches – instead, we have attempts to intimidate them into silence through the abuse of the state's police powers'. The *Tribune* newspaper paid a mere Rs 500 (US\$7.83) to an agent who created a gateway with a login and a password and was thus able to get the full details of each of the one *billion* people enrolled. For a further Rs 300 (US\$4.70), he could print official cards for any number of them, including photos and addresses. For another Rs 300, he supplied the newspaper with the tools to do likewise. The Aadhaar Authority initially responded by insisting – incorrectly – that no breach had been achieved. Then, instead of conducting an internal investigation, it initiated criminal proceedings against the reporter who wrote the story. The Editors' Guild of India condemned this as an attempt to 'browbeat' the journalist.¹⁷ After a complaint from the authority, New Delhi police registered a criminal case against a man merely for saying on social media that the 'Aadhaar ecosystem is flawed, vulnerable, has very poor security, and can be easily hacked'.¹⁸

Coercive action by a government in a hurry

Prime Minister Modi gives the impression he would like to be perceived as a dynamic leader who lets little stand in the way of the changes he forces through. That is how he has managed Aadhaar. His enabling law for the programme was passed by the lower house of parliament in just three hours, and all amendments were rejected. Then the speaker controversially designated it a money bill so that the upper house could not amend it. An attempt was made to revise a clause that allows any individual or private or public organisation to use Aadhaar data. Members – anticipating the problems noted above – tried to restrict this to government agencies, but all proposals from the upper house were rejected.¹⁹

Thereafter, Modi undertook a forceful drive to maximise enrolment. The government's main tactic was to require citizens to register in order to obtain benefits from welfare programmes. This occurred despite two 2015 Supreme

15 Scroll.in, 14 September 2017.

16 Scroll.in, 5 March 2017.

17 *The Tribune*, 4 January 2018 and Scroll.in, 7 January 2017.

18 *The Asian Age*, 28 February 2017.

19 Scroll.in, 24 March 2016.

Court rulings – which in India are ‘law declared’, they have the force of law – that Aadhaar could not be made mandatory. In one case, the government asked the Court to require it for 88 social welfare programmes, but justices made it voluntary for only four.²⁰ In August 2015, the Court directed that a major publicity campaign be launched to explain that Aadhaar ‘is not mandatory’. It was ignored.²¹ In March 2016, the Court reiterated that officials could not make it mandatory for any subsidies, benefits or services, and stressed the point again in a ruling in October. In July, it sent letters to all state governments in this federal system – which were under heavy pressure from the national level to make Aadhaar mandatory – reminding them of its earlier judgments, and it re-emphasised the point in October. But those governments pressed ahead regardless, on many fronts.²²

In March 2016, parliament required enrolment for an array of government subsidies and benefits.²³ By early 2017, many schools – for example, 2,700 in Delhi – began demanding that both parents and prospective students have Aadhaar ID numbers and that students must open bank accounts before they could gain admission and receive educational benefits. That enabled schools to exclude slum dwelling and migrant children. In March 2017, the central government took a further, gravely damaging step: insisting that children would not receive free midday meals – on which a vast number of hungry families depend since pupils can take food home from school – unless they presented Aadhaar numbers.²⁴ This posed a grievous threat since Aadhaar enrolment among destitute families was very low – for example, only 17 per cent in Meerut District of Uttar Pradesh. Officials overseeing midday meals were exasperated but said that they were helpless to resist.²⁵

The government then announced that pregnant women would not receive maternity entitlements at state-run health centres unless they were registered with Aadhaar. Then, in its headlong drive for enrolment, the government began requiring newborn infants ‘within minutes of birth’ to sign on. In some cases, officials made this mandatory before issuing birth certificates.²⁶ The government also required Aadhaar certification for 11 other welfare programmes, including the Public Distribution System that provides subsidised food to over 800 million poor people, subsidised gas for homes, the National

20 Scroll.in, 22 October 2015.

21 Scroll.in, 9 April 2017.

22 On at least one occasion, the government also misrepresented a Supreme Court ruling. It made a serious error in allowing banks and telecom companies to threaten customers into enrolling with Aadhaar. Scroll.in, 15 November 2017.

23 Scroll.in, 8 September 2016.

24 Scroll.in, 11 January and 16 and 23 April 2017.

25 *The Hindu*, 1 June 2017.

26 Scroll.in, 8 March, 1 May and 19 September 2017.

Social Assistance Programme, the massive MGNREGA and a programme for disabled people.

Education activists stressed that making Aadhaar mandatory for access to schools was a violation of the Right to Education Act. Others referred to Supreme Court orders covering many services. But the drive for enrolment continued apace. A legal scholar, Usha Ramanathan, stated that ‘They are making it clearer and clearer that the Unique Identification project is not about including or reaching one’s entitlements, but coercion and exclusion’.²⁷

The government responded to such concerns with blatant falsehoods. On 21 December 2017, an official reply to a question in the upper house of parliament stated that it was ‘not mandatory for a beneficiary to avail subsidies, benefits or services’.²⁸

Given the risks to destitute human beings and, not incidentally, to the government’s popularity, one commentator asked, ‘What is the extreme urgency?’ The rush was apparent from the silence of senior officials in state-level food departments who saw the damage being done but felt powerless, as a result of ‘tremendous pressure’ from New Delhi.²⁹ The commentator then asked: ‘Why the sneaky tactics?’ The latter included – again in violation of Supreme Court rulings – a last-minute amendment to the Finance Bill 2017, making Aadhaar mandatory for filing taxes. The commentator concluded that by requiring enrolment in so many areas, the government sought to force the Supreme Court to permit the mandatory use of Aadhaar because the system would become too big to fail, and to undo.³⁰

‘Inhuman and illegal’:³¹ malfunctions and denials of services cause hardships

The aggressive pursuit of mass enrolment has not been matched by efforts to simplify the process of ‘seeding’ – that is, linking the system with documents that demonstrate a person’s eligibility for services and benefits. Studies by the eminent economist Jean Dreze have shown the process to be ‘cumbersome’ and prone to error – for example, if a person’s name is spelled slightly differently in different databases. Even middle-class people have struggled with seeding when they tried to link Aadhaar numbers with numbers used to pay taxes. The vast majority of poor people – many of them illiterate and nearly all of them

27 Scroll.in, 5 March 2017.

28 Government of India, Ministry of Social Justice and Empowerment, Rajya Sabha, starred question no. *75.

29 *The Hindu*, 17 January 2018.

30 Scroll.in, 22 March 2017.

31 The distinguished economist Jean Dreze used these words to describe the mandatory imposition of Aadhaar, in *The Hindu*, 17 January 2018.

unskilled in navigating bureaucratic minefields – have found the process too complex and thus have been excluded.³²

To make matters worse, glitches in computer networks, inadequate training for those implementing the process, poor connectivity to internet signals, power failures (a particularly serious problem)³³ and more have prevented many from achieving ‘seeding’. Even people whose data have been ‘seeded’ have found that these problems stopped them from obtaining vital services and resources. The machines intended to identify beneficiaries by fingerprint – or much more unusually iris – recognition often fail to work, leading to people being turned away. The gnarled, cracked fingerprints of manual labourers and the elderly – who are often the people in the most urgent need – have been especially hard to recognise.³⁴ Even Nandan Nilekani, an iconic figure in India’s IT sector who stoutly defended Aadhaar during a long spell as head of its authority, eventually conceded that action was needed to address these problems.³⁵

As an exercise in records management, the Aadhaar system has been a woeful disappointment. It is expected to run before it can walk, but the feverish drive for universal implementation rolls on unabated. Shopkeepers who take pity on poor people who are excluded by malfunctions in the system, by using manual overrides, face retribution from higher officials. The grandson of one in Rajasthan said, ‘The signal is patchy and (when they get the signal) the internet stops working’. But when overrides were used, officials issued a ‘show cause’ notice, threatening punishment.³⁶

The prime minister’s pressure for quantifiable results has persuaded or forced some bureaucrats into drastic actions that have caused immense suffering. For example, the head of the Jharkhand state civil service claimed that ration cards and MGNREGA job cards that had not been successfully seeded were ‘fakes’ – even though the complexity and unreliability of the Aadhaar system caused much and probably most of this problem. He therefore barred those holding ‘fakes’ from receiving such benefits as subsidised food, pensions and work

32 *Indian Express*, 21 November 2017. See also, J. Dreze, N. Khalid, R. Khera and A. Somanchi, ‘Aadhaar and food security in Jharkhand: pain without gain?’, *Economic and Political Weekly*, 16 December 2017, pp. 50–9; and R. Khera, ‘Impact of Aadhaar on welfare programmes’, *Economic and Political Weekly*, 52 (2017): 61–70.

33 This is an extremely serious problem. Consider, for example, that at the 2014 election in the state of Odisha (Orissa), over half of the polling stations (school buildings, village halls, government offices, and so on) had no electricity supply. In underdeveloped Kalahandi District, the figure was over 97 per cent. J. Manor, ‘An Odisha landslide buries both national parties: assessing the state and parliamentary elections of 2014’, *Contemporary South Asia*, 23 (2015): 198–210.

34 Scroll.in, 15 November 2017.

35 NDTV report, 28 January 2018. (Full disclosure: in 2008, this writer briefly advised Nilekani on a book that he was writing.)

36 Scroll.in, 10 April 2016.

opportunities under MGNREGA. That caused ghastly suffering among huge numbers of poor people.³⁷

Prime Minister Modi has celebrated such ‘discoveries’ of ‘fakes’ as victories in the fight against corruption. In February 2017, he proudly told parliament that major gains had been made against ‘bogus’ ration cards. One of his ministers then informed the house that 700,000 ‘fake’ cards had been discovered in the state of Odisha. But an official reply by that state’s food department to a Right to Information petition indicated this was not the case. If they were excluded by the Aadhaar system, which seems very likely, another gross injustice had been done.

In a chapter of this length, it is only possible to refer to sources on the epidemic of glitches and malfunctions.³⁸ But to grasp the appalling scale of these problems, consider evidence provided from official sources in just two of India’s 29 states. In Andhra Pradesh, a state government study of access to subsidised grain found that ‘technical hiccups are depriving the poor of their access to food’. The problems identified included ‘glitches, lack of training (of staff in fair price shops in how to operate the point of sale devices) and mismatches’. As a result, 58.6 per cent of beneficiaries ‘couldn’t access their ration quota. The scale of the problems is astounding’.³⁹ Similarly, in Rajasthan, a government official stated that machines used to confirm beneficiaries’ eligibility worked for only 45 per cent – excluding the rest.⁴⁰ Many millions were excluded in these two states, and vastly more suffered across the rest of India.

The results of all this are alarming. Five people who were denied food from the Public Distribution System starved to death in Jharkhand state. The government responded by announcing that enrolment would not be mandatory for access to food, but in many parts of the country, nothing actually changed.⁴¹ An eminent analyst reported that in Rajasthan, people were ‘facing hunger and starvation’.⁴² Three deaths from starvation also occurred in Karnataka, where people denied rations were eating roots and leaves.⁴³

Officials’ reactions to these cases were astonishing. In Jharkhand, they claimed that one of the five deaths – of a young girl – was the result of malaria, even though ‘right to food’ campaigners found that she had not eaten for eight days. Her mother was pressed to say that malaria was the cause. Her refusal

37 *Indian Express*, 21 November 2017.

38 See e.g., Scroll.in, 22 October 2015; 10 April and 1 June 2016; 8 March, 1 May and 15 November 2017.

39 *Hindustan Times*, 7 October 2017.

40 Scroll.in, 1 June 2016. See also, on Jharkhand, Scroll.in, 2 February 2017 and *The Hindu*, 17 January 2018.

41 Scroll.in, 2 January 2018.

42 This was Nikhil Dey, Scroll.in, 1 June 2016.

43 *The Hindu*, 10 January 2018.

triggered a crass campaign of victim blaming. The state chief minister claimed that by saying her daughter had starved to death, she had given the state a bad name. Others claimed that she had brought shame to the nation. Her family was subjected to a social boycott. No one would employ them or sell them anything, and she was heckled when she appeared in public.⁴⁴

Other officials also behaved outrageously. After the three deaths from starvation in Karnataka, civil servants in the district where they had occurred, facing heavy pressure to push ahead with Aadhaar, cancelled 40,000 ration cards, claiming that they were 'fakes'⁴⁵ even though they were mainly the result of failures in the Aadhaar process.

Reports of deaths from starvation are alarming, but a similarly ghastly outcome has been the malnutrition suffered by vast numbers of poor families. These are people who have been denied subsidised food, and/or whose children have been barred from school enrolments or, for those who attend school, from midday meals there. Even before Aadhaar, malnutrition was at appalling levels. India ranks 100th out of 119 countries on one hunger index⁴⁶ – below Myanmar, Nepal and Bangladesh. Denials of subsidised food grains under Aadhaar have compounded this ghastly problem.

The most shocking aspect of this is its impact upon children. Chronic malnutrition has caused 38.7 per cent of Indian children under five to suffer from stunting (low height-for-age) – a higher proportion than in most sub-Saharan African countries. If they suffer malnutrition during their first two years, their bodies and brains do not develop fully. The damage is *irreversible*. This means, to put it bluntly, that over a third of Indian children are less intelligent than they might be.⁴⁷ That is both a grotesque injustice and – to put it in terms that might attract Modi's attention – the squandering of a major national resource. Here again, Aadhaar has made things worse. It did so not just by denying beneficiaries food, but also by denying young mothers access to ante- and post-natal care if they had not successfully enrolled.⁴⁸

Hundreds of leprosy patients in Andhra Pradesh have been impeded by Aadhaar from obtaining rations and modest government pensions. One who had no fingers had her iris scan rejected. Others who still have fingers failed fingerprint tests.⁴⁹ Across India, many HIV patients are dropping out

44 Scroll.in, 16 October 2017.

45 Scroll.in, 20 October 2017.

46 The figure is from Global Hunger India, *Indian Express*, 11 January 2018. See also N. Choudhary, 'India's slip on global hunger index', *Economic and Political Weekly*, 52 (2017): 23–5.

47 For details, see D. Maiorano and J. Manor, 'Poverty reduction, inequalities and human development in the BRICS: policies and outcomes', *Commonwealth and Comparative Politics*, 55 (2017): 278–302.

48 Scroll.in, 8 March 2017.

49 *The Wire*, 26 December 2017.

of programmes supplying life-prolonging medicines because they fear that the system may leak information about their condition – leading to shaming and social exclusion.⁵⁰ The government claims that Aadhaar enrolment is not required in antiretroviral therapy centres, but in practice many demand it.⁵¹ In the drive to reduce ‘fake’ documents, many poor villagers have been denied the right to work under the MGNREGA – severing a lifeline for destitute people. And many who have performed manual labour under it have not received wages due to them as a result of Aadhaar malfunctions.

The Supreme Court has asked how homeless people – of whom there are 1.77 million in Uttar Pradesh state alone – will access Aadhaar since they cannot provide proof of address. One justice asked ‘Does this mean that they do not exist for the Government of India?’ The Aadhaar Act clearly states that an address is not required as proof of identity, but in practice, it is often demanded. The government has been reluctant to assist the homeless by clarifying this, because it might slow down the pace of enrolment.⁵²

An immense number of vulnerable people – the elderly, widows, the disabled – depend on modest government pensions. Many have had nightmarish experiences. In Rajasthan the government boasted of a major saving in expenditures after stopping pensions for a vast number of pensioners who could not achieve enrolment. Some who succeeded in linking Aadhaar numbers to bank account numbers (as is required) found that coding errors sent their pensions to others. Many were declared dead although they were very much alive. An investigation by a leading newspaper found that over a third of the state’s 297,000 pensioners had been incorrectly declared dead. A senior journalist said that ‘In village after village, we found that more than half of those declared dead were still living’. The state government paid arrears to people who showed that they were wrongly denied benefits in only 33 out of more than 29,000 cases.⁵³ Jean Dreze argues that pensions – together with ration cards and MGNREGA job cards – have been discontinued ‘just to meet the “100% seeding” targets’ set from on high.

Food provisions for impoverished families have been reduced in three other ways. First, if only three of five members of a family are registered, their provision is cut by 40 per cent. Second, in some states, a so-called ‘Direct

50 The threat posed by Aadhaar to privacy is a major concern. For an overview of privacy issues, see Medianama, 25 July 2016. For anxieties expressed by a professor at the Indian Institute of Management-Bangalore, see Scroll.in, 29 September 2016. Some civil society leaders view Aadhaar as a potential surveillance device. Suhrith Parthasarathy, a Madras High Court advocate, argues that it ‘flagrantly infracts fundamental rights, granting, in the process, enormously invasive powers to the state’, *The Hindu*, 16 January 2018.

51 Scroll.in, 17 November 2017.

52 *The Hindu*, 10 January 2018.

53 Scroll.in, 6 August 2016.

Benefit Transfer' programme disburses funds for food to bank accounts, but that requires the poor to buy grain at 32 times the subsidised rate. In many cases people are not told that those funds have been transferred to bank accounts.⁵⁴ Finally, glitches reduce the amounts of food provided. For example, one former bonded labourer who had previously received 35 kilograms of wheat per month for his family found the provision cut to only between five and ten.⁵⁵

Curbing – and enabling – corruption

Aadhaar is intended to reduce corruption, but its record is at best mixed. Some gains have been made. The authority's blacklisting of 49,000 operatives for malfeasance was noted above.⁵⁶ The system also detected a scam in one sub-district of Karnataka which received 42,000 litres of kerosene for distribution, as against an average in others of 80 to 100 litres.⁵⁷

Ministers' celebrations of gains in the struggle against corruption are based on a breathtaking misperception. They overlook the fact that identity fraud represents a tiny proportion of overall corruption in the Public Distribution System. The main problems are the provision of less than prescribed quantities of food (with the surpluses sold at market prices), and the substitution of poor-quality food (with better quality products being sold privately). The government has avoided the implementation of grievance procedures through social audits (local-level hearings) which are supposedly mandatory features of multiple social programmes, and which could reveal quantity and quality fraud.⁵⁸ Modi and his colleagues are either unaware that these are the main sources of corruption, or they fight shy of tackling them. Instead, we hear statements – based on the aggregate numbers of allegedly 'bogus' ration cards – that Aadhaar has saved the government US\$9 billion 'by eliminating fraud in beneficiary lists'.⁵⁹

As Jean Dreze's analyses demonstrate, in the main, Aadhaar has increased corruption. Since many beneficiaries have been denied subsidised food because they have been unable to complete the complex process of linking ration cards to Aadhaar, because those linkages have been faulty, or because of widespread glitches in the system, proprietors of ration shops have been left with substantial surplus food. Some of them pass it to those excluded in exchange for bribes. Many of them sell it illegally for higher prices on the open market. He also found that many are able to distribute less than 100 per cent of the rations due

54 *The Hindu*, 17 January 2017.

55 Scroll.in, 1 June 2016.

56 Scroll.in, 14 September 2017.

57 *Times of India*, 19 July 2017.

58 *Indian Express*, 11 January 2018.

59 This statement was made by Nandan Nilekani, former head of the Aadhaar Authority, at a World Bank conference. See *Hindustan Times*, 13 October 2017.

to beneficiaries without being found out by Aadhaar. Some of them have told beneficiaries who are fully registered that it is ‘Modi’s wish’ that they undergo one ID exercise without receiving rations – and then they sell that food. In all of these cases, the system has provided new opportunities for corruption.

Dreze notes that the state government website in Jharkhand has shown that 10 per cent of beneficiaries have been unable to obtain food. He believes that the real number is ‘probably more’, but even if the figure is only 10 per cent, it means that 2.5 million mostly vulnerable people have been victimised. The number for the whole of India is far more massive – as evidence noted above from other states indicates – and so are the opportunities for shopkeepers to make illicit profits. He thus concludes that the new system has entailed a ‘revival of corruption’.⁶⁰

Under the Congress-led government (2004–14), the inept drafting of enabling laws for certain programmes prevented them from having the constructive impact that was intended.⁶¹ That was not true of the two Acts under discussion here. Both were carefully drafted and the aims of their architects were apparent in the two texts. But the aims differed, and so did the consequences that ensued.

The MGNREGA has demonstrated that a meticulously designed law that includes an IT system that is carefully crafted to enable poverty reduction can make an immensely constructive impact. The drafters had a realistic understanding of political, logistical and infrastructural constraints, and took steps to minimise them. They also knew that an IT system could not, on its own, ensure adequate transparency, a core element in the programme. So they supplemented it with other transparency mechanisms. When shortcomings were discovered, enlightened civil servants in New Delhi tackled them. This explains why it is harder to siphon funds from the MGNREGA than from nearly all other government programmes, and why it has enabled such vast numbers of poor people to earn funds that they badly need. Finally, the energy that drove the MGNREGA came mainly from below – from those poor people in this demand-driven programme – who were responding to important new opportunities and not to coercion from above.

The damage that Aadhaar has done is not the accidental result of implementation that diverged from its architects’ original aims. It has emerged from the intent of the Act which is apparent from its text: most notably, to promote radically centralised control. It is also a consequence of the forceful

60 Scroll.in, 8 September 2016; *The Hindu*, 17 January; Catchnews, 5 August; *Indian Express*, 21 November 2017.

61 See, e.g., J. Manor, ‘The Forest Rights Act’, in J. Chiriyankandath, D. Maiorano, J. Manor and L. Tillin, *The Politics of Poverty Reduction in India: The UPA Government, 2004 to 2014* (New Delhi: Orient BlackSwan, 2020), pp. 63–85.

pursuit of those aims by Modi and his government, and their haste to maximise enrolment even before they had provided adequate training to implementers and adequate infrastructure – reliable machinery, internet connectivity and electricity supplies. If the MGNREGA is demand-driven from below, Aadhaar is command-driven from the apex of the political system.

Haste and the drive for control are apparent on several fronts. The Act was rushed through parliament after severely limited debate in which all amendments were brushed aside. Only the Aadhaar Authority is empowered to take action when violations occur, or when citizens experience injustices. The authority and the government often fail to reveal such violations and injustices – even to citizens who suffer them. Right to Information petitions have been ignored on dubious grounds of ‘national security’. Critics of the system have been threatened with criminal charges. Heavy pressure on civil servants and shopkeepers in the Public Distribution System to ensure swift implementation has forced or persuaded many to join in the headlong drive for enrolment. Others who see the damage being done have been prevented from stopping it – and some who have sought to help victims by disregarding Aadhaar procedures have been punished.

All of this has been done in patent violation of multiple Supreme Court rulings that Aadhaar must be voluntary. At this writing in February 2018, the court has not found the government in contempt. Such a finding would carry serious penalties, but it would also entail a constitutional confrontation. The court has threatened charges of contempt on one previous occasion under the Modi government. In September 2017, after over two years in which Hindu extremist ‘cow protection’ vigilantes had indulged in an epidemic of beatings and murders – mainly of Muslims, but also of government officials and policemen – the court gave state governments one week to crack down. They complied, and the atrocities largely ceased.⁶² It remains to be seen whether the Supreme Court will take similar action on Aadhaar.

Aadhaar has failed to achieve the goals of its architects in one sphere: curbing corruption. It has had that effect at times, but for the most part, it has facilitated an increase in familiar forms of thievery and enabled the emergence of new ones. But the government has either failed to recognise this or has chosen to tolerate it as a price worth paying.

These two cases – MGNREGA and Aadhaar – demonstrate that advances in IT have made digital records a more formidable force than ever before. But they also indicate that records can serve both constructive and destructive purposes. Excitement about its positive impact is plainly warranted, but so are anxieties about its Orwellian potential.

62 See, for example, Manor, ‘The Forest Rights Act’.

5. Statistical accuracy and reliable records: a case study of mortality statistics in The Gambia*

Andrew Griffin

This chapter examines the relationship between data, statistics and records in the context of mortality statistics in The Gambia. The aim is to illustrate the complexities and practical challenges in acquiring reliable information as a basis for national development planning, as well as for measuring the achievement of the Sustainable Development Goals (SDGs), especially in low resource environments. The first section provides a brief overview of The Gambia's economic challenges. The second deals with the complexities of birth and mortality data, how they are estimated and discrepancies between different data sources. It argues that, ultimately, accurate statistics can only be derived from verifiable data and that in the case of mortality statistics, reliable data must be based on the verifiable records of individual deaths captured as part of a prescribed process. The third section focuses on challenges for collecting reliable mortality statistics in The Gambia. The fourth section explains how The Gambia Bureau of Statistics (GBS) and the National Records Service (NRS) intend to address issues relating to the quality, integrity, completeness and ongoing availability of data, statistics and records. A concluding section suggests that if the two organisations could work together to maximise their strengths and resources, the quality and integrity of the data, statistics and records required to monitor the SDGs and support development planning in general could be significantly strengthened.

Background

The Gambia is a small country in West Africa with a population of less than two million. It occupies a long narrow strip of land on either side of the Gambia River, extending eastwards inland from the Atlantic Coast for approximately 210 miles/338 km. UNICEF and the World Bank estimate that 48 per cent of The Gambia's population live below the poverty line (US\$1.25

* I am grateful to Elizabeth Bahoum and Bartholomew Marong, respectively director and deputy director of the Government of The Gambia's National Records Service, for arranging introductions and meetings, providing documents and supporting research for this chapter.

a day), with large disparities within the population in accessing basic social services.¹ Many of the poor depend on subsistence agriculture, while informal jobs, such as street selling, are predominant on roadsides and in urban areas. As the population has grown, many Gambians, especially young people, have moved from rural to urban areas or left the country to seek a better life. The Gambia relies heavily on its tourist industry and on remittances sent home from abroad.

Mortality rates in The Gambia

How are mortality rates calculated?

Obtaining accurate mortality statistics in The Gambia is not feasible at present, for significant practical reasons that will be explored later in the chapter. However, before examining the causes, it is useful first to look at some of the consequences of inconsistent statistics for the reliability of the available data. It is important to note that the World Bank, WHO and other international organisations constantly update their statistics and online information with new URLs. The information provided in the footnotes that follow was checked in August 2020.

Internet searches for The Gambia's mortality rates tend to produce confusing results, which illustrate the problem. IndexMundi, 'a data portal that gathers facts and statistics from multiple sources and turns them into easy to use visuals', provides annual Gambian mortality rates per 1,000 for the years 2000 to 2017:² IndexMundi gives the source of these statistics as the CIA World Fact Book, a reference resource produced by the US Central Intelligence Agency. According to this source, the estimated death rate for 2017 (currently the latest year for which statistics are provided) is seven deaths per 1,000 population³ and shows a large drop (36 per cent) in the mortality rate between 2010 and 2011 (see table and highlighted figures below). All other years show a gradual overall reduction.

2000	2001	2002	2003	2004	2005	2006	2007	2008
13.21	12.92	12.63	12.35	12.08	11.81	12.25	11.99	11.74
2009	[2010]	[2011]	2012	2013	2014	2015	2016	2017
11.49	[12.03]	[7.65]	7.50	7.38	7.26	n/a	7.10	7.00

The World Bank, however, gives very different figures for the number of deaths per year per 1,000 population, rounded to two decimal places in the table below.⁴

- 1 <https://reliefweb.int/report/gambia/wfp-gambia-country-brief-may-2018>; <https://data.worldbank.org/indicator/SI.POV.NAHC>.
- 2 <http://www.cia.gov/library/publications/resources/the-world-factbook/geos/ga.html>.
- 3 <https://www.cia.gov/library/publications/the-world-factbook/geos/ga.html>.
- 4 <https://data.worldbank.org/indicator/SP.DYN.CDRT.IN?locations=GM>.

In particular, the figures for 2010–11 are very different from those provided by IndexMundi:

2000	2001	2002	2003	2004	2005	2006	2007	2008
11.57	11.29	11.01	11.73	10.46	10.19	9.93	9.68	9.45
2009	[2010]	[2011]	2012	2013	2014	2015	2016	2017
9.23	[9.03]	[8.84]	8.67	8.51	8.35	8.20	8.06	n/a

The World Bank's estimate for 2000 is 12 per cent lower than IndexMundi's, whereas its estimate for 2016 is 13 per cent higher. In terms of estimated deaths for the population as a whole, the World Bank crude death rate of 8.67 in 2012 (when the population of The Gambia was estimated to be 1,857,181)⁵ provides an estimated 16,102 deaths for the whole population, whereas the IndexMundi death rate provides an overall death rate of 13,093, a difference of over 3,000 deaths, making IndexMundi's figure 19 per cent lower. Still other sources give the mortality rate in a date range and show further disparities. For example, Wikipedia provides the 'crude death rate' as an average over six years, giving the source as 'The Population Department of the United Nations':⁶

2000–5	2005–10
10.9	9.8

The average death rates in these date ranges from the statistics provided by IndexMundi and the World Bank are again significantly different:

IndexMundi		World Bank	
2000–5	2005–10	2000–5	2005–10
12.50	11.88	11.04	9.58

UNICEF demographic indicators give the crude death rate for 2012 as 9.8, which is significantly higher than the IndexMundi figure of 7.5 but, coincidentally, the same as the six-year average for 2005–10 provided by Wikipedia.

Finally, the World Health Organization's African Health Observatory gives an adult mortality rate per 1,000 for 2008 as 276,⁷ which, taken at face value, further confuses the picture. Adult mortality refers to the probability

5 This figure is provided in the GBS Mortality Report, 2013. However, the World Bank provides a lower figure of 1,802,125 and IndexMundi also gives a lower figure of 1,840,454 which will lead to different estimates of the crude death rate.

6 https://en.wikipedia.org/wiki/Demographics_of_the_Gambia.

7 <https://apps.who.int/gho/data/node.country.country-GMB>.

that those who have reached age 15 will die before reaching age 60 (shown per 1,000 persons). A value of 276 means that of 1,000 persons who have reached the age of 15, 276 are expected to die before reaching age 60. Direct comparison between this statistic for 'adult mortality rates' and the 'crude death rate' used by other authorities would be difficult if not impossible, at least for the lay person.

The United Nations defines civil registration as 'the continuous, permanent, compulsory and universal recording of the occurrence and characteristics of vital events ... provided through decree or regulation in accordance with the legal requirements of each country'.⁸ Reliable civil registration, including issuing relevant certificates, requires proof of such registration. Estimates of death rates can be derived from censuses and surveys, but only a well-developed and fully functioning civil registration system is able to collect this information on a continuous basis and to ensure that there are records of 'vital events', including deaths. Gambian mortality statistics have to be estimated because there is no systematic recording of deaths. Registration of deaths is not a legal requirement and is not feasible for many of the population. All the figures published by the World Bank, World Health Organization, UNICEF and other sources in The Gambia are, by necessity, estimates only.

Challenges for collecting reliable birth and death statistics in The Gambia

How are deaths recorded?

Although The Gambia is the smallest country in mainland Africa, some rural areas are still relatively inaccessible. Transport and communications are a problem for many Gambians, not only for the rural poor but for the urban population in the capital Banjul and the surrounding areas. The majority of the population are Muslims, and burials of the deceased take place as soon as possible, usually within 24 hours of the death. Introducing compulsory registration of births, marriages (many of which take place according to custom) and deaths would present huge challenges for a country with limited financial resources.

The Ministry of Health and Social Welfare is currently responsible for registering deaths in The Gambia, although, as noted, this is presently an unrealistic task for the country as a whole. Deaths are registered formally only for those who die in hospitals, where the cause of death is ascertained by a

8 *Principles and Recommendations for a Vital Statistics System, Revision 2* (New York: United Nations, 2001).

medical doctor. The Gambia has four public hospitals and a network of smaller health facilities throughout the country. Death certificates are only prepared by the Births and Deaths Registry Unit at Medical Headquarters in the capital, Banjul.⁹ As most deaths occur at home, formal registration of death is not common. The information available from registered deaths in the urban area does not enable the extrapolation of accurate national patterns of disease and mortality, especially among adults.

Deaths that occur in health facilities, such as clinics, are reported in the monthly health statistical reports through a national surveillance system. Where government health facilities have access to a power supply and the internet, the staff enter data in DHIS2, a web-based open-source information system.¹⁰ Facilities without internet access use a manual system to record data, which are then entered in the DHIS2 information system. Data from DHIS2 are consolidated in the government's Health Management Information System and should be published on the Ministry of Health website.¹¹ Unfortunately, at the time of writing, there are no statistics available on the website. The data are reported to the World Health Organization and the West African Health Organisation (WAHO), an agency of the Economic Community of West African States (ECOWAS).¹² However, data from private health care facilities, accounting for 25–30 per cent of the country's health care system, are not included in the surveillance and reporting system, though it is understood that they may be included in future. If conducted reliably by all health facilities, the surveillance system has the potential to provide more accurate national mortality statistics, but it is not clear to what extent or how they will include deaths occurring at home.

How are death rates estimated?

Systematic birth and death registration require an infrastructure (offices, equipment, systems, policies and standards) and the human resources needed to register the events. However, at present, in The Gambia, funding is inadequate to cover these costs. Without a civil registration system, death rates must be estimated from data collected by censuses, surveys and sampling. Organisations such as the World Bank derive data from reports gathered by in-country management units and other independent sources. The Gambia currently uses a combination

9 Status of Civil Registration and Vital Statistics in the Republic of The Gambia. United Nations Economic Commission for Africa [undated article published at http://www.uneca.org/sites/default/files/images/crmc_status_of_crvs_in_gambia.pdf].

10 For more information, see <http://www.dhis2.org>.

11 <http://www.moh.gov.gm/>. At the time of writing DHIS2 data was not available to the public.

12 Information supplied by Sana M. Sanbou, Ministry of Health and Social Welfare statistician.

of surveys (censuses) and sampling (using a sub-set of the population to derive information about the entire population) as illustrated below.

The most recent nationally representative demographic survey in The Gambia is the 2013 Population and Housing Census, which was conducted by the GBS in collaboration with the Ministry of Health and Social Welfare and the National Population Secretariat Commission. This was the fifth census to take place since The Gambia achieved independence in 1965 and the first to take place under the worldwide Demographic and Health Survey programme funded and supported by the United States Agency for International Development. The main objective of the survey was 'to provide comprehensive data on fertility and mortality, family planning, maternal and child health and nutrition, as well as information on maternal mortality and domestic violence'.¹³

Over a three-month period in 2013, a total of 6,217 Gambian households were contacted: 10,233 women aged 15–49 and 3,821 men aged 15–59 were interviewed.¹⁴ Fourteen separate data analyses and reports were derived from the census data, including, reports on children, economic characteristics, education, housing and household characteristics, the elderly, mortality and access to ICT.¹⁵ The Mortality Analysis and Evaluation, published by the GBS,¹⁶ included reports of deaths in the households 12 months prior to the census, including sex and age at death. For every deceased female aged 15 years or older, further enquiry was made to establish whether the death occurred during pregnancy, while giving birth or within six weeks after the birth. Any death where there was an affirmative response to any of these three questions was classified as a pregnancy-related death. Women aged 12 years and over were also asked to state the total number of children they had ever given birth to, as well as how many were alive at the time of the census and how many had died.

The data gathered were used to estimate the infant mortality rate for the decade prior to the census as well as to develop an estimation of the crude death rate and the age- and sex-specific death rates experienced by the entire population. Various direct and indirect estimation techniques were used to provide reasonably acceptable mortality indicators for the country, although inevitably, as noted earlier in this chapter, there were discrepancies. The discrepancies, together with the fact that the survey is becoming outdated with time, have undermined the usefulness and relevance of the data. Plans to

13 The Gambia Demographic and Health Survey 2013. GBS, Banjul, The Gambia ICF International Rockville, Maryland USA, September 2014. The next Demographic and Household survey was due to be carried out between October 2019 and February 2020; survey data is not yet available. <https://dhsprogram.com/pubs/pdf/FR289/FR289.pdf>.

14 <https://dhsprogram.com/pubs/pdf/FR289/FR289.pdf>.

15 <https://www.gbosdata.org/downloads>.

16 <https://www.gbosdata.org/downloads>.

conduct a similar survey have yet to be developed and in any case, it is doubtful that the required resources would be secured.

The reliability of birth dates

Birth registration is an essential part of a functioning civil registration system and complements the registration of deaths. The existence of accurate and complete birth records facilitates the calculation of the ages of individuals at the time of their deaths. Most countries have mechanisms in place for registering births. However, the coverage, the type of information obtained and the use of resulting data differ, based on the country's infrastructure, legal frameworks, administrative capacity, barriers to accessing services, availability of funds, accessibility to the population and technology for data management. Levels of registration vary substantially across countries due to these and other factors and to the availability of data on birth registration, which is highly uneven across countries.

In The Gambia, registration of births is a more common practice than is registration of deaths, but it is hampered by inadequate human and material resources, particularly at the health clinic level. Birth registration is supposedly mandatory, and a decentralised system operates at health facilities and outreach stations. Gambians are far more likely now than in the past to have a formal birth certificate, or they can apply for one if the original is lost, so the government and its development partners should have accurate birth statistics as a basis for planning. However, at present, there is no means of determining the reliability of The Gambia's birth records.

Efforts to strengthen official statistics in The Gambia

The Gambia Bureau of Statistics

In February 2018, the Government of The Gambia launched a National Development Plan, 2018–2021.¹⁷ One area of focus is reducing maternal and newborn mortality. The Gambia's second National Strategy for the Development of Statistics, 2018–2022, published in 2017,¹⁸ is aligned with the National Development Plan and the UN SDGs. The demand for improved statistics has grown across many subject areas. Timelier and more reliable statistics are seen as critical in informing policies and planning, in monitoring and evaluation of programmes and services, and more generally in public sector management. A key aim of the National Strategy for Statistics is to develop a

17 The Gambia National Development Plan (2018–2021): an abridged version, December 2017, <https://mofea.gov.gm/downloads-file/national-development-plan>.

18 National Strategy for the Development of Statistics (NSDS II) 2018–2022 (GBS, 2017).

civil registration and vital statistics system, including the official recording of all births and deaths.

The ability to capture these statistics reliably is crucially important for development planning and protecting citizens' rights. It establishes legal identity and civil status and enables access to services such as health care, education and social protection; it contributes to timely and accurate estimates for effective planning and resource allocation; and it influences policy development and research and the ability to monitor and evaluate. In particular, it enables accurate monitoring of progress towards achieving the SDGs.

The strategy has six overarching goals:

- Goal 1: An efficient and well-coordinated national statistical system
- Goal 2: Quality and sustainable human resources across the national statistical system
- Goal 3: Enhanced physical, ICT and statistical infrastructure
- Goal 4: Improved data production and dissemination
- Goal 5: Assured quality of data
- Goal 6: Sustainable funding arrangements and partnerships

Goals 4 and 5 are of particular interest for this chapter because of the focus on data quality and integrity.

The aim of strategic goal 4 (improved data production and dissemination) is to increase the frequency of data collection, conduct new surveys, develop civil registration/vital statistics systems, strengthen administrative records (including civil registration records) and assure proper data dissemination. The goal embraces the three main sources of official statistics across the national statistical service in The Gambia:

- surveys and censuses, which are conducted periodically (such as Population and Housing Censuses, the Integrated Household Survey and the Multiple Cluster Indicators Survey)
- administrative records, including civil registration
- vital statistics systems, which provide continuous data.

Recognising that The Gambia lacks an effective civil registration and vital statistics system, the GBS and the national statistical programme in general aim to scale up existing data management information systems and implement new systems where needed. In order to meet the increasing demand for statistics, the strategy states that 'data production and dissemination should take advantage of using innovative statistical technologies employed in data management processes, such as mobile devices and advanced imagery technology'.¹⁹

19 National Strategy for the Development of Statistics (NSDS II) 2018–2022, p. 32.

The initiatives under strategic goal 5 (assured quality of data) are designed to enhance the quality of statistics through the development of a Statistical Compendium of Concepts and Definitions, a household sampling frame with a master sample, a business establishment sampling frame and a business registry. There are also plans for a metadata handbook and a customised National Quality Assurance Framework that adapts international classifications and nomenclatures to The Gambia's needs.

These ambitious plans are commendable, but it remains to be seen whether they will be fully implemented, bearing in mind The Gambia's lack of resources²⁰ and the fact that practices for gathering statistics are not coordinated at present. For instance, the national strategy notes that information management systems derive most of their data from administrative records and that these systems are 'generally weak and require substantive development and improvement'. Furthermore, 'the key suppliers of data require a lot of motivation in order to provide data and records for the compilation of the needed statistics'.²¹

The significance of records for mortality statistics and the contribution of the National Records Service

Weaknesses in information management systems that derive data from administrative records underscore the role records play in ensuring the quality and integrity of data and statistics for development planning. Their importance is often poorly understood and not well addressed in data quality initiatives that focus on improving the quality of processes generating data and statistics. Understanding the characteristics of records and the role they play is an important prerequisite to understanding the challenges that organisations face in managing their data and statistics.

Records are created by any given process for collecting, verifying or manipulating data and producing and disseminating statistics. Records document decisions concerning the design and management of the process as well as the reliability of the data and statistics themselves. When well-managed, records serve as evidence that the data and statistics were managed properly and that they have the required level of quality and integrity. In the case of death registration records, for instance, they tell the story of how, when, where and by whom a given death was identified and recorded, how data from the form

20 Although health expenditure in The Gambia increased by 61 per cent between 2002 and 2013, bringing the annual per capita health expenditure to an estimated \$28, this is still well below the minimum WHO recommendation of \$34–40. It is therefore not surprising that, in turn, The Gambia lacks reliable and comprehensive records for individual Gambian citizens in hospitals, clinics and health facilities and for health care planning, and that health care data is incomplete.

21 National Strategy for the Development of Statistics (NSDS II) 2018–2022, p. 13.

capturing the data was verified and migrated to a master file or database, how it was manipulated and processed to enable statistics to be produced, how the statistics were published and disseminated, and how the data and statistics, not to mention the records themselves, were retained.

Based on their content, records can serve as sources of qualitative and quantitative data and can be used for multiple purposes beyond the evidentiary purpose that led to their creation. For instance, various records in the form of emails, correspondence, procedures and other documents may document and support the management of processes generating births and deaths data and producing statistics. These records, together with the forms and registries capturing the instances of births and deaths, serve, in themselves, as sources of data that can be used to produce a wide range of statistics. The quality and integrity of statistics are based on the quality and integrity of the data input to the statistics. This, in turn, reflects the quality and integrity of the processes for collecting, processing, analysing and maintaining the data and producing and reporting the statistics. The ability to demonstrate this quality and integrity is dependent on the availability of complete, authentic and accurate records.

In The Gambia, the NRS is responsible for facilitating the management of records across the government. Created as part of a series of records management reforms in the 1990s, the NRS controls standards and guidelines for records management in government records offices and carries out regular monitoring and training activities aimed at maintaining and strengthening the standards. All ministries and departments have 'restructured' records offices, and the NRS continues to restructure records in parastatals. Recently, a new National Records Service Act was drafted to strengthen, and to state explicitly in law, NRS responsibility for records in all formats, including data held in information systems in public offices. It would also make the NRS responsible for implementing policies and standards to support public records and data management. The Act has yet to be enacted, but in any case, the NRS will be responsible for records and recordkeeping systems created as part of any future civil registration and vital statistics system.

The significance of the NRS role is becoming ever more fundamental for national development as the government increases the use of new technology to create and manage its information and as the challenges of managing information in digital form become more apparent. Two examples of NRS contributions relate to the quality of government data and statistics in the areas of health on the one hand and finance on the other.

The first example, involving health data, relates to clinical coding or medical classification, which is the process of converting descriptions of medical diagnoses and procedures into universal medical code numbers in order to

provide the raw data for analysis. Information about diagnoses and procedures, converted to codes, is usually taken from a variety of patient records, for instance, doctor's notes, laboratory results and radiological results.²² In the past, records at The Gambia's Royal Victoria Hospital (now the Edward Francis Small Teaching Hospital) were poorly managed, incomplete or inaccessible. Clinicians often had to treat patients without a full case history, which added to the difficulty of getting complete birth and mortality data.

The NRS introduced a unified patient record system that brought together the records of inpatient and outpatient episodes relating to the same individual. Patients were registered and given a unique patient number; their records were kept together by this number in a newly constructed registration block. The unified patient record system was also introduced at the Bansang and Farafenni Hospitals. The NRS team sought to capture and analyse statistical information about patients and treatments in the expectation that the new system would also improve clinical coding. The lack of resources and skilled staff has made it difficult to maintain clinical coding, but the potential contribution remains.

The second example relates to the integrated financial management information system (IFMIS) that the Ministry of Finance and Economic Affairs has been implementing in The Gambia in recent years. The aim has been to manage the government's financial transactions, using the IFMIS human resources management module to manage civil servants. There is an opportunity to link this to reliable national registration data. Under the dynamic leadership of its director, the NRS has established an IT unit and is currently implementing an electronic records/content management system (compliant with international standards), initially for the NRS's own records but with the aim of rolling out the system to four other pilot sites.

These and other examples demonstrate that the NRS should be a key player in strengthening the quality and integrity of the data and statistics used by the government. If records systems were linked to strengthening data for national development, they could provide valuable evidence needed to support statistical integrity. Without NRS input, the controls for managing records that document how decisions are taken and implemented in regard to statistics will remain weak, making it difficult to carry out effective audits and quality assurance. However, while the NRS could make an invaluable contribution strengthening data and statistics, it does not have the resources and the range of skills needed to play its potential role.

22 The International Classification of Diseases (ICD) is the standard coding system used by member states of the World Health Organization. Most of the 117 member countries use the system to report mortality data, a primary health indicator.

The benefits of shared responsibility for the quality of statistics

The NRS and the GBS share an important role in ensuring that the data, statistics and records supporting development planning are complete, accurate, authentic and reliable. As in many countries, however, they tend to work in isolation from one another, with the GBS focusing on data and statistics and the NRS on records. Each organisation supports its own set of policies, procedures, standards and practices, and the initiatives they manage are not coordinated. While each organisation is concerned about quality and integrity issues, they approach the issues independently based on media (data and statistics or records). They have yet to recognise that because the data, statistics and records generated by a given process, such as the process for registering births and deaths, are part of a comprehensive whole, their management needs to be coordinated.

Both organisations have strengths that can be brought to bear on a given process. The NRS team lacks resources, but it understands how records should be managed. The GBS understands data and statistics, and while it, too, is hindered by a lack of resources, its team understands how data and statistics should be managed. Both organisations have a vested interest in ensuring that the processes that generate the data, statistics and records are complete, well designed and reflect an appropriate level of quality and integrity. Much could be gained by both organisations if they could coordinate their human and financial resources to improve the quality and integrity of not only mortality data and statistics but development more generally across The Gambia. As joint leaders, they could draw in other relevant organisations, for instance those responsible for IT security and audit, as well as the users of the data and statistics. They could maximise scarce resources and present a set of unified strategies that would produce results of far greater impact and relevance than if each continues to work on its own.

Summary and conclusion

In the absence of a legal requirement to register all deaths and to create an official record of each death, mortality statistics must be estimates only. Official mortality statistics for The Gambia are based on estimates derived from periodic censuses and surveys. The most recent Population and Housing Census, conducted in 2013, provides the best available data for estimating the death rate in the year preceding the census and it is a basis for projections for the following years. Other methods of collecting death statistics, such as reporting by health facilities through a national surveillance system, do not provide complete nationwide coverage.

Independent sources have estimated a higher crude death rate than that provided by the GBS, implying that deaths in the year preceding the 2013 census were under-reported, that independent estimates were too high, or possibly a combination of both explanations. Statistics published by international organisations for The Gambia's mortality rate vary widely, in some cases by up to 20 per cent. The GBS acknowledges that the country lacks an effective civil registration and vital statistics system, but compulsory civil registration, which could generate an accurate record of every birth and death, would require a huge investment in infrastructure and human resources.

At present, The Gambia lacks the funding and expertise to manage the quality, integrity and completeness of data, statistics and records to meet both current and future requirements. The GBS has established a strategic plan for addressing data quality issues, but it lacks the resources to execute the plan effectively. The NRS provides standards, common procedures, training and oversight for managing the government's paper records and is currently piloting an electronic records management system that complies with international standards, but its role in setting policies and standards for data and digital information systems, including for preservation, is limited and its services and capacity remain a low priority for the government.

These two organisations, despite their statutory responsibility for the management of data and statistics on the one hand and records on the other, operate in isolation from one another. A working relationship between them, to formulate and agree standards for managing data and records as well as the content of information systems from which official statistics are derived, would maximise the use of limited resources and improve the quality of the data, statistics and records that the government needs to support national development planning and to monitor the achievement of the SDGs.

6. Mainstreaming records and data management in sustainable development: lessons from the public and private sectors in Kenya

Justus Wamukoya and Cleophas Ambira

Kenya is one of more than 190 countries supporting the United Nations Sustainable Development Goals (SDGs) initiative, also known as the Global Goals for Sustainable Development. Successful implementation of the SDGs will complement Kenya's socio-economic development blueprint for the period to 2030. All 17 SDGs address global challenges and social issues, including those related to poverty, climate change, environmental degradation, prosperity, peace and justice. Moreover, they are intended to bolster government efforts to implement national development plans and to revitalise the global partnership for sustainable development.

In Kenya, as in most other countries, implementing the goals successfully depends on contributions from various stakeholders, including the government, the private sector, civil society, international organisations and technical experts in a variety of fields. Data are generated through surveys and through operational and administrative systems, and statistics are collected and aggregated to measure multiple indicators for each goal. Records are essential to documenting the processes that generate data and statistics. If well-managed, they can help to ensure the integrity and trustworthiness of the processes by making it possible to hold those who create the data and produce the statistics accountable for their decisions and actions. The reliability of the measurements depends fundamentally on the quality of the data, statistics and records produced.

The purpose of the chapter is twofold. The first purpose is to use examples from the public and private sectors in Kenya to illustrate the distinct perspectives that each brings to the quality and integrity of the data, statistics and records needed to support the SDGs. The second is to suggest that the quality and integrity issues the various sectors are facing can best be addressed by drawing on their different strengths and by working together to develop strategies for ensuring that each contributes to measuring the SDGs reliably in a manner that is consistent, comprehensive and multisectoral.

The public sector experience in Kenya

Across Africa, including Kenya, public demand for information and access to official government records is at a new high. This can be attributed to many factors, including, but not limited to, new access to information laws, which have encouraged citizens to seek legal access to information held by government agencies in the form of datasets or official government records. It is also a consequence of the growing widespread use and affordability of computers and other related technologies, a fast-growing elite of young middle-class professionals hungry for information, an increasingly educated and informed general public, an aggressive and proactive civil society and an active media network.

Unfortunately, not all countries take the management of public records and/or public information as seriously as others. Western countries such as the United Kingdom, United States, Canada, Australia and New Zealand have tended to recognise that information must be well-managed if it is to remain useful through time and have embraced information and communication technologies as modern tools for managing records. Records, in particular, have tended to be neglected in Kenya and across Africa. Many African countries still are burdened with challenges associated with poor filing systems and with a lack of capacity and skills in records management. Often, they must deal with wasteful and cumbersome traditional paper systems that tend to result in volumes and volumes of poorly organised records.

In Kenya, officials tend to regard records management as a routine clerical function associated with lower cadre staff. Most officials do not yet recognise it as an essential business function that provides a basis for developing and implementing reliable and trustworthy policies that inform planning, provide services for citizens and support organisational efficiency and effectiveness. Nor do they realise that digital records, just as do paper records, should provide essential evidence to protect the rights of stakeholders, including citizens, government employees and the government itself.

Other challenges include dwindling resources from government, the absence of a champion within government to drive forward an agenda for records management reforms, declining opportunities for professional training in core competencies, the lack of organisational plans for managing records, the tendency of public servants to willfully destroy records to hide crucial evidence and lacklustre performance by the Kenya National Archives and Documentation Service, which has yet to provide informed guidance to ministries and departments on the requirements for creating and keeping high-quality records. As a result, the management of public sector records has deteriorated to worrying levels in recent years.

Despite these significant challenges, there are signs that the situation may be changing. The government of Kenya recognises that well-managed records and information are key resources for efficient and effective public services, transparency and accountability. It is well aware of the challenges presented by cumbersome manual filing systems, inadequate staffing and lack of top management support, and it is exploring strategies to address these challenges. Recently, faced with issues of corruption that have been exacerbated by the poor management of records, the government has taken steps towards improving public sector records by developing a national records management policy.

According to the foreword to the Draft Public Records and Information Policy (June 2019), adhering to the guidelines will go a long way towards streamlining the management of public service records. If approved by the cabinet and the national assembly, the policy will provide guidance and direction on how current and semi-current records are managed across government, throughout the country. The government envisages that the policy will support strategies for improving the quality of records and data generated and maintained in public offices and for using them efficiently to support objectives that include sound decision-making, improved service delivery, management planning and protecting rights, transparency and accountability.

In support of the policy, the government is developing a change management strategy that will involve hiring new staff at entry level, redesigning existing staffing structures and training, sensitising all staff to key records management issues, reviewing the current scheme of service for records managers and introducing new budget categories for records management. It is expected that this will help to ensure that records management units take their rightful place in the hierarchy of government bureaucracy and assume full responsibility for all current and semi-current records, regardless of the form of media, in line with internationally agreed standards and good practices. Ultimately, this should have a significant impact on the quality of the information available to measure the SDG indicators.

Moreover, the Kenyan government is emphasising the importance of computerising and digitising records across a number of key sectors to deal with the slow pace of retrieving hard-copy records. Currently, it is digitising hard-copy records in departments including lands, civil registration, immigration and motor vehicle registration. This is seen as a first step towards fully automating registries concerned with identity cards, registration of births and deaths, passports, motor-vehicle logbooks, driving licences, among others.

There is another significant challenge for the records management community where action has yet to begin. If the SDGs are to be measured accurately, it will be important to build bridges between the data, statistics and records communities and between the different sectors involved in order

to strengthen the quality and coverage of the data. In the example of banking data, the banks generate source data needed to produce statistics for the Ministry of Finance. The ministry then manipulates the data, creates new data and produces statistics to monitor and measure the SDG indicators. The entire process, from collecting banking data to producing statistics, crosses private and public sector boundaries.

The challenge is for relevant organisations in both sectors to come together to design and implement comprehensive, consistent and relevant solutions across the sectors and to document the steps involved. If the quality and integrity of banking data and statistics are to be trusted, the processes for collecting, manipulating and producing them must be trustworthy. The degree to which they can be trusted will depend on how well the various communities involved work together to design and implement integrated policies, standards and technologies and to manage and document the overlap.

Addressing quality and integrity issues in relation to measuring the SDGs involves not only multiple media, such as data, statistics and records, but also multiple sectors such as the private and public sectors. The more detailed case study of mobile banking in Kenya that follows illustrates the importance of recognising that the data, statistics and records of the public and private sectors are parts of a whole.

Mobile banking in Kenya

Mobile banking, one of the first aspects of the mobile applications upon which the Kenyan economy increasingly depends, began as a private sector development but has become an increasingly important aspect of Kenya's complex banking structure, which cuts across the private and public sectors. The entire mobile banking function is regulated by the Central Bank of Kenya and the Communication Authority of Kenya, so its success is enabled and facilitated by the government. The Safaricom mobile network itself, one of the most profitable companies in East and Central Africa, has 35 per cent government ownership. Moreover, some of the government-owned commercial banks have mobile banking infrastructures, and mobile technology is being used increasingly to deliver relevant government services, for instance payment for parking in Nairobi, driving licence renewal, change of car logbooks and the registration of businesses.

The case study examines the scale of mobile banking in Kenya, its impact on society and the kinds of records generated through mobile computing devices. Issues associated with managing the data, statistics and records generated by mobile banking are highlighted, and ways in which the government and the private sector can start to work together to address them are explored. Through time, smart phones, and even less sophisticated mobile phones, will form a

central hub that will allow Kenyan society to access a wide range of public and private sector services, blurring the lines between the two sectors. If the SDGs are to be measured effectively and to have the intended impact, it will be essential that policies, standards and practices developed to address the quality and integrity of data, statistics and records are multisectoral, both in their development and their application.

The Kenya Bankers Association (2014) describes mobile banking as:

the provision and availment of banking and financial services through the help of mobile telecommunication devices. The scope of offered services may include facilities to conduct bank and stock market transactions, administer accounts and to access customized information.¹

Mobile banking is simply the use of a mobile phone to perform banking services. It does not necessarily mean access to a bank account, even though access to a bank account via a mobile is mobile banking. Banking is not necessarily just about financial transactions but also includes access to information on banking matters. Access to information is a two-way street between the banking institution and the customer. Some authors argue that mobile banking requires smart phones, but, in reality, in countries like Kenya, mobile phones do not necessarily need to be smart phones. Non-smart phones use the unstructured supplementary service data (USSD) technology to carry out a transaction. In Kenya, mobile banking services are predominantly offered by commercial banks, mobile telephone network providers, mobile loan providers and savings and credit co-operatives. Ordinarily, commercial banks create an interface between customers and their bank accounts using mobile technologies.

In 2008, only 19 per cent of Kenya's 35 million people had bank accounts.² When M-PESA, the first money-transfer service, was launched in March 2007, there were only 1.5 bank branches per 100,000 people and only one Automated Teller Machine (ATM) per 100,000 people. As of March 2019, there were 50.36 million mobile banking accounts in Kenya, with a total of 161.38 million transactions being conducted through these accounts.³ The total value of these transactions was US\$3.68 billion. Each transaction was captured by the various Kenyan mobile banking platforms and reported to the Central Bank, the regulator. Today, mobile banking in Kenya is a predominant mode of transactions, especially in microtransactions.

These statistics imply that, on average, every Kenyan has an active mobile money account that he/she uses for the day-to-day activities of sending and

1 Kenya Bankers Association, *The Mobile Banking Survey 2014* (Nairobi: Kenya Bankers Association, 2014).

2 Kenya Bankers Association, *The Mobile Banking Survey 2014*.

3 The Central Bank of Kenya in collaboration with the Kenya National Bureau of Statistics and FSD Kenya has launched the 2019 FinAccess Household Survey Report.

receiving cash and making payments for goods and services. It is important to note that a mobile money account does not mean a bank account, but rather a mobile-based wallet that in some cases interfaces with bank accounts. The low cost of mobile phones has made it possible for almost every Kenyan to own a mobile phone. This makes it easier for the banks to penetrate the market through mobile banking. Mobile banking can be used even in the most remote areas because all one needs is a mobile phone. People in the rural areas have benefited greatly from this service and are now able to pay bills and make purchases without needing hard cash and without the need for long journeys to get it.

Relationship to the SDGs

The Kenya Bankers Association has noted that the use of mobile banking in Kenya is an increasingly important component of national and regional economic development. In this respect it has a direct impact on the achievement of several SDGs. As mobile banking has enhanced the ability of people at all levels of society to access financial services, even in the most remote areas of the country, banking has ceased to be the preserve of the well-to-do, whether through conventional banking or through mobile money wallets. Even the poorest in society have equitable access to financial services, a development that is viewed as a major contributor to the achievement of SDG 10 (reduced inequalities).

Mobile banking also plays a role in achieving SDGs 1 (no poverty) and 2 (zero hunger). M-Shwari offers a good example. A paperless banking service that enables account holders to open and operate bank accounts through their mobile phones, M-Shwari was launched by the Commercial Bank of Africa and Safaricom in 2013. In 2013, 19 per cent of M-Shwari users were below the national poverty line; a figure that increased to 30 per cent by the end of 2014. According to FSD Africa (2016), by the end of 2014, M-Shwari boasted 9.2 million savings accounts, having disbursed 20.6 million in loans to 2.8 million borrowers.⁴ What this implies is that the M-Shwari platform is affording more and more access to financial services to the poor. Out of the 9.2 million account holders, 7.2 million were individuals. Thus, through M-Shwari alone, millions of Kenyans are now able to tap into mobile banking for savings, credit and transactions to improve their livelihoods.

Finally, mobile banking has had a direct impact on the achievement of SDG 9 (industry, innovation and infrastructure). Mobile banking has become a major contributor to economic development and to establishing an infrastructure that is both innovative in design and effective in promoting the country's economic

4 Central Bank of Kenya, Kenya National Bureau of Statistics and FSD Kenya, *The 2016 FinAccess Household Survey on Financial Inclusion* (Nairobi: FSD Kenya, 2016).

growth. In fact, statistics indicate that almost half of Kenya's GDP in 2018 was moved through mobile phones: a total of 3.98 trillion shillings or 10.92 billion shillings per day, or the equivalent of 44 per cent of Kenya's GDP.⁵

These developments would not be possible without the resilience of technological innovations in the country. Safaricom's M'Pesa has won Kenya many accolades and is rated the best mobile money platform worldwide. Continued innovations by Safaricom in mobile money technologies and capabilities have fostered significant innovation by other financial sector players, including commercial banks, mobile loan providers and other telecommunication companies, the results of which are evident in the economy and people's livelihoods. They also have catalysed the Kenyan government's investment in ICT infrastructure across the country, which has enabled 43.3 million Kenyans, out of a total population of about 51.58 million, to have access to the internet.⁶ Digital connectivity is playing an important role in transforming and improving lives by opening doors to employment, financial opportunities and access to knowledge, and the information professions must respond to this development.

How do data and records management support mobile banking?

Mobile banking is anchored entirely by data management. The fact that 161.38 million transactions were conducted by 50 million mobile accounts as at March 2019 illustrates the amount of data generated on these mobile banking platforms. Moreover, the fact that these 161.38 million transactions had a value of US\$3.68 billion, shows the significance and potential risk levels associated with the data and the importance of managing it carefully. The statistics generated using the data that have been transacted over the years have been fundamental to supporting government development plans. This is why entities like the Communication Authority of Kenya, the Central Bank of Kenya and the Kenya National Bureau of Statistics consistently provide trend analysis reports to help inform development policies and strategies.

Records management and data intersect at this point. What would happen if the data were not well-managed? Would records extracted from the data be available to present in court to support litigation? According to a 2016 report on cybersecurity in Kenya in 2016 by Serianu,⁷ mobile money in Kenya has

5 Soko Directory Team, 'Half of Kenya's GDP moved through mobile phones in 2018: the power of mobile money transactions in Kenya', 25 January 2019.

6 Kevin Namunwa, 'Kenya leads Africa in smartphone usage', *Business Today*, 11 March 2019.

7 Serianu, *Kenya Cybersecurity Report* (Nairobi: Serianu, 2016). Serianu is an award-winning pan-African based cybersecurity and business consulting firm that enables organisations to extract value from their information assets. It helps its customers collect, protect and analyse critical business information, specialising in new and emerging technology areas, including information security, data analytics, network security, application security, cloud security and cybersecurity.

been attacked repeatedly through social engineering, the use of malware and identity theft; hackers are exploiting the weak security controls around the mobile money platform to steal millions. The Central Bank has observed, in a 2017 report, that cybercrime targeting the financial sector is a significant risk that is expected to increase in sophistication and frequency. According to another report by Serianu, in 2018, KSh230 million was lost through computer fraud, KSh100 million through business emails, KSh70 million through fake cheques and KSh66 million through identity theft.⁸ In early 2018, there were social media complaints that customer accounts at the Kenya Commercial Bank, Kenya's largest bank by assets, were under attack and were incurring unauthorised deductions.⁹

Once a fraud is reported and forensic investigations are sanctioned, the data have to be extracted, usually by court order, certified as authentic and admitted in court to support the litigation. The data must then be preserved as evidence of the court proceedings, in whichever form the court admits it – digital or physical. In addition to its importance in securing customer financial resources, data generated through mobile banking help authorities in criminal investigations to identify criminals. For instance, in early 2019, during investigations in the 'Dusit D2' terror attack in Nairobi, data from M-Pesa transactions were used by the investigating agencies to unravel the identity of the terrorists.¹⁰

These cases illustrate the need for robust mechanisms for managing data safely so that they can serve the purposes for which they were captured and maintained, including for security purposes and to support investigations. The mechanisms must be carefully designed if the quality and integrity of the data as evidence is to be ensured. Kenya's Evidence Act (Chapter 80 of the Laws of Kenya) provides for the admissibility of electronic/digital evidence in litigation. The Act, under Section 78A, Subsection 3, states that in estimating the weight to be attached to the evidence, regard shall be given to the reliability of the manner in which the electronic/digital evidence was generated, the reliability of the manner in which the integrity of the evidence was maintained, as well as the manner in which the originator of the evidence was identified.

8 M. Wangui, 'Billions lost to cybercrime in 2018: report', *The Kenya Wall Street*, 18 May 2019; Serianu, 'Sacco cybersecurity report 2018: demystifying cybersecurity for Saccos' (Nairobi: Serianu, 2018).

9 'Panic as customers allege KCB mobile accounts being hacked', *Business Today*, 7 March 2018.

10 M. KaKah, 'DTB official charged in Dusit attack case', *Business Daily*, 20 February 2019; R. Munguti, 'City M-Pesa dealer who "transferred" Sh34.7m to Dusit attackers to see a shrink', *Nairobi News*, 1 March 2019; F. Karanja, 'Over Sh100 million received by terror suspects, court heard yesterday', *Standard Digital*, 24 January 2019.

These requirements under the Evidence Act help to illustrate the relationship between data management and records management. There is a growing appreciation that banking data cannot be viewed in isolation from other forms of information generated by the processes supporting mobile banking. For instance, the processes controlling the transactions for depositing funds in a bank account generate data that represent money being deposited; the result of the deposit is reflected in the bank account. Statistics can be produced from the bank account data to describe, for instance, the amount of money deposited over a given time period. Records documenting the transactions involved in depositing the money are captured to enable both the individual and the bank to prove that the deposit was made. If evidence of the deposit is required, then the data, the records and even the statistics may need to be presented. These data, statistics and records all need to be managed based on a framework of policies, standards, practices, technologies and qualified professionals to provide evidence of sufficient quality and integrity to support accountability.

The overall design of the processes supporting mobile banking is similar to those supporting other processes in other sectors, such as the public and academic sectors. A transaction initiates a process (for instance completing a form to initiate the deposit of funds to an account or applying for a licence), and subsequent transactions carry out the process (such as depositing the funds or processing the application). The last transaction confirms that the process has been completed (notification that the deposit has been made, notification that the application has been approved or rejected).

The framework of policies, standards, practices, technologies and qualified professionals for ensuring the quality and integrity of banking data, statistics and records should not be restricted to a single sector, such as the private sector, nor should it be specific to a single discipline such as data management. It should apply to all sectors of society, including the public sector, and its design should draw on multiple disciplines, such as data management, the management of statistics and records management.

Processes should be the reference point for measuring the SDG indicators. Processes can extend beyond a single sector, as in the case of the close relationship between processes that the banking sector uses for managing banking transactions and processes that the government follows to manage the review of the banking sector. Measuring the relevant SDG indicators reliably will require an understanding of the processes followed in both the banking sector and the government, and a common framework of policies, standards, practices, technologies and qualified people will be needed to address the data, statistics and records that these processes generate. The challenge is to bring together representatives of the disciplines involved to develop such a framework and to ensure that it is in place.

Building bridges between the sectors

The examples from the government of Kenya and online banking in Kenya demonstrate that the lines between the public and private sectors are blurring in relation to the processes for measuring the SDGs. Many, if not most, of the domains covered, such as the environment, social and economic development, poverty reduction and health are supported by not one but multiple sectors. Frameworks for managing the quality and integrity of data, statistics and records need, therefore, to cut across sectors and across the information professions.

An important first step will be to bring together like-minded individuals from different disciplines, such as data management, the management of statistics and records management, as well as from different sectors, with each bringing different strengths, weaknesses and biases. Those from the banking sector in Kenya, for instance, would bring their knowledge of the management of financial data and statistics but probably do not have substantial knowledge about how to capture and maintain records documenting the processes that generate data and statistics or that document their characteristics. They may also lack knowledge and expertise about how data and statistics can be preserved through the long term, for instance through the life of the SDG initiative. Here, gaps in knowledge and expertise would be likely to be filled by records managers.

Conversely, records managers might lack sufficient knowledge of what it means to ensure the quality and integrity of data and statistics, especially those generated in complex financial information systems, such as those supporting online banking. This is where the knowledge and expertise of data management specialists and statisticians would become important in filling the gaps. In the case of online banking, for instance, data management specialists from private sector organisations working to protect valuable banking data from criminal activity could be brought together with records managers from relevant ministries, such as the finance ministry, to develop strategies for blending the knowledge and expertise of both sets of professionals. By collaborating, the different disciplines would be able to:

- define the scope and characteristics of the processes supporting banking activities across private and public sector boundaries
- identify where data, statistics and records are being generated in the processes
- identify points in the processes where records should be created to document processes as well as data and statistics themselves
- analyse the threats and risks to the quality and integrity of both the processes and the data, statistics and records
- develop a framework of policies, standards, practices and technologies to address the threats and risks.

The framework should reflect the integration of policies, standards, practices and technologies developed in support of the management of data, statistics and records. The result would be a set of processes that generate data, statistics and records with sufficient quality and integrity to meet a wide range of requirements, including measuring the SDG indicators through time. It would respect the cross-sector scope of the processes and present an integrated view of data, statistics and records.

Achieving this is no longer just a nice idea to consider: it is becoming an imperative. Poor quality information can seriously undermine the success of national and international development initiatives, as the SDG initiative demonstrates. It is not just a question of not meeting the goals; it is a question of not knowing whether or not the goals have been met. Not knowing, as a result of poor-quality data, can leave a nation vulnerable to poor resource allocation decisions, missed opportunities and corruption.

Building on the public and private sector examples set out in this chapter, the following steps could provide a practical way forward:

- assign responsibility for coordinating private and public sector strategies for enhancing the quality and integrity of data, statistics and records to one organisation. This would mean that resources could be used efficiently, and it would provide private and public sector organisations with one point of reference for strengthening the quality and integrity of data, statistics and records across the sectors
- convene a series of stakeholders' meetings to formulate a national policy on the management of data, statistics and records in relation to ICTs and automated information systems. This would build on the efforts to develop a records policy for the Kenya government
- develop policies, standards and practices for managing data, statistics and records that can be integrated in the strategic and operational plans for ICT projects
- develop guidance materials, standards and policies to be used in public and private sector institutions. These could be developed by consortia of public and private sector organisations, drawing on expertise in both sectors and on international standards for managing the quality and integrity of data, statistics and records
- formulate guidelines and standards for managing records converted to digital form through digitisation and related initiatives. These could be produced by the Kenya National Archives and Documentation Service in collaboration with other private and public sector stakeholders and the Kenya Bureau of Standards
- align systems for managing data, statistics and records with business processes

- develop strategies for measuring compliance with policies, standards and practices spanning public and private sector organisations to ensure that organisations, regardless of their type, can be held accountable for acts of negligence involving records.

Conclusion

This chapter has drawn on examples from the public and private sectors in Kenya to illustrate the extent to which issues associated with managing the quality and integrity of data, statistics and records are common across different sectors. Common issues need to be addressed through a common framework of policies, standards, practices, technologies and people. Rather than build solutions separately, the chapter suggests that organisations from the various relevant sectors could work together to build common solutions and strategies that both use resources efficiently and support quality and integrity.

This is especially important in the case of the SDGs, because measuring, and ultimately achieving, many, if not most of the goals, will depend on how effectively the relevant organisations can manage their data, statistics and records as a whole. Collaboration based on a shared understanding of the strengths and weaknesses of each sector and on a willingness to work together to develop shared strategies and solutions, will contribute enormously to achieving the SDGs.

7. Open data and records management – activating public engagement to improve information: case studies from Sierra Leone and Cambodia

Katherine Townsend, Tamba Lamin, Amadu Massally and Pyrou Chung

Open data initiatives support transparency, innovation, the promotion of a knowledge-based society and the advancement of democratic principles. Data in the hands of citizens can facilitate empowerment and support improved government efficiency and accountability. Open data promotes transparency by enabling citizens to reanalyse data underpinning government decisions and to monitor the impact of government policies. Citizens with access to the same government data used by policy-makers are more informed and better able to participate in and contribute to policy-making. Through their access to the administrative data generated by government, they are also able to identify incidents of corruption. For instance, citizen advocacy groups could potentially draw upon and analyse data derived from government payroll records, budgetary records, teacher employment records and other sources to assess the level of corruption in hiring teachers.

Open data provides an excellent vehicle for civic engagement, for information sharing, for rapid response and for supporting citizens' rights. The quality and integrity of the data underpinning open data initiatives and the ability to trace decisions documenting how the data were collected, processed and manipulated is key to achieving these benefits through time. The data management community is committed to the goal of achieving high-quality data, especially in the context of the Sustainable Development Goals (SDGs), but the role and importance of records management in providing evidence of the quality and integrity of the data through time has not yet received adequate attention.

In an open data environment, citizens should be able to trust that their governments are providing data with sufficient quality and integrity through an open data initiative that they can use it confidently. If it cannot be demonstrated that the processes used to produce the data and the data themselves are trustworthy, citizens' trust of the government can be damaged. Records, if well-managed, can provide the evidence needed to have confidence in the data. The role of records, and the steps required to manage them effectively, are

only beginning to be appreciated by those implementing open data initiatives. This chapter draws on examples from Sierra Leone and Cambodia to illustrate the positive impact of open data initiatives for citizens and, at the same time, the role that records management can play in ensuring that data quality and integrity can be demonstrated through time.

The example from Sierra Leone focuses on the goal of achieving free and fair elections, while the Cambodian example concentrates on successful land investment mapping. Each of these examples begins with an overview of an open data initiative and its role in advancing democratic values, knowledge dissemination and accountability. Both then go on to explore the extent to which records management can help strengthen the processes and the data produced and how a high level of quality and integrity can be sustained through time. A concluding section uses the two examples to consider the nature of the potential relationship between the open data and records management communities and the benefits for maximising the value of open data initiatives for citizens and governments alike.

Sierra Leone

Open data in support of free and fair elections

Sierra Leone is a small country in West Africa with seven million people. The nation has held elections every five years since 1996. In addition to the National Electoral Commission (NEC), there is a civil society organisation, the National Elections Watch, that aims to represent the people and to watch over NEC actions. Sierra Leone also has a Right to Access Information Commission (RAIC) that is responsible for making data and government transparent, including elections data. RAIC hosts an Open Data Council, comprising representatives from various private sector organisations, government agencies, academia and NGOs to make public sector data available and useable. Support for developing the RAIC and establishing the Open Data Council has been driven in large part by the need of the people of Sierra Leone to access government data and records and hold the public sector accountable.

At the time of the 2017 national elections in Sierra Leone, the government's open data portal had been down for months. The data were not being updated, and concerned and frustrated citizens and private sector organisations had decided to work together to start a parallel open data portal for anyone to use. This duplication caused some contention, with the government perceiving that its role had been overtaken. Nevertheless, discussion and debate occurred on a diversely represented WhatsApp group of approximately 250 professionals, journalists, government representatives and international actors.

Initially, the government wanted the alternative website taken down on the grounds that it should be the sole arbiter of open data. However, ultimately, the need for a consistent portal prevailed. The government had to bow to strong public opinion; as it was not providing the service itself and could not identify the harm being done by someone else playing this role, the site should continue. One company, LAM-TECH Consulting, which has now rebranded itself as TpISENT (SL) Limited, took on the primary role of hosting data and updating it with new datasets. After running the portal for several years, the team decided to focus on election monitoring and established the Sierra Leone Open Elections Data Portal (SLOEDP).

The portal is a resource that makes it possible for anyone to collect, aggregate, share and socialise elections data in an open format. The data adhere to the National Democratic Institute's open elections data format, which follows nine principles, namely that the data are timely, granular (at the finest level of detail possible), available for free on the internet, complete and in bulk, analysable, non-proprietary, non-discriminatory, licence-free and permanently available.¹ Unusual among many open data initiatives in developing economies, this movement began without any global donor funding but rather through people coming together through volunteer effort, self-funding and small-scale crowdfunding. The result was a solution sponsored and provided by Sierra Leoneans, for Sierra Leoneans.

In Sierra Leone, traditionally predictions of election results have been made by individual parties and candidates, and election results also are announced by the candidate and party. Misinformation, mismatched results and confusion had led to charged debate and even violence and bloodshed. SLOEDP's primary goal was to reduce violence by making information more accessible and easier to understand and trust. Its team included individual volunteers and organisations, all of them invested in election monitoring in Sierra Leone with the aim of preventing or at least reducing violence during the election cycle. They were well aware that violence would reduce the prospects for long-term, durable peace and stability and would undermine economic growth by limiting the purchasing power of citizens.

A strong WhatsApp group developed, including members of the Federation of Civil Society and Media Organisations (NaFCSMO-SL), Democracy and Development Associates – Sierra Leone (DADA-SL), the Open Government Initiative (OGI), the Women's Situation Room – Sierra Leone (WSRSL) and Njala University students. For instance, WSRSL, a women-led approach to preventing and reducing violence during the electoral cycle, was committed to reducing cases of violence, particularly sexual violence, and increasing the

1 <http://www.openelectiondata.net/en/guide/principles/>.

number of women in electoral processes. Even after the election, discussions among the group continued to be open and transparent, with little or no indication of censorship. Tamba Lamin, an experienced Sierra Leonian business analyst, site builder, trainer, content manager and passionate supporter of open data, described the group as a place, 'where we talk publicly about what has happened and share what we feel'. All parties are fully aware of each other's activities and of SLOEDP's efforts to make the data public, including converting PDFs to machine-readable data.

A major example of open data driving public recordkeeping involved the availability of the list of candidates running for office. The list was not publicly available online. According to SLOEDP, the NEC had advised, via Twitter, that the majority of the population was illiterate and did not have access to the internet. The more effective approach, the NEC advised, would be to post the list on a wall at polling stations. However, SLOEDP, the civil society platform, then took the initiative to submit what is believed to be the first freedom of information (FOI) request ever made in Sierra Leone to discover NEC's election records.

When the response was slow, Tamba Lamin went to Twitter and asked a CNN reporter to raise the issue with the NEC directly. Two days later, following a series of publicly viewed tweets, the NEC did respond to SLOEDP's FOI request with the count and voter roll of each of the stations. SLOEDP posted the information on its own website and shared it on Twitter as well as through shared WhatsApp groups. Shortly afterward, the NEC posted the information on its own website.

To effectively monitor the election, the team at SLOEDP introduced the ingenious process of training and paying motorcyclists to take photographs of each polling station and post them via multiple WhatsApp groups. Managing 16 WhatsApp groups, one for each district, it aggregated more than 10,000 snapshots of the actual results of each station. SLOEDP's methodology and capacity meant that it could cover more than presidential elections and could, moreover, produce results within 24–48 hours, as opposed to the week that the NEC required. The team at SLOEDP suspected that the NEC favoured a single political party, and that NEW (National Election Watch), which should have been an independent representative of civil society needs, had aligned too closely with the NEC. On election day, the NEC launched a new website with most of its historical content gone.

The value of public spaces for discourse on contentious issues cannot be overstated. Sierra Leone's Open Data Collaboratives WhatsApp group is immensely popular. It has reached its capacity of 250 persons, with many more waiting to join. The forum is a true marketplace of ideas, which draws together open data players as one forum to explore civic issues and how to solve them.

This group has been the most engaged and active of Sierra Leone's WhatsApp groups, with the most robust discussions. The smaller WhatsApp groups, which are focused on different geographic divisions within Sierra Leone, have also been vital in providing support, answering questions and facilitating coordination among the group monitors. SLOEDP has made its platform and methodology available for anyone to use via an open source licence, which has meant that anyone can use the platform, provided that they make their findings and any improvements and modifications publicly available.

In 2018, 23 elections were held across Africa. Similar efforts and initiatives for civil society-driven monitoring and data collection to help in election monitoring have occurred across the continent, from Nigeria, to Tunisia, to Kenya and more.² With a greater commitment to open data, ideally driven by government but with leadership and initiative from civil society, historical records of elections can be produced to help support fair elections and better systems for running them. With election monitoring more transparent, more public involvement during the election cycle, greater knowledge about candidates and issues, and larger voting turnout, more peaceful, trustworthy outcomes should become the norm.

The potential records management contribution

SLOEDP has had significant success as a grassroots initiative emerging from citizens' efforts, and WhatsApp groups have flourished because individuals and various interest groups have seen the value of collaboration and used available technology to support their common communication objectives. They have tended to assume that the elections data and the WhatsApp communications can be trusted because they were designed and managed by individuals and groups with a stake in their success; the high level of trust in the quality and integrity of elections data makes sense in relation to the existing data. What is open to question, however, is the extent to which such a high level of trust can be sustained through the long term. It will be important for those managing the applications to be able to demonstrate, not just now, but at any given point in the future, that the data generated and used in an open data application such as that developed by SLOEDP can be trusted.

Examples of questions that can be addressed from a records management perspective are:

- trust in the government's portal was eroded considerably and irrevocably when citizens discovered the NEC's portal was down and the data had not been updated. Have steps been taken to ensure this doesn't happen in the case of SLOEDP?

2 <http://www.eisa.org.za/calendar2018.php> and <http://www.ifes.org/news/elections-watch-2018>.

- is the new citizen-driven portal based on generally accepted standards that focus on ensuring the quality and integrity of the data and the processes supporting the portal? Can the quality and integrity of the data be sustained through time? Do the standards address the kinds of records that will be needed to document the data and supporting processes so that evidence of data quality and integrity can be assessed through time?
- what steps will those managing SLOEDP take to address the National Democratic Institute's principle on 'permanently available'? How will the elections data and the records documenting their characteristics be preserved in an accessible manner through time in spite of changes to the technology? What policies and standards will be needed to preserve the WhatsApp communications given that they will provide an important resource for future research?
- how do the SLOEDP data relate to the data generated by the NEC? Are records in place to document the relationship so that future users will be able to discern the difference?
- what training will be needed for SLOEDP volunteers and organisations to equip them to manage the quality and integrity of the data and records effectively through time?
- are records in place to document the methods used to capture, organise and maintain the photos that SLOEDP volunteers have taken and to demonstrate their quality and integrity?
- are governance structures in place to ensure that accountability is assigned for the quality and integrity of the data generated and collected by the WhatsApp groups and by SLOEDP?

The answers to these and related questions will help guide what needs to be in place to ensure that the data are sustainable through time. The availability of authentic, complete and accurate records that can serve as evidence of the quality and integrity of the data will be fundamental to the answers.

Lower Mekong, Cambodia: land investment mapping

The open data initiative

As in Sierra Leone, Cambodia's approach to improving the availability of its public information illustrates how progress can be driven by an ambitious civil society. Open Development Cambodia (ODC) established a website in 2011,³ to compile as many public resources as possible about government activities and international organisations that contribute to Cambodia's development.

3 <http://www.opendevdevelopmentcambodia.net>.

The site pulls information from academia, newsrooms, the private sector, local and international non-government organisations, and government resources. As more and more academics and international institutions cite ODC for research and for work, pressure has increased on the Cambodia government to provide more consistent information on its own activities. When applied to a targeted issue where there is a recognised need for accountability, and when these documents are open and available to the public, the effect is transformational, as has been the case with land investment information across the Lower Mekong.

The Lower Mekong Region, comprising Cambodia, Myanmar, Lao PDR, Vietnam and Thailand, has been experiencing rapid and unfettered development that is transforming the region and these countries' economies, while fundamentally changing the region's environmental landscape. The majority of the population is composed of smallholder farmers, fishermen and Indigenous forest communities who depend for their livelihoods on the Mekong River, the adjacent land and the rich natural resources of the Mekong ecosystem. The governments in the area rely heavily on an economic development model that depends on exploiting the land and natural resources for economic gain, which places economic development at odds with the local communities as they lose access to their traditional resources. The situation has been exacerbated by land tenure systems that are in transition from customary and communal use based on possession rights to various titling schemes in the different countries. These fluctuations make the poor especially vulnerable. Civil conflict in some of the countries has added a layer of complexity to the already fragile institutional and social framework that supports land-focused development.

When the ODC platform was launched in 2011, there was poor access to information, and ineffective public participation processes, where they existed at all, had intensified the situation. Publicly available data relevant to development were limited and difficult to access or to track systematically, which created difficulties, both for citizens and for the decision-makers themselves. Even in countries where some data on economic development were available, they tended to be generated and controlled by the governments, donors or the private sector, so that decision-making was not transparent. The result was rapid environmental changes with significant implications for both local communities and biodiversity.

Today, the initiative has six sites: five national level sites, one each in Cambodia, Lao PDR, Myanmar, Thailand, Vietnam, and a regional level site for the Mekong. The Open Development platform aggregates, organises and presents a wide range of information, while conforming to open data principles.⁴ The

4 <http://www.opendevdevelopmentcambodia.net>; <http://www.openelectiondata.net/en/guide/principles/>.

data's usefulness is enhanced through maps, infographics and other visualisations and by being juxtaposed with related data. The Open Development platform has fostered an increase in public demand for information and has influenced the governments to provide it. In part, this has been due to the platform's objectivity. Data presented by advocacy groups have tended to be perceived as biased and to be discredited. The Open Development platform, however, provides the necessary combination of content, training and infrastructure to engender credibility. Objective information is presented by recognised and impartial sources, including governments in the Mekong Region, despite the fact that historically they have impeded access to information.

The ODC platform has targeted a wide public audience with the aim of developing greater awareness about the work of the Cambodia government and the actions of the international community invested in Cambodia. There has been a real effort to present the information in a way that anyone can access and understand, so that people can become more engaged in decisions that affect their own lives and welfare. ODC, which is the most mature site of the platform, provides good examples of how the open data initiative has functioned in practice. It has worked with a variety of stakeholders, including the government and NGOs, to pioneer open access to data on economic land concessions, mining and hydropower in the country.

The usefulness of these data, which are all associated with natural resource development contracts, has been enhanced by the way they have been presented. For instance, census data have been displayed across a period of years against a map showing the locations of economic land concessions. This has allowed users to see where local communities have declined or disappeared in relation to economic development. ODC has also reached out to local communities, journalists, university students and human rights NGOs and has trained them in how to use the datasets on the site. Through these activities, it has received valuable feedback from users on how to improve the site's usability and on new datasets that could be relevant to natural resource development, for instance data on environmental protection.

Data sourced from the government are presented in the same way as all other data presented on the site – openly and with context. Some of ODC's followers include government technocrats, who, in the past, often had limited access to information. ODC allows them to see how their plans relate to one another, not infrequently across siloed ministries with related goals. The site encourages ministries to be more forthcoming in sharing their data. For example, Cambodia's Ministry of Agriculture, Forestry and Fisheries (MAFF) increased its online information on economic land concessions from a few dozen to almost 100 after ODC and others set an example by publishing wider datasets.

ODC seeks to present data with context but without editorial comment. The intention is not to support analysis and opinions, but rather to provide resources for the public, for data journalists or for experts on particular issues so that they can provide their own perspectives and draw their own conclusions. Data are shared, whether sourced from the government or elsewhere, and are perceived as credible and objective across sectors, without being provocative or biased. The website has remained available online, regardless of changes in government or policy.

ODC has developed resources to make the data easier to understand and more attractive to a wider audience. For example, it developed an animated map showing the rate of the decrease in forest cover in Cambodia through time. The data used were not new, but the method of presentation, in the Cambodian national language, Khmer, as well as in English, was. It shows clearly the discrepancy between policy and reality. The launch of the map was covered by two major Cambodian newspapers, and within weeks, it reached almost 2,000 users, with almost a third of them able to access it in Khmer. The release of this information triggered action by local, national and international organisations, and as a result, the government was required to account for its decision-making. Eventually it began working with ODC to create an updated forest cover map.

The potential for a records management contribution

As was the case in Sierra Leone, the ODC platform was developed in response to shortcomings in the government's ability to provide easy and timely access to information that had real value to citizens – in this case, environmental data for the Mekong Region. The ODC's success is reflected in the steady growth of its holdings, the high-quality and highly relevant data that it collects, maintains and makes available, and the increasingly diverse audience of users and contributors.

Nevertheless, ensuring the quality and integrity of the data and being able to prove their trustworthiness through time will inevitably present challenges, especially given the growing diversity in the types of data being collected, the potential for mashing up data from related and diverse sources, and the increase in the number of organisations participating in the initiative. In the future, there are likely to be questions about the ODC platform's ability to demonstrate through time, the quality and integrity of the data as well as the processes for supporting data collection, use and maintenance. The kinds of records management questions that could helpfully be raised are:

- how can the Open Data platform demonstrate through time that it is able to present objective information by recognised and impartial sources? What policies, standards, practices are required to support the

quality and integrity of ODC data now and in the future, and what records need to be in place to provide evidence of the level of objectivity and impartiality through the long term?

- one of the open data principles is 'permanently available'.⁵ How will the data managed on the Open Data platform and the records documenting their characteristics be preserved in an accessible manner through time in spite of changes to the technology?
- disseminating data beyond a core audience, as for instance to the MAFF, requires that the data can be understood and that their integrity can be demonstrated to the new audience. How can the level of data integrity be documented reliably?
- combining or 'mashing' data, as in the case of census data being mapped onto digital economic land concession maps, must be handled carefully since dissimilar data sets will be based on different standards. Are records in place to document how the data were assembled, manipulated, mashed up with other data and displayed?
- does the governance structure in place for the ODC platform assign accountability for the quality and integrity of the data and the records that document the data and the supporting processes?

As in the case of Sierra Leone, the answers to these and related questions will help guide what ought to be in place to ensure that the integrity of the data can be sustained through time. The availability of authentic, complete and accurate records can serve as evidence of the quality and integrity of the data as well as the processes that support its collection, processing and dissemination.

Key issues from the two case studies

The initiatives in Sierra Leone and in Cambodia that are described in this chapter illustrate the power of open data to promote democratic principles, enhance a knowledge-based society, stimulate the economy and fight corruption. At the same time, open data initiatives are not stand-alone projects. The data used in these initiatives have often been derived from, or are based on, data generated to support the administrative and operational activities of a government agency or some other participating organisation.

In the case of elections data in Sierra Leone, the results of the election and the election rolls generally will have been produced through a defined process carried out by an organisation mandated to administer elections, such as the NEC. The process would typically involve a sequence of steps, beginning with collecting the data and proceeding through data manipulation steps to produce the election rolls and election results. Various versions of the data, such as

5 <http://www.opendevelopmentcambodia.net>; <http://www.openelectiondata.net/en/guide/principles/>.

input data, verified and cleaned data, master edited data and published data will usually have been produced as a result of the process. Some data might be retained for short periods of time, while other data, with greater significance, might be retained far longer because of their significance.

The entire process should be supported by policies, procedures and technologies set up to administer the elections and manage both the process and the various forms of data that it generates. Records should be generated throughout the process to document the data, the steps involved and any decisions about how the process and the data were managed. Records will be necessary if evidence (or proof) is required regarding how the steps were carried out and that the data are reliable.

In the case of Cambodia, and based on the way that mapping systems are typically designed and managed, the process guiding map production would involve a sequence of steps beginning with collecting, processing, verifying and manipulating mapping data collected through field observations, satellite measurements or other processes and used to produce the digital base maps and economic, demographic, environmental and other data. Subsequent steps would involve merging the data, verifying their integrity and producing a range of digital and analogue maps on a variety of media (such as web, paper, digital media) to support a range of government objectives and respond to queries made by a wide range of users.

The process would normally generate an array of diverse data files. These would generally include raw cartographic data, source economic, demographic and environmental data, merged cartographic data files, verification and edit files, and analogue and digital files or products designed for access by the public, government officials and other interested groups. Throughout the process, records should be created to provide evidence that both the data and the process were managed properly, for instance that they were properly described, classified, retained, protected and preserved. The data and the process should be supported by effective quality and integrity controls.

The examples from Sierra Leone and Cambodia illustrate the importance of records in documenting the data, the sequence of process steps generating the various forms of the data and the decisions about how the data and the process were managed through time. When well-managed, records document the entire process, provide evidence of how the various steps were carried out, make it possible to assess the quality and integrity of the data generated at each stage and identify accountability for the data generated. The evidence that they provide should make it possible to demonstrate and manage the quality and integrity of the data and processes that produce them. By serving as authoritative, trusted sources of information, records can augment an open data initiative. They are an information source in themselves and a complementary component of any open data initiative.

Establishing a comprehensive approach to managing the quality and integrity of data in open data initiatives is challenging, especially when there are multiple players and disciplines involved (for instance, open data, data management, records management). A useful starting point would be to recognise that just as human and financial assets are managed by rigorously designed management frameworks, data provided through open data initiatives are a corporate asset that needs to be managed. From an asset management perspective, this will require laws and policies, standards and practices, systems and technologies, and qualified people, all geared to ensuring the quality and integrity of the data, the records and the processes.

Is such a framework too cumbersome and bureaucratic given the relatively small scale of a given open data initiative? Would it be enough to simply address quality and integrity issues in the context of the open data initiative? The answer depends on the level of acceptable risk: if the data generated to support the online mapping or the management of elections are flawed, then the mapping application or the Open Elections Data Portal will be flawed. What consequences would this have for the initiative, for its users or for the trust between the data providers and the data users? Such flaws can go undetected and can undermine what might otherwise be a healthy trust relationship. Once trust is eroded it is difficult to bring it back. A comprehensive framework for managing the data and records generated in the context of an entire process rather than just the process supporting the open data initiative would, through time, greatly reduce the risk that consumers will not trust the data that they access or receive.

Conclusion

There is no doubt that open data initiatives can empower communities, equipping them with knowledge of key issues that affect them. It is important that the data they provide should be accurate and trustworthy, not only in the present but that the information should remain reliable and accessible through time. Open data initiatives can serve as important catalysts for galvanising organisations to address long-standing data quality and integrity issues, presenting valuable opportunities for the data management and records management communities to work together to address not only the quality and integrity of open data but of the data generated by administrative and operational activities of the government itself.

Designing a comprehensive management framework for ensuring the quality and integrity of data and records and the processes that support an open data initiative needs considerable care. An interdisciplinary approach with common strategies can have substantial benefits for both communities and for the citizens that they serve. Ultimately, a coordinated approach can serve the dual purpose of activating public engagement to improve the use of information and protecting its quality, integrity and accessibility through time.

8. Assuring authenticity in public sector data: a case study of the Kenya Open Data Initiative

James Lowry

Measuring the Sustainable Development Goals (SDGs), and ultimately the success of the whole SDG initiative, will depend on the availability of authentic public sector data. By the time it reaches policy-makers, either as baseline data or as comparative data indicating progress, it will have been assembled through one or more methods from sources as varied as paper or digital records, management information systems, people and scientific instruments. After collection, it will be curated, cleaned, analysed, augmented and remixed. It will be subjected to formulas in spreadsheets, algorithms in apps and the intervention of people in various roles with various priorities and agendas. It will be distributed, published, cited and (it is hoped) preserved through numerous channels, platforms and systems.

At every stage, the authenticity of these data is potentially at risk. If public sector information is going to be available and authentic, there must be a regulatory environment and information culture that supports openness. From at least the 1960s, the international open government movement has been working towards laws, policies and standards that have supported the availability of public sector information, for instance through freedom of information laws. What it has not yet addressed is the need for technical and procedural controls to establish authenticity. This chapter argues that the principles and techniques developed over centuries in the field of recordkeeping for the purpose of assuring the authenticity of the records documenting decisions and actions also can be used to improve data quality, so that the information needed for implementing and monitoring the SDGs is not only available but authentic.

This chapter presents a case study of the Kenya Open Data Initiative (KODI). It examines the level of control presently in place for establishing and maintaining the authenticity of information released through KODI. While KODI is noteworthy as the first government open data portal in sub-Saharan

Africa, it is studied here only as an example; problems identified with KODI data can be seen in open government datasets worldwide, to varying extents.¹

This chapter begins by defining authenticity in information before providing an overview of KODI (<http://www.opendata.go.ke>). To identify issues relating to the authenticity of information released via KODI, the chapter then examines a KODI dataset relating to land use. This analysis is discussed in relation to Kenya's provisions for managing land information. The chapter goes on to describe the processes of preparing, publishing and using the land dataset. It maps the lifecycle of the data and identifies strengths and weaknesses in the controls for protecting the data's authenticity. Questions about authenticity raise questions about the data's contribution to implementing and monitoring the SDGs.

Data authenticity

In records and archives management literature, an authentic record is one 'that is what it purports to be and is free from tampering or corruption'.² Authenticity depends on 'integrity', which is the 'wholeness and soundness' of a record, and on 'identity' – 'the attributes of a record that uniquely characterize it and distinguish it from other records'.³ Over many centuries, records and archives professionals have developed principles and techniques for assuring information authenticity – from medieval chancery practices to technical standards for digital information management systems.

What these controls have in common is an emphasis on documenting the provenance and custodianship of records through metadata captured in auditable systems. For instance, the registry system used to control paperwork in the British empire required specific officers to capture specific information in registers and on file covers and flyleaves in such a way that traces of the registered record existed in multiple places and could not easily be erased or doctored. In this way, records became wrapped in metadata that described their management and their movement within and between government offices. The obligations on custodians and the system requirements were mutually reinforcing, so that custodians were encouraged to comply with and give effect to the system through its oversight mechanisms. These same principles can be

- 1 J. Lowry, 'Addressing information asymmetry in the social contract: an archival-diplomatic approach to open government data curation', unpublished PhD thesis, University College London (2019).
- 2 InterPARES 1 Authenticity Task Force, 'Appendix 2: requirements for assessing and maintaining the authenticity of electronic records', in InterPARES 1 Project, *The Long-Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (InterPARES Project, 2002), pp. 1–2.
- 3 InterPARES 1 Authenticity Task Force, 'Appendix 2'.

seen in standards for digital records management systems, such as MoReq and ICA-Req,⁴ which require that metadata be captured to document all actions in relation to records.⁵

The Kenya Open Data Initiative

In July 2011, the KODI portal was launched by President Kibaki to provide public access to Kenyan public sector information.⁶ This was a landmark moment in the history of state secrecy and openness in Kenya. Nevertheless, KODI experienced a number of challenges in its first years of operation, as the World Wide Web Foundation's Open Data Barometer (ODB) reports showed. The ODB ranks countries on three criteria:

- readiness: how prepared are governments for open data initiatives? What policies are in place?
- implementation: are governments putting their commitments into practice?
- impact: is open government data being used in ways that bring practical benefit?⁷

The ODB methodology draws on government self-assessment, peer-reviewed expert survey responses, detailed dataset assessments and secondary data.⁸ The ODB Regional Report for Africa (3rd edition, 2016) examined 21 sub-Saharan African countries, including Kenya.⁹ Key findings were:

- 1 Performance across the continent is relatively poor in comparison to leading countries in the Global South and globally. The report noted that Kenya 'does not publish a single, fully open dataset – health,

4 These are standards that set out functional requirements for digital records management systems. MoReq is the 'Model Requirements for the Management of Electronic Records' published in 2001 by the DLM Forum – a European network of government archives and information professionals – with funding from the European Commission. The current version of MoReq (MoReq2010) was published in 2011. ICA-Req is the International Council on Archives' Principles and Functional Requirements for Records in Electronic Environments, published in 2008. In 2010, ICA-Req was adopted by the International Standards Organization as ISO 16175.

5 Lowry, 'Addressing information asymmetry'.

6 T. Davies, *Open Data Policies and Practice: An International Comparison* (European Consortium for Political Research, 2014), p. 14, <https://ecpr.eu/Filestore/PaperProposal/d591e267-cbee-4d5d-b699-7d0bda633e2e.pdf>.

7 World Wide Web Foundation, Open Data Barometer: ODB Methodology – v1.0 28 April 2015, p. 3.

8 World Wide Web Foundation, Open Data Barometer, p. 3.

9 World Wide Web Foundation, Open Data Barometer, Third Edition, Regional Report, Africa, May 2016, p. 6, <http://opendatabarometer.org/doc/3rdEdition/ODB-3rdEdition-AfricaReport.pdf>.

- education and legislation data are open licensed but fall short of being fully open because the data are not available in bulk’.
- 2 A downward trend is common in the overall Barometer scores from 2013 to 2015. The assessment showed a drop in ten points in Kenya’s ODB score in this period.
 - 3 Open data initiatives lack long-term commitment and resources, resulting in short-term gains that are unsustainable. ODB data shows a drop in scores between 2014 and 2015 in all African countries except Nigeria and Cameroon. However, the report notes ‘While there is a net decline in the scores for Kenya and Mauritius between 2013 and 2015, there is a recovery in their overall scores between 2014 and 2015’.
 - 4 ODB implementation scores are lower than readiness scores. The ODB’s comparison of readiness and implementation scores show a consistent gap in the case of Kenya, so that both scores increase and decrease in parallel, with no sign of the gap closing.
 - 5 There is no stand-out performer in Africa. The report states that ‘Africa is the only region without a clear open data champion ... In previous editions, Ghana and Kenya looked likely to assume this role, but the data show that the performance of these countries is erratic’. The ODB Global Report acknowledges that Kenya and Ghana were in ‘a holding pattern as they try to revamp their initiatives’.¹⁰

Within the ICT Authority, which is responsible for KODI, there is a good deal of enthusiasm for improving and expanding the open data programme.¹¹ Members of the KODI team participate in the Data Science Africa research network and its annual conferences.¹² The KODI team has a network of ‘fellows’ across the public sector who identify relevant datasets, and there is a Data Science Team that cleans data released by government agencies.¹³ Although the African Data Consensus had not yet been ratified by Kenya at the time of writing,¹⁴ the staff of the ICT Authority recognised the challenges it identified, which informed the priorities for KODI’s development.¹⁵

KODI staff are aware of problems with the data they receive for publication through the portal. They cite many of the same problems reported by civil society and the ODB, including the questionable accuracy and completeness

10 World Wide Web Foundation, *Open Data Barometer*, 3rd edn, p. 30.

11 Interview with Sifa Mawiyoo, open data specialist and GIS technologist, Kenya Open Data Initiative (KODI), ICT Authority, and Prestone Adie, data analyst, KODI, ICT Authority, Nairobi, 20 September 2016.

12 Data Science Africa, <http://www.datascienceafrica.org/>.

13 Interview with Sifa Mawiyoo, 20 September 2016.

14 Interview with Sandra Musoga, senior programs officer – Transparency, Article 19, Nairobi, Kenya, 16 September 2016.

15 Interview with Sifa Mawiyoo, 20 September 2016.

of the data, the risk of introducing errors during the cleaning and curation of data, problems of timeliness and infrequent updates, and ongoing resistance to data release within the public sector because of the long-standing culture of secrecy in the Kenyan government. Several of the staff noted in an interview in September 2016 that ‘the quantity of data received does not seem to reflect the enactment of the FOI law’; though the law had only just received Presidential Assent, there was a sense that it should have resulted in an increase in proactive information release.¹⁶ At the time of our interview, KODI team members had not considered the issue of data authenticity and the significance of contextual information, but they noted that they ‘hardly ever get data with metadata’.¹⁷

Nevertheless, KODI staff have put in place a number of quality control measures. They take a snapshot of the portal with every new data upload. They keep copies of original datasets received from ministries, departments and agencies, so that sources can be checked in the event that the curated data are queried. In addition, they have created a ‘data release calendar’ that they use to schedule and monitor updates, and they have created templates for datasets. All of this encourages more complete and consistent data.¹⁸

Land data

Examining a sample dataset from KODI should help to illustrate what controls and processes for authenticity are in place. In view of the significance of land in Kenyan political and economic life, a dataset relating to land use is examined here in relation to Kenya’s approach to land information management.

Land information management

Kenya, which was among the ‘first [independent African] countries to experience comprehensive land reform’,¹⁹ has experienced problems in managing information about land and related resources. For instance, in October 2015, the Thomson Reuters Foundation reported, in relation to the Kenya Groundwater Mapping Programme:

One key problem is lack of data ... According to the Kenya Water Industry Association, not one of the country’s several water regulation agencies, including the Water Resources Management Authority, has reliable data that captures the distribution, quantity and quality of available groundwater.²⁰

16 Interview with Sifa Mawiyoo, 20 September 2016.

17 Interview with Sifa Mawiyoo, 20 September 2016.

18 Interview with Sifa Mawiyoo, 20 September 2016.

19 J. Herbst, *States and Power in Africa: Comparative Lessons in Authority and Control* (Princeton: Princeton University Press, 2000), p. 185.

20 M. Waruru, ‘To arm against drought, Kenya maps its water resources’, Thomson Reuters Foundation, <http://news.trust.org/item/20151030082053-5dgn/>.

Since the colonial period, Kenya's land management system has functioned largely through the creation, transmission and exchange of paper records. Under the Registration of Titles Act (1982), a Central Registry was established in Nairobi and a Coastal Registry in Mombasa, for managing paper land registration records. Under the Registered Land Act (1989), a registry was established in every land registration district. The 2012 Registration Act aimed to rationalise and devolve the land registration process but, at the time of writing, regulations were still being developed. Recordkeeping in Kenya has faced numerous challenges, including a lack of cohesive policies, lack of compliance with procedures, ad hoc systems and lack of adequate staffing and other resources.²¹

In 2010 and 2011, during research into the readiness of Kenyan government recordkeeping for e-government and freedom of information, Justus Wamukoya and I found that what was then the Ministry of Lands (now the Ministry of Lands and Physical Planning) had experienced a period of recordkeeping reform.²² At the time of our study, its registries were well-functioning and monitored, and sanctions were imposed for infringements:

Records management is not audited, but when breaches of records management procedure are identified, they are investigated. At the time of the interview a member of the registry staff was on suspension for removing a file that he was not permitted to access.²³

At that time, a digitisation project was underway, motivated by a sense that computerisation and digitisation would, among other benefits, reduce delays in work processes. These, according to one member of staff, accounted for the great majority of the written complaints that the ministry received.²⁴ The digitisation project was led by the Land Management Systems Technical Working Group in the Lands Reform Unit, which oversaw the ministry's target

- 21 Administrative histories of recordkeeping in sub-Saharan Africa are limited, but for an overview of the problems that are common across many of those countries, see J. Wamukoya, 'Records management and governance in Africa in the digital age' and N. Mnjama, 'Anne Thurston and record-keeping reform in Commonwealth Africa', in J. Lowry and J. Wamukoya (eds), *Integrity in Government through Records Management* (Farnham: Ashgate, 2014). The roots of these problems are to be found in the colonial period. Again, this history is under-researched, though the colonial origins of more recent recordkeeping problems are noted by M. Musembi, 'Development of archive services in East Africa', in *Historical Development of Archival Services in Eastern and Southern Africa: Proceedings of the 9th Biennial General Conference* (Mbabane; Rome: ESCARBICA, 1986), p. 116.
- 22 International Records Management Trust, *Aligning Records Management with ICT, e-Government and Freedom of Information in East Africa*, Kenya Country Report, p. 13, <http://www.irmt.org/portfolio/managing-records-reliable-evidence-ict-e-government-freedom-information-east-africa-2010-2011>.
- 23 International Records Management Trust, *Aligning Records Management*.
- 24 International Records Management Trust, *Aligning Records Management*.

under the government-wide Vision 2030 strategy in relation to improving land title acquisition.²⁵

Wamukoya and I found that while records staff were confident that the organisation of paper records was adequate for land title process improvements, members of the working group had identified significant gaps in the paper records being digitised. Digital surrogates were comprehensively and regularly backed up, but no digital preservation measures had been developed and no consideration had been given to the need to eventually transfer digital records to the Kenya National Archives and Documentation Service (KNADS). Moreover, there seemed to be no planning for moving from creating and scanning paper records to managing born-digital records.²⁶ When I again visited the Ministry of Lands and Physical Planning in September 2016, paper records were still being digitised, including maps. The staff could conduct online searches for Nairobi, and the ministry was setting up a central database of land titles as part of the Kenya National Spatial Data Infrastructure, which was expected to expand the ministry's capacity to conduct online searches.²⁷

Today, recordkeeping practices in the ministry continue to be guided by the standard public service manual on records management, with no separate guidance on managing land title records.²⁸ The Records Management Procedures Manual for the Public Service provides procedures for registering and managing records, with provisions for mail management, filing, indexing, cross-referencing, classification, file tracking, 'bring-up', storage, survey, appraisal and disposal (transfer to archives or destruction).²⁹ It also discusses disaster management, capacity building and the institutional framework for recordkeeping, including ministerial responsibility for compliance. The 'Security of Records' chapter states that access to classified records should be on a 'need to know' basis, reflecting the long-standing civil service bias towards secrecy. However, it also notes that 'Access to public records shall be provided within the existing legislative and regulatory framework', which, since 2016, includes freedom of information legislation.³⁰ The manual warns staff to 'guard against a natural tendency to over classify documents'.³¹

25 International Records Management Trust, *Aligning Records Management*.

26 International Records Management Trust, *Aligning Records Management*, p. 5.

27 Interview with Edward Kosgei, head of lands administration, and Emily Ndungi, principal records management officer, Lands Department, Ministry of Lands and Physical Planning, Nairobi, Kenya, 19 September 2016.

28 Interview with Edward Kosgei and Emily Ndungi.

29 Republic of Kenya, *Records Management Procedures Manual for the Public Service*, Office of the Prime Minister, Ministry of State for Public Service, May 2010.

30 Republic of Kenya, *Records Management Procedures Manual*, p. 50.

31 Republic of Kenya, *Records Management Procedures Manual*, p. 51.

There is also a chapter on digital records management that sets out responsibilities and provides guidance on naming conventions, media handling and storage. It empowers the National Archives (KNADS) to set procedures for digital records management and to authorise the destruction of digital records. It states that all public institutions are required to install the Integrated Records Management System (IRMS) developed and introduced by the Personnel Office in the Ministry of Public Service. Although the IRMS has not been audited against international standards such as ICA-Req or MoReq, it covers core records system functionalities relating to registration, file tracking, 'bring up' and reporting.³² It does not, however, address the need for a system to preserve born-digital records or digital surrogates. Kenya's second Open Government Partnership National Action Plan included a commitment to establish 'a central digital repository to provide lasting access to government records and data and all information of public interest' by June 2018.³³ However, this has not yet happened.

Examining the land dataset

Although authentic information is crucially important for the SDG initiative, difficulties in accessing or understanding data can make it virtually inaccessible to potential users, including government and civil society actors involved in SDG work. Kenyan open data specialists, Leonida Mutuku and Christine Mahihu, have reported that 'data quality' is a key issue for open data in Kenya. They have highlighted low relevance of data to citizens, the irregularity of data updates or dataset releases and the questionable utility of the data (for instance, incomplete data and data that are poorly structured or formatted) as key issues.³⁴ However, their study did not examine authenticity when considering data quality.

In the field of recordkeeping, authenticity requires that metadata, documented provenance and custodianship and auditable systems work in concert. For the purpose of examining how these three prerequisites are applied in producing and publishing data released through KODI, I studied the dataset 'Proportion of Parcels Using Fertiliser 2006'. Not only is the dataset relevant to the contentious issue of land use, but the fact that it was accompanied by some contextual information suggested that it might be possible to determine if the prerequisites for authenticity were addressed in its production and publication. The dataset was uploaded by a user called 'Knoema' in 2015 and the original

32 International Records Management Trust, *Aligning Records Management*, p. 9.

33 Republic of Kenya, OGP National Action Plan 2.

34 Mutuku and Mahihu, *Open Data in Developing Countries* (iHub, 2014), p. 47.

data source was identified in its metadata as the Kenya National Bureau of Statistics.

I first looked at the dataset in October 2016. Returning to KODI on 11 January 2017, the site rendered a login interface (see Figure 8.1). Creating an account and signing in brought me to an error message (Figure 8.2) indicating that permission was now required to access the data.

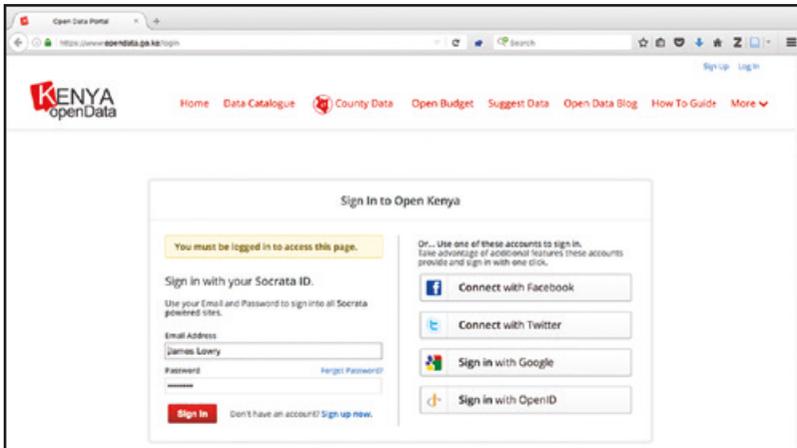


Figure 8.1. KODI interface on 11 January 2017



Figure 8.2. KODI sign-in error message

On querying the staff of the ICT Authority, I was directed to use a mirror site while the portal was migrated between service providers. A search of the interim platform for the title of the dataset on 16 January 2017 produced no results. A search for ‘parcels’ produced a short list of results that included a dataset called ‘Proportion of Parcels Using Fertiliser County Estimates 2005/6’. The metadata for the dataset indicated that the dataset was uploaded by ‘kodipublisher’ rather than ‘Knoema’, with a publication date of 20 December 2016, which was probably the date of migration to the interim platform. There was nothing to indicate that the dataset was first published in 2015.

The dataset, viewed as a CSV file, comprised seven columns:

- A – object ID (a sequence of ascending numbers from 1 to 47)
- B – ‘county_name’ (a list of county names)
- C – ‘proportion_of_parcels_using_f’ (the same list of county names)
- D – ‘proportion_of_parcels_using_1’
- E – ‘proportion_of_parcels_using_2’
- F – ‘proportion_of_parcels_using_3’
- G – ‘proportion_of_parcels_using_4’

Columns D to F provide figures ranging from 0 to 0.94. Column G provides figures ranging from 0 to 428581.8. There is no data or metadata within the CSV file to help interpret these figures, such as the meaning of ‘using_1’, ‘using_2’, etc., and no formulas that provide a key to the relationship between the figures in columns D through F and the figures in column G. Columns D through F for Wajir County each have a value of 0, as does column G. Columns D through F for Mombasa County each have a value of 0, but column G has a value of 2425.2. This means that the content, context and structure of the dataset are insufficient for the average user to interpret the data.

Instead, with regard to this particular dataset, KODI does the work of interpretation for its users by visualising the data as a map. The portal includes three tabs: ‘Overview’, ‘Data’ and ‘Visualization’;³⁵

- ‘Overview’ provides basic metadata about the dataset: title, publisher, last modified date (but no other dates related to events in the life of the data), licensing information; it attributes the dataset to the Kenya National Bureau of Statistics
- viewing the ‘Data’ tab produces a generic error message (“There was an error”)
- the ‘Visualization’ tab allows users to generate a map of Kenya’s counties coloured to show the proportion of parcels using fertiliser. To the general user, this visualisation feature is necessary to understand the dataset.

35 Kenya Open Data Initiative (KODI), <http://www.opendata.go.ke>.

The ‘Proportion of Parcels Using Fertiliser County Estimates 2005/6’ dataset lacks essential assurances of authenticity. In terms of metadata, the content of the CSV file itself lacks sufficient metadata to enable users to interpret the data. The fact that human interpretation of the CSV is not possible does not necessarily undermine an assumption of authenticity. However, since visualisation is necessary to understand the data, questions need to be asked about the mechanisms of visualisation. What formulas and algorithms are used to render the dataset as a map? The formulas and algorithms that render the data interpretable to users are a significant component of the system for managing the information and should be capable of being audited to support authenticity. However, there is no technical information available through KODI to help users understand the process that produces the map. This effectively constitutes a gap in the chain of custody.

Of more significance for demonstrating authenticity, KODI offers little metadata about the file itself. Looked at in isolation, this dataset appears to have ‘identity’ (the attribute of a record that distinguishes it from other records).³⁶ This ‘identity’ is supplied by its metadata, in particular its unique title, author and dates. However, when compared with the dataset first viewed in October 2016, the discrepancy in publishers (Knoema and kodipublisher) and dates (uploaded 2015 and last modified 20 December 2016, without reference to an upload date) raises questions about the identity of the dataset. Is it the same dataset? If identity is called into question through inconsistent metadata, one of the two fundamental elements of authenticity (identity, with integrity) is absent.

In addition, the provenance of the dataset is obscured by the lack of metadata documenting the custodianship of the data from the point of collection. Working backwards, there is metadata about the publisher of the data (though some ambiguity about the identity of the publisher when the two versions of the dataset are considered together) and about the source of the dataset, which is given as the Kenya National Bureau of Statistics, but it is not possible to see the sources that the Bureau used.

In this dataset, there are two types of data:

- information about the boundaries of the land parcels, which depends on information generated by the land registration process conducted by the Ministry of Lands and Physical Planning
- estimates of fertiliser use, which are likely to come from the Ministry of Agriculture, Livestock and Fisheries or one of its agencies.

36 InterPARES 2, Terminology Database, http://www.interpares.org/ip2/ip2_terminology_db.cfm.

To be assured of the integrity of the dataset it must be possible to know the sources of the data. In this case, assurance would require users to seek this information from the Bureau of Statistics. The recordkeeping system of the Ministry of Lands and Physical Planning functions well and conforms with the system set out in the manual, which, being government-wide, is also likely to guide the recordkeeping of the Ministry of Agriculture, Livestock and Fisheries.³⁷ The recordkeeping system documents custodianship throughout the lifecycle of the ministries' records, but KODI users do not have a way to be aware of this. Therefore, there is a break in the documentation of the chain of custody between the creation or capture of the data by the ministries and the aggregation of the data by the Bureau of Statistics. Moreover, the user does not know what processes and controls the Bureau followed to prepare the data and assign and document responsibility for the data.

The data pass through a series of systems, and these need to be auditable if authenticity is to be assured. From the point that it is created, and throughout its use within the ministry, land parcel information is managed through an auditable recordkeeping system. The data may then pass to the Ministry of Agriculture, Livestock and Fisheries, where they serve as a parameter for the ministry's estimates of fertiliser use. There is no metadata that would help KODI users understand this. Even at the point of aggregation, when the data passes into the Bureau of Statistics' systems, this is not transparent to KODI users.

In preparation for publication, the data then pass into the custody of KODI and its systems. Unlike the previous transitions, this is documented in publicly available metadata. As outlined above, KODI staff follow defined measures for protecting the integrity of data, and, while these measures are not brought together in a formal system that allows each action and custodian to be audited, they would, theoretically, enable published datasets to be compared with datasets as received by KODI. In this way, there is a basic accountability mechanism in place for KODI's data custodianship. However, there is no information about these controls on the KODI portal. Rather than assurances of integrity, KODI users have ambiguous information about the sources and treatment of the data.

Although 'Proportion of Parcels Using Fertiliser County Estimates 2005/6' may be authentic in the sense that it 'is what it purports to be', the user does not have the information needed to determine its authenticity. Partial metadata, opaque provenance, undocumented custody (particularly during aggregation)

37 International Records Management Trust, *Aligning Records Management*, pp. 12–13.

and the lack of information about the systems for its management, including its visualisation, introduce doubts about the data's identity and integrity.

Conclusion

The development of open data in Kenya has not yet been linked to records management. Connecting the two would help facilitate trust in opened datasets and support their survival through time. At present, there are problems with the data being released through KODI, and the ability to rely on their authenticity remains limited.

This study has charted the lifecycle of a specific dataset and found that, in general, strengthening the recordkeeping practices of public sector bodies lays a foundation for assuring authenticity in public sector data. This would be enhanced by the publication of recordkeeping audit reports documenting processes from the beginning of the data lifecycle. Later in the lifecycle, when data are aggregated and then published on KODI, authenticity is called into question by partial metadata about data's provenance and management. Leaving aside problems that may be unique to KODI's transition between platforms, there are unanswered questions about the management of the data throughout their lifecycle.

Authenticity is at the greatest risk at the point that data are aggregated. The data aggregation practices of the Bureau of Statistics are not documented for the public. In this space between creation and publication, provenance is obscured, and custody cannot be audited. After publication, most users will need to rely on KODI's visualisation of the dataset in order to interpret the data, making KODI an essential component of the data management system. Accountability requires that information management systems must be transparent, and custodianship must be public. At present, KODI's mechanisms for visualisation are not published. The chain of custody of the dataset appears to be broken in several places.

These issues all undermine assurances of the authenticity of the data. Not only is there a negative impact on government openness and community and commercial reuse, but there are repercussions for the ability to pursue the SDGs. Implementing and monitoring the goals depend on access to authentic data.

KODI is taking steps to build controls into its processes. It takes snapshots of the portal with every new upload, retains original datasets received from government and standardises templates for datasets. All of this contributes to building an audit trail of the accuracy and completeness of data and metadata. If these steps can be brought together, documented and published, KODI will have taken important steps towards assuring authenticity.

The lack of technical controls for information authenticity is not unique to Kenya – it is widespread, and arises from the disconnect between the communities of practice involved in open data on one hand and records management on the other.³⁸ To improve data quality generally and data authenticity in particular, records management principles and know-how need to be incorporated into open government data curation. Records management has supported previous efforts to establish government openness (particularly freedom of information), by organising and making information available for release. The major contribution of records management to open data is, however, not in providing records for data mining but in offering techniques for improving data quality by making datasets more like records.

38 International Records Management Trust, *Aligning Records Management*.

9. Preserving the digital evidence base for measuring the Sustainable Development Goals

Adrian Brown

The role of good recordkeeping in providing evidence for measuring the Sustainable Development Goals (SDGs) has been discussed in depth in Chapter 1 of this volume, ‘Records as evidence for measuring sustainable development in Africa’, as has the fact that data, documents and other forms of recorded information are increasingly digital in form. The need for digital preservation will only continue to grow, an issue that cannot be avoided if SDGs are to be met. Furthermore, the SDG agenda requires that these records remain accessible to 2030 and beyond, which, in the digital world, can only be achieved through continuous active management.

Digital preservation is not just a technical issue. It also requires an ecosystem of organisations, policies and standards, resources and people. Demonstrating whether or not SDGs have been achieved requires the management of the evidence base, which in turn is dependent on the application of digital preservation methodologies. The significance of developing digital preservation capacity for maintaining reliable SDG measurements over time, as a fundamental component to the SDG agenda, cannot be overstated.

The main challenges of collecting and preserving information specified by the SDGs in ways that will allow it to be combined and compared will be discussed further in David Giaretta’s chapter on ‘Preserving and using digitally encoded information as a foundation for sustainable development’ ([Chapter 10](#) in this volume).

This chapter considers the practical implications for developing digital preservation capabilities. It begins by considering the component elements of such a capability, and it examines how the concept of maturity models can be used to help organisations define models for digital preservation that are appropriate to their needs, as the chapters by Shepherd and McLeod and McDonald also emphasise. It then looks at the variety of models available for delivering a digital preservation service and concludes with a summary of the operational implications.

Elements of a digital preservation capability

There are many definitions of digital preservation, but that provided by the Digital Preservation Coalition is a good starting point:

Digital Preservation refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. [It]... refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technological and organisational change. Those materials may be records created during the day-to-day business of an organisation; “born-digital” materials created for a specific purpose (e.g. teaching resources); or the products of digitisation projects. This [definition] specifically excludes the potential use of digital technology to preserve the original artefacts through digitisation.¹

In an archival context, we can refine this definition to encompass the concept of authenticity which, as set out in ISO 15489,² depends on three essential characteristics:

- *reliability*: the record must be a full and accurate representation of the cultural or business activity to which it attests, for instance the management of finance, land or other resources. This depends upon trust in the regimes within which the record has been managed throughout its lifecycle and on the continuing ability to place it within its original context or the circumstances in which it was created. For digital archives, reliability is supported by adopting transparent and fully documented preservation strategies and processes, as well as by the existence of metadata (data describing the record’s content, its context and/ or the circumstances of its creation) and its provenance (or origin)
- *integrity*: the record must be protected against unauthorised or accidental alteration. For a digital archive, integrity is maintained through the process of bitstream preservation (see below), and through the existence of an audit trail for every action relating to the record’s management and preservation
- *usability*: the record must be capable of being accessed in meaningful form through time. It must therefore be discoverable and retrievable by authorised users, accessible, and interpretable within the current technical environment. Usability is ensured within a digital archive through the process of logical preservation (see below), and the existence of metadata sufficient to allow the record to be located, retrieved and interpreted.

1 Digital Preservation Coalition, *Digital Preservation Handbook*, 2nd edn (2015), <http://dpconline.org/handbook>.

2 ISO 15489-1: 2016 – Information and Documentation – Records Management – Part 1: Concepts and Principles.

Fundamentally, therefore, digital preservation involves acquiring authentic digital records, storing them and making them accessible to users for as long as they are required. Given this definition, we can begin to identify the requirements through which this goal can be achieved. At its most fundamental level, all digital information is encoded as a sequence of 1s and 0s – a bitstream – which may be written to a storage medium or transmitted across a network. These bitstreams have no intrinsic meaning but require layers of technology to transform them into meaningful information. Preserving meaningful digital information requires us to maintain both the underlying bitstream and the means to correctly interpret it. Digital preservation is, therefore, often considered to require two fundamental activities:

- *bitstream preservation*: the series of activities required to maintain the integrity of a bitstream, ensuring that a demonstrably bit-perfect copy can be retrieved on demand, for as long as required
- *logical preservation*: the series of activities required to ensure that bitstreams can continue to be rendered as meaningful information through time.

However, this is not a purely technical exercise. For any organisation to undertake digital preservation, in practice, it also needs broader capabilities, encompassing:

- *organisational viability*: the organisation must have the necessary organisational and governance structures in place, and commensurate resources to deliver a digital preservation service. It must also be able to demonstrate that it can expect to maintain that capability over the lifetime of the archives in its custody, or that it has a credible strategy to transfer those responsibilities to another organisation in future
- *stakeholder management*: it must be able to identify, understand and engage with its stakeholders, both within and beyond the organisation, including funders, service partners, content depositors and end users
- *legal basis*: the organisation must have an appropriate legal basis within which to operate, with the means to manage contracts, licensing and its other legal rights and responsibilities
- *policy framework*: the organisation must have suitable policies, strategies and procedures in place to govern its digital preservation operations. These should be subject to regular review
- *acquisition and ingest*: it must have the means to acquire and ingest (import) authentic digital content in accordance with a defined collections development policy, bringing it fully within its control
- *metadata management*: it must have the means to create and maintain all metadata required to support the management and use of that digital content, including preservation and reuse

- *dissemination*: it must provide the means for its designated user community to locate and use the digital content in its custody, in accordance with the applicable conditions of use
- *infrastructure*: the organisation must have access to the necessary physical and technical infrastructure to deliver the above capabilities, whether it owns and manages this directly or outsources it to a third-party provider.

While any organisation must have these capabilities in order to preserve digital content over the long term, digital preservation is not a one-size-fits-all enterprise – every organisation, large or small, whether well-funded or underfunded, is unique and will have its own requirements, opportunities and constraints. Any organisation wishing to develop the capability to preserve digital records through time must therefore develop an understanding of its current capability, design a blueprint for the future that meets its specific requirements, identify the gap between the latter and the former, and introduce a strategy for building the necessary capacity to bridge that gap.

Maturity models are well-established tools that can help an organisation assess its capabilities in a particular area against a benchmark standard as a means of articulating its current and desired future operating models. They typically describe organisational capacity in two dimensions: first, they define the component parts that together constitute the specific service or function to be addressed, and second, they define a scale for measuring capability against each of these requirements.

The key elements of digital preservation capability, including organisational factors such as governance, resources and policy, and technical or functional aspects, have been discussed above. In the language of maturity models, these are often referred to as ‘process perspectives’. A typical scale for measuring capability will comprise five or six steps, spanning the stages from developing awareness to building increasing capability, such as set out in [Table 9.1](#).

These steps can be mapped to the records management capacity levels in Julie McLeod and Elizabeth Shepherd’s chapter, which also identifies the role and responsibilities of the major players involved. These also correspond with the maturity levels defined in John McDonald’s chapter.³ McDonald sets out capacity levels as:

- *Level 1*: poor-quality data, statistics and records make it virtually impossible to measure SDGs reliably

3 John McDonald, ‘The quality of data, statistics and records used to measure progress towards achieving the SDGs: a fictional situation analysis’, [Chapter 13](#) in this volume.

Table 9.1. Maturity levels

<i>Stage</i>	<i>Maturity step</i>	<i>Description</i>
Awareness	0 – no awareness	The organisation has no awareness of either the need for the process or basic principles for applying it.
	1 – awareness	The organisation is aware of the need to develop the process and understands basic principles.
	2 – roadmap	The organisation has a defined roadmap for developing the process.
Capability	3 – basic process	The organisation has implemented a basic process for capturing and preserving digital records.
	4 – managed process	The organisation has implemented a comprehensive, managed process that reacts to changing circumstances.
	5 – optimised process	The organisation undertakes continuous process improvement, with proactive management.

- *Level 2:* data, statistics and records are adequate to measure the SDGs to basic levels in some sectors
- *Level 3:* the quality of data, statistics and records makes it possible to measure SDGs effectively and supports government programme activities
- *Level 4:* well-managed data, statistics and records make it possible to measure SDG implementation effectively and consistently through time; data and statistics are of high enough quality to support government programme activities at the strategic level
- *Level 5:* processes generating data, statistics and records, and the framework for managing them are designed to make it possible to exploit data, statistics and records in new and innovative ways.

Achieving Level 4 will require a basic level of digital preservation capability, while Level 5 corresponds to the more advanced, managed and optimised capabilities. Taken together, we can begin to develop detailed definitions of each maturity level against every process perspective. To illustrate the principle, the ‘basic’ level of maturity (that is, Step 3) might be defined as shown in [Table 9.2](#).

Table 9.2. A basic preservation capability

<i>Process perspective</i>	<i>Step 3 definition</i>
A – organisational viability	<ul style="list-style-type: none"> • staff have assigned responsibilities and the time to undertake them • a suitable budget has been allocated • staff development requirements have been identified and funded
B – stakeholder engagement	<ul style="list-style-type: none"> • key stakeholders have been identified • objectives and methods of communication have been identified
C – legal basis	<ul style="list-style-type: none"> • key legal rights and responsibilities, together with their owners, have been identified
D – policy framework	<ul style="list-style-type: none"> • a written, approved digital preservation policy exists
E – acquisition and ingest	<ul style="list-style-type: none"> • some individual tools are used to support accession and ingest • an acquisition policy exists that defines the types of digital content that may be acquired • a documented accession and ingest procedure exists, including basic guidance for depositors
F – bitstream preservation	<ul style="list-style-type: none"> • there is dedicated storage space on a network drive, workstation or removable media • at least three copies are maintained of each object, with backup to removable media • basic integrity checking is performed • virus checking is performed • existing access controls and security processes are applied
G – logical preservation	<ul style="list-style-type: none"> • basic characterisation capability exists (or the capability to identify and describe a file and its defining technical characteristics) allowing at least format identification, such as file formats and technical attributes • ad hoc preservation planning takes place • ad hoc preservation actions can be performed if required • the ability to manage multiple manifestations of digital objects exists. That is, all of the different renderings of the same object (in different file formats) can be identified, described and managed

Table 9.2. (continued)

<i>Process perspective</i>	<i>Step 3 definition</i>
H – metadata management	<ul style="list-style-type: none"> • a documented minimum metadata requirement exists • a consistent approach to organising data and metadata is implemented • metadata is stored in a variety of forms using spreadsheets, text files or simple databases • the capability exists to maintain persistent links between data and metadata • persistent, unique identifiers are assigned and maintained for all digital objects
I – dissemination	<ul style="list-style-type: none"> • basic finding aids exist for all digital content • users can view or download data and metadata, either online or on-site
J – infrastructure	<ul style="list-style-type: none"> • sufficient storage capacity is available, and plans exist to meet future storage needs • IT systems are documented, supported and fit for purpose

This should be considered the minimum standard for any organisation to measure SDGs and provide a genuine digital preservation service. For many, this may be an achievable target, sufficient to meet their objectives. It should not be assumed that every organisation should strive to reach Step 5 in every process, which could be excessive and infeasible, particularly in lower resourced countries. The power of maturity models is that they make it possible to define nuanced and proportionate levels of service capability. In practice, most organisations will wish to define different target levels for different process perspectives.

In many cases, a relatively modest level may be entirely appropriate. The value of maturity models lies primarily in providing a means of thinking about digital preservation as a broad spectrum of capabilities, rather than a single, and almost-certainly unobtainable, ideal. This should help organisations to think about what ‘good enough’ preservation looks like in their own particular circumstances, and therefore to ensure a proportionate and appropriate level of investment.

The approach summarised here is intended to provide a flexible and realistic methodology.⁴ However, other maturity models for digital preservation

⁴ Based on the maturity model defined in A. Brown, *Practical Digital Preservation* (London: Facet, 2013), pp. 86–91.

exist. For example, the National Digital Stewardship Alliance has developed its own framework of levels of digital preservation,⁵ while the EU-funded E-ARK project has created a model which has been tested across a range of institutions.⁶ Most recently, the Digital Preservation Coalition has developed a Rapid Assessment Model based on that described in this essay.⁷ Whichever model is used, it is essential for any organisation that is planning to develop a digital preservation capability to have a clear and pragmatic definition of the level of capability it wishes to build. Having done so, it will then need to consider how to establish that capability in practice.

Implementation options

Many different models are possible for implementing a practical digital preservation capability.

This means that options are available to meet the needs of a wide variety of sizes and types of organisation in many different contexts with very different levels of resources at their disposal. The market for digital preservation solutions is still comparatively young, but its vast potential size (one study estimated its potential value in 2011 as in excess of \$1 billion,⁸ and this figure has surely only increased in the intervening period) has encouraged a growing number of increasingly mature products to emerge to meet a wide range of requirements.

It is very important to recognise that digital preservation is achievable not only by large institutions with substantial budgets – practical solutions are possible with much more modest means. The main options are summarised below, with a brief discussion of their advantages and disadvantages. Mention of specific products or tools does not constitute an endorsement or recommendation.

Doing nothing

Any analysis of options should always include the status quo, as this allows a true comparison to be made with other, more positive, options, and it highlights the implications of not taking action. Doing nothing is not a cost-free option; actually, through time it is likely to be very expensive because of the continuing burden of maintaining archival data on inappropriate storage infrastructure and the actual or lost opportunity costs of having to recreate digital records that will inevitably be lost through inaction.

5 See <https://nds.org//activities/levels-of-digital-preservation/>.

6 See <http://www.eark-project.com/resources/project-deliverables/95-d75-1/file>.

7 See <https://www.dpconline.org/our-work/dpc-ram>.

8 Y. Au, R. Kandalafi, M. Kuang and S. Nair, *Digital Preservation and Long-Term Access Functionality* (Cambridge: Judge Business School, 2010), <http://www.scribd.com/doc/45412331/Cambridge-Judge-Business-School-Market-Research-Digital-Preservation>.

Without active management, data loss is inevitable over the longer term, and the associated costs are likely to be much higher than the costs of establishing appropriate control systems. The reactive, ad hoc rescue of digital information will almost inevitably be much more expensive than proactive management. Doing nothing will also, of course, incur significant risks and will not meet the requirement of maintaining access to vital records. It will have major consequences for the delivery of government programmes and for the ability to measure the SDGs reliably.

Using open source software

It is possible to develop a digital preservation solution entirely from open source tools. A number of complete open source digital repository management systems are available, although they vary in the level of preservation functionality that they offer directly. The most widely used systems include Archivematica,⁹ DSpace,¹⁰ EPrints,¹¹ Fedora,¹² LOCKSS¹³ and RODA.¹⁴ In addition to these complete systems, a number of toolkits and individual utilities have been developed that can be used to add preservation functionality to existing repository systems. These include characterisation tools,¹⁵ such as DROID¹⁶ and JHOVE,¹⁷ forensic tools, including BitCurator,¹⁸ and web archiving tools, such as Heritrix.¹⁹

Open source solutions offer a number of attractions. They are often free to use, may have thriving user and developer communities, and can offer very flexible solutions. The fact that users have complete access to, and control over, the source code, can also be very attractive from a sustainability perspective. However, open source solutions are not cost-free: resources are likely to be required to adapt and configure the software, either from in-house staff or procured from a third party. Also, the organisation will have to bear all the risks, rather than sharing them with suppliers.

9 See <http://www.archivematica.org/en/>.

10 See <https://duraspace.org/dspace/>.

11 See <http://www.eprints.org/uk/>.

12 See <https://duraspace.org/fedora/>.

13 See <http://www.lockss.org/>.

14 See <https://demo.roda-community.org/#welcome>.

15 Characterisation tools (together with format identification tools) aim to automate the process of identifying the format of a digital object based on its extrinsic and intrinsic elements and by extracting metadata about its properties that are significant to its ongoing preservation.

16 See <http://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/file-profiling-tool-droid/>.

17 See <http://jhove.openpreservation.org/>.

18 See <http://bitcurator.net/>.

19 See <https://github.com/internetarchive/heritrix3/wiki>.

It is possible to develop a functioning digital archive capacity very affordably, by using simple, cheap and readily available open source tools and existing infrastructure. Compromises may be needed to keep costs down – limited integration will involve a lot of manual or semi-manual processes, and this approach probably won't be appropriate for managing larger volumes of records or high numbers of users. It may also be difficult to support or sustain as a service over time. Nonetheless, for many smaller organisations or those with limited resources, this may be a good starting point. Investment in limited configuration of these tools to meet local needs, and with good support arrangements, will be very worthwhile, and may compare favourably with the costs of commercial products.

Developing a bespoke solution

Developing a bespoke, or tailor-made, solution is an option, especially for organisations with significant in-house software development capacity, or with the resources to commission external developers. A bespoke solution can be developed to meet the organisation's specific requirements, but software development is an expensive, complex, uncertain and time-consuming option. Given the availability of a number of mature third-party solutions, this is unlikely to be an economical approach except for organisations with very unique requirements and substantial resources.

Procuring a commercial solution

Commercial off-the-shelf (COTS) digital preservation solutions are a common approach, thanks to a growing and comparatively mature market for such products. They typically command a relatively high price, including one-off licence fees, annual support costs and potentially expensive customisation and configuration. They can also create a degree of dependency on an external supplier, and on proprietary software. At the same time, they can offer a high level of flexibility and support, usually have well-established user communities, and can therefore provide a comparatively low-risk approach. Building a strong relationship with the supplier is often vital.

There is a small, but now well-established, community of commercial digital preservation products that caters not only to the library, archive, museum and gallery sector but increasingly to customers in the private sector with long-term data retention requirements. It includes industries such as banking, pharmaceuticals, aerospace, gas and oil exploration, government and the nuclear energy sector. Examples of commercial products include Preservica²⁰ and Rosetta.²¹

20 See <https://preservica.com/>.

21 See <http://www.exlibrisgroup.com/products/rosetta-digital-asset-management-and-preservation/>.

Outsourcing the service

One of the most recent developments in the digital preservation market has been the growth of preservation-as-a-service, whereby third-party suppliers provide an end-to-end solution hosted on their own infrastructure. This option minimises the direct impact on the organisation, including the need to host and support significant infrastructure, and can be easily scaled up or down to meet demand. It also has a very low barrier to entry, making it particularly attractive to smaller organisations or those with limited resources. These services are typically priced on the basis of actual usage, removing the need for significant up-front capital investment, which can make them financially attractive. However, it is always important to consider the long-term costs when comparing the economics of different models.

A growing number of suppliers offer a full range of digital repository services. In some cases, these have emerged to meet the needs of specific communities. For example, the UK Data Archive (UKDA)²² and the Inter-University Consortium for Political and Social Research (ICPSR)²³ both provide data archive services to the international social sciences community. The international library community is served by services such as Portico, which preserves e-journals, e-books and digitised historical collections on behalf of publishers and libraries,²⁴ and by the OCLC CONTENTdm digital collections management service.²⁵ A number of cloud-based commercial services have also emerged, such as ArchivesDirect,²⁶ DuraCloud,²⁷ Preservica Cloud Edition,²⁸ while services are also available for specific functions such as web archiving, such as Archive-It.²⁹

Partnership approaches

A very collaborative community has developed around digital preservation. This can include the actual provision of services, with a group of organisations that share a common set of requirements establishing a partnership to develop and share services. Given the long timeframes involved and high level of confidence required, such arrangements are generally formalised in some way, whether through a contract, consortium agreement, or less legally binding instruments such as memoranda of understanding or letters of agreement. A partnership may even be set up as a separate legal entity representing the

22 See <http://www.data-archive.ac.uk/>.

23 See <http://www.icpsr.umich.edu/icpsrweb/>.

24 See <http://www.portico.org/digital-preservation/>.

25 See <http://www.oclc.org/en/contentdm.html>.

26 See <https://duraspace.org/archivesdirect/>.

27 See <http://www.duracloud.org/>.

28 See <https://preservica.com/digital-archive-software/products-editions>.

29 See <http://www.archive-it.org/>.

shared interests of the partners. This might be a non-profit entity, such as a charity, trust, foundation or private company limited by guarantee.

The precise forms of non-profit organisation allowed vary from country to country, but most enjoy tax-exempt status. For instance, an existing legal entity might host a partnership entity that exists independently from its members but is not itself legally constituted. Partnerships can establish their own shared infrastructure, which may use a distributed model such as LOCKSS (see above) to jointly procure a service from a third party, or they can designate one partner to operate the service on behalf of the others.

A partnership approach can offer significant economies of scale and allow better deals to be negotiated with suppliers than individual partners could achieve on their own. However, they depend on a strong and ongoing alignment of objectives between partners that can be complex to establish and require compromise. There are some excellent examples of the partnership model in practice, many based on LOCKSS technology. These include CLOCKSS (Controlled LOCKSS), an international, not-for-profit community partnership between libraries and publishers using a private network based on LOCKSS technology to provide a distributed archive for electronic scholarly content;³⁰ the Alabama Digital Preservation Network (ADPNet), which provides a low-cost distributed LOCKSS-based digital preservation service for cultural heritage organisations in Alabama;³¹ and the MetaArchive Cooperative, a community-based digital repository solution, serving over 20 libraries, archives and other cultural memory institutions in three countries.³²

In the context of SDGs, there are obvious opportunities for partnerships, both between national archives and national statistical agencies and between countries. Such partnerships could facilitate the establishment of shared digital preservation capabilities.

Hybrid approaches

It is always possible to adopt more than one of these approaches for different elements of the solution. Such a hybrid approach can be very flexible and cost-effective, but it may also increase complexity.

Using consultancy services

Consultancy can provide important support for organisations in understanding their preservation needs. Typically, this may involve specific projects, for example auditing existing holdings, defining requirements, developing policies and procedures, or advising on standards. To benefit from a consultancy – which

30 See <https://clockss.org/>.

31 See <http://www.adpn.org/>.

32 See <https://metaarchive.org/>.

can be costly – it is essential to have a clear, focused brief and to choose both the project and the consultant with great care. However, at its best such consultancy can bring an impartial and expert perspective to issues that must ultimately be implemented internally.

Implementation and operational implications

Implementing a digital preservation service

The task of establishing facilities for managing and preserving digital records and data can vary enormously, depending on the complexity of the requirements and the approach taken. In some cases, this will involve a major technical and organisational change programme; for others it will be much more straightforward. However, in all cases, success will depend on careful planning and organisation. The steps involved, from galvanising organisational support to delivering an operational service, have been described in detail elsewhere,³³ and are summarised briefly here.

No initiative of such importance can succeed without first securing the organisation's support, not only in terms of allocating the necessary resources but also of achieving the cultural change that is essential to achieving the benefits of digital preservation and ensuring its long-term viability. In turn, such support can only be achieved by first understanding the specific concerns that will trigger action for a particular organisation or in a specific context. This must also be rooted from the start in an understanding of the needs of all relevant stakeholders, from data creators to those who will use the information now and in the future. An early objective should be to develop and agree a digital preservation policy and strategy, including a commitment to regularly review and update them to address evolving issues. These policies will help to build an institutional commitment, laying the foundations for developing a detailed business case to secure the time and resources needed to make that commitment a reality.

There must be a convincing argument for the long-term value of digital records and data, based on evidence such as case studies and data about, for example, the economic or societal value of specific information or data, and the costs that would be incurred if it were to be lost. In the case of SDGs there is ample evidence that without reliable information, the goals cannot be measured reliably. It also will be important to consider the various available options, such as those discussed above, to identify the most appropriate and economically achievable approach, taking account of the benefits and risks, and any dependencies on other projects or activities. It will take time to create

33 See, e.g., Brown, *Practical Digital Preservation*.

and gain approval for such a business case, but the reward will be a practical and achievable path to developing a digital preservation capability that is appropriate and proportionate to the need. The Digital Preservation Coalition has developed a toolkit that is thoroughly recommended for anyone creating a business case.³⁴

Once the case has been approved, the next step will be to define detailed requirements, which again must be firmly rooted in the needs of stakeholders, including content creators, information managers and those with curatorial responsibilities, end users, decision-makers and funders, and technology suppliers. All key business processes will need to be modelled, showing, for instance, how information is used in the organisation. Having a comprehensive requirements catalogue is one of the most important building blocks for developing a successful digital preservation capability. A wide range of organisations have completed similar exercises. The digital preservation community is very open and collaboration-minded, and many excellent examples are available to draw upon.

The requirements catalogue should inform the design and implementation of the solution, whether this be through a commercial procurement exercise, building a solution in-house, or some form of partnership. The solution should cover all the key digital preservation functions, as outlined above, from acquisition and ingest, metadata management, bitstream and logical preservation, and dissemination. Implementation will always involve designing, building and testing both the technologies and the procedures for using them, often through a number of iterations. In parallel, the other aspects of the service will need to be put in place, as illustrated below.

Governance

Organisational and governance structures needed to deliver the service must be established, including staffing, which is discussed separately below. This must include legacy planning in the event of organisational change.

Roles and responsibilities

While the precise types and numbers of staff required to operate a digital repository will vary considerably, a number of generic roles are likely to be needed, including:

- *repository manager*, to oversee all key digital preservation functions, including ingest, preservation and access. This will require suitably trained curatorial staff, or a specialist digital archivist

34 See http://wiki.dpconline.org/index.php?title=Digital_Preservation_Business_Case_Toolkit.

- *ingerster*, to manage the accession and ingest into the digital repository of individual records from start to finish, including liaising with depositors. This will usually be performed by librarians or archivists with suitable training. In smaller operations, it might be performed by the repository manager
- *cataloguer*, to ensure that descriptive metadata is created and captured to appropriate standards, either during or after ingest. This role will usually be undertaken by existing cataloguing staff. In some cases, it may be combined with the ingerster role
- *system support for users*, which is often referred to as first-line support (more complex issues may be referred to as second-line support provided by system administrators or suppliers). Where possible, this should be integrated with existing IT helpdesk support. In a small organisation, it might instead be combined with the repository manager role
- *system administrators*, to manage IT systems and infrastructure, including second-line support, database administration, managing storage and managing user accounts. This role will normally be performed by IT staff

In all cases, some redundancy is highly desirable, to avoid any single-points-of-failure. If a failure does occur in a digital preservation process, other processes (duplicate or otherwise) can negate the impact of the failed process.

Training

Although digital preservation draws on a wide range of skills, including many that will already be part of the core expertise of librarians, archivists and other information management professionals, it is a specialism in its own right and therefore needs to be supported by high-quality training in digital preservation theory and practice. Digital preservation awareness and expertise is increasingly recognised as being a core part of information management professionals' core skill set, and it is now explicitly addressed within relevant graduate and postgraduate training courses, particularly in well-resourced countries, where most library and archives courses now cover digital preservation in some depth.

Training is, of course, also essential for existing staff who need to develop new expertise. A number of established face-to-face and online training courses are available and are highly recommended for anyone seeking to develop practical skills and knowledge. Notable examples of online training and related resources that are suitable for and accessible to an international audience (albeit primarily in English only), include:

- *MIT Digital Preservation Management Tutorial*: currently provided by MIT Libraries, but based on the seminal training programme developed by Cornell University and subsequently ICPSR. This is an award-winning online tutorial [English, French and Italian]³⁵
- *'Digital Preservation in a Box'*: an online training toolkit developed by the US National Digital Stewardship Alliance's Outreach Working Group, which provides a portal to a wide range of training resources [English only]³⁶
- *Digital Preservation Coalition Handbook*: although not a training course, the DPC handbook is a wonderful source of guidance and good practice, as well as links to further resources [English only]³⁷
- *ICA Digital Preservation resources*: the International Council on Archives, in partnership with the InterPARES project has developed an online educational initiative called 'Digital records pathways: topics in digital preservation' [English only].³⁸ Separately, and in conjunction with the International Records Management Trust, it has published two online training modules [English only]³⁹

Policies and procedures

All the policies and procedures necessary to underpin the service need to be developed and then maintained. These should cover issues such as collections development, collections management, documentation, preservation, business continuity and access.

Conclusion

Preserving digital information is essential both for programme management and for measuring the SDGs. It requires sustained active management, which in turn necessitates specialised skills and technologies, as well as institutional will to establish and maintain new capabilities. The operational demands of digital preservation, for instance setting up governance structures and determining training needs, are just as important to consider as the technical ones. The long-term success of any service will depend on both being met appropriately.

Digital preservation is, and will continue to be, a growing challenge for organisations across the world, but it is a challenge that can and must be

35 See <https://dpworkshop.org>.

36 See https://wiki.diglib.org/index.php/NDSA:Digital_Preservation_in_a_Box.

37 See <https://dpconline.org/handbook>.

38 See <http://www.ica-sae.org>.

39 See <http://www.ica.org/en/digital-preservation-training-modules>.

addressed. Practical solutions are available to suit a wide range of needs and are achievable by organisations of all sizes and with widely varying resources at their disposal. Many different approaches to implementation are possible, and it is essential to carefully choose the right option to suit the needs of the organisation. For organisations in lower resource environments, the open source, preservation-as-a-service or partnership models may offer the most practical way forward. The need for digital preservation must be addressed if SDGs are to be met-careful planning for this fundamental aspect of global development from the outset will have wide benefits for transparency and accountability.

10. Preserving and using digitally encoded information as a foundation for achieving the Sustainable Development Goals

David Giaretta

The health, wealth and happiness of a great many people worldwide, now and in the future, will depend upon the accuracy, intercomparability and preservation of the Sustainable Development Goal (SDG) measurements, information that is of global significance. This chapter offers a view on how to manage the complexity of gathering comparative data to monitor progress towards the SDGs in the years leading to 2030 from the perspective of collecting, using and preserving digitally encoded information, in particular scientific data. The aim is to help improve the way that information relevant to the SDGs is collected and used, so that the conclusions and actions arising from them are based on information that can be regarded as authentic, the results from which can be trusted, and comparisons between nations can be made sensibly through time.

The extract that follows, quoted from *Transforming Our World: The 2030 Agenda for Sustainable Development*, Section 74,¹ notes that follow-up and review processes at all levels will be guided by the following principles:

- (f) They will build on existing platforms and processes, where these exist, avoid duplication and respond to national circumstances, capacities, needs and priorities. They will evolve over time, considering emerging issues and the development of new methodologies and will minimize the reporting burden on national administrations.
- (g) They will be rigorous and based on evidence, informed by country-led evaluations and data which is high-quality, accessible, timely, reliable and disaggregated by income, sex, age, race, ethnicity, migration status, disability and geographic location and other characteristics relevant in national contexts.
- (h) They will require enhanced capacity-building support for developing countries, including the strengthening of national data systems

¹ *Transforming Our World: The 2030 Agenda for Sustainable Development*, http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E.

and evaluation programmes, particularly in African countries, least developed countries, small island developing States, landlocked developing countries and middle-income countries.

- (i) They will benefit from the active support of the United Nations system and other multilateral institutions.

These statements make it clear that data are to be collected at the national and subnational levels for every nation and that there may be differences in the way they are collected and recorded. The data are to be disaggregated, but the exact method and level of disaggregation may differ. All these differences make comparisons between nations and regions difficult or even impossible, and the results may have political and financial ramifications. Additional levels of proof of the validity of the data may be necessary.

A vast amount of information is needed to monitor progress towards achieving the SDGs. More challengingly, this information is hugely varied, covering all aspects of life, including health, wealth, nutrition, education, industry, society and the natural environment. Moreover, the volume of information being collected is growing at an ever-increasing rate, and an ever-greater portion of this information is in digital form. Of particular concern here are sets of information, including data, statistics and records, that provide the foundation for measuring the achievement of the SDGs. This digital information is fragile; the bits can decay, or the information encoded in those bits can be lost, which is difficult to guard against.

The chapter looks first at the challenges of measuring the SDGs and at the ideal solutions from the perspective of the international standards with which the author has been involved. It then examines representative SDGs, considers some of the realities for measuring them and explores what is required to preserve them. Finally, it explores potential ways of reaching the ideal, in line with the principles set out above, given the difficult realities faced. The chapter should be of use to those responsible for collecting the information for measuring the SDGs, those responsible for the use of SDG data and those whose interest is in deriving lessons from combining information that is relevant to the SDGs.

Requirements for SDG data to be fit for purpose

If the SDG measurements are to guide decisions that could affect the lives of billions of people, the information used to measure them must be authentic and verifiable, with clear health warnings in terms of its applicability and accuracy. This section looks in general at what must be addressed to achieve these aims and outlines some of the complexities that must be considered.

Authenticity

If SDG results lead to unpalatable conclusions, the authenticity of the data on which those conclusions are based is likely to be questioned, as has happened in the case of climate change. Evidence to support authenticity must therefore be collected; it should be possible to verify all the information related to the SDGs as authentic.

To achieve this, the provenance of the information needs to be recorded, which includes a record of who created it, when it was created, how it was created and what has happened to the information subsequently. This also should include the procedures, methodologies and algorithms used to preserve it. The importance of authenticity and provenance can be illustrated by considering questions about evidence, which could easily be distorted, with far-reaching consequences for countries and individuals.² Questions that should be asked include:

- are these the actual measurements made?
- were the measurements made correctly and in the location claimed?
- was the process used to reach the conclusion correct?

To answer these questions, it is important to be able to provide reliable evidence, for instance:

- the hash codes or the Transformational Information Properties (TIPs; discussed below) of the original measurements (so that those used can be checked against them)
- a verifiable record of how the measurements were made and by whom
- the documentation and software used (so that the whole process can be checked)
- the process used to reach the result and a record of how the process can be repeated and checked.

All these answers constitute the types of provenance that should be captured as part of SDG data.

For individual measurements, there are simple ways of proving that the values have not changed and are authentic, namely by using digests or hashes. Hashes are created by taking the bits that make up a file and applying an algorithm (or a set of procedures), such as: divide by a certain number, chop the sequence of bits into smaller pieces and multiply together in a certain way. This creates a sequence of digits and characters much shorter than the original file, known as the hash, which is like a fingerprint for that file. Keeping the original hash for a

2 Michael Grubb, 'Climate researchers' work is turned into fake news', *Scientific American*, January 2018, <http://www.scientificamerican.com/article/climate-researchers-rsquo-work-is-turned-into-fake-news/>.

file makes it impossible to make a change in the file without detection, because recalculating the hash of the ‘imposter’ will produce something that does not match the original hash.

Between now and 2030 and beyond, it may be necessary to transform digital objects in order to preserve them, for example from MS Word to PDF (if the required version of MS Word is no longer supported), or from CSV to XML. In such cases a hash cannot be used, and other evidence must be collected. The OAIS³ Reference Model introduced the concept of ‘Transformational Information Property’ (TIP) to capture evidence demonstrating that there is sufficient similarity between the original and the new objects. Transforming the file will almost certainly result in the loss of some information, and the TIP chosen should ensure that what is lost is not important. The TIPs are an explicit statement of what aspects should not be lost. Examples could include the pagination of a document, which may be important for legal documents, or the colours in an image, or the numerical differences which are allowed when making changes to scientific data. The TIPs should be agreed early on, and when the transformation is carried out, they should be checked.

There are other complexities that must be addressed when seeking to measure the SDGs, as described below.

Longitudinal studies

It is clear that several of the SDGs will take some time to achieve. For example:

SDG 2

End hunger, achieve food security and improved nutrition and promote sustainable agriculture.

SDG 10

Reduce inequality within and among countries.

SDG 17

Strengthen the means of implementation and revitalise the global partnership for sustainable development.

Monitoring progress between now and 2030 will require reliable means of measuring nutrition, inequality and implementation; the information will need to be captured consistently through the months and years. Each of these types of measurement can then be compared, month by month and year by year, to check that progress has been made in the desired way. This type of ‘longitudinal’ study is common to many disciplines.

3 The OAIS Reference Model is an ISO standard (ISO 14721), which forms the basis of essentially all work done in digital preservation. It can be downloaded from: <https://public.ccsds.org/Pubs/650x0m2.pdf>.

Consider the following SDGs, for which it will be difficult to quantify and collect data that are immediately relevant:

SDG 4

Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.

SDG 15

Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, halt and reverse land degradation and halt biodiversity loss.

SDG 16

Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels.

The question is how these goals can be monitored, as they will require complex analyses combining information from many areas of activity. For instance, if something is done to promote peaceful and inclusive societies, how will it be possible to check whether societies have become more peaceful and inclusive? If societies have not become more peaceful and inclusive, should some other course of action be taken?

Many SDGs, namely 6, 7, 8, 9, 12 and 14, seek sustainability of certain things:

SDG 6

Ensure availability and sustainable management of water and sanitation for all.

SDG 7

Ensure access to affordable, reliable, sustainable and modern energy for all.

SDG 8

Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all.

SDG 9

Build resilient infrastructure, promote inclusive and sustainable industrialisation and foster innovation.

SDG 12

Ensure sustainable consumption and production patterns.

SDG 14

Conserve and use the oceans, seas and marine resources sustainably.

Evidence of sustainability must in some way involve a longitudinal study to show that what is being sustained is unchanged through time and indeed for a significant time beyond 2030. Longitudinal studies will have to be started well before positive results can be expected.

Combining data

Longitudinal studies can, in principle, be relatively simple, involving comparing raw measurements from one point in time to the next. However, it may be necessary to combine a number of measurements to achieve a meaningful response. For example, SDG 2 requires progressively improved land and soil quality. One way of measuring soil quality is to use the USDA measurement of soil quality which combines measurements of 'soil respiration, infiltration, bulk density, electrical conductivity, pH, nitrates, aggregate stability, slake, earthworm, water quality and observations of soil structure'.⁴ Each of these is a separate measurement in itself, each with its own procedures and processes. All must be recorded alongside individual results.

The measurements are in general processed according to a specific algorithm, or some specific procedure such as 'divide measurement of this quantity by the measurement of that quantity and then add the result to the measurement of a third quantity'. Alternative algorithms may be used by different sets of people to produce specific results, because the groups follow different theories or make different assumptions about how the measurements were made. The different algorithms will almost certainly produce different results.

When there are large datasets, the algorithm is normally encoded in software that takes in the various datasets and combines them. The results of such algorithms affect our everyday lives,⁵ and these algorithms can change.⁶ Therefore, when comparing results computed using numerous inputs reported by different countries, we should ensure that the algorithms used are the same, or, if this is not possible, that the algorithms are well-documented so that the methods used by different countries can be properly compared.

If it is found to be useful to perform meta-analyses using SDG data from a number of countries, then it will be even more important to be sure that like is being combined with like.

4 USDA Soil Quality Test Kit, http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/health/assessment/?cid=nrcs142p2_053873.

5 'How algorithms rule the world', 1 July 2013, <http://www.theguardian.com/science/2013/jul/01/how-algorithms-rule-world-nsa>.

6 'Why Facebook's news feed is changing: and how it will affect you', 1 July 2013, <http://www.theguardian.com/technology/2018/jan/12/why-facebooks-news-feed-changing-how-will-affect-you>.

Errors

Scientific measurements always have random and/or systematic errors. The reader will probably be familiar with a number of statistical methods and perhaps standard deviations of measurements. But information about the SDGs is likely to require more detailed considerations because it will involve the combination of diverse sources of information, each of which will have some kind of errors associated with it.

The aim in designing a good measurement is, first, to understand the potential sources of errors and, second, to minimise the errors wherever possible. It is also important to distinguish between various kinds of errors and to understand the difference between precision and accuracy. The word precision is related to the random error distribution associated with a particular experiment or even with a particular type of experiment. The word accuracy is related to the existence of systematic errors, such as differences between laboratories. For example, one could perform very precise but inaccurate timing with a high-quality pendulum clock that had the pendulum set at not quite the right length.⁷

Some scientific data can be checked by repeating the measurement, ideally by different people and at different places. Measurements of physical constants, such as the speed of light, can be repeated to reduce errors. Random errors can be reduced by statistical means, assuming the distribution of errors is normal. When lots of measurements are combined, the errors tend to cancel in a predictable way, reducing the overall error.

However, this may not be the case for SDG-related data. To give a concrete example of the effect of errors, let us imagine that one measures temperature, which happens to be constant. A random error would be like rolling a dice for each measurement and adding the number shown on the dice, minus 3. That means that one will see the temperature jittering up and down, but *on average* the temperature will be constant. On the other hand, if the dice is weighted so that it always lands with the number 6 showing, one will see an average temperature 3 degrees higher than the actual temperature. The former are called random errors, while the latter are systematic errors.

To give an SDG-related example, if the Gross Domestic Product of a country is measured but an important part of the economy is omitted, the error will be repeated each time the exercise is repeated. Systematic errors can be reduced or identified by repeating measurements with different set-ups, for example using different measurement techniques.

7 E.M. Pugh and G.H. Winslow, *The Analysis of Physical Measurements* (London: Addison-Wesley, 1966).

Combining measurements of different things, for example when determining soil quality, will combine and propagate errors in the data. Adding measurements of two things together, each of which has a random error, one can estimate the combined error, which will be larger than either separate error. For example, if one has errors which are both systematic then the combined error may be the sum of the errors or they may cancel each other out.

Consider, for example, SDG 5: gender equality, which has nine targets and 14 indicators. It is reported that, too often, women are not identified separately in datasets,⁸ which must affect conclusions to be drawn about gender equality as a systematic error. Even if there are no systematic errors, there may be random errors. It may be necessary to combine the errors in an appropriate way, which could require significant work to decide or even to estimate what the overall error in the results might be. These issues should be considered in looking at trends over the years covered by the SDGs.

Collecting and preserving data for SDGs

Having looked at the challenges and solutions from a general point of view, this chapter now examines representative SDGs and considers some of the realities for measuring them. The Global Indicator Framework for the Sustainable Development Goals and targets⁹ sets out 244 measures, which can be roughly categorised in diverse ways, as discussed below.

Semantic issues

The SDGs are measured in various ways, with particular issues related to different classes of measurements. Some of the issues are generic and concern numbers, especially where similar measurements are to be combined or compared. It is important to capture a description of how the measurement was taken.

Proportions

One hundred and eleven measures, or about 45 per cent, are proportions of one measure with respect to another, for instance, measure 1.4.1: proportion of population living in households with access to basic services. Proportions have the advantage of not involving units. At the same time, a proportion may be expressed as a percentage (a number between 0 and 100) or as a fraction (a number between 0 and 1). Each one needs to be specified, as does the way it is encoded.

8 'Measuring the UN's Sustainable Development Goals: an update', September 2017, <http://www.statslife.org.uk/news/3556-measuring-the-un-s-sustainable-development-goals-an-update>.

9 Annex of the resolution adopted by the General Assembly on 6 July 2017, Work of the Statistical Commission pertaining to the 2030 Agenda for Sustainable Development (A/RES/71/313), http://ggim.un.org/meetings/2017-4th_Mtg_IAEG-SDG-NY/documents/A_RES_71_313.pdf.

Unclear metrics

About 54 measures, or 22 per cent, are very unclear. For example, 2.b.1: agricultural export subsidies and 2.c.1: food price anomalies, are non-specific in terms of units and even about what is to be measured.

Rates

Twenty-two of the measures, or 9 per cent, are expressed in terms of rates. Some are expressed fairly clearly, for example, 8.1.1: annual growth rate of real GDP per capita. However, 3.3.4: hepatitis B incidence per 100,000 population does not specify the time period, and it may be that the time period assumed in one country is different from that assumed in another.

Number of countries

Nineteen metrics, or 8 per cent, are simply numbers. For example, 1.5.3: number of countries that adopt and implement national disaster risk reduction strategies in line with the Sendai Framework for Disaster Risk Reduction 2015–2030, which countries may interpret differently.

Money

Eleven of the measures, or 4.5 per cent, are expressed in terms of monetary values. For some, the currency is explicitly expressed in units of US dollars, but for others no currency is specified, making comparisons impossible. Comparisons are difficult between countries or even within a single country in a given range of years because of currency variations and inflation.

Prevalence

Four of the measures, or less than 2 per cent, are expressed in terms of prevalence and are not always clear. For example, 2.1.1: prevalence of undernourishment, is not clear, whereas 2.1.2: prevalence of moderate or severe food insecurity in the population, based on the Food Insecurity Experience Scale (FIES), is more specific. However, it is still not clear what prevalence means here.

Structural issues

Information such as the proportions or amounts of money tends to be encoded in some sort of table, for instance:

2016, 7.4

2017, 7.7

2018, 9.2

This could be encoded as a simple text file or as a complex binary file. Alternatively, it could be in an XML file, such as Microsoft Excel uses, or else something like:

```
<year>2016</year><value>7.4</value>
<year>2017</year><value>7.7</value>
<year>2018</year><value>9.2</value>
```

There are of course many variations and many possible XML schema. Alternatively, the data could be stored in a database in some internal format. Again, these variations make comparisons very difficult.

Virtual data

Whether they are in a spreadsheet or a database, the values of the data may not be defined explicitly anywhere, and this adds another level of complexity. The value may be calculated from other data values through a formula. For example, a value shown as a proportion may only exist in an Excel spreadsheet as a formula '=100*A1/B1'. The value of B1 itself may be calculated from other values. In the future, appropriate software may be available to access the data, as may be the case, for instance, for Excel. However, even then there is no guarantee that the formulae will be applied correctly.¹⁰ Lack of data in a spreadsheet cell may be indicated by a blank or by a zero or '999', which can produce uncertain results.

Input data

Many of the measures discussed above are the result of a combination of other measurements, as illustrated earlier in relation to SDG 2, zero hunger. The USDA measurement of soil quality includes 'soil respiration, infiltration, bulk density, electrical conductivity, pH, nitrates, aggregate stability, slake, earthworm, water quality and observations of soil structure'. Each of these measures requires a separate test in itself, each with its own procedures and processes, all which should be recorded, as should the individual results. Even a simple proportion requires that the units of the two quantities are the same and are calculated year after year in the same way. For example:

1.a.3 Sum of total grants and non-debt-creating inflows directly allocated to poverty reduction programmes as a proportion of GDP.

¹⁰ See, for example, the report of errors in the spreadsheet formulae of Harvard's Carmen Reinhart and Kenneth Rogoff who are two of the most respected and influential academic economists active today, at <http://www.theconversation.com/the-reinhart-rogoff-error-or-how-not-to-excel-at-economics-13646>.

The way that funds are identified as being ‘directly allocated to poverty reduction programmes’ needs to be expressed consistently if the measure is to be accurate.

The same point applies to many of the measures used.

Digital preservation and exploiting digital data

In addition to understanding how to manage the complexity of gathering comparative data as a basis for monitoring progress towards the SDGs, it is vital to think about and plan for digital preservation, at least over the timescales relevant for the SDGs. Whereas printed documents can be used for hundreds of years, digital data are different. The things we rely on to use data, such as technology, software and know-how, quickly evolve and change and even become unavailable.

Basic concepts in digital preservation

While there are many different factors influencing the use and longevity of digital information, such as software dependency or users’ knowledge, in order to evaluate preservation issues, it is necessary to understand how the information is actually recorded and how easy it is to distort what it means.

Consider the meaning of these bits:

```
01001110 01001101 01010001 01001101 01010000 01001010
00100000 00100000
```

They could mean, among other things:

Two IEEE 754 32-bit real numbers: 8.6116461x10⁸ 1.35644119x10¹⁰

Two 32-bit integers: 164211241 168379396

Eight 7-bit ASCII characters: ‘NMQMPJ’

In fact, in this instance it was the last of the three. The characters were my flight reference for a recent trip – quite important for me at the time, but not really of interest now. This illustrates the point that just keeping the bits is not enough.

Types of digitally encoded information

There are many types of digital objects, including documents that can be rendered on a screen or on paper to be viewed by a human. Other types of digital objects that can be rendered are images, sounds or movies. Data, and in particular measurements relevant to the SDGs, are not normally simply rendered and viewed, and they are therefore referred to below as ‘non-rendered’.

Data, such as encoded numbers and text, are normally processed in a number of ways, and then the results of these processes are rendered and viewed, for example as a graph. The numbers and text could simply be printed

and viewed, but often this is not really useful, especially if the dataset is very large. The information in rendered objects may be ‘combined’ in the tens or hundreds of ways in the minds of individuals. Non-rendered objects (data), on the other hand, can be combined using computers with millions or billions of other pieces of data to be analysed and evaluated.

Evidence in the form of data, such as population statistics or economic indicators, must be collected to guide and monitor the SDG work. These data will be summarised in words or graphs, but only a limited number of analyses will be included. However, in many cases, where the measured progress could be challenged, datasets will need to be preserved and maintained so that they will be useable and capable of serving as evidence until at least 2030 and probably beyond.

Digital preservation

Digital preservation has been described as ‘interoperability with the future’. A fundamental aim of information preservation is to make it possible to trust and use it in the future. Preservation through time involves making sure that future users, who may have different technologies, formats, languages and understandings of words, can still use that information and can be confident that it has not been altered. Digital preservation also includes interoperability with the current time – preservation techniques should help to make information accessible and trustworthy right now. Use between communities can present the same challenges as use through time; the same techniques will be useful.

Today we need the capacity to use information from many sources, many disciplines, many people and many software applications. Coping with a large amount of data in a timely and repeatable manner requires that the data be digital and that each type of information should be encoded digitally. Since these measurements may provide the basis for making important decisions in the future, each should be preserved.

The widely accepted way to preserve digital information is to follow the Reference Model for an OAIS, or ISO 14721:2012. The standard was developed to facilitate broad consensus on the requirements for a repository capable of providing long-term, discipline independent preservation of digital information, or digital archives. The group that developed the standard brought together people from space agencies, national archives and libraries, commercial organisations and many other domains with an interest in the long-term preservation of digitally encoded information. The standard defines a set of responsibilities that an OAIS archive must fulfil, making it possible to distinguish it from other uses of the term archive. It also was intended to support the development of additional digital preservation standards.

Since being adopted as an ISO standard, the OAIS Reference Model has been welcomed and widely adopted by virtually all types of digital preservation communities. Most modern digital preservation initiatives refer to the OAIS Reference Model standard, and it has also been widely used by organisations to inform their implementations of new or upgraded preservation systems. The term ‘open’ in OAIS is used to imply that this standard was developed in open forums. It does not imply that access to the archive is unrestricted.

OAIS defines a number of important concepts for successful digital preservation. One central concept is that Archival Information Packages (AIPs) should be created to capture all the metadata needed for preserving the data. An AIP includes representation information (semantic, structural and other types, such as the software that makes it possible to understand and use the data) as well as preservation description information, which includes:

- provenance, including the date it was created, why it was created and what happened to it subsequently
- location reference
- fixity (to ensure the information has not been altered, for example by calculating access rights and how they are controlled)
- the way all this is linked together
- a description of the whole package.

Active data management plans

Once it is determined what metadata is needed for preservation in any given situation, it is important to capture this metadata as soon as the data are created rather than waiting to collect them at the end or attempting to collect them later. Here, two resources are valuable: Project Management Body of Knowledge (PMBOK)¹¹ is the entire collection of processes, good practices, terminologies and guidelines accepted as standards within the worldwide project management industry; the Data Management Body of Knowledge (DMBOK),¹² is a collection of processes and knowledge areas that are generally accepted as good practices within the data management discipline.

A new standard is now being developed that will bring together ideas from PMBOK, DMBOK and OAIS. The new standard, Information Preparation to Ensure Long Term Usability (IPELTU), is being prepared by the same international working group that wrote OAIS: the International Standards

11 *A Guide to the Project Management Body of Knowledge* (PMBOK® Guide), 6th edn (2017), see <https://www.pmi.org/pmbok-guide-standards/foundational/pmbok>.

12 *The DAMA Guide to the Data Management Body of Knowledge* (DMBOK Guide), 1st edition, 2009, <http://www.dama.org/content/body-knowledge> and DMBOK Version 2 see <http://damadach.org/dmbok2-DMBOK-version-2/>, final version available from <http://www.amazon.co.uk/DAMA-DMBOK-Data-Management-Body-Knowledge/dp/1634622340>.

Organization Technical Committee on Space Data and Information Transfer Systems. Drawing upon PMBOK and DMBOK, it breaks the project down into types of activities. At each stage, from conception to planning the collection of data and preserving the data, checklists are provided to illustrate additional information that will be needed to ensure the data can be used now and into the future.

For every activity there are areas of information that should be considered for collection. This makes it possible to draw up a table, with each element representing a particular activity and a corresponding area of information that should be considered for collection. IPELTU uses the term ‘collection groups’ for data creation and collection activities:

- *initiating*: the reason for creating the data and the initial definition of the data project
- *planning*: planning for the data creation and encoding
- *executing*: creating/collecting/encoding the data (at each point there may be deviations from the planned results, including instrument effects and unexpected influences)
- *closing*: completing the data creation/collection/encoding to satisfy the requirements of the project, phase or contractual obligations, and, at the end of the project, turning the information over to the long-term preservation organisation
- *controlling*: tracking, reviewing and orchestrating the progress and performance of the activities.

‘Additional information areas’ include:

- content information: content data object and representation information
- preservation description information (PDI)
- provenance information
- context information
- fixity information
- access rights information
- package description
- packaging information
- issues outside the OAIS Information Model: publications and related datasets.

The table that follows shows a sample of activities and can serve as a checklist for the data and metadata that need to be captured. This should help to ensure that everything necessary for preservation and future use is available. Each column describes a ‘collection group’, while other types of information that should be collected are described in the corresponding ‘additional information area’. For

example, for the ‘initiating’ collection group, the additional information that should be collected about the ‘data object’ is estimates of the volume of data to be produced and ideas about the potential value of the data.

Representation information should also be collected. In particular, this should include the standards that are expected to be used as well as the OAIS Information Model components, including provenance, access rights, fixity, reference and context information, and so on down the rows.

This does not mean that each project must be broken down into only one initiating collection group, one planning collection group, and so on. Rather, a project may be carried out in multiple phases, and the process may be repeated in each phase. For example, in a longitudinal study, the aim may be to collect information in one country about one SDG over the whole period to 2030. This is, in principle, repeated year after year. One can then look at each year as a project phase. However, practical considerations may mean that there are changes from one year to another. Applying the IPELTU checklist is a reminder to capture necessary information (data as well as metadata) at each project phase.

Is it really being preserved? The importance of certification

Care of the information relating to measuring the SDGs is of global significance. Having collected all the information needed for preservation, it is important to ask where and how the information will be preserved and to ensure that it remains useable for as long as it is required (at least until 2030 and very possibly beyond), along with evidence to support claims of authenticity.

ISO certification based on ISO 16363¹³ requires third-party verification that the information holdings are being preserved securely. Conducting audit and certification under ISO accreditation¹⁴ has real benefits because the process requires continuous improvements to the repository and regular checks to ensure that everyone involved has up-to-date skills, including the auditors, certification organisations and accreditation organisations – all are checked repeatedly and consistently. Moreover, cross-checks between countries help to guarantee consistency. One of the aims of ISO certification is to facilitate international trade in services and products by allowing measurements certified in one country to be accepted in other countries. In this way, many systems on which our health, wealth and happiness depend, such as medical and food products, can be audited and certified as following the correct procedures.

13 Audit and Certification of Trustworthy Digital Repositories, 2011, CCSDS 652.0-M-1 and ISO 16363:2012. Available from <http://www.public.ccsds.org/Pubs/652x0m1.pdf>.

14 See <http://www.iso16363.org>.

Table 10.1. Information that should be captured to support preservation and future use

<i>Collection Group →</i> <i>Additional Information Areas ↓</i>	<i>Initiating</i>	<i>Planning</i>	<i>Executing</i>	<i>Closing</i>
Data object	<ul style="list-style-type: none"> • produce estimate of volume of data to be produced • develop ideas about the potential value of the data 	<ul style="list-style-type: none"> • update additional information from the initiating step, based on more detailed plans • identify types of data, for instance raw or processed, that should be preserved • identify categories of data, for instance images, tables and any generic interfaces • identify quality constraints • plan the rate of data production • expand and add detail 	<ul style="list-style-type: none"> • update additional information from planning phase based on what really happens 	<ul style="list-style-type: none"> • finalise additional information from executing phase • create inventory of data produced that should be preserved • determine the volume that would require preservation • quality checks may be performed on the data by non-experts • define information properties that may be useful • check for (and create logs of) any missing data
Representation information	<ul style="list-style-type: none"> • standards expected to be used • the OAIS Information Model 	<ul style="list-style-type: none"> • update additional information from initiating phase on more detailed plans • review applicable standards • refine information model • choose data format • identify hardware and software dependencies • identify relationships between data items 	<ul style="list-style-type: none"> • collect semantics of the data elements, e.g. data dictionaries and other semantics • collect format definitions and formal descriptions • create other data documentation • calibrate and test tools and system test data to be delivered 	<ul style="list-style-type: none"> • finalise additional information from executing phase • finalise representation information networks to reasonable level • identify other software that may be used on the data • create suggestions for the designated community and the representation information needed

Provenance information

- record of origins of the project, e.g. in a Current Research Information System (CRI)
 - update additional information from initiating based on more detailed plans
 - define processing workflow, processing inputs and processing parameters
 - define system testing required
 - documents from system development milestones
 - update additional information from planning based on what really happens
 - documentation about the hardware and software used to create the data, including a history of the changes in these over time
 - update documentation of processing workflow, processing inputs and processing parameters
 - record who was responsible for each stage of processing
 - record when each stage was performed
 - record any special hardware needed
 - record calibration
 - processing logs
 - record checking of fixity
 - finalise additional information from executing
 - identify related data which may in the future be combined with these data
-

Getting to where we need to be

Having looked at the challenges and solutions from an ideal point of view and the realities of measuring information related to the SDGs, we can now explore potential ways of making the difficult realities approach the ideal. Data collection has already begun across many countries, and it may not be practical to make changes in order to approach the ideal. However, there are things that can be done to strengthen the way that data are collected. Notably, we can:

- ensure that all the metadata, as required by the OAIS AIP, is collected in order to fill in the gaps that should be identified in a data management plan
- ensure that the information is preserved. At the very least, an ISO 16363 audit will identify opportunities for improving the operations of a repository.

It is important to take immediate steps to clarify what data should be collected (including clarifying the units being measured and the specific measurements to be made). It is also important to improve the way data are collected (for example by identifying women separately in datasets). If a greater level of disaggregation becomes possible, and the data can be separated into a finer level of detail, then future measures can be intercompared in greater depth. However, even with only aggregated measures, they can be compared immediately.

Having considered the guiding principles for the SDG initiative quoted at the beginning of this chapter, it is worth adding further commentary here:

- (f) They will build on existing platforms and processes, where these exist, avoid duplication and respond to national circumstances, capacities, needs and priorities. They will evolve over time, considering emerging issues and the development of new methodologies, and will minimize the reporting burden on national administrations.

A collective effort should be made to draw up more detailed guidelines for the algorithms needed to process the information for each SDG so that the processes evolve in such a way that they converge and the results are compatible between countries.

- (g) They will be rigorous and based on evidence, informed by country-led evaluations and data which is high-quality, accessible, timely, reliable and disaggregated by income, sex, age, race, ethnicity, migration status, disability and geographic location and other characteristics relevant in national contexts.

The guidelines should include enough detail to help the data collectors capture information as evidence of authenticity and as far as possible errors in the results can be estimated. For example, changes through time can be said

to be real rather than due to random errors and the information can be re-processed/re-purposed in the future.

- (h) They will require enhanced capacity-building support for developing countries, including the strengthening of national data systems and evaluation programmes, particularly in African countries, least developed countries, small island developing States, landlocked developing countries and middle-income countries.

If the concepts described in this chapter can guide this support, then the data systems in different countries can converge and training and common software systems can be shared.

- (i) They will benefit from the active support of the United Nations system and other multilateral institutions.

Staff of the UN and other multilateral institutions will benefit from becoming familiar with the concepts in this chapter, which will provide a standards-based blueprint for capacity building and implementation.

Such activities will result in improvements in:

- existing platforms and processes to ensure validity, authenticity and intercomparability of the information gathered for the SDGs
- specification of the information to be gathered and the detail in which to gather it in order to improve the quality of the responses to each SDG
- ability to compare the results year on year
- ability to compare the results between countries.

Preservation requires resources, so it makes sense to share the costs (facilities, human resources) with other organisations. Sharing the techniques and knowledge needed to capture, encode and preserve information will be a valuable start. A more advanced step would be to use an ISO 16363 certified repository to preserve the information. Such a repository will need to be inspected by expert independent auditors and certified as being capable of preserving information.

Conclusion

This chapter has taken a pragmatic approach to viewing, monitoring and measuring SDG information as a large data project designed to support measuring progress towards achieving the SDGs and to focus attention on key issues that need to be resolved if the goals are to be achieved. If UN and other multilateral institution staff can draw on the concepts set out, as part of the planning process for meeting the SDGs, there will be real benefits in terms of data management and the longevity of digital materials.

The chapter is directly relevant to the issues that the key SDG working groups responsible for monitoring developments and issues relating to the indicators and their metadata have worked on, namely: metadata exchange, geo-spatial information and interlinkages between the SDGs. Incorporating the contributions that ISO standards can make to improving data quality and building the framework for preservation will strengthen the whole complex web of sustainable development and solidify the efforts underway by the United Nations and its partners.

11. Transparency in the 21st century: the role of records in achieving public access to information, protecting fundamental freedoms and monitoring sustainable development*¹

Victoria Lemieux

A reliable and accessible evidence base is vital for all aspects of transparent and open government, particularly in an age of propagation of fake news and misinformation. Now more than ever, we need reliable facts about government decisions and actions for public accountability and economic and social development. Unless digital data and documents are created in the first instance, and then managed and protected as reliable evidence, they cannot serve the ends of transparency, openness and accountability, nor can they be used to track progress on goals for sustainable development.¹

This chapter primarily draws upon a programme of research² on transparency and information management conducted at the World Bank by the author and her colleagues from 2014 to 2016 to explore the issue of records and their relationship to 21st-century transparency. Although the research did not focus singularly on records and information management, it did encompass a number of findings on this theme. The main findings in relation to records and information management were:

- 1 The effectiveness of current transparency initiatives globally is being undermined by weak records and information management.
- 2 Weakness in records and information management is a widespread and persistent problem.

* The author would like to thank Stephanie Trapnell and Anne Thurston for their contributions to the research this chapter has drawn on. A version of this chapter was originally presented as a paper at the Conference on Transparency in the 21st Century, organised by the Information Commissioner of Canada, 21–23 March 2017.

1 <https://unstats.un.org/sdgs/>.

2 V. Lemieux and S. Trapnell, *Public Access to Information for Development: A Guide to the Effective Implementation of Right to Information Laws, Directions in Development–Public Sector Governance* (Washington, DC: World Bank, 2016), <https://openknowledge.worldbank.org/handle/10986/24578>.

- 3 New digital forms of communication and conducting government business have exacerbated earlier weaknesses in records and information management.
- 4 Weak control over digital records and information weakens transparency and public accountability mechanisms, such as right to information laws and open data initiatives.
- 5 Persistent cultures of secrecy also lead to oral government and avoidance of record-making and keeping.

Given these findings, the final section of the chapter proposes steps that can be taken to strengthen records and information management, as follows:

- 1 Develop indicators of effective administrative recordkeeping in support of transparency policies and laws, identify a baseline for all countries, at least at the national level, and track progress.
- 2 Strengthen laws and policies governing digital records management that affect transparency initiatives.
- 3 Strengthen records and information governance frameworks to enable an alignment with transparency policies and laws.
- 4 Strengthen the role of independent oversight bodies such as information commissioners in relation to monitoring and oversight of records and information management effectiveness.
- 5 Encourage more collaboration between public offices responsible for records and information management and those responsible for transparency and open government initiatives (for instance, right to information officials).

Current transparency initiatives are undermined by weak records and information management

Despite the fact that records – defined as ‘[i]nformation created, received and maintained as evidence and as an asset by an organization, or person, in pursuit of legal obligations or in the transaction of business³ – provide an essential foundation for transparency in the 21st century, there is plenty of evidence to suggest that the state of records and information management in public agencies in countries around the world is problematic. For at least the last 15 years, there have been regular warnings about the impact of the loss of

3 International Standards Organization, ISO 15489-1:2016. Information and Documentation – Records Management. Part 1: General. Geneva: ISO.

control of records relating to a range of government functions coming from the press, auditors, academic researchers and records professionals.⁴

Studies of barriers to effective implementation of right/access to information laws consistently emphasise that poor records and information management prevents governments from responding to requests or, if able to respond, slows down the timeliness of the responses.⁵ A recent survey of information commissioners points to an increasing number of denials of requests for information on the grounds that the information cannot be found or that it is too costly to produce.⁶ For instance, a 2016 shadow report from South Africa, compiled from statistics on requests for information (August 2015 to July 2016) under the Promotion of Access to Information Act (2000), confirms the trend: 46 per cent of requests (n=369) submitted to government were refused – that is, no information was provided. Of these, 58 per cent of the requests were ignored.⁷

Nor is the effect of weak records and information management limited to the effectiveness of implementation of right to information laws. According to World Bank Open Data Readiness Assessments, in many countries (for example, Kazakhstan, Serbia and Sierra Leone) proactive disclosure and open data initiatives are impeded by weak records management. In the United Kingdom, poor data quality was found to be hindering the government's Open Data programme. The authors of the report conducted an analysis of 50 spending-related data releases by the Cabinet Office since May 2010. They found that the data were of such poor quality that using them would require advanced computer skills. Far from being a one-off problem, research suggests that this issue is ubiquitous and endemic. Some estimates indicate that as much as 80 per cent of the time and cost of an analytics project is attributable to the need to clean up 'dirty data'.⁸ In addition to data quality issues, data

4 V. Lemieux, 'One step forward, two steps backward? Does e-government make governments in developing countries more transparent and accountable?', World Development Report Background Paper (Washington, DC: World Bank, 2016), <https://openknowledge.worldbank.org/handle/10986/23647>.

5 L. Neuman and R. Calland, 'Making access to information laws work: the challenges of implementation', in A Florini (ed.), *The Right to Know* (New York: Columbia University Press, 2007).

6 ICIEN (Information Commissioners International Exchange Network), In the Experience of Information Commissioners: The Information Commissioners' International Exchange Network Survey, Centre for Freedom of Information, University of Dundee, 2014, http://www.centrefoi.org.uk/news/images/98364000_1415912545.pdf; T. Taillefer and N. Elliot, Promoting Legislated Duty to Document for Government Accountability, Open Discussion Forum on Transparency and Information Management, World Bank, 4 June 2015, <http://www.web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPUBLICSECTORANDGOVERNANCE/0,contentMDK:23585462~pagePK:148956~piPK:216618~theSitePK:286305,00.html>.

7 South Africa, Access to Information Network, 2016, 'Shadow Report', <http://www.r2k.org.za/wp-content/uploads/CER-Shadow-Report-2016-Final.pdf>.

8 T. Dasu and T. Johnson, *Exploratory Data Mining and Data Cleaning*, Vol. 479 (New York: John Wiley & Sons, 2003).

provenance can be difficult to determine. Knowing where data originate and by what means they have been disclosed is key to being able to trust them. If end users do not trust data, they are unlikely to believe they can rely upon the information for accountability purposes.⁹

Weak records and information are hindering efforts to monitor progress on the Sustainable Development Goals (SDGs) as well. Administrative records are a primary source of government statistical information. When the systems that generate administrative records do not create trustworthy evidence of government actions, quality statistical data are not available for reporting purposes. As the report from the Workshop on Managing Digital Information as Evidence to Underpin Global Development Goals notes:

Governments and donors worldwide tend to believe that information produced in computerised systems will offer the basis for planning, monitoring and measuring national and international development goals. Most do not realise that IT systems create records but lack the full functionality needed to keep them reliable and authentic for as long as they are needed. As a result, IT systems have been developed without the supporting framework of policies and systems needed to protect, preserve and make digital evidence available through time.¹⁰

A study of disclosure information under right to information laws for the period 2011 to 2013 provides a summary of data about requests, complaints and appeals published by central reporting bodies in eight countries (Brazil, India, Jordan, Mexico, South Africa, Thailand, the United Kingdom and the United States). The study authors found that practices were far from standardised, and data were often unavailable or incomplete. They pointed out that, when good data are available:

such data can ... reveal information about high-performing agencies, which generates positive incentives for better performance (as evidenced by the work of the La Alianza Regional por la Libre Expresión e Información in Latin America) ... while poor performance can be identified and addressed. Reporting of performance data by oversight bodies is also crucial to the principle of openness, especially in cases where the oversight body is autonomous. Access to the performance statistics of various agencies can not only encourage civil society organizations to analyze and disseminate findings but can also facilitate collaborative engagement with government over possibilities for improvement or scale-up.¹¹

9 V. Lemieux, O. Petrov and R. Burks, 'Good open data ... by design', World Bank blog post, 2014, <http://blogs.worldbank.org/ic4d/good-open-data-design>.

10 'Managing digital information as evidence to underpin global development goals', University of London, 20–21 April 2017, unpublished report.

11 V. Lemieux, S. Trapnell, J. Worker and C. Excel, 'Transparency and open government: reporting on the disclosure of information', *JeDEM – eJournal of eDemocracy and Open Government*, 7 (2015): 75–93, <http://www.jedem.org>.

Issues related to missing data, data quality and reliability of government administrative systems have made it difficult to use the reporting data to assess the effectiveness of right to information laws. Efforts to draw conclusions about the effectiveness of RTI law from alternate data sources, for instance reports of civil society groups or expert reviews, can lead to disputes between different groups about the accuracy and reliability of the facts.

Weakness in records and information management is a widespread and persistent problem

As noted earlier, weakness in records and information management is not a new phenomenon. Pino Akotia highlighted it explicitly in 1997 in relation to the management of public sector financial records in Ghana and the implications for government transparency and accountability. His study demonstrated how poor records management can undermine public sector reforms and financial controls.¹² His findings have been supported by those of the International Records Management Trust (IRMT), which has regularly conducted research into the relationship between good governance and record-keeping since 1989. For instance, an IRMT study in 2011 on the management of public sector financial records in sub-Saharan Africa found that:

Poor records management threatens all government programmes and processes, including e-government and other service delivery activities, economic development initiatives, health care programmes, land reform initiatives, environmental projects and initiatives designed to enhance citizen rights. At the core of these issues is the erosion of trust in government programmes and decision-making where records cannot be found, the accuracy of the information in the records cannot be trusted, or the records are lost or destroyed.¹³

A more recent study conducted by the World Bank on drivers of effectiveness in the implementation of right to information laws also found that weak records and information management can prevent effective operation of laws governing public access to information.¹⁴ A follow-up survey of six countries (Albania, Jordan, Scotland, South Africa, Thailand and Uganda) using the World Bank's Right to Information Drivers of Effectiveness (RIDE) indicators found that records management scored poorly in all of the countries studied with the exception of Scotland, suggesting that this issue is an overlooked

12 P. Akotia, 'The management of public sector financial records: the implications for good government', PhD thesis, University of London (1997).

13 'Managing records as reliable evidence for ICT/e-government and freedom of information', White Paper for Senior Management, International Records Management Trust, London, 2011.

14 Lemieux and Trapnell, *Public Access to Information for Development*.

factor that undermines right to information implementation.¹⁵ Anecdotal cases continue to lend weight to the general findings, even in countries with fairly effective right to information systems. For example, in India in 2014, government officials in the Union Home Ministry were unable to respond to a request for information having destroyed the approximately 11,000 files that would have enabled them to respond. They were also unable to produce so much as a list of the files that had been destroyed.¹⁶

New digital forms of communication and conducting government business have exacerbated earlier weaknesses in records and information management

Digitisation and the introduction of e-government is often viewed as a panacea to government records and information management weaknesses. Even as technology makes it easier to initiate transparency reforms, the lack of attention to the quality and management of public sector data and documents can undermine the impact of right to information and open data initiatives. This trend can be observed across the board – virtually all countries are currently struggling with digital records management challenges to varying degrees.

For example, a survey on records management in Côte d'Ivoire, Ghana, Kenya, Liberia, Malawi, Sierra Leone, South Africa and Tanzania conducted by the World Bank in 2016 indicates that: 85 per cent of participating countries have digitised their public records, but only 16 per cent are storing digitised records and information in secure, professionally managed digital repositories that will ensure they will have access to good quality information through time.¹⁷ Seventy-one per cent of countries recognise that email is being used to conduct government business but, in an equal number of countries, public officials are using their personal email accounts and there are no policies in place to capture these types of records.¹⁸ Eighty-five per cent felt that they did not have sufficient policies and procedures in place to manage records in digital format in support of right to information and open government.¹⁹ And,

15 S. Trapnell and V. Lemieux, 'Report on a Pilot Study for Right to Information Indicators on Drivers of Effectiveness (RIDE)', World Bank, Washington, DC (hereinafter referred to as the RIDE Report), 16 April 2015.

16 'RTI reply suggests Union Home Ministry destroyed 11,100 files in July casually, without caring for rules', *Counterview*, 1 September 2014, <http://www.counterview.net/2014/09/rti-reply-suggests-union-home-ministry.html>.

17 A. Thurston and V. Lemieux, 'African countries come together to address gaps in managing digital information for open government', World Bank blogpost, 5 May 2016, <http://blogs.worldbank.org/governance/african-countries-come-together-address-gaps-managing-digital-information-open-government>.

18 Thurston and Lemieux, 'African countries'.

19 Thurston and Lemieux, 'African countries'.

more than half the countries reported that their staff had had no training in managing and preserving digital records and recognised an urgent need for technical assistance to provide such training.²⁰ They also pointed to the need to raise awareness across civil society organisations about records and information management challenges and their link to open government.

Weak control of digital records and information weakens transparency and public accountability mechanisms

Though much has been written about the positive power of technology and information to support greater transparency and accountability²¹ and, by extension, development, there have been unintended consequences and downside risks for transparency and accountability associated with the way recorded information is produced and managed in digitally enabled developing country public sector contexts. As a result of e-government initiatives and increasing digitalisation of government operations, public sector authorities have come to rely upon a growing array of communications technologies to create, exchange and store information – from traditional paper-based filing systems, to structured databases, ‘unstructured’ content management systems, social media platforms, web technologies, mobile platforms and cloud technology.²²

Managing the information created and contained in these systems has been another matter. Generally, there has been an inverse relationship between the age of the technology used to create, exchange and store information and the capability of public sector authorities to manage and preserve the information in a trustworthy and accessible form. Data and documents are often stored on personal drives, on personal accounts on commercially available email or social

20 Thurston and Lemieux, ‘African countries’.

21 See, e.g., J. Bertot, P. Jaeger and J. Grimes, ‘Using ICTs to create a culture of transparency: e-government and social media as openness and anti-corruption tools for societies’, *Government Information Quarterly*, 277 (2010): 264–71.

22 S. Katuu, ‘Enterprise Content Management (ECM) implementation in South Africa’, *Records Management Journal*, 22 (2012): 37–56; P. Van Garderen, ‘Electronic Records Strategy: Final Report’, 2002, http://www.interpares.org/display_file.cfm?doc=ip1-2_dissemination_rep_van-garderen_world-bank_ers_2002.pdf; J. Wamukoya and S.M. Mutula, ‘Capacity-building requirements for e-records management: the case in East and Southern Africa’, *Records Management Journal*, 15 (2005): 71–9; T. Peterson, ‘Temporary courts, permanent records’, Wilson Center, <https://www.usip.org/sites/default/files/resources/sr170.pdf>; D. Luyombya, ‘ICT and digital records management in the Ugandan public service’, *Records Management Journal*, 21 (2011): 135–44; G. Mampe and T. Kalusopa, ‘Records management and service delivery: the case of department of corporate services in the Ministry of Health in Botswana’, *Journal of the South African Society of Archivists*, 45 (2013): 2–23.

media platforms, on isolated computers, or on unmanaged network drives, making them difficult to access and unlikely to survive through time.

Nor are these issues unique to developing countries. In Canada, it is possible to point to examples of the impact of weak records and information management on transparency and public accountability. These include:

- the triple deletion of email messages,²³ specifically in connection with the murdered and missing indigenous women and generally in connection with freedom of information requests, as investigated by the former Information and Privacy Commissioner of British Columbia
- the use of personal devices by Nova Scotia public servants, as highlighted by the Information and Privacy Commissioner in that province, which can make it difficult, if not impossible, to locate relevant provincial information.²⁴

Persistent cultures of secrecy lead to oral government and avoidance of record-making and keeping

Many politicians and public officials do not truly embrace values of openness and transparency but continue to hold on to secrecy. Traditional cultures of secrecy can remain resistant to the introduction of right to information laws. Qualitative data from a recent study by the Africa Freedom of Information Centre show that in Kenya, Malawi and Uganda government/public officials continue to treat public information as ‘secret’ and therefore not accessible by anybody.²⁵ Journalists are considered by public officials to be the main beneficiaries of right to information legislation and not those they represent, namely the public, thus the need to enhance public officials’ knowledge of the right to information as being a right for all, and not only journalists.

When there are few staff incentives for openness and transparency (Table 11.1), this can leave an opening whereby public officials circumvent transparency and openness laws. This is most noticeable in the recent widespread use of private email servers and ‘out of band’ communication channels (for example,

23 E. Denham, *Access Denied: Record Retention Practices of the Government of British Columbia* (Victoria BC: Office of the Information and Privacy Commissioner, Investigation Report F15-03, 2015). Public officials were found to have triple deleted an email by moving it to the computer system’s ‘deleted’ folder, removing the email from the folder itself, and then manually overriding a backup that allows the system to recover deleted items for up to 14 days to avoid responding to freedom of information requests.

24 J. Laroche, ‘Personal email and government work should never mix, says Nova Scotia watchdog’, CBC News, 26 September 2016, <http://www.cbc.ca/news/canada/nova-scotia/foipop-email-texts-access-information-catherine-tully-1.3779161>.

25 *Enhancing Good Governance through Citizen Access to Information in Kenya, Malawi and Uganda* (African Freedom of Information Centre, 2017).

Table 11.1. RIDE indicators pilot survey results on staff incentives for six countries

<i>Staff incentives indicators</i>	<i>Albania</i>	<i>Jordan</i>	<i>Scotland</i>	<i>South Africa</i>	<i>Thailand</i>	<i>Uganda</i>
Protection from penalties for disclosure of information	Moderate	Weak	Excellent	Very good	Weak	Weak
Appropriate job demands	Moderate	Weak	Very good	Moderate	Moderate	Very weak
Clarity of rules	Moderate	Weak	Very good	Weak	Weak	Moderate
Performance evaluation	Weak	Weak	Not available	Weak	Moderate	Very weak
Lines of accountability	Moderate	Weak	Very good	Moderate	Weak	Moderate
Career prospects	Weak	Moderate	Very good	Very weak	Moderate	Weak

Source: Lemieux and Trapnell, *Public Access to Information for Development*.

BlackBerry Messenger) by public officials. When not addressed, such practices may erupt into a government scandal that undermines public trust, as was the case in British Columbia in Canada's triple-deleted emails scandal.²⁶

These cultures of secrecy can lead to bypassing of record controls and avoiding creating and keeping records altogether as a deliberate strategy to frustrate transparency laws. According to some, this 'off the record' practice of government is leading to more denial of access to information under right to information laws due to missing records.²⁷ For example, the Office of the Information Commissioner of Canada has seen an increase in the number of complaints received about missing records: of 51 per cent between 2011/2012 to 2012/2013 and 66 per cent from 2012/2013 to 2013/2014. This trend has led to a growing call to establish a legislative duty to document. Indeed, in a number of jurisdictions, such a duty already exists.

26 Denham, *Access Denied*.

27 N. Haughey, 'Leave no trace? How to combat off the record government', Open Government Partnership Blog, 2017, <http://www.opengovpartnership.org/stories/leave-no-trace-how-to-combat-off-the-record-government/>.

Good data are needed on records and information management implementation in support of transparency

Much of the evidence of negative side-effects of weak records and information management and, by extension, government transparency and accountability, derives from professional literature in the field of records and information management and archives administration, with some further evidence coming from transparency research. This body of literature has the advantage of being written by practitioners regarding what is happening on the ground within the public sector. However, it remains largely anecdotal in nature and geographically uneven in coverage. Consequently, it is very difficult to generalise from the findings: robust claims about effects and causality are difficult to make. To advance the discussion, good comparative data are needed about the state of records and information management in countries that could be used to analyse the relationship between the quality of recordkeeping and the ability to gather statistical data for reporting on SDGs.

Currently, no such up-to-date comparative data exist. In 1995, the International Council on Archives (ICA) undertook an 'International Survey of Archival Development', gathering statistical and qualitative information from national archives around the world.²⁸ This survey is now nearly 20 years out of date, and was completed before many countries began digitisation programmes or began to create and store records in digital form. There is therefore a need to establish a baseline picture of the current state of records and information management and archives administration as a first step towards identifying how the introduction of information and communication technology has changed the equation and the effect this has on public accountability and transparency.

The existence of good data and comparative indicators helps drive improvements and progress towards realisation of the SDGs. The Global Right to Information Rating developed jointly by the Centre for Law and Democracy and Access Info Europe, for example, has helped governments design new right to information laws, framed discussions between government and civil society about such laws, and in a number of cases (for instance, Morocco) has helped improve what was ultimately passed into law. There is an opportunity to achieve similar progress through development and application of records and information management indicators.

Recognising the need for a set of indicators to assess the state of records and information and to track progress, the programme of research on transparency and information management included the development of a tool for use in high-level assessments of the strength or weakness of systems of record, as

28 M. Roper, 'The present state of archival development world-wide', *Janus Archival Review* (1995): 11–47.

defined in international standards and in relation to ICT implementation. The tool identifies essential factors that predict whether the records created and held in these systems will be available and trustworthy through time to support development goals, and it provides simple tests to determine whether or not these factors are in place. The tool is structured as follows:



The three main goals that good records and information management support are transparency, accountability and the rule of law. In turn, these goals deliver progress on other SDGs, such as reduced corruption or gender equality. Records are key enablers for achieving these goals, but only if they are available and have integrity. If records cannot be found, contain errors or have been manipulated, they cannot fulfil their purpose of providing an evidence base to measure economic and social progress. To test whether recordkeeping regimes are sufficiently robust, the project team devised a series of simple questions as follows:

Policy

- has an information governance policy been developed to mandate the creation, management and preservation of digital records and associated metadata across the government or the organisation?
- has the policy been approved at the highest level of government?
- has it been disseminated at all levels of government?

Standards

- are international or national standards in place to provide consistent guidance on the structure and management of records and on metadata capture?
- are the standards understood and applied?

Roles and responsibilities

- has responsibility for the management of records been defined and assigned?
- is there a central government agency or authority with oversight for the delivery of an information governance programme across all media?

- do ICT and records authorities collaborate to define, implement and audit good practices for managing records and metadata?
- do system planners define and document systematic processes for records and metadata capture when ICT systems are defined?

Systems and practices

Can the five categories of information listed below be retrieved to provide evidence of actions and transactions one year old? Three years old? Five years old?

- is there an access control list specifying who may access the records and for what purpose?
- is there a robust event history that provides an audit trail of how the record has been viewed, accessed and used through time?
- are disposition requirements for records defined and applied in relation to legal, regulatory, fiscal and operational needs?
- is authority for disposition documented whenever records are destroyed?
- are there measures and structures in place to ensure the secure preservation of records and their metadata across time and across technological change?

Capacity

- are records professionals trained, in universities or management institutes, to understand national policy and international standards for creating, managing and preserving digital records as evidence?
- are they trained to implement good practice for managing records as defined in national or international standards?
- are they trained to communicate effectively with relevant government stakeholders, particularly ICT professionals and auditors?

Once the assessment is complete and areas of weakness have been identified, these can be addressed through the following strategies:

Policy

- define, disseminate and implement an information governance policy to mandate the creation, management and preservation of digital records and associated metadata across the government or the organisation.

Standards

- implement international or national standards to provide consistent guidance on the structure and management of records and on metadata capture.

Roles and responsibilities

- define and implement roles and responsibilities for creating, managing and preserving digital records and associated metadata as defined in the policy
- identify a central government agency or authority with oversight for the delivery of an information governance programme across all media
- establish a platform for collaboration between ICT and records authorities in defining, implementing and auditing good practices for managing records and metadata as part of ICT systems development and design.

Systems and practices

- develop and implement compliance audit provisions for metadata capture in line with those used for other asset management resources, such as human and financial resources, as a basis for ongoing assessment of gaps in systems and practices for managing records as evidence
- audit provisions for maintaining access control lists, event histories, disposition requirements and measures and structures to ensure the secure preservation of records and their metadata through time and across technological change.

Capacity

- develop education and training programmes that equip records professionals to understand national policy and international standards for creating, managing and preserving digital records as evidence; to implement good practice standards; and to communicate effectively with ICT professionals, auditors and other relevant government stakeholders.

Steps that can be taken to strengthen records and information management*Strengthen laws and policies governing digital records management*

Although a comprehensive and up-to-date dataset on the status of records and information management in governments worldwide does not yet exist, our programme of research indicates that in many countries, laws governing the management of records and information management need to be updated to improve the governance of digital records. Even countries with strong records management capacity struggle to keep up with the challenges to effective

records management brought about by technological change. Efforts must be made, therefore, to clarify the status of new forms of documents, such as emails and social media, and to bring legal and regulatory frameworks for the management of records and right to information regulations into alignment with technological realities.

A recent example comes from the United States, where laws and policies were updated in 2016 to clarify guidance on how digital records, such as emails, must be handled in order to support the public's right to information.²⁹ Scotland, too, has introduced recent reforms to its public records law requiring 250 public bodies to introduce records management plans approved by the national Keeper of the Records.³⁰ These reforms were introduced as a result of a public inquiry into problems with missing records.

Introduce independent records and information management oversight

Most public accountability mechanisms, including the right to information, operate best when there is some form of independent monitoring and oversight. In the case of right to information laws, this role is most often filled by an information commissioner who reports directly to the legislature. Given the importance of records in providing the evidence base for public accountability, it is surprising that in general, no such independent monitoring and oversight exists for the records and information management function. Instead, most public officials responsible for government recordkeeping report up the chain of command within the executive branch of government. This can weaken records and information management because records management staff may have a tendency to default to the values of secrecy in the same way that other public officials may do.³¹ Even when they are supportive of transparency and openness, they may have no recourse to draw attention to practices or conditions that undermine records and information management effectiveness.³² For this reason, there is a need to establish the same independent monitoring and oversight for records and information management as exists for other public accountability mechanisms within government.

In many jurisdictions, information commissioners have been taking up this role informally, but generally they have no power in relation to recordkeeping

29 See, e.g., United States, The Freedom of Information Act, 5 USC, Chapter 552 and the FOIA Improvement Law, Public Law No 2016, 114–85, <http://www.justice.gov/oip/freedom-information-act-5-usc-552>.

30 Public Records (Scotland) Act (2011), Acts of Parliament, 2011, asp 12, <https://www.legislation.gov.uk/asp/2011/12/contents>.

31 See, e.g., R. Jimerson, 'Archives for all: professional responsibility and social justice', *American Archivist*, 70 (2007): 252–81.

32 Postmedia News, 'Destroying gun registry records a "terrible precedent": archivists', *National Post*, 2011, <http://news.nationalpost.com/news/canada/destroying-gun-registry-records-a-terrible-precedent-archivists>.

practices. Often, they can only make recommendations and level criticism when records are not created or effectively managed. An update to right to information laws, for example to grant information commissioners more power to investigate records and information management practices, to issue penalties and fines, to require remedial action and to provide records and information management training to information commissioners and their staff, would strengthen records creation and management as public accountability mechanisms.

Align incentives of public officials with RIM principles and transparency policies and laws

Transforming cultures of secrecy and oral government into open and transparent government, with effective records and information management practices, will require going beyond legislative and policy changes to address values and incentives structures. Values of openness and transparency need to be supported at the level of policy-makers. Senior executives and public officials at all levels need to be incentivised to shift their values into alignment with these values. This can be achieved through internal training, setting appropriate job demands, clear rules, clear lines of accountability, performance management systems and strong career prospects for officials responsible for records and information management and for right to information implementation on the front lines. For example, in British Columbia, following an incident where a public official was found to have destroyed documents to avoid making them available in response to a request under the province's access to information law, the Chief Records Officer is leading a government-wide programme to provide access to information training and establish new lines of accountability, including setting access to information-related performance objectives for all staff.³³

Encourage collaboration

Staff working on right to information requests and those working on records and information management would benefit from working together closely, since records and information management staff are familiar with the nature and location of data and documents that may be requested under right to information laws. For example, the World Bank Group's chief archivist has responsibility for implementing the access to information policy.³⁴

33 See Government of British Columbia, A Practitioner's Guide to the Information Management Act, 2016.

<http://www2.gov.bc.ca/assets/gov/british-columbians-our-governments/organizational-structure/crown-corporations/central-agencies/practitioners-guide-information-management-act.pdf>.

34 See <http://www2.gov.bc.ca/gov/content/governments/organizational-structure/ministries-organizations/central-government-agencies/corporate-information-records-management-office/chief-records-officer>.

Collaboration among professional records managers and archivists and those engaged in right to information work will lead to more opportunities to work collaboratively within government as well. In that regard, it was heartening to see that the 2016 Open Government Partnership Summit included a session on records management and how to combat ‘off the record’ government. A blog on the session pointed to an observation arising from the discussions that, when it comes to open government reformers, ‘there are not always strong synergies between, on the one hand, Information Commissioners and right to know activists, and on the other, records-management professionals and public sector employee bodies who face records keeping challenges on a day-to-day basis’.³⁵ On the other hand, there was not one records manager or archivist on the panel of official speakers. Surely meaningful roundtable discussions about how to transform records and information management in support of transparency must begin by inviting records managers and archivists to sit at the table.

Conclusion

Although it is only one of many components of achieving effective implementation of right to information laws and open data, effective records and information management is critical. It has received little attention relative to its impact upon the effectiveness of transparency initiatives. The effectiveness of these initiatives, and by extension public accountability and trust, will depend upon making improvements to current practices of record creation, management and preservation in governments around the world.

Ultimately, this will require the ability to gather good data about the current state of the records and information management needed to benchmark initial conditions and make it possible to track progress. It will also rely upon updated laws and regulations that adequately address the new digital reality of government, the growing trend towards oral government/destruction of records and the need to align incentives in the public service to encourage behaviour that supports the goals of transparency and public accountability.

Finally, it will require independent monitoring and oversight of records and information management practices within government agencies, including more meaningful dialogue and closer ongoing collaboration between those responsible for transparency and those with responsibility for records and information management. With stronger records and information management across governments, the data needed to derive accurate, reliable and authentic statistical information relating to progress on the SDGs are achievable.

35 Haughey, ‘Leave no trace?’.

12. Information management for international development: roles, responsibilities and competencies

Elizabeth Shepherd and Julie McLeod

This chapter addresses the roles, responsibilities and competencies needed to manage information for international development, particularly in connection with measuring the Sustainable Development Goals (SDGs) reliably. We specifically focus on the management of records, which are information carriers and can contain or be used to create data and statistics. However, our approach can equally be applied to the management of data, statistics and information more generally.

Quality information for international development

An important theme running through the chapters in this volume is the role of records as evidence for accountability and transparency in civil society and for organisational decision-making. Information, data and records are crucially valuable for both national and international development generally, and for achieving the UN SDGs in particular. The ability to share information is of paramount importance for sustainable development in all areas.¹ Indeed, the importance of information is stressed in many of the SDGs. For example, SDG 16.10 specifically advocates ensuring public access to information. Information is also recognised as being vital to reviewing progress in implementing the SDGs and targets and is embedded in the UN Sustainable Development Agenda: 'Quality, accessible, timely and reliable disaggregated data will be needed to help with the measurement of progress and to ensure that no one is left behind. Such data is key to decision-making. Data and information from existing reporting mechanisms should be used where possible.'² Records

- 1 G. Chowdhury and K. Koya, 'Information practices for sustainability: role of iSchools in achieving the UN Sustainable Development Goals (SDGs)', *Journal of the Association for Information Science and Technology*, 68 (2017): 2128–38. <https://doi.org/10.1002/asi.23825>.
- 2 United Nations, *Transforming Our World: The 2030 Agenda for Sustainable Development*, point 48, 2015. A/RES/70/1, <https://sustainabledevelopment.un.org/post2015/transformingourworld/publication>.

document reporting mechanisms and other organisational processes. However, to realise the value and the role of records and the data they contain, records need to be managed effectively.

One of the significant challenges of managing digital information in the context of supporting international development is the quality of information and, therefore, the quality of data, records and statistics. Quality information is an essential requirement. If we make decisions, develop strategies or policies based on poor quality information then, at best, they will be flawed or inadequate, at worst, catastrophic. This is explicit in the SDGs. For example, SDG 14.5 states that the conservation of coastal and marine areas will ‘be consistent with national and international law and based on the best available scientific information’. To take another example, SDG 12.6 refers to sustainable consumption and production patterns and encourages companies to ‘adopt sustainable practices and to integrate sustainability information into their reporting cycle’.

But what is quality information? Scholars who have studied information quality have approached the question by developing sets of attributes, or criteria, that information should have (or should meet) for it to be quality information. For instance, Miller identified ten attributes of quality information as: relevance, accuracy, timeliness, completeness, coherence, format, accessibility, compatibility (with other information so that it can be combined), security and validity.³ Validity, Miller said, was ‘resultant rather than a causal dimension of information quality’. The ten attributes reflect four dimensions of information or data quality that are important to data users. Wang and Strong refer to these as:

- *intrinsic quality*: the information/data should have quality in their own right, such as correctness, consistency, validity and completeness
- *contextual quality*: data quality must be considered in the context of the task at hand, for instance, relevance and timeliness
- *representational quality*: for example, the amount of information and its format
- *accessibility*: for example, availability and access security can be restricted and provided securely.⁴

The last two dimensions emphasise the importance of the role of systems in supporting information quality. In other words, and slightly rewriting the authors, ‘high-quality information [data] should be intrinsically good,

3 H. Miller, ‘The multiple dimensions of information quality’, *Information Systems Management*, 13 (1996): 79–82.

4 R.Y. Wang and D.M. Strong, ‘Beyond accuracy: what data quality means to data consumers’, *Journal of Management Information Systems*, 12 (1996): 5–34.

contextually appropriate for the task, clearly represented, and accessible to the [data] consumer'.⁵ Information that lacks some of these attributes has resulted in flawed decision-making that has been identified as a contributory cause of major disasters.⁶ Quality information is not necessarily 'perfect' information but is rather the best quality we can have in the particular context or circumstances. It is vital for all stakeholders, including information creators, managers and users, to be discerning. They should adopt a degree of scepticism and evaluation, seek to evaluate the risks involved and question the degree of trust that can be placed in information when using it to underpin decision-making and action. Foster et al. suggest an approach to information governance that helps to balance risk and value by asking questions about people, processes and value.⁷ They identify organisational conditions such as ICT infrastructure, capability and culture and, at a micro level, structural, procedural and relational (which includes education and training) conditions that will support better information governance.⁸

Delivering information quality comes from good design of systems and processes together with good governance, including policy, standards and the audit of those systems and processes. International standards provide guidance on the design and operation of effective information systems, for instance standards on data quality and its management (ISO 8000 series) and on the management of records (to protect their characteristics of authority, authenticity, reliability, integrity and usability – ISO 15489:2016; ISO 30300:2020; ISO 30301:2019).⁹

Delivering quality information requires not only system capability and financial resources, but also human capacity. Human capacity covers three main groups of people. The first group comprises what we might broadly call the information professionals – those with direct responsibility for managing the quality of information, including information managers, records managers and those with responsibility for delivering the infrastructure that supports the management of information, such as processes and context for information security, data protection, information and communication technologies. ISO

5 Wang and Strong, 'Beyond accuracy', p. 6.

6 C.W. Fisher and B.R. Kingma, 'Criticality of data quality as exemplified in two disasters', *Information and Management*, 39 (2001): 109–16.

7 J. Foster, J. McLeod, J. Nolin and E. Greifeneder, 'Data work in context: value, risks, and governance', *Journal of the Association for Information Science and Technology*, 69 (2018): 1414–27. <https://doi.org/10.1002/asi.24105>.

8 Foster et al., 'Data work', p. 1424.

9 ISO 15489:2016, Information and Documentation – Records Management – Part 1: Concepts and Principles; ISO; ISO 30300:2020, Information and Documentation – Management Systems for Records – Core Concepts and Vocabulary; ISO 30301:2019, Information and Documentation – Management Systems for Records – Requirements.

30301:2019 identifies the people responsible for implementing management systems for records as including 'professionals in the areas of risk management, auditing, records, information technology and information security'.¹⁰ The second group comprises senior managers within the organisation who make decisions, determine policies and provide resources that affect the ability of the information professionals to work effectively.

The third group comprises information users. The users may be internal to the organisation, both at operational and strategic levels, including, for instance support functions such as HR and finance, as well as specialists in particular operational areas and statisticians. They also include users external to the organisation. In the context of international development, external users can include policy-makers, governments, citizens, advocacy groups, third sector and charitable organisations and independent oversight authorities.

The first group, professionals concerned with ensuring information quality, are people with the knowledge and skills needed to ensure that data, records and information are managed from creation through to destruction or preservation and remain accessible and useable. Foster et al. identify groups of professionals who are critical to data work as including 'IT professionals, legal specialists, risk and security professionals, health and business users, along with data and information professionals'.¹¹ This group also includes data scientists and statisticians, who have the analytical knowledge and skills needed to design algorithms, analyse, link, extract, visualise and present data for the users.

This first group needs the support of the second key group – senior managers – or those 'who make decisions regarding the establishment and implementation of management systems within their organisation',¹² who are in a position to advocate high-quality and effectively managed information, as highlighted in ISO 30301:2019. Managers provide the contextual infrastructure that is essential for the work of information professionals, the necessary resources, and the communication and policy systems that enable the creation, management and use of data and information, both internally and externally. Senior managers are also information users.

The third group comprises data and information users, who need to be satisfied with the quality of the data or information and aware of their own role in assessing quality. This involves being able to trust the systems and organisations that provide the information.¹³ Information users need to ask questions about how the information was produced. Their judgement about the

10 ISO 30301:2020, p. vi.

11 Foster et al., 'Data work', p. 1424.

12 ISO 30301:2019, 0.4.

13 A. Sexton, E. Shepherd, O. Duke-Williams and A. Eveleigh, 'A balance of trust in the use of government administrative data', *Archival Science*, 17 (2017): 305–30. <https://doi.org/10.1007/s10502-017-9281-4>.

reliability and quality of the information involves assessing the risks involved. For instance, if the information is based on an analysis of data or statistics, then the analytical methods, approaches to data linkages, models or algorithms must be transparent so that they can be understood clearly. Increasingly, researchers are seeking to develop models and approaches for Human Explainable Artificial Intelligence, which should make it easier to understand algorithms and to improve transparency and accountability.¹⁴

A senior manager or decision-maker is unlikely to fully understand all the questions that need to be asked in order to have full confidence in the reliability and quality of the information in order to trust it. Judgement about information quality involves assessing the risks involved. There is a need to adopt a risk-based approach, much as a statistician does in presenting the results of a statistical analysis with a greater or lesser 'degree of confidence'. It is also important to realise that quality information for one person may not be perceived as quality information by another.¹⁵ Users' needs are different, may change through time or may depend on the particular context.

It is clear that decisions and actions are only as good as the information on which they are based. A balanced approach to managing quality information requires skilled records and data management professionals, supported by managers who advocate systems for quality information management. It also demands discernment on the part of users in assessing and trusting the quality of the information they use.

Key players in records management, their roles and responsibilities

Managing records is not just the remit of a specialist group of information professionals; rather it is a shared responsibility of multiple players who fall into the three groups outlined above. The first group (information professionals) includes records professionals, legal and information technology professionals. The second group comprises managers (including leaders and senior managers, as well as project and programme managers) who enable the work of the information professionals. The third group (the users) includes all organisational personnel and staff, together with external stakeholders, including contractors, with whom business processes and records are shared.¹⁶

14 D. Gunning, 'Explainable artificial intelligence', 2018, <http://www.darpa.mil/program/explainable-artificial-intelligence>; UK Information Commissioner's Office, 'Automated decision-making and profiling', 2018, <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/>.

15 H. Miller, 'The multiple dimensions of information quality', *Information Systems Management*, 13 (1996): 79–82.

16 ISO 15489:2016; ISO 30300:2020. ISO 30301:2019.

This section identifies the respective responsibilities of these three groups of players for managing records and information for development.

Group 1: professionals with the necessary technical skills and qualifications (such as records, IT) to ensure information quality

Organisations need to focus on making professionals aware of how their knowledge, professional skills and information competencies can be used effectively to support sustainable development. ISO 30301:2019 sets out the broad operational responsibilities to be carried out by ‘a specific records operational representative who shall have a defined role, responsibility and authority’ that includes ‘implementing the M[anagement] S[ystem for] R[ecords] at the operational level, reporting to the top management on the effectiveness of the MSR for review, including recommendations for improvement, and establishing liaison with external parties on matters relating to the MSR’.¹⁷

Group 2: managers (senior, programme, functional) who enable or facilitate the work of the professionals

Managers need to understand the importance of managing records, the impact of their decisions on the organisation’s capacity to manage and protect records’ quality and the implications for decision-making and actions in a development context. ISO 30301 recommends that ‘top management shall ensure the responsibilities and authorities for relevant roles are assigned and communicated within the organization’ and that those responsibilities ‘shall be appropriately allocated to all personnel at relevant functions and levels within the organization ... who create and control records as part of their work’, to support and enable the work of those professionals.¹⁸ In particular, ‘they shall assign the responsibility and authority for: a) ensuring that the MSR conforms with the requirements of this document and b) reporting on the performance of the MSR to top management’.¹⁹

Group 3: all other stakeholders and users of the information, inside and outside the organisation

Organisations need to recognise the value of quality information, the contribution that good records and information management makes and the need to be able to access relevant and timely information for development purposes. They need to be able to assess the quality of information and to ensure that it is used appropriately to inform development decision-making, policy, processes and operations. ISO 30301:2019 states that there should

¹⁷ ISO 30301:2019, 5.3.

¹⁸ ISO 30301, 5.3.

¹⁹ ISO 30301, 5.3.

be a 'periodic review of the competencies and training of those personnel' to ensure that 'these persons are competent on the basis of appropriate education, training, and experience'. The requirement for managers to 'take actions to acquire the necessary competence and evaluate the effectiveness of the actions taken' implies a thorough programme of skills development and training which goes beyond the information professionals and managers.²⁰

The competency of these three groups of key players directly affects an organisation's capacity for managing its information and records and, as a consequence, its sustainable development capacity.

Capacity for managing records

One of the targets in SDG 17 (strengthen the means of implementation and revitalise the Global Partnership for Sustainable Development) is capacity building (SDG 17.9). The AHRC Network, from which this book emerged, is an example of 'international support for implementing effective and targeted capacity-building in developing countries to support national plans and implement all the Sustainable Development Goals'.²¹ John McDonald identifies five capacity levels, with Level 1 being the lowest level and least developed capacity, and Level 5 being 'an ideal state for a country that wants to ensure that data, statistics and records used to measure the SDG indicators are of a high enough quality to measure and implement the goals'.²² These levels, he states:

reflect diminishing levels of risk, with Level 1 representing the highest risk of loss and inaccuracy and Level 5 being the least risk. They also reflect increasing levels of sophistication in terms of the way data, statistics and records can be used to support implementation of the SDG goals and, more broadly, the government's operational and strategic goals.

Inevitably, building capacity for managing records is constrained by the resources available. ISO 30301 states that 'the organization shall determine and provide the resources needed for establishment, implementation, maintenance and continual improvement of the MSR', and therefore careful consideration is needed in the context of short-term priorities, longer-term strategy and

20 ISO 30301:2019, 7.1, 7.2.

21 United Nations, 2015, Goal 17.9 Capacity Building.

22 J. McDonald, 'A matter of trust: records as the foundation for building integrity and accountability into data and statistics to support the UN Sustainable Development Goals. Concepts, issues and potential strategies' (Institute of Commonwealth Studies, 2018), pp. 18–19.

an assessment of the concomitant risks.²³ However, Level 2 is the minimum capacity level that all organisations should aim to reach. McDonald warns that ‘achieving Level 5 or even Level 4 will be challenging’.²⁴ There may be pockets of good or excellent practice in organisations, but they should strive to reach the best level possible given the resources available in all key functions, in order to minimise risks.

McDonald’s five-level model has resonance with a well-established Information Governance Maturity Model, developed by ARMA International, based on a high-level framework of good practice, the Generally Accepted Recordkeeping Principles (GARP).²⁵ The GARP Information Governance Maturity Model descriptors are used as a benchmark in the section that follows. As that model reflects, as an organisation develops its information governance and management programme, the people involved ‘will likewise progress through a spectrum of increasing competence and effectiveness’, ultimately achieving a transformational and sophisticated skill set.²⁶

This chapter considers, for each of the five capacity levels, the skills and competencies of the three groups of people identified in the previous section: the professionals (Group 1), the managers (Group 2) and other stakeholders and users of information (Group 3).

Across all five levels, users and other stakeholders need to be aware of what records exist, how they can be used to support the organisation’s work, and how the quality and value of information for development purposes can be judged. We set out some descriptors and examples of the skills, competencies and knowledge needed by each group at each level. However, we focus mainly on the professionals and the managers, who are the key players most actively involved in ensuring records and information quality. Professionals and managers need to be able to recognise the value of quality records and information and to follow systems that are capable of producing good records and information management if they are to play their role in ensuring the ‘use-value’ of the data they contain for SDGs.

Capacity Level 1

(Poor quality records undermine SDG implementation)

Organisations whose capacity is at Level 1 do produce some statistics for SDGs, but these are unreliable; the systems for managing them are unaccountable

23 ISO 30301:2019, 7.1.

24 McDonald, ‘A matter of trust’, p. 18.

25 ARMA International, Information Governance Maturity Model, <https://www.arma.org/page/PrinciplesMaturityModel>.

26 ARMA International, Generally Accepted Recordkeeping Principles®, Information Governance Maturity Model, 2013, p. 3.

and lack transparency. In general, at Level 1, organisations are unwilling or unable to commit resources to records management systems, and they do little or nothing to assure data quality. They therefore take the risk that they will be unable to measure progress towards SDGs. The GARP Information Governance Maturity Model describes this level as ‘Sub-Standard’.²⁷ It is ‘an environment where information governance and recordkeeping concerns are not addressed at all, are addressed minimally or are addressed in an ad hoc manner ... programs will not meet legal or regulatory scrutiny and may not effectively serve the business needs of the organisation’.

Group 1: professionals

At this level, the organisation may have no records professionals or, where there are records staff, they lack the knowledge and skills needed to develop a reliable framework of policies, standards, practices and systems for managing records. The information management role is largely non-existent or treated as a purely administrative role without the need for any specialist knowledge. As a result, there is little or no metadata for records; metadata standards are not implemented systematically, and systematic control processes are largely or entirely lacking or unreliable. Version controls are not implemented and information requests cannot be fulfilled. Records are preserved in an ad hoc manner and there are no staff skilled in implementing preservation standards appropriately. Staff lack understanding of basic information security and controlled access processes. Paper records are at risk, although they may survive, but digital records are very likely not to be preserved if there are no staff with digital curation and preservation expertise to implement active preservation strategies. The necessary ICT skills and knowledge are lacking.

Group 2: managers

At this level there is no senior leadership for effective information management. There may be no recordkeeping processes or systems, or they may be ad hoc and undocumented. There is no application of existing standards. Managers fail to provide central oversight or guidance to ensure consistent information practices and lack an awareness of the need for such guidance; they fail to ensure that retention processes are developed and implemented and do not understand the need for timely destruction or preservation of records. Personnel are not trained to document their decisions. Staff are unaware of their responsibilities as information creators and users, and there are no processes or procedures for

27 ARMA International, Generally Accepted Recordkeeping Principles®, Information Governance Maturity Model, p. 2.

managing information that needs to be shared with external stakeholders such as contractors, collaborators and citizens.

Group 3: other stakeholders and users

The needs and skills requirements of this group who, whether internal or external to the organisation, should be able to access and use records and data relating to SDGs, are neglected. At this lowest level, awareness of the existence of data and records is completely or almost completely lacking. Records and data are not available or accessible to users and other stakeholders, who generally do not know about relevant information that might help them do their work. They take decisions, make policy and carry out actions without consulting or using records and data. Often they do not have skills in resource discovery or information literacy that would enable them to access and use data. Development of policy and work that takes place in ignorance of relevant records and data is poorly informed, based on irrelevant or poor-quality information, or purely on personal or very localised understanding. Decisions lack transparency and accountability. Decision-making tends to be idiosyncratic and inconsistent across time and space.

The consequences for the organisation of poor, inconsistent or inadequate staff skills and knowledge for managing information are the loss of evidence of activities and actions, and inadequate authoritative, quality and reliable information to underpin decision-making. There is a high level of risk that decisions will be inappropriate or poor. There is a loss of organisational memory for planning and development, and, in the wider context of international development, SDG measurements and implementation are undermined.

Capacity Level 2

(Records enable SDG implementation at a basic level)

At Capacity Level 2, a basic framework of laws, policies, standards, procedures and people is in place to ensure that data and statistics are gathered and analysed to measure the SDGs with some accuracy and reliability. Level 2 is the minimum acceptable level for meeting basic needs for records to support the achievement of SDGs. The GARP Information Governance Maturity Model describes this level as 'In Development', 'an environment where there is a developing recognition that information governance and prudent recordkeeping have an impact ... however ... its practices are ill-defined, incomplete, nascent, or only marginally effective'.²⁸

28 ARMA International, Generally Accepted Recordkeeping Principles®, Information Governance Maturity Model, p. 2.

Group 1: professionals

At this level, the role of records and information professionals is recognised within the organisation, but staff are only skilled sufficiently to administer existing information programmes. They lack the skills needed to develop policies for managing the records that document processes for collecting and processing data and producing statistics; they lack the expertise needed to manage the complex interrelationships among data, statistics and records, especially those that need to be preserved through time and in digital formats. Emails, reports, logs and other records documenting the design and conduct of data collection cannot be related to records documenting processes for extracting and analysing data and producing statistics. Digital skills lie with ICT staff, who lack understanding of how to apply them to data management and analysis. Holistic approaches needed to ensure that records are managed coherently, and through time, are lacking. Typically, no staff have expertise in digital preservation strategies and approaches. Information practices, such as metadata management, retention scheduling and approved record destruction are applied at best in localised processes and often inconsistently.

Group 2: managers

Senior managers at Level 2 are generally aware of their responsibility for ensuring that data and statistics, with their supporting documentation, are stored properly. However, no control framework is applied universally, and some managers maintain poorly documented records. Managers realise that some degree of transparency and accountability in information asset management is needed, but they are not sufficiently experienced to be able to ensure that this is implemented widely. They do not provide training or guidance for employees in a formal or systematic fashion, which results in patchy and variable practices and a lack of universal policy. Senior managers may be aware of some compliance issues, but they are insufficiently knowledgeable about the details or of good information practices. Compliance is not given the profile it should have by senior managers, which opens the organisation up to risks.

Group 3: other stakeholders and users

At Level 2, users and external stakeholders are not given sufficient guidance and training to be able to understand the records and data they use, which typically lacks metadata or contextual information, or if metadata is available, users are not skilled in interpreting it accurately. Exchange of information between internal and external users is not properly regulated. Therefore, some users may access records that others cannot, and data may be shared in illegal or unethical ways. Users need training in proper information handling practices that respect legal and regulatory requirements. Legal discovery and access requests by third parties do not receive consistent responses.

Capacity Level 3

(The quality of records makes it possible to measure SDGs effectively and supports government programme activities)

At Capacity Level 3, a more comprehensive framework of policies, standards and practices, systems and technologies, and qualified staff exists, which means information and records can be trusted. The GARP Information Governance Maturity Model describes this level as 'Essential'.²⁹ Organisations that have achieved the minimum requirements are 'characterised by defined processes and procedures ... the key basic components of a sound program in place'. This makes the organisation 'at least minimally compliant with legal, operational and other responsibilities'.

Group 1: professionals

At this level, information professional roles are recognised and staff are competent to apply clear, consistent standards and practices. Records, information and other professionals work effectively with data management, ICT and other professional staff to ensure that requirements for identifying, describing, classifying, protecting and retaining data, statistics and records are integrated in the design of processes for collecting data and producing and using statistics. Professional staff are trained to apply clear, consistent standards and practices, at least to paper records and, to a more limited extent, to digital data. Preservation is generally not addressed adequately, so that data or statistical comparisons over long periods are not assured. Information staff put organisation-wide policies and standards in place as a strategic basis for protecting records of decisions.

Generally, staff have not developed detailed retention requirements, nor metadata standards for records. Staff generally lack the necessary knowledge of digital preservation standards, procedures and technologies. ISO 30301:2019 states, 'This family of standards is intended to be used in support of: a) top management who make decisions regarding the establishment and implementation of management systems within their organization; b) people responsible for the implementation of MSR, such as professionals in the areas of risk management, auditing, management of records, information technology and information security.'³⁰ In order to achieve Capacity Level 3, professionals need to be able to develop and implement coherent records and information systems.

29 ARMA International, Generally Accepted Recordkeeping Principles®, Information Governance Maturity Model, p. 2.

30 ISO 30301:2019, 0.4.

Group 2: managers

Senior managers at Capacity Level 3 know that they are responsible for ensuring that the records generated are authentic, reliable, accessible and understandable and can be retrieved when needed. ISO 30300 provides guidance on the requirements for Management System Standards (MSS), which 'provide tools for top management to implement a systematic and verifiable approach to organizational control in an environment that encourages good business practices'.³¹ The standards on management systems for records are 'designed to assist organizations of all types and sizes, or groups of organizations with shared business activities, to implement, operate and improve an effective management system for records ... The MSR directs and controls an organization for the purposes of establishing a policy and objectives with regard to records and achieving those objectives'.³² ISO 30301:2019 sets out four activities to achieve this: 'defined roles and responsibilities; systematic processes; measurement and evaluation; review and improvement'.³³ At Capacity Level 3 or better, senior managers should understand the requirements for effective management systems for records and should provide the policy and resource infrastructure needed to develop and implement them.

Senior managers understand the requirements for information compliance and take responsibility for ensuring that compliance has a sufficiently high profile in the organisation to justify the allocation of resources to formal systems and processes in order to implement compliance policies. They understand data privacy, legal issues and confidentiality, and training for all staff in understanding these issues is available. The management of risk and the need for compliance (and its costs) are assessed by skilled people, so that organisational risks are balanced. Managers and employees across the organisation are trained and knowledgeable about information policies, and they understand personal and organisational responsibilities for records.

Group 3: other stakeholders and users

At this level, record users are aware of 'the records policy; their contribution to the effectiveness of the MSR, including the benefits of improved records processes and systems performance; the implications of not conforming with the MSR requirements'.³⁴ They have sufficient knowledge of the creation context and sufficient understanding of the quality of the information they require so that they can make fairly informed decisions about the data they access for development questions and how to reprocess them appropriately.

31 ISO 30301:2019, 0.2.

32 ISO 30301:2019, 0.2.

33 ISO 30301:2019, 0.2.

34 ISO 30301:2019, 7.3.

Data requests and data sharing across third parties are, in the main, systematic and legally compliant.

Capacity Level 4

(Well-managed records make it possible to measure SDG implementation effectively and consistently through time; data and statistics are of high enough quality and integrity to support government programme activities at the strategic level)

At Capacity Level 4, data and statistics generated to measure SDG indicators are reliable and can be linked and combined with other data sources to support other activities. The GARP Information Governance Maturity Model describes this level as ‘Proactive’, with an ‘information governance program throughout its operations ... routinely integrated into business decisions’.³⁵ The organisation is ‘substantially more than minimally compliant with good practice and easily meets its legal and regulatory requirements’. It is able to mine its information for better services and is ‘transforming itself through increased use of its information’.³⁶ ISO 30301:2019 suggests that ‘Top management shall demonstrate leadership and commitment with respect to the M[anagement] S[ystem] for R[ecords] by ensuring that the records policy and records objectives are established and are compatible with the strategic direction of the organization’.³⁷ Well-managed and properly resourced organisations should aspire to achieve Capacity Level 4 across the organisation.

Group 1: professionals

At Level 4, records and information roles are assigned to senior appointments. Recruitment is fair and open to ensure the necessary skills and experience in the post holders; records management is embedded in the strategic operation of the whole organisation. Records professionals have the skills to deliver accountability requirements through consistently applied records management policies and standards. Preservation standards ensure that records are stored properly and migrated to take account of changes in technology. Staff are trained to deliver a preservation and management programme that ensures continued accessibility and authenticity of records in all formats through

35 ARMA International, Generally Accepted Recordkeeping Principles®, Information Governance Maturity Model, p. 3.

36 ARMA International, Generally Accepted Recordkeeping Principles®, Information Governance Maturity Model, p. 3.

37 ISO 30301:2019, 5.1.

time. Professional staff have the skills needed to implement information access regimes that are compliant with legal, regulatory and ethical practices. They are also trained trainers, so that they can develop information skills training for all organisational staff.

Group 2: managers

The role of senior management in a compliant organisation is to ‘set an organization’s direction and communicate priorities to employees and stakeholders.’ Senior managers need the skills to lead the organisation with respect to information and data, creating an environment in which managers can ‘establish a records policy that: a) is appropriate to the purpose of the organization; b) provides a framework for setting records objectives’. In addition, ‘The records policy shall include the high-level strategies with regard to the creation, capture and management of authentic, reliable and useable records capable of supporting the organization’s functions and activities’.³⁸ Resources and staff competencies are available to maintain, review and develop the information infrastructure. At this level, project or programme managers should have the skills to analyse information trends through time and make comparisons from year to year using reliable records, because changes to formats, coding schemes and data collection and analysis methods are well-documented. Managers understand the audit and compliance requirements and are able to implement and oversee them.

Group 3: other stakeholders and users

External users of data can be assured that what they are using is quality information and that the compliance and access processes are robust and reliable, so that data requests are fulfilled in a timely, complete and transparent way. Managers ensure that information professionals can develop and deliver staff training programmes that formally train all information users to have the correct skills of data analysis, information literacy and understanding of the context of the data they use. All staff receive training related to their information and records handling needs, which could cover classification and metadata tagging of information, retention and disposal rules, access and privacy regimes and destruction processes across all formats and media.

Capacity Level 5

(Processes generating records, and the framework for managing them, are designed to make it possible to exploit data, statistics and records,

38 ISO 30301:2019. 5.2.

including the information used for measuring SDGs, in new and innovative ways)

At Capacity Level 5, organisations responsible for managing data and statistics enable innovative thinking about implementing and going beyond SDGs. These organisations are international leaders in good practice; they enable government and international policies to be developed and new thinking to emerge. Standards makers, and setters of high benchmarks, are hallmarks of these organisations. They influence practice on SDGs data worldwide. The GARP Information Governance Maturity Model describes this level as ‘Transformational’, where the organisation has integrated information governance into its infrastructure and processes, making compliance ‘routine’.³⁹ It both recognises the value of information as a critical asset and implements ‘strategies and tools to achieve these gains’. As a result, the risk of inappropriate information disclosure and data loss is low. However, Level 5 represents the best possible world of information management for SDGs and may be costly to implement holistically. Organisations need to evaluate the return on investment and come to a judgement about the correct balance between risk and cost.

Group 1: professionals

At this level records professionals have the knowledge and expertise needed to design comprehensive management frameworks, covering multiple organisations and technology environments that encourage information in the records to be exploited to the greatest possible extent. IT professionals work to support innovative and advanced technologies, ensuring that information is published in new forms to meet the needs of a wide range of individuals and groups and to give citizens access, regardless of location. Open access data should meet high interoperability standards, for instance the Five Star Open Data Scheme.⁴⁰

At Level 5, information professionals are world-leading in their skills and knowledge, which goes well beyond the boundaries of the organisation, taking a wide and well-informed view. They are able to develop information systems that can be adopted within and outside their own organisation and that provide benchmarks for good practice. For example, records preservation in all formats, including digital, is highly developed and fully implemented; staff are engaged in continual improvement across the whole organisation and across all functions and processes.

39 ARMA International, Generally Accepted Recordkeeping Principles®, Information Governance Maturity Model, p. 3.

40 Five Star Open Data Scheme, <http://5stardata.info/>.

McDonald suggests that a leader in this field should have ‘background in data management, statistics, information technology or records management, the capacity to bridge these disciplines and the ability to communicate with a variety of stakeholders, including senior management’.⁴¹ He or she should be supported by a steering committee, ‘made up of representatives from government programmes supporting the SDGs as well as programmes where the quality and integrity of data, statistics and records is particularly important. Specialists in managing data, statistics, records and information technology, as well as legal experts and auditors should also be included’.⁴²

Group 2: managers

Senior managers of SDG initiatives at Level 5 must understand the benefits of sharing and exploiting data, statistics and records for stimulating innovative thinking on implementing the SDGs, as well as for achieving the operational goals of individual programme activities and the strategic goals of the organisation. The organisation’s governing council should include the chief information officer in order to ensure high-level support for information goals. Teams of managers should be assembled to review the nature and quality of the information needed to meet SDG targets; these teams should include specialists in managing records, data and statistics, and information technology, legal experts and auditors.

Group 3: other stakeholders and users

At this level, information users have full and free access to the data they need, which is fully interoperable across systems, reliably reproducible, and can be easily reprocessed. Comprehensive training is available, which enables data users to develop their skills and understanding of the records and data needed to measure SDGs.

Determining and achieving the desired capacity level

Organisations should make risk assessed and strategic decisions about the desirable capacity level for the organisation – both for the organisation as a whole and in part. This is based on what is realistic and affordable and what provides the required return on investment in terms of reducing information risks. Benchmarking, good practice, the organisation’s mandate, stakeholder expectations and trust are all issues that affect the decision about desirable

41 McDonald, ‘A matter of trust’, p. 20.

42 McDonald, ‘A matter of trust’, p. 20.

capacity level. Capacity Level 1 is not desirable for international development and the measurement of SDGs as it does not meet even the lowest threshold of quality. As McDonald suggests, Level 2 should be the minimum level that is acceptable in an organisation with development goals in its mandate.⁴³ However, Level 5 may not be appropriate either, at least not organisation-wide, as the costs of compliance and policy engagement, whilst desirable, may not be affordable. Organisations should assess and measure the risks they are willing to take, with the resources at their disposal, to achieve an acceptable quality of information with the people needed to operate, manage and use the records needed to measure the SDGs.

Once an organisation commits itself to building systems for an identified capacity level, it will need appropriately skilled, experienced and trained people to move from its current state to the desired state. Normally, that will be done a step at a time. The organisation may also have short-term goals for developing certain aspects of the information systems and longer-term goals to implement a more comprehensive system to deliver quality information. ISO 30301:2019 identifies three key steps: 'determine the necessary competence of person(s) doing work under its control that affects the performance of its records processes and systems; ensure that these persons are competent on the basis of appropriate education, training, and experience; [and] where applicable, take actions to acquire the necessary competence and evaluate the effectiveness of the actions taken'.⁴⁴ It also states that evidence of competence should be documented and retained.

The standard identifies actions that an organisation can take, such as 'the provision of training to, the mentoring of, or the reassignment of currently employed persons; or the hiring or contracting of competent persons'.⁴⁵ Similarly, McDonald recommends a wide range of strategies to develop the necessary skills and competencies.⁴⁶ These include:

- defining the activities needed to manage records associated with measuring SDGs
- defining the staff competencies required
- designing and implementing staff training
- appropriate workforce development and recruitment
- measuring staff performance appropriately to develop competency

43 McDonald, 'A matter of trust', p. 18.

44 ISO 30301:2019, 7.2.

45 ISO 30301:2019, 7.2.

46 McDonald, 'A matter of trust', p. 16.

- establishing partnerships and secondments in order to enhance skills and expertise across the organisation and beyond
- ensuring the development of educational programmes (in partnership with universities) to ensure the provision of professional skills and qualifications across the records management framework.

In order to develop the required competencies to meet these information system goals, organisations should focus on five key approaches. First, they should seek to employ staff with formal qualifications, taught and accredited by universities and professional bodies. Second, they should develop a training programme for existing staff to develop their skills. Such training might be provided, for instance, by internal expert professionals, freelance trainers and consultants, relevant professional bodies that provide training courses and universities. Third, consultants and contracted staff with appropriate knowledge and expertise can be employed as change makers to boost the knowledge needed to make significant improvements. Fourth, organisations should consult national, regional and international standards that provide information and guidance about the skills needed and ensure that recruitment and development of people is in line with good practice. Finally, organisations should refer to relevant competency frameworks to benchmark the skills and knowledge needed by staff.

Employ staff with formal qualifications

In order to ensure that professional staff (Group 1) meet the needs of the desired capacity level, organisations commonly recruit staff with professional qualifications in relevant disciplines. Such qualifications are accredited and taught by universities and educational colleges or by the professional bodies in the field. Professional bodies sometimes accredit qualifications offered by universities rather than delivering qualifications directly, for example, an undergraduate or postgraduate level degree in records and information management, information security, or information science. This includes, for instance, programmes offered at Northumbria University in information governance and at University College London in archives and records management (both UK); at the School of Information Sciences at Moi University in Kenya; at the University of Botswana in information and knowledge management; at Universiti Teknologi Malaysia Faculty of Technology and Informatics in records and archives management; and at Monash University (Australia) in data science, information technology or social informatics. There are many others, too many to list here, around the world that teach in different languages and with different specialisations in the broad field of records and information.

Train existing staff

Many organisations already employ staff in information roles, but in many cases, the staff lack some of the skills and knowledge needed to develop their roles further. A programme of targeted training and continuing professional development for information staff can be developed. This might include, for instance, short courses offered by external providers, including professional associations, such as ARMA International⁴⁷, the International Council on Archives⁴⁸ or the learning materials by the International Records Management Trust⁴⁹. It could involve secondments to more advanced organisations or different functional areas, or training offered by national or regional institutions, such as the national archives and records service. Professional staff with leading-edge skills can also offer training for managers and information users covering their responsibilities for quality information and resources; the ICA's 'Training the Trainer' pack can support this.⁵⁰ Such training would respond directly to SDG 4 – 'Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all'.⁵¹

Contract expert staff short term as change makers

Expert contract staff and consultants are helpful to organisations who lack specific skills and who need an expert opinion and some specific guidance in order to improve systems for information. Employing a consultant enables an organisation to boost expert capacity for a short period and to deal with an identified and targeted issue. Consultants can also provide training for staff at all levels in the organisation according to need. Professional bodies may be able to provide contacts with suitably skilled consultants.

Use standards to guide practice and inform staff recruitment

International standards can provide a good practice threshold or benchmark that organisations can use to evaluate their practices and systems. Standards can also be used to identify gaps and omissions and inform staff recruitment or training needs. Relevant standards have been cited in this chapter, including ISO 15489:2016 Information and documentation – Records Management; ISO 30301:2019 Information and documentation – Management Systems for Records; and the ISO 8000 series on Data Quality. Practice standards are also published nationally or regionally or for specific functional areas, such

47 <http://www.arma.org/>.

48 <http://www.ica.org/en/training-programme>.

49 <http://www.irmt.org/>.

50 <http://www.ica.org/en/training-trainer-resource-pack>.

51 International Council on Archives, SPA [Section of Professional Associations] Report on Competency Accreditation, 2014, <http://www.ica.org/en/spa-report-competency-accreditation-certification>.

as national health services; in the UK, for example, the Data Security and Protection Toolkit has been developed by NHS Digital.⁵² Recruitment and person specifications are useful here. ARMA International, for instance, has linked publications providing a set of core competencies for records and information management⁵³ and a set of job descriptions based on those competencies.⁵⁴ Beyond that, job adverts and specialist recruitment agencies provide useful support.

Benchmark staff skills and knowledge against competency standards

Professional associations have developed competency standards to identify 'the skills, knowledge and behaviour required to work within a profession. These are acquired both through initial education and training and participation in a program of continuing professional development'.⁵⁵ The ICA's 2014 report identified a range of competency standards, for instance in Australia, where the Australian Library and Information Association has provided standards on core knowledge, skills and attitudes, and work level guidelines (in 2005), while the Australian Society of Archivists and the Records Management Association has developed an Australasia Statement of Knowledge for Recordkeeping Professionals (also in 2005).

In Canada, the Information Management Forum produced the Information and Records Management – Competency Profile (2000), and in the United States, ARMA International produced a standard on Core Competencies (2007, with a second edition published in 2017). In the UK, the Information and Records Management Society offers a statement of individual competency through its accreditation programme,⁵⁶ while the Archives and Records Association has developed a detailed competency statement for records staff at a range of levels of experience, from Level 1 Novice, Level 2 Beginner, Level 3 Competent, Level 4 Proficient, to Level 5 Expert/Authoritative, grouped under three areas of work: organisational, process and stakeholder/customer.⁵⁷

52 NHS Digital, 2019, <http://www.dsptoolkit.nhs.uk/>.

53 ARMA International, *Records and Information Management Core Competencies*, 2nd edn (2017).

54 ARMA International, *Job Descriptions for Information Management and Information Governance*, 2nd edn (2017).

55 International Council on Archives, SPA [Section of Professional Associations] Report on Competency Accreditation, 2014, <http://www.ica.org/en/spa-report-competency-accreditation-certification>.

56 <https://irms.org.uk/page/Accreditation>.

57 <http://www.archives.org.uk/cpd/competency-framework.html>.

Conclusion

Human capacity and competence are essential for delivering quality information for effective decision-making and organisational efficiency and, in the specific context of this book, to support international development and the measurement of the UN SDGs. In this chapter, we have identified three main groups of people with roles and responsibilities for delivering quality information by managing an organisation's records. These are: information professionals (including records, legal and information technology professionals), managers (including leaders, senior managers, project and programme managers) and users (including all organisational staff, external stakeholders and contractors). Their competence directly affects an organisation's capacity for managing its information, records and data.

We use the five capacity levels for delivering quality data, statistics and records identified by John McDonald to determine the competence (knowledge, skills, expertise) required of each of these three groups of people at each level. In doing so it is clear that Level 1 is not desirable within the specific context of the SDGs or more generally. Moving to higher levels can contribute directly to improving the way the organisation carries out its regular operational functions and its ability to achieve the SDGs. Achieving Level 5, a position of transformation, enables the organisation to use its information resources in new and innovative ways to support its operations, achieve its strategic goals and, most important, the ability to alter and even fundamentally change its business. However, this may be unrealistic for many organisations. The approach set out here will enable an organisation to assess its current capacity level, agree the desirable capacity level based on a risk assessment, and identify strengths and gaps in its human capacity for managing its information, data and records. We offer a range of strategies for building capacity to address the gaps, from employing staff with formal qualifications, to benchmarking staff skills and knowledge against competency frameworks. While we have specifically focused on records, our approach is equally applicable to building capacity for managing data and statistics.

13. The quality of data, statistics and records used to measure progress towards achieving the SDGs: a fictional situation analysis

John McDonald

This is a fictional situation report from the fictional nation of Patria that aims to illustrate the issues associated with managing the data, statistics and records used to measure the Sustainable Development Goals (SDGs) in lower- and middle-income countries. This approach should make it possible to present key issues without identifying individual countries and it should allow them to assess the extent to which the fictional situation matches their own realities. A close match will suggest greater relevance. It is hoped that this will increase awareness and understanding of issues and their consequences, leading to concrete action.

Background

Patria is one of 193 countries that signed on to the SDGs initiative led by the United Nations. Within the government, the Ministry of Government Affairs (MGA) is responsible for coordinating its commitments under the initiative and for submitting the statistics to the UN Statistics Division. Generated by the Patrian National Bureau of Statistics (NBS), they are based on its own in-house surveys, as well as on surveys and other data sources supported by individual ministries and are used to measure the SDG indicators developed by the UN.

The government's efforts to implement the Millennium Development Goals (MDGs) have revealed significant weaknesses in the data and statistics used to measure their achievement. In a number of cases, it has been found that the statistics produced to measure the indicators were flawed, often because the data used to generate them were also flawed. In some cases, it was possible to identify where and why this was so, but in too many others it was impossible because there was a lack of supporting documentation. The records that should have documented the processes tended to be fragmented or missing altogether. It was embarrassing to the government when it was discovered that certain goals had not been achieved or when it was clear that the government's statistics

could not be trusted. This caused several development partner organisations and private sector investors to question whether to trust the statistics when deciding on the level of donor support to provide.

In order to avoid the same issues emerging in the SDG initiative, the MGA commissioned a situation analysis to assess the quality, completeness and integrity of the data and statistics used to measure SDG indicators. The government recognised that while issues associated with the quality and integrity of data and statistics were at least partially understood, the role of records was poorly defined. This helped shape the scope of the study, the analysis of issues and the development of suggested strategies. The following key issues were identified:

- the quality and integrity of *statistics* are based on the quality and integrity of the data input to the statistics
- the quality and integrity of *data* input to the statistics relies on the quality and integrity of the processes for collecting, processing, analysing and maintaining the data, as well as on the processes for producing and reporting the statistics
- the quality and integrity of the *processes* can be demonstrated by complete, authentic and accurate records of sufficient quality and integrity to provide evidence of decisions and actions supporting the processes.

The MGA retained an information management expert to undertake the situation analysis. In addition to improving the quality and integrity of the data and statistics used to measure the SDGs, the government expects that the analysis will also improve the data and statistics that support operational and administrative programmes. The result should be more accurate, complete, authentic, relevant and trustworthy data, statistics and records.

Organisation of the report

This report describes the results of the situation analysis. It defines the methodology for the study and the terms used, and it analyses the quality and integrity of the processes followed in collecting and manipulating data and producing statistics used to measure the SDG indicators, with an emphasis on the quality and integrity of the records that document the processes. The implications for achieving the SDGs are highlighted at both ministry and NBS levels.

The report suggests strategies for addressing the issues that have been identified by establishing a framework of policies, standards, systems and people supported by an effective management structure. The last section describes a series of maturity levels to help the government establish milestones for implementing the strategies, and it recommends immediate first steps.

Methodology

The consultant's activities focused on:

- conducting research to identify current relevant initiatives underway at the international level
- conducting interviews and reviewing documentation to identify and describe the work processes and management frameworks for collecting, analysing and presenting data and statistics needed to measure the SDG indicators¹
- conducting interviews and reviewing documentation to identify and describe the characteristics of the records (correspondence, documents, completed forms, data files, logs, and so on) needed to collect data and produce statistics as well as the supporting management frameworks
- analysing and assessing the level of authenticity, completeness, accuracy and integrity of data, statistics and records documenting the supporting processes
- identifying areas where the authenticity, completeness, accuracy and integrity of the data, statistics and records are at risk and why
- preparing an overview of the implications of risks for the government's ability to deliver on SDG commitments
- proposing strategies to address the issues and reduce or eliminate risks
- defining a roadmap describing the way forward based on a set of progressively more sophisticated maturity levels
- providing a set of immediate next steps that can be taken to strengthen the management of data, statistics and records in Patria.

Definitions

When government officials were interviewed, including those responsible for data, statistics, records management, IT, audit and programme management, it was clear that their understanding of basic concepts such as data, statistics, records and processes differed. For some, the concept of data embraced all recorded information, regardless of physical form, from information recorded on paper or in electronic form, to the highly structured information recorded digitally on computer readable media. For these individuals, records in digital form, including email and reports, contained data that could be manipulated and exploited just as readily as the data recorded in highly structured computer-based data files. Records were just another form of data.

1 Information for this step was derived from the SDG Indicators and Metadata Repository, United Nations, 2017, <https://unstats.un.org/sdgs/metadata/>.

Others made a clear distinction between records and data. According to them, data were highly structured codified information recorded in computer readable form for processing and manipulation by computers. Records, whether in paper or electronic form, were information recorded with the primary purpose of documenting actions and decisions and serving as evidence to meet various accountability requirements. Records, for these individuals, were static, never-changing documents, rather than data that could be manipulated. Some were even more focused in their views, believing that records were the paper files that they use, while data were what are held in databases that IT people use.

Given this range of views, it was decided to use definitions that reflected a balance, respected the scope and objectives of the study and were, as far as possible, based on authoritative sources. The following definitions resulted from considerable discussion among representatives from the various disciplines:

- *data* are, technically, recorded information regardless of physical media; for the purposes of this study they are defined as the set of values of qualitative or quantitative variables (recorded in multiple physical forms) that are generated, manipulated and analysed to support the production of statistics
- *statistics* are the results of manipulating and analysing data. They are a type of data. For the purposes of the study, they are the instruments used to measure the SDG indicators²
- *records* refers to recorded information produced or received in the initiation, conduct and/or completion of activities and that document those activities. They are a special form of recorded information or data. When well-managed, they comprise content, context and structure sufficient to provide evidence of the activities.³ Records are not simply correspondence or other documents generated to oversee management of the activity. They include all forms of recorded information, including data and statistics, that can serve to document the activity. This is why a data file can serve as both input to a set of statistics and, at the same time, as part of a series of records documenting the activity. It is all data,

2 According to the Inter Agency and Expert Group on Sustainable Development Goal Indicators, 'official data' refers to a set of values of qualitative or quantitative variables, which are produced and/or disseminated by an official source such as the National Statistical Office or another governmental agency or department including non-traditional types of data. 'Official statistics' means a part of official data, which is produced in compliance with 'Fundamental Principles of Official Statistics'. See: 'Guidelines and Best Practices on Data Flows and Global Data Reporting for Sustainable Development Goals', 9 November 2017, p. 4, https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-06/20171108_Draft%20Guidelines%20and%20Best%20Practices%20for%20Global%20SDG%20Data%20Reporting.pdf.

3 Derived from definitions provided by the International Council on Archives, <http://www.ica.org/en>.

but the data take on different forms depending on the context and their purpose

- *metadata* refers to data that provide context for data and statistics used to measure the SDG indicators and the supporting processes. They are also an important attribute of the records that document an SDG activity, such as the conduct of a survey, the analysis of data or the production of statistics. Metadata describe the relationships among the records, which provides a documentary trail of the activity, and places them in the context of their creation, management and use. In short, metadata make it possible for the information in data, statistics and records to be understood, verified and used in context
- *process* refers to a collection of related, structured steps or tasks needed to achieve a specific service, product or goal.⁴ For the purposes of this study, it refers to the structured steps or tasks involved in collecting, processing and manipulating data to produce the statistics that are used to measure SDG indicators. These include, for instance, the steps involved in planning and approving the survey, designing and testing the survey methodology, conducting the survey, collecting the data, processing and analysing the data, producing and reporting on the statistics, and performing an evaluation of the entire exercise. Data, statistics and records are generated continuously throughout the process
- *records management* is the management function responsible for efficient and systematic control of the creation, receipt, maintenance, use and disposition of records.⁵ It enables ongoing capture and continued accessibility of high-quality, authentic, reliable, accurate, complete, relevant and timely records. This includes data files which, as part of a given documentary trail, must share these characteristics if they are to be trusted
- *records management framework* refers to the policies, standards and practices, systems and technologies, and governance structures for managing records. Just as policy frameworks govern the management of personnel, finances and security, a records management framework must be based on a government-wide policy. The records management framework is designed such that records in all of their different forms can play multiple roles.

4 Derived from Wikipedia, https://en.wikipedia.org/wiki/Business_process.

5 For a complete definition, see ISO Standard 15489, Records Management, which states that: 'records management is the field of management responsible for the efficient and systematic control of the creation, receipt, maintenance, use and disposition of records, including processes for capturing and maintaining evidence of and information about business activities and transactions in the form of records'.

The roles of records include:

- providing evidence of how a ministry or a person conducted business including decisions, actions, non-decisions and inactions.⁶ *For instance, records could be used to prove that a statistical survey was deliberately manipulated to show a favourable outcome*
- enabling organisations to hold themselves accountable in relation to laws and practices. *For instance, records documenting the process for extracting statistical data from a government database could be requested under an access to information law; the same records could also support an audit of a government programme responsible for extracting the data*
- supporting individual rights and freedoms. *For instance, records documenting processes for producing land settlement statistics could be used to locate original survey forms completed by individuals seeking to substantiate their claims*
- being the source of qualitative and quantitative data that can be used for multiple purposes beyond those that led to the records' creation. *For instance, the master data file produced as a result of the annual health survey could be merged with census data and data from other sources to perform analyses not possible using the master data file alone.*

Analysis

This section describes issues that present high risks for the quality of data, statistics and records, based on an analysis of:

- *processes* that support collecting and analysing data used to produce statistics for measuring the SDG indicators and for disseminating the statistics themselves
- *data* and *statistics* generated by the processes
- *records* documenting the processes, the data and the statistics.

The government of Patria and the SDGs

Ministries are required to submit final versions of data files and statistics to the NBS as the basis for measuring specific SDG indicators. The Bureau, which serves as a coordinating hub, incorporates the data and statistics in reports that it submits to the MGA. It also collects, verifies, analyses and produces its own

6 Derived from State of Queensland – Department of Public Works, Glossary of Archival and Recordkeeping Terms (Queensland: Queensland State Archives, 2010), as described in International Council on Archives Multilingual Terminology.

data and statistics. In some cases, it merges data from a ministry with data from other ministries as well as its own data to produce statistics covering multiple SDG indicators. It may also undertake additional data processing to ensure that statistics reported to the MGA are presented in a consistent format. In turn, the ministry reviews the reports, confirms their acceptance and produces a summary report that it submits to the UN Statistics Division according to a predefined schedule.

Data collection and analysis at the ministry level

Twelve of 23 government ministries are responsible for collecting data used to measure the SDG indicators. A ministry may be responsible for one or several SDGs, or it may only support one of several indicators associated with a given SDG. The ministries may use one or more of the following methods to collect data and produce statistics.

Survey data

Survey data are collected through longitudinal (repeated observations of the same variables over short or long periods of time) or one-time surveys using questionnaires or interviews. Some are large, such as the census; others are small, such as household surveys. In some cases, ministries have designed entirely new surveys to meet the requirements of a specific SDG.

Surveys are generally designed and administered by research divisions in ministries. Typically, data are collected on survey forms by mail or distributed by contract survey staff to a sample of a target population. In a few cases, data are submitted online. The data then are transferred to coding sheets and input digitally to a raw data file. Several process files may be produced as the data analysis moves through various stages. Where the data contain personal information, anonymised versions may be created.

The resulting statistics are formatted into tables and embedded in various reports for distribution to a wide range of audiences in paper form or, in a few cases, in digital form via a ministry website. In some cases, especially for longitudinal survey results, the data file might be input to a database of data files from previous surveys. A customised report, together with a copy of the master data file produced as a result of the analysis, is forwarded to the NBS, which produces a standardised report containing the statistics and submits it to the MGA to report on progress towards meeting specific SDGs.

Records documenting decisions and actions relating to planning, designing and conducting surveys may be in various forms. These include emails, paper-based correspondence and reports about a given survey; process and master files created as a result of the survey; survey documentation, such as completed

coding sheets, survey design reports and descriptions of the methodology; and records describing the business context for planning, designing and carrying out the survey. Together, these records, when well-managed, provide evidence that can substantiate the integrity and trustworthiness of the data and statistics used to measure relevant SDG indicators.

Examples of SDG indicators supported by survey data are:

1.4.1

Proportion of population living in households with access to basic services

5.1.1

Legal frameworks are in place, or not in place, to promote, enforce and monitor equality and non-discrimination on the basis of sex

5.2.2

Proportion of women and girls aged 15 years and older subjected to sexual violence by persons other than an intimate partner in the previous 12 months, by age and place of occurrence

6.1.1

Proportion of population using safely managed drinking water services

10.1.1

Growth rates of household expenditure or income per capita among the bottom 40 per cent of the population and the total population

11.1.1

Proportion of urban population living in slums, informal settlements or inadequate housing

16.6.2

Proportion of population satisfied with their last experience of public services

16.7.2

Proportion of population who believe decision-making is inclusive and responsive, by sex, age, disability and population group

Registration and administrative data

This type of data results from administrative activities, such as personnel and finance or operational registration activities, such as licensing. Personnel and finance data tend to be generated in relation to ministry-wide standards and procedures or to workflows associated with hiring and retaining staff, processing expenditure and preparing budgets. Work processes established for registration activities, such as licensing, vary depending on the process, and most are well defined. For instance, in the case of a typical licensing process, licensing applications are received by the responsible ministry and reviewed for completeness and suitability; applicants are notified of whether or not they are

accepted, the information is processed and included in a database, and licences are issued to applicants.

Records documenting these processes may be in multiple forms. For instance, records documenting a licensing process might include emails, paper-based correspondence and reports about a given application for a licence; completed application forms; copies of notifications; completed data conversion forms or logs documenting entry of the data into a database; review and analysis documentation (to verify qualification for a licence); copies of approval and notification documents; and reports documenting review and renewal actions. Together these records provide evidence that verifies the integrity and trustworthiness of the data and statistics used to measure relevant SDG indicators.

Examples of SDGs supported by registration and/or administrative data are:

2.3.1

Volume of production per labour unit by classes of farming/pastoral/forestry enterprise size

3.6.1

Death rate due to road traffic injuries

5.5.2

Proportion of women in managerial positions

8.1.1

Annual growth rate of real GDP per capita

9.1.2

Passenger and freight volumes, by mode of transport

12.4.2

Hazardous waste generated per capita and proportion of hazardous waste treated, by type of treatment

16.1.1

Number of victims of intentional homicide per 100,000 population, by sex and age

17.1.1

Total government revenue as a proportion of GDP, by source

Scientific data

These include instrument readings measuring natural or physical phenomena, such as weather (for example, temperature, rainfall), geology (for example soil composition, erosion) and hydrology (for example, water levels, pollutants). Data generated from instruments are stored within the instrument or transmitted to receivers that store the data separately. In the case of weather data, for instance, readings are made on a regular basis from instruments located around the country. These are transmitted to a satellite, which transmits the data to ground stations

supported by computers that automatically convert the readings and merge them in a master database holding not only the readings but also the processed data that underpins weather reports. There is very little human intervention.

To take another example, water acidity measurements are taken by staff and volunteers for the natural resources ministry using instruments that take water quality readings, including acidity levels, on an annual basis in selected areas of the country. These are recorded on coding sheets, submitted to the ministry, converted to digital form, analysed, and used to produce a wide range of statistics including average marine acidity statistics used to measure SDG indicator 14.3.1. Some of the statistics also are combined with land use data to measure the impact of agricultural land use on the levels of water pollution.

Again, records documenting these processes may be in multiple forms. From the water quality example, these might include emails, paper-based correspondence and reports about a given water quality activity; completed water quality measurement logs; completed data input forms; data verification logs; extract files (data files created from the master database); and report files describing statistics resulting from analysis of the data. Documentation on planning, designing and operating the water quality measurement process, the database, the programme administering the process and the database all form part of the documentary trail.

Examples of SDGs supported by scientific data are:

2.4.1

Proportion of agricultural area under productive and sustainable agriculture

6.4.2

Level of water stress: freshwater withdrawal as proportion of available freshwater resources

14.3.1

Average marine acidity (pH) measured at agreed suite of representative sampling stations

15.1.1

Forest area as a proportion of total land area

Data and records issues at the ministry level⁷

The findings that follow are based on interviews with selected staff and on-site observations in the 12 ministries responsible for measuring the SDGs. Quotes from some of those interviewed for the study are included to illustrate the practical issues involved in measuring the SDG indicators reliably using official data and statistics.

7 Information for this section was inspired by M. Jerven's *Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It* (Ithaca, NY: Cornell University Press, 2013).

In general, ministry staff tend not to recognise the need to build a documentary trail to support the processes of collecting and processing data and producing statistics. Often, they are unaware of:

- the kinds of records that need to be in place
- how the records can be related to one another
- where and how they should be organised and stored
- how ongoing accessibility should be managed.

In most of the ministries responsible for conducting longitudinal or one-time surveys, data files are not well described; documentation on data structures, coding and formats are fragmented; and data verification and quality control procedures are weak and often non-existent. Little care has been taken to ensure that a documentary trail is in place to provide evidence of how surveys are designed and conducted, how data are collected and processed and how statistics are produced.

Manager in a research division:

The ministry wants us to document our surveys, but I don't know what this means. The minister was worried about a sensitive data file that had errors and we couldn't explain where the errors came from. It's not because of us in the Research Division. We've tried to follow some data management standards and survey guides we found online. Now the ministry says we need to document things like why the surveys were done and how they were managed. That information is mostly with other people in emails and memos that I don't see. Action officers in other divisions have that information on their desktops.

Large operational databases in the participating ministries tend to be well-managed, but data extracted from the databases to measure SDG indicators are often poorly documented. Records documenting data extraction tend to be fragmented or non-existent, and procedures for managing the data after they are extracted and used are generally poorly defined. In some cases, the lack of metadata makes it hard to understand the relationship between the extracted data and the source data in the database. Without records documenting changes made to the structure of the extracted data or to the definition of key fields, it is often difficult to know to what extent statistics are inaccurate or misleading.

Manager, IT division:

We had a senior managers' meeting, and someone said we should store all our data with an outside service bureau that has better storage conditions than we do. Other managers agreed and said storing in the 'cloud' was the answer to storing the government's data. But it's too risky. We don't know if it's secure. I think we must keep our data inhouse. Anyway, we would still have data quality issues. Storing outside is not the answer.

In a number of ministries, staff managing large databases are being asked to generate statistics to support measuring the SDGs. This is a new task, and many staff do not have the expertise needed to document the processes generating the statistics or to ensure quality control.

Database manager:

Senior management asked me to get statistics from our immigration database. I was told to send them to the NBS because they needed them for the SDG indicators. I can write a program to extract the data in a report, but I was told I needed to produce the statistics according to industry standards. I don't know what that means. What are the right industry standards? I've talked to other IT managers, but no one seems to know anything about industry standards. We need training.

Given the need for rigorous standards for collecting and analysing scientific data, the quality of the documentary trail is somewhat better than for administrative and survey data. However, data reliability is undermined by the failure to keep records of changes in the instruments used to make scientific measurements, by changes in sampling methods or by failure to update metadata schema.

Staff member, environmental monitoring division, environment ministry:

We monitor marine acidity at stations along the coast and take manual samples. But we don't have trained staff to take samples, and equipment has been stolen from some sites. At others it has broken down. How are we supposed to generate good statistics? The ministry still wants us to use the data we have for the SDG indicator 14.3.1 – that's the one about marine acidity. I've told the minister that our data are not good enough to do the analysis, but he wants us to try anyway.

Several of the surveys used to measure the SDG indicators lack sufficient documentation about the metadata schema supporting the surveys. There are inadequate definitions of key terms, which has led to confusion when interpreting some of the statistics generated from the surveys.

Official in the social development ministry:

The labour ministry uses a different definition of 'employment status' from us. I think we should include more people, like part-time street traders and part-time farmers, even children. How can we report statistics for employment if we are using different definitions? I've searched in our files to find out why we use our definition, but I can't find any records. Maybe there are no records. I've asked the labour ministry where their definition comes from, and sent reminders, but I haven't had a reply.

Problems in finding, retrieving and understanding data held in older data files for trend analysis purposes are hindering the government's ability to regularly measure SDG indicators.

Official, agriculture ministry:

The chairman of one of our farmers' association asked us for data on crop production. He wanted the information for his members so that they could look at trends. He also asked the NBS, but we can't find any data files earlier than four years ago. We don't have any record of where the data files are stored, and all the staff that were involved have left.

Metadata describing the context for many data and statistical files tend to be incomplete. This makes the analysis of trends very difficult and it also makes it difficult to respond to access to information requests or court challenges.

Government lawyer:

NOPA, that's the National Oil Producers' Association, says the government has sent the UN incorrect statistics. This is in connection with SDG 7 on energy. They've asked the government to provide the documentation on how the statistics were produced, but we can't find the records. We've asked the records office and the action officers involved but no one can find anything. To be honest, I'm not even sure the methodology was properly documented.

Many organisational units across government are involved in developing statistics that support the SDG indicators. Multiple organisations may be involved in developing any given SDG indicator, from the initial planning for a survey or the extraction of data from a database, to the final submission of the statistics to the UN Statistics Division. In many ministries, it is practically impossible to bring together the complete story of measuring an indicator because each unit takes its own approach to capturing and classifying the records documenting its activities.

Official, labour ministry:

We've had a big problem with statistics for the SDGs initiative. We receive data from two other ministries and merge them with our own data to produce the statistics. Now there is an expert looking at how we produce the statistics. We gave him copies of the records we send to the NBS. He says the records are not good enough to document the processes in the other two ministries. The quality of the data can't be trusted. He's right. We can't match their data to the records we keep, so the statistics can't be trusted.

When one organisational unit passes data to another unit, if the units take different approaches to capturing and managing records documenting the processes they follow, it can be challenging if not impossible to bring together the complete story of how the indicators are measured.

Assistant secretary, social development ministry:

Our Research and Statistics Division (R&S) has complained that its statistics have been altered. R&S sent the statistics to our Communications Division for submission to the NBS and somehow the statistics were changed. The NBS sent the statistics on to the MGA. The ministry was

supposed to send the statistics to the UN for SDG indicator 10.1.1, but that hasn't happened because we don't know why or how the statistics were altered. I'm trying to sort things out and get to the bottom of this, but no one seems to be able to find any record of why the statistics were changed.

Most ministries have assigned accountability internally for producing statistics to support the SDGs. However, often no one is accountable for classifying records that should document processes for collecting and analysing data and producing statistics or for ensuring that they are complete and accessible through time. Changes in methodology (for instance, in the sample size) and in definitions of key concepts (for instance the target object being measured) tend not to be well-documented. In a few ministries with long involvement in generating statistics there is documentation on survey methodologies (such as coding schemes and analytical techniques) and on conducting surveys (data verification checks, evaluations and audits) somewhere in the ministry, such as in the library. Even in these cases, however, there is seldom a link to the records, such as emails and correspondence, that document the conduct of the survey itself. As a result, the quality and completeness of the documentary trail for individual surveys and for the survey programme varies considerably.

Records management programmes do not exist in most ministries. The one exception is the Ministry of Health, which has a small records management unit with responsibility for managing all of the ministry's records and ensuring that they are accessible through time. Unfortunately, the unit does not yet support the Health Statistics Division, which is responsible for generating statistics measuring several SDG indicators. The staff are on their own in managing records documenting their surveys.

Records manager, health ministry:

I only have three staff and none of us have professional qualifications. We have some training, but it only covers paper records. We keep asking for professional training or training in electronic records management. We see other people going for training, but not us. The ministry think we are here to manage the paper files and that's our job, but I can't deal with the records of the Health Statistics Division without more training. I don't know anything about data files.

The government of Patria does not have a digital preservation strategy. Most IT staff believe that digital preservation means storing data securely but don't recognise the importance of managing the metadata that will make it possible to access and understand the data through time. Nor do they realise that there is a need to convert data to new formats that new software can read or to generate and maintain complete and accurate records documenting these changes. Many look to the NBS for direction and guidance, and some have suggested that it should become a centre of expertise or even a storage centre

for data files with long-term value. However, the Bureau lacks the necessary resources and expertise, as does the National Archives.

Head of IT in a large ministry:

I think our archiving strategy is sound; we back everything up on tape.

Senior officials in several ministries suggested that some or all activities involved in producing statistics for measuring the SDG indicators should be outsourced. They feel that if the resources and expertise aren't available in-house, outside service bureaus and contractors should fill the gap. Others argue that the government needs to control its own production processes, assess the quality of its own data and be able to prove that the statistics it provides to the UN Statistics Division can be trusted. In their view, most companies don't have the necessary expertise in any case.

Interviews in several ministries participating in the SDG indicators process revealed that sometimes the numbers are changed as the result of political pressure before statistics are provided to the NBS. This doesn't seem to happen often, but when it does, it usually isn't recorded. The combination of poor recordkeeping practices and corrupt actions on the part of government officials has undermined significantly the quality and trustworthiness of the statistics used to measure the SDGs.

Data and records issues at the NBS

The NBS maintains a large database that describes the demographic characteristics of the population, including sex, geographic location, education, employment status and income level. Many of the data are collected through the surveys managed by the NBS, with some provided by ministries based on their own surveys. In some cases, the NBS amalgamates data provided by several ministries to generate statistics on cross-cutting topics. In these cases, data submitted by the ministries are converted to formats and structures that can be matched with specific sets of demographic data from the demographic database and matched with other survey data files.

Regardless of their source, data held in the NBS are used to produce statistics that are then incorporated into report files and submitted to the MGA before being transmitted to the UN Statistics Division. The reports are in both digital and hard-copy form. Hard-copy reports are held in filing cabinets managed by the administrative assistant in the office of the director responsible for the demographic database. Digital versions of the report, together with any master data files, are held in the data library 'forever' and managed by the head of the IT area. Copies of the data from the ministries are also maintained in the library but disposed of after five years on the assumption that if the files are needed, they can be accessed through the respective ministries.

Statisticians and IT staff in the NBS understand the importance of providing quality data to support the SDGs. However, the lack of resources and of a records management infrastructure make it difficult to document processes for collecting and processing data and producing statistics. This, in turn, makes it hard to ensure that data and statistics are of high enough quality and integrity to be used effectively.

Records documenting the design of the demographic database and the management of the data (data collection, processing, analysis and reporting) are poorly maintained, and there are no documentation standards.

Staff member, socio-economic statistics division, NBS:

Three ministries send us data for indicator 8.1.1. We convert the metadata to a standard format before we merge it with the census data. If we didn't convert the metadata to a common standard, it wouldn't match up. The problem is there are so many differences in the data, like spellings and names, that the statistics we produce are not very reliable. Also, the ministries are always changing their staff and how they do things.

The NBS assumes that ministries are submitting data files and statistics of appropriate quality and integrity.

Staff member, socio-economic statistics division, NBS:

We sent some incorrect data to the MGA for indicator 2.3.1, but it was not our fault. The ministry said we must check the data before sending them, but it's the ministries' responsibility to check their own data. It's not our job. Even if it was, we don't have the documentation to verify the data. I am not sure if even the ministries have the documentation.

Sometimes the documentary trail is broken when statistical files are transferred from ministries to the NBS. Each participating ministry uses its own classification standards, which makes it difficult to get a complete picture of any given survey/data collection and analysis activity. The lack of evidence of the quality and integrity of statistics increases the risk that they could be flawed.

Senior official at MGA:

Some months ago, the environment ministry changed its definition of hazardous waste to make it wider. This meant that they had to make changes to their surveys and databases, and the way they produce statistics. Now the environment ministry has found out that the NBS has not been using the new definition in its reports for the UN. We have not been able to find any records about why the NBS is not using the new definition and none of the staff can explain it. We don't know how this affects the data used to measure SDG indicator 12.4.2 and the related indicators.

Although the professional staff responsible for the demographic database are concerned about preserving the data, they do not feel equipped to tackle this complex issue.

Manager of the household surveys division:

I've read all of the literature and I think I know how I would go about developing a digital preservation strategy, but I don't have the resources. I have too many other problems to address, and, in any event, the data and their supporting documentation are in a mess.

There tend not to be formal retention and disposition schedules.

Staff member, the socio-economic statistics division:

I am worried about our policy for deleting data. The last director general made up rules for how long we keep data in our division. We are supposed to keep anonymised master files and summarised versions forever. But input and process files must be deleted one year after we create the master files. I don't know why he came up with this idea. I've raised it at management meetings and asked if we can look at it again. If we don't keep the raw data, how can we demonstrate how we measured the SDG indicators, especially over time? It worries me.

Implications of the failure to establish a management framework

The implications of these issues for the government's inability to achieve the SDGs are:

- poorly managed records make it hard to verify the quality and integrity of data generated to measure SDG indicators; this will undermine the government's efforts to report on progress to the UN and jeopardise its ability to make good use of the findings
- data can be flawed, but without a reliable documentary trail to reveal the flaw, it can go unrecognised. Without records as evidence, the government will find it difficult to demonstrate the data's integrity or to trace where a flaw occurred
- flawed data from one source could skew the statistics provided to the MGA, even when the quality of the data from all other sources can be proven by the existence of properly managed records. This could lead to flawed statistics being inadvertently provided to the UN Statistics Division by the MGA
- the government could waste resources taking action to implement SDG findings based on data that lacks integrity
- the quality of data collected through time may be eroded as more and more flawed data join the database. This could have significant consequences for the quality of the data and statistics used to measure SDG indicators in the future

- the loss of credibility due to flawed data could bring the quality of other data into question, which could be problematic without records to prove the quality of the processes followed
- in addition to the implications for measuring the SDG indicators and implementing the SDGs themselves, the impact of poor recordkeeping is likely to affect the government's ability to carry out its mandated responsibilities
- individual rights can be compromised if individuals who provided data as part of a data collection activity (such as a survey) cannot be accessed, or if data or records documenting decisions about the collection and use of the data cannot be found
- national economic interests could be threatened if government policy and direction are based on flawed data and statistics or if the level of quality and integrity cannot be confirmed.

Strategies for sustainable solutions

This section focuses on strategies for developing a comprehensive and sustainable framework for managing data, statistics and records. The issues identified in the previous sections reflect weaknesses in the overall framework for managing data, statistics and records. Just as there are frameworks for managing human and financial resources, this framework should provide an integrated combination of laws and policies, standards and practices, systems and technologies, and people, supported by management and governance structures. A focus on symptoms, without considering the broader causes, will result in fragmented and ineffective strategies and offer only short-term temporary solutions. Inevitably, there will be issues needing urgent and immediate attention, but the focus should be on establishing a comprehensive and sustainable framework for managing the completeness, authenticity and trustworthiness of data, statistics and records.

The section is organised according to the components of the framework. The key issues and relevant strategies are described below for each component.

Laws and policies

Issues

- there is no law requiring the government to set up a records management programme. The access to information law provides the right of access to a wide range of government records, but it does not require the government to ensure that its records are authentic, accurate, complete and accessible. The Privacy Act requires that personal information be protected and retention standards applied, but there is no public pressure for this to be enforced

- apart from the Ministry of Health and the MGA, none of the ministries participating in measuring the SDGs, including the NBS, has a records management policy. At the health ministry, the policy focuses on managing paper records and does not yet address records in digital form. The policy for the MGA is limited to managing the paper records of the secretary and the executive committee
- there are some policies in place for managing data and statistics and conducting surveys, but they do not address the role of records in providing evidence to document survey and other data collection and processing activities.

Strategies

It is important that existing laws, such as a national archives act, data protection legislation, statistics act or other relevant legislation, should support the effective management of information needed to measure the SDGs and enable the government to achieve its operational and strategic goals and meet a wide range of accountability requirements, notably:

- ensure that the freedom of information or right to information law enables citizens to have the right of access to the data, statistics and records generated to support measuring the SDGs
- ensure that the Privacy Act gives citizens the right of access to their personal information as recorded in the data, statistics and records generated to support measuring the SDGs
- develop a government-wide policy on managing records as evidence that embraces data and statistics as high-quality sources of information for decision-making and for verifying the integrity of the processes involved
- strengthen policies for managing data and statistics to ensure that responsibility and accountability for documenting relevant processes are clearly defined and that there are provisions for managing data and statistics as part of the documentary trail of surveys and other data collection and analysis activities
- develop policies and guidance to protect personal information in relation to the data, statistics and records generated for measuring the SDGs
- ensure that in all contracts with private sector firms conducting surveys on behalf of the government, the contractor is obliged to document its activities, protect the data and statistics it generates, respect the government's ownership of the data and statistics, and transfer all data, statistics and supporting records to the government when the contract is completed.

Standards and practices

Issues

- there is no guidance on how to document processes for collecting and processing data and producing statistics. Ministries establish their own practices for creating and managing documentary trails, which often are not complete, accurate and authentic. Variations in how these processes are designed and managed makes it difficult to establish standard approaches to documenting them
- the NBS follows standards for managing survey documentation, such as code books and survey methodology documentation, but generally these standards and practices are not in place in the ministries. Even when survey documentation standards are applied, there is no way to link the documentation to the records, which are often in the form of emails and attachments that document decisions and actions about the management of the survey itself. Establishing a complete and comprehensive documentary record of the survey is impossible
- procedures are not in place for converting and sharing data across ministry boundaries. Achieving interoperability when there are multiple recording formats and diverse technologies is virtually impossible. For instance, the NBS must convert data and statistics it receives from ministries in order to provide statistics to the MGA in a standard format. There has been little effort to document these conversion activities, which means that flaws in the data that emerge at this stage are difficult to trace
- retention standards for data and statistics are rarely in place, and even when they have been assigned, they are not consistent across the data, statistics and records associated with a given process. Final statistical data files may be kept 'forever', but records documenting the circumstances of their creation may be destroyed much earlier. Digital preservation presents a huge challenge for any organisation, but it is possible to take preliminary steps, such as researching possible strategies, and assessing needs. At present, there is little evidence that this is happening
- especially when several ministries are involved and records documenting a given process are in multiple forms, the difficulties of bringing together the complete story, make it nearly impossible to establish a digital preservation plan covering all the records associated with the process.

Strategies

- develop criteria for identifying records that should be in place to document processes for collecting and processing data and producing statistics

- develop procedures to ensure that records documenting data and statistics activities are captured, managed and integrated with procedures for conducting surveys, analysing data, merging data and reporting statistics
- develop metadata standards and guidance for managing individual processes, linking records, data and statistics, and accessing data, statistics and records within and across processes and different media
- establish retention and disposition standards and guidance for all forms of records that document collecting and processing data and producing statistics
- monitor and draw from international work on digital preservation strategies and implementation plans for the long-term accessibility and integrity of data, statistics and records.

Systems and technologies

Issues

- technologies for managing data and statistics are usually specific to the unit responsible and the kinds of data being managed. For instance, the technology for managing data in a database may be different from technologies for extracting data from the database and processing them as statistics to support SDGs. Once the statistics are passed to the NBS for further processing, other technologies may be used. Documenting the changes that take place from one technology environment to another is a significant challenge
- custom-designed databases are in place for managing survey documentation, but technologies have yet to be developed to manage the records of decisions and actions taken regarding surveys or other data collection activities. Nor are there systems for tracking how data are collected and processed and how statistics are produced. Records generated by these activities are not being identified, classified and managed.

Strategies

- use generally accepted IT project management standards to plan, design, test, implement and maintain systems for managing the authenticity, integrity and continued accessibility of statistics, data files and records across space and through time
- use internationally approved standards to develop functional requirements for managing statistics, data files and records and incorporate them into the requirements for designing IT systems

- develop audit and evaluation tools for assessing the quality and integrity of data, statistics and records supporting the SDGs; integrate them into standards and practices for systems and into management audits and evaluations.

People

Issues

- the NBS has some staff with professional expertise in managing data and statistics, but they do not have the records management expertise needed to manage records documenting the processes that generate data and statistics. Few people in ministries have this expertise. Records management staff in some ministries, such as the health ministry and the MGA, are generally only responsible for paper records. The NBS recently introduced a training programme for ministry staff responsible for collecting and processing data and producing statistics to support the SDGs. This will help, but at present there are no training materials on managing records in relation to data and statistics
- in some ministries there is a wide gulf between those responsible for technical aspects of the data (such as IT), and those responsible for processes that generate the data, statistics and records (such as programme managers); often each assumes that the other is looking after the requirement. This gap has serious consequences for the integrity and quality of data and statistics.

Strategies

- define the work involved in managing data, statistics and records used to support measuring the SDGs
- define competencies associated with the work
- design and implement appropriate training programmes
- design and implement appropriate recruitment programmes
- enhance tools and techniques for measuring performance so that competencies for managing data, statistics and records can be assessed
- establish programmes for allocating staff with the required expertise between ministries to fill competency gaps
- establish partnerships, including with organisations outside of the government, to pool human and financial resources for developing the framework
- work with relevant university programmes to enhance existing courses or develop new ones to address the management of data, statistics and records.

Management and governance

Issues

- although there are accountability frameworks for managing personnel and finance, they have not been introduced for managing records documenting how data are collected and processed and statistics are produced. Accountability has not been assigned for ensuring that a complete and accurate documentary trail is in place. Audit units in ministries measuring the SDGs don't yet cover this issue in management and systems audits.

Strategies

- establish accountability and assign roles and responsibilities⁸ for staff at all levels to ensure the quality and integrity of data and statistics used to measure the SDGs
- establish an authority at a senior level of government with responsibility for ensuring that records are managed to support high-quality data and statistics across government.

Awareness

Issues

- some senior managers are beginning to recognise the importance of preserving data files and statistics, but few understand the crucial role that records play. Records documenting processes by which data files were created and used and documenting the data files themselves, such as coding schemes and storage formats, must be preserved if the data files are to be accessed in the future
- through time, as the demand for historical data to analyse trends grows, this lack of awareness will have greater implications. Few recognise that the issue needs to be addressed now, rather than in the future, when data files generated early in the SDG initiative may already be inaccessible. The initiative, which asks governments to measure indicators over a 15-year period, is bringing the issue into sharp focus
- few citizens are aware of their rights in relation to data collected about them in connection with the SDGs, and few have challenged the way the data are used. Government ministries have not yet felt the pressure to

8 It is important to note the difference between accountability and responsibility: accountability is always upward *to* someone; responsibility is *for* something (to be done).

ensure the completeness and accuracy of the data, statistics and records for which they are responsible. However, there is growing citizen concern about these issues and growing awareness of the government's inability to manage the personal information it holds, especially in digital form.

Strategies

- ensure that senior managers responsible for programmes and processes supporting measuring the SDGs are aware of key concepts, issues, implications and possible strategies
- develop tools and techniques for enhancing awareness, for instance briefings and brochures for relevant staff at all levels
- incorporate these tools and techniques in training and awareness programmes, including orientation programmes for staff, management seminars and workshops.

The Ministry of Public Administration (which manages the civil service) could be an appropriate agency to take the lead in establishing a framework to address the quality and integrity of the data, statistics and records used to measure the SDGs and, at a more general level, to support the requirements of government programmes for authentic, complete, accurate and relevant data, statistics and records for decision-making and accountability.

Implementing the strategies

Capacity levels to guide the way forward

A roadmap, in the form of capacity levels, will enable the government to move incrementally through defined stages to build the capacity needed to manage data, statistics and records in line with available resources. Five capacity levels are described below, the fifth level being an ideal state for a country that wants to ensure that data, statistics and records used to measure the SDG indicators are of a high enough quality to measure and implement the goals. For most organisations, achieving Level 5 or even Level 4 will be challenging.

The levels reflect diminishing degrees of risk, with Level 1 representing the highest risk of loss and inaccuracy and Level 5 being the least risk. They also reflect increasing levels of sophistication in terms of the way data, statistics and records can be used to support implementation of the SDGs and, more broadly, the government's operational and strategic goals. The roadmap for moving forward will support an objective and systematic approach. Examples are included in the maturity level descriptions, drawn from the targets and indicators supporting SDG 5:

- SDG 5: achieve gender equality and empower all women and girls
- SDG target 5.5: ensure women's full and effective participation and equal opportunities for leadership at all levels of decision-making in political, economic and public life
- SDG indicator 5.5.2: proportion of women in managerial positions.

Level 1: poor-quality data, statistics and records undermine SDG implementation

The organisation produces statistics to measure the SDGs, but they are unreliable. Professionals responsible for data, statistics and records lack the knowledge and skills needed to develop a reliable framework of policies, standards, practices and systems.

Example

Annual labour force data is collected through survey forms sent to companies across the country. Data are collected and analysed, and the resulting statistics serve multiple purposes, including measuring the proportion of women in management positions in support of SDG 5. The lack of metadata standards and the absence of records documenting how data were collected and processed and the statistics produced makes it impossible to relate statistics from year to year. Data files from previous years are poorly organised and documented, so records of decisions, including changes in survey design, data-processing methods and data formats are fragmented and scattered in multiple locations. The implications will not be known for some time, but without a reliable evidence base or the expertise to prove the trustworthiness of the data, the annual statistics cannot be relied upon as an accurate measure of SDG indicator 5.5.2.

Level 2: data, statistics and records enable basic SDG measurement

A framework of laws, policies, standards, procedures and people is in place to ensure that data and statistics are gathered and analysed to measure the SDGs. Managers are generally aware of their responsibility for ensuring that data files and statistics, with their supporting documentation, are stored properly. However, the framework is not applied universally, with some managers providing poorly documented data and statistics. There are no standards for documenting surveys and other data-gathering and analysis activities, nor have policies been developed for managing the records that should document processes for collecting and processing data and producing statistics. Records and data management professionals do not have the expertise needed to manage the interrelationships among data, statistics and records, especially those that need to be preserved through time.

Example

Annual labour force data, including data extracted from the labour force database to produce statistics for measuring SDG indicator 5.5.2, are collected and processed based on approved standards and procedures. However, emails, reports, logs and other records documenting the design and conduct of the survey, including changes in survey methodology, cannot be related to records documenting processes for extracting and analysing data and producing statistics. Records management professionals in the Labour Force Statistics Division responsible for measuring SDG 5.5.2 do not have the expertise needed to ensure that records, data and statistics are managed as a whole. The lack of a digital preservation strategy increases the likelihood that trend data needed to measure SDG 5 from now until 2030 will not survive.

Level 3: the quality of data, statistics and records makes it possible to measure SDGs effectively and supports government programme activities

Data, statistics and records generated to measure SDGs are managed through a comprehensive framework of policies, standards and practices, systems and technologies, and qualified people. Records management staff work effectively with data management and other professional staff to ensure that requirements for identifying, describing, classifying, protecting and retaining data, statistics and records are integrated in the design of processes for collecting data and producing and using statistics. Managers know that they are responsible for ensuring that the data, statistics and records generated are authentic, reliable, accessible and understandable and can be retrieved when needed. Professional staff apply clear, consistent standards and practices. However, preservation is not addressed adequately; retention requirements have not been established, metadata standards for data, statistics and records have not been developed, and preservation standards, procedures and technologies are not in place.

Example

All processes for generating statistics to measure SDG 5 are supported by the same framework of policies, standards and practices, systems and technologies, and people. For instance, data, statistics and records generated to measure SDG indicator 5.2 (the proportion of women in management positions) are well described, organised and managed to provide a comprehensive documentary trail of evidence. The statistics can be trusted because the comprehensive management framework itself can be trusted. Unfortunately, the lack of a digital preservation strategy means that while statistics measuring the participation of women in management positions can be compared for the past two years, the government cannot ensure the integrity of the statistics over the 15-year life of the SDG initiative.

Level 4: well-managed data, statistics and records make it possible to measure SDG implementation effectively and consistently through time; data and statistics are of high enough quality and integrity to support government programme activities at the strategic level

Data, statistics and records generated to measure the SDG indicators can be reliably merged or combined with other data sources to support programme activities, including those supporting the organisation's strategic goals. Organisation-wide policies and standards are in place to protect records of decisions, and accountability requirements, for instance under access to information legislation, are supported by consistently applied records management policies and standards. Trends can be analysed through time, and comparisons can be made from year to year because changes to formats, coding schemes and data collection and analysis methods are well-documented. Preservation standards ensure that data, statistics and records are stored properly and migrated to take account of changes in technology. The preservation programme ensures continued accessibility and authenticity of data, statistics and records through time.

Example

Gender equality is a government strategic priority. Labour force data used to produce statistics for measuring the proportion of women in management positions (SDG 5) is being merged with statistics from the Ministry of Industry on female participation in various industry sectors to support the strategic goal. This is possible because of the way the data from both sources were formatted and described. The resulting database can be used to measure progress toward gender equality, while at the same time contributing to the statistics needed to measure SDG indicator 5.2. The comprehensive framework of policies, standards and assigned accountability ensures the integrity and trustworthiness of the data, statistics and records. A preservation programme dedicated to ensuring the authenticity and completeness of the increasing volumes of data and statistics makes it possible to perform complex analyses through time.

Level 5: processes generating data, statistics and records, and the framework for managing them, are designed to make it possible to exploit data, statistics and records, including those measuring SDGs, in new and innovative ways

Managers of SDG initiatives understand the benefits of sharing and exploiting data, statistics and records for stimulating innovative thinking on implementing the SDGs and achieving the operational goals of individual programme activities and the strategic goals of the organisation. Professional staff have

the knowledge and expertise needed to design comprehensive management frameworks covering multiple organisations and technology environments that encourage information in the data, statistics and records to be exploited to the greatest possible extent.

Example

Employment data from several large private enterprises have been merged with the government's labour force data and employment data to create a government-industry database. The complex interjurisdictional processes are well-documented, data are well-managed, and the statistics produced from the database can be trusted because the management framework can be trusted. Staff have the confidence to look for new and innovative ways to exploit the data, even as its volume and complexity grows. Innovative and advanced technologies are applied, and information is published in new forms to meet the needs of a wide range of individuals and groups and to give citizens access regardless of location. A wide range of statistical products serve multiple purposes, including not only the measurement of SDG 5 but also the management of the government's commitments in support of the Open Government Partnership's agenda on gender equality.

First steps

Rather than trying to work on everything at once, it is suggested that the government should start by identifying and defining solutions for a few processes where weak management of data, statistics and records has significant implications for achieving the SDGs. This experience will then inform the development of the framework.

Identify a leader and assemble a team

Given the MGA's leading role in the SDG initiative, a senior official in the ministry should oversee the initiative. This person should have a background in data management, statistics, information technology or records management, the capacity to bridge these disciplines and the ability to communicate with a variety of stakeholders, including senior management.

A steering committee should be appointed, made up of representatives from government programmes supporting the SDGs as well as programmes where the quality and integrity of data, statistics and records is particularly important. Specialists in managing data, statistics, records and information technology, as well as legal experts and auditors, should also be included. The committee should help select the SDG processes to be covered, to identify issues and strategies, and to explore how to extend the results to other SDGs.

Some government officials have argued that attempting to build a comprehensive framework is creating a 'mountain out of a mole hill' and that the focus should be on addressing immediate issues associated with specific SDG initiatives. Others have realised that systemic issues need to be addressed across government as a whole. This tension between the need to address immediate and critical problems and the goal of developing comprehensive and sustainable solutions needs careful management. One way to address the tension is to focus on specific carefully identified processes in order to gain knowledge and skills that can be extended to other processes or used in developing a comprehensive management framework.

Identify processes as examples

For each of the three process types (survey, registration/administrative and scientific), identify one or two processes that present significant challenges for measuring one or more SDG indicators and for using data, statistics and records for operational and programme delivery. These are likely to be processes where undocumented flaws or inaccuracies in data, statistics and/or records have led to embarrassment, bad decisions about the use of government resources, missed opportunities or increased risk and costs.

Describe the selected processes

The description should cover the stages of generating the data, statistics and records and of managing a given process:

- the stages of a *survey process* are likely to include planning and approving the survey, designing the survey methodology, designing the data collection tools and techniques (such as survey forms), testing the survey methodology, conducting the survey, analysing the results, reporting the findings and reviewing how the survey was conducted
- in the case of a *registration/administration process*, the stages are likely to reflect the stages of the systems development life cycle, including planning the system, defining functional requirements, designing the system and database, testing the design, implementing the system and the database, maintaining the system and database, and evaluating the extent to which the system and database follow the stated requirements
- in the case of a *scientific process* the stages would include planning the project, assessing data collection methods and technologies, designing the process, testing data collection and measurement tools, procedures, analytical techniques and statistical reporting methods, implementing and maintaining the process, and reviewing/evaluating the project.

It should be possible to identify the data, statistics and records created at each stage. The aim is not to describe every single stage and every piece of data

and statistics and every record for a given process but to identify key stages of the process and the associated data, statistics and records that are significant in terms of measuring an SDG and providing a complete and authentic documentary trail of the process.

Finally, the overall framework for managing both the process and the data, statistics and records should be reviewed. Policies and standards are particularly important, as is the governance structure (who is accountable to whom for what). This will provide a template for analysing the quality and integrity of the process itself and the data, statistics and records it generates.

Identify issues and implications

It should then be possible to analyse the issues, distinguish between symptoms and causes and identify solutions. For instance, a poorly documented data file input to a set of statistics that turned out to be flawed is a *symptom*. The *cause* was the failure to establish metadata and documentation standards at the planning and design stages of the process and to assign accountability for implementing them as part of the management framework for the process. In identifying the issues, it is important to distinguish between immediate issues particular to measuring a given SDG indicator and issues related to the broader management framework for the organisation as a whole.

Finally, issues should be explained in a way that programme managers responsible for generating SDG statistics can understand. A key idea that should be reinforced continually is that where data, statistics and records are flawed, where their accuracy cannot be established and where they are lost or destroyed, the credibility of the manager responsible for the data, statistics and records will be undermined irrevocably. By extension, society's trust that the government is capable of carrying out its obligations, including achieving the SDGs, will be eroded significantly.

Develop strategies for resolving issues

Understanding where symptoms and causes are located on the roadmap and how they relate to one another will be helpful in developing integrated strategies within the context of the overall management framework. Most of the strategies should focus on the planning and design stages of the survey, system or other data collection and analysis activity when the steps in the strategy can be integrated more easily and more cost-effectively. For instance, the need to develop and apply enhanced metadata standards and procedures for enabling data, statistics and records to be related to one another should be acknowledged and addressed at the planning stage with subsequent stages incorporating the testing, implementation and assessment of the standards and procedures themselves. This approach to developing and implementing strategies can be applied to any process, from small one-time surveys to large

IT systems supporting continuously updated databases from which data and statistics supporting SDGs are extracted.

Apply the experience to other processes and to the framework for managing data/statistics/records

The approach should result in strategies that can be applied to all processes, not just for those measuring the SDGs but for any process supporting the government's programmes and services where data, statistics and records essential to decision-making and the ability to meet accountability requirements are being placed at risk. In parallel and over the longer term, the results will be invaluable in developing a comprehensive, policy-driven standards-based framework for managing data, statistics and records, regardless of the process or business function.

Ultimately the goal is to build a comprehensive management framework to cover all government programmes and services and to allow the government to demonstrate that the data, statistics and records it generates can be trusted. The outcome should be that the government is able to demonstrate, through the availability of complete, accurate and relevant data, statistics and records, a high level of credibility, both to Patrian people and to international partners, investors, development agencies and other international organisations, including the United Nations.

Index

- Aadhaar initiative, 4
- Addis Ababa Plan of Action for Statistical Development in Africa, 40
- ADPNet, 146
- Africa Freedom of Information Centre, 180
- African Census Programme, 24
- African Conference on Archival Policies and Programmes in Africa (1994), 11
- African Development Bank, 46
- African Union (AU), 43
- AIP, 165, 170
- Airtel, 71
- Akotia, Pino, vii, 15, 177
- Alianza Regional por la Libre Expresión e Información, 176
- Ambira, Cleophas, 4, 97
- Archivematica, 143
- ArchivesDirect, 145
- ARMA International, 196, 197, 202, 204, 208, 209
- artificial intelligence, 53, 65
- Australia, 98, 207, 209
- Society of Archivists and the Records Management Association, 209
- Automated Teller Machine (ATM), 101
- Banjul, 86–8
- BBC, 14
- Bearman, David, 56
- Bell, Alan, 56
- big data, 29, 40, 53
- birth and death statistics, 17, 28, 30, 63, 73, 83, 87–9, 93, 95
- BitCurator, 143
- Botswana, 13, 25, 30, 179, 207
- Brazil, 176
- Bretton Woods, 10
- British Columbia, 180, 181, 187
- Information and Privacy Commissioner, 180
- Brown, Adrian, 5, 17, 56, 135, 141, 147
- Cain, Piers, 52, 54
- Cambodia, 4, 110, 114–19
- Ministry of Agriculture, Forestry and Fisheries (MAFF), 116, 118
- Open Development Cambodia (ODC), 114
- Canada, 14, 98, 173, 180, 181, 209
- census data, 23, 24, 27, 43, 88, 116, 118, 216, 226
- Centre for Law and Democracy, 182
- certification, 17, 73, 167
- Chandigarh, 70
- characterisation tools, 143
- Chung, Pyrou, 4, 109
- coders, 51

- Comma Separated Value (CSV),
130, 131, 135, 156
- Commercial Bank of Africa, 102
- computerisation, 10, 12, 13, 126
- corruption, 13
- Cukier, Kenneth, 53
- cultures of secrecy, 174, 180, 181, 187
- data authenticity, 122
- datafication, 53, 54
- Democratic Republic of Congo
(DRC), 22, 27, 32
- Demographic and Health Surveys
(DHS), 29–39, 43, 48
- Devarajan, Shantayanan, 29, 46
- digital preservation, 5, 43, 127,
135–50, 156, 163, 164, 165, 199,
200, 224, 227, 230, 231, 236
- Digital Preservation Coalition, 136,
148, 150
- digitisation, 15, 29, 54, 107, 126,
136, 182
- Domesday Book*, 60
- Dreze, Jean, 74, 75, 78–80
- DROID, 143
- DSpace, 143
- DuraCloud, 145
- E-ARK, 142
- Economic Commission for Africa,
24, 40, 46, 87
- ECOWAS, 87
- elections, 110–14, 118–20
- EPrints, 143
- Fedora, 143
- Five Star Open Data Scheme, 204
- forensic tools, 143
- freedom of information laws, 121
- Gambia
Bureau of Statistics (GBS), 83
DHIS2 information system, 87
integrated financial management
information system (IFMIS), 93
Ministry of Finance and
Economic Affairs, 93
Ministry of Health and Social
Welfare, 86–8
National Development Plan, 89
National Records Service, 83, 92
- gender equality, 160, 183, 235,
237, 238
- Generally Accepted Recordkeeping
Principles (GARP), 196
- Ghana, 11, 12, 15, 25, 33, 124, 177,
178
- Giaretta, David, 5, 135, 153
- Global Indicator Framework for
the Sustainable Development
Goals, 160
- Global Partnership for Sustainable
Development Data, 2, 195
- Google, 70
- Griffin, Andrew, 4, 53, 83
- gross domestic product, 159
- hashes, 155
- Heritrix, 143
- HIV/AIDS, 41, 44, 47
- IndexMundi, 84, 85
- India, 4, 67–80, 176, 178
Congress Party, 68
cow protection, 81
Editors' Guild, 72
National Rural Employment
Guarantee Act, 4, 67, 68
National Social Assistance
Programme, 74

- New Delhi, 68, 69, 72, 74, 80
 Supreme Court, 73, 74, 78, 81
 Unique Identification project, 74
 United Progressive Alliance, 68
 Information Governance Maturity Model, 196–8, 200, 202, 204
 information technology, 11, 67, 192, 193, 205, 207, 210, 238
 input data, 119
 Institute of Commonwealth Studies, 195
 Inter Agency and Expert Group on SDG Indicators, 1
 International Council on Archives (ICA), 8, 10, 13, 16, 123, 150, 182, 208, 209, 214, 216
 International Monetary Fund (IMF), 10
 International Records Management Trust, 7, 12, 13, 16, 52, 126, 128, 134, 150, 177, 208
 International Research on Permanent Authentic Records in Electronic Systems (InterPARES), 16
 International Standards Organization (ISO), 166
 Inter-University Consortium for Political and Social Research (ICPSR), 145

 Jerven, Morton, 1, 20, 21, 29, 46, 220
 Jharkhand state, 71, 75, 76
 JHOVE, 143
 Jordan, 176, 177, 181

 Kandala, Ngianga-Bakwin, v, xi, 4, 19
 Kanpur, 71

 Karnataka, 76, 77, 79
 Kazakhstan, 175
 Kenya
 Bankers Association, 101, 102
 Bureau of Standards, 107
 Central Bank, 100–4
 Communication Authority, 100, 103
 Evidence Act, 104, 105
 Groundwater Mapping Programme, 125
 Integrated Records Management System (IRMS), 128
 Mau Mau, 8
 Ministry of Lands and Physical Planning, 126, 127, 131, 132
 Moi University, 207
 Nairobi, 100, 102, 104, 124, 126, 127
 National Archives and Documentation Service, 98, 107, 127
 National Bureau of Statistics, 102, 103, 129–31
 Open Data Initiative (KODI), 121–5, 128–33
 Open Government Partnership National Action Plan, 128
 Registered Land Act (1989), 126
 Registration Act (2012), 126
 Registration of Titles Act (1982), 126
 Keorapetse, Dithapelo Lefoko, 13
 Khmer language, 117
 Kibaki, Mwai, 123
 Komba, Paul, 4, 19

 labour force data, 235, 236, 238
 Lamin, Tamba, 4, 109, 112
 Lao PDR, 115

- Lemieux, Victoria, [xii](#), [5](#), [16](#), [173](#),
[175–8](#)
- leprosy, [77](#)
- LOCKSS, [143](#), [146](#)
- Lowry, James, [4](#), [17](#), [49](#),
[121–3](#), [126](#)
- Mahihu, Christine, [128](#)
- Malawi, [14](#), [25](#), [34](#), [178](#), [180](#)
- Management System Standards
(MSS), [201](#)
- Manor, James, [4](#), [67–9](#), [75](#), [77](#),
[80](#), [81](#)
- Mayer-Schönberger, Viktor, [53](#)
- Mazikana, Peter, [11](#)
- McDonald, John, [5](#), [14](#), [135](#), [138](#),
[195](#), [196](#), [205](#), [206](#), [210](#), [211](#)
- McLeod, Julie, [135](#), [138](#), [189](#), [191](#)
- Mekong Region, [114–17](#)
- metadata, [122](#)
- Mexico, [176](#)
- Millennium Development Goals
(MDGs), [24](#), [211](#)
- MIT Digital Preservation
Management Tutorial, [150](#)
- Mo Ibrahim Foundation, [27](#)
- mobile banking, [4](#), [100–5](#)
- Modi, Narendra, [69](#), [71](#), [72](#), [76](#), [77](#),
[79](#), [80](#), [81](#)
- Monash, [207](#)
- mortality data, [4](#), [83–9](#), [93–5](#)
- Mozambique, [26](#), [36](#), [40](#)
- M-Shwari, [102](#)
- Multiple Indicators Clusters Surveys
(MICS), [29](#)
- Mutuku, Leonida, [128](#)
- Myanmar, [77](#), [115](#)
- Namibia, [25](#), [36](#), [40](#)
- National Digital Stewardship
Alliance, [142](#), [150](#)
- National Statistics Institutes, [44](#)
- NHS Digital, [209](#)
- Nigeria, [11](#), [25](#), [28](#), [36](#), [42](#), [45](#),
[113](#), [124](#)
- Nilekani, Nandan, [75](#), [79](#)
- Northumbria University, [207](#)
- OAIS Reference Model, [136](#), [156](#),
[165](#)
- Odisha, [75](#), [76](#)
- Open Archival Information System
(OAIS), [17](#)
- open data initiatives, [109](#), [110](#), [111](#),
[118](#), [120](#), [123](#), [174](#), [175](#), [178](#)
- Open Government Partnership
Summit (2016), [188](#)
- open source software, [143](#)
- Organisation of African Unity
(OAU), [43](#)
- Preservica, [145](#)
- Project Management Body of
Knowledge (PMBOK), [165](#)
- Public Record Office, [61](#) (*see also* UK
National Archives)
- Rajasthan, [75](#)
- Ramanathan, Usha, [74](#)
- random error, [159](#), [160](#)
- recordness, [56](#)
- RODA, [143](#)
- Rosetta, [144](#)
- Safaricom mobile network, [100](#),
[102](#), [103](#)
- scientific data, [5](#), [153](#), [156](#), [159](#),
[220](#), [222](#)
- Searle, John, [50](#)
- Serbia, [175](#)
- Shepherd, Elizabeth, [5](#), [135](#), [138](#),
[189](#), [192](#)

- Sierra Leone, 4, 13, 14, 37, 110–12, 114, 117–19, 175, 178
- Democracy and Development Associates (DADA-SL), 111
- Federation of Civil Society and Media Organisations (NaFCSMO-SL), 111
- National Elections Watch, 110
- National Electoral Commission (NEC), 110
- Open Data Council, 110
- Open Elections Data Portal (SLOEDP), 111–14
- Open Government Initiative (OGI), 111
- Right to Access Information Commission (RAIC), 110
- TpISENT (SL) Limited, 111
- Women's Situation Room, 111
- smart phones, 101
- South Africa, 25–8, 37, 40, 43, 44, 175, 176, 178, 179, 181
- Promotion of Access to Information Act (2000), 175
- Strengths, Weaknesses, Opportunities and Threats (SWOT) model, 21, 40–2
- structural adjustment programmes, 29
- Sustainable Development Goals (SDGs), 1–5, 7, 17, 18–21, 23, 24, 26, 27, 42, 48–51, 55, 59, 64, 66, 83, 89, 90, 95, 97, 99–102, 106, 108, 109, 121, 122, 133, 135, 139, 141, 146, 147, 150, 153–7, 159, 160, 163, 167, 170–2, 176, 182, 183, 188, 189, 196–8, 200, 204–6, 210–12, 216, 217, 219, 220, 222–9, 231–8, 240–1
- Sustainable Development Solutions Network, 2
- Thailand, 115, 176, 177, 181
- Thomson Reuters Foundation, 125
- Thurston, Anne, 4, 7, 19, 21, 52, 54, 126, 173, 178
- Townsend, Katherine, 4, 109
- United Kingdom, 176
- Cabinet Office, 175
- Data Archive (UKDA), 145
- High Court, 8
- National Archives, ix (*see also* Public Record Office)
- Open Data programme, 175
- United Nations, 1, 7, 20, 24, 40, 85, 86, 87, 97, 154, 171, 172, 189, 195, 208, 211, 213, 241
- Population Fund, 24
- Statistical Commission, 1
- UNESCO, 8
- United States, 88, 98, 176, 186
- Universiti Teknologi Malaysia Faculty of Technology and Informatics, 207
- University of Botswana, 13
- unstructured supplementary service data (USSD), 101
- Uttar Pradesh, 73, 78
- Vietnam, 115
- virtual data, 162
- Vision 2030, 27, 127
- Wamukoya, Justus, 4, 97, 126, 127, 179
- web archiving tools, 143

- West African Health Organisation (WAHO), 87
- WhatsApp, 110–14
- Wikipedia, 85, 215
- World Bank, 5, 10, 13, 16, 20, 30, 44, 46, 79, 83–7, 173, 175–8, 187
- Open Data Readiness Assessments, 175
- Right to Information Drivers of Effectiveness (RIDE), 177
- World Health Organization, 85, 93
- World Wide Web Foundation Open Data Barometer (ODB), 123
- Yeo, Geoffrey, 4, 49, 51, 55, 59, 66
- Zambia, 10, 25, 39
- Zimbabwe, 11, 39

SUSTAINABLE DEVELOPMENT GOALS

The United Nations Sustainable Development Goals initiative has the potential to set the direction for a future world that works for everyone. The 17 goals build on the successes of the Millennium Development Goals, but also include new priority areas, such as climate change, economic inequality, innovation, sustainable consumption, peace and justice.

Each goal presents considerable challenges in terms of collecting and analysing relevant data and producing the statistics needed to measure progress. Ultimately, the quality of the data and statistics depends on the availability, completeness and integrity of the records that document them. Most governments in lower resourced countries do not yet have the systems and controls in place to produce high quality, reliable data, statistics and records, and it is questionable whether the quality and integrity of the available information is adequate to support meaningful decisions and to set direction for the future. There are substantial implications: where progress cannot be measured accurately because of inadequate or flawed statistics, the result can be misguided decisions, doubts about achievement of the goals and significant wasted resources.

This book explores the substantial challenges for assembling reliable data, statistics and records to address pressing development challenges, particularly in Africa. Hopefully, by highlighting the enormous potential value of creating and using high quality data, statistics and records, and describing how this can be achieved, the book will contribute to effective global and national development strategies in the critical period to 2030.



**UNIVERSITY
OF LONDON
PRESS**