

Claudia Engelhardt et al.

# How to be FAIR with your data

A teaching and training  
handbook for higher  
education institutions



Georg-August-Universität Göttingen



Claudia Engelhardt et al.  
How to be FAIR with your data

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Published by Göttingen University Press 2022

---

Claudia Engelhardt

with

Katarzyna Biernacka, Aoife Coffey, Ronald Cornet,  
Alina Danciu, Yuri Demchenko, Stephen Downes,  
Christopher Erdmann, Federica Garbuglia, Kerstin  
Germer, Kerstin Helbig, Margareta Hellström, Kristina  
Hettne, Dawn Hibbert, Mijke Jetten, Yulia Karimova,  
Karsten Kryger Hansen, Mari Elisa Kuusniemi, Viviana  
Letizia, Valerie McCutcheon, Barbara McGillivray, Jenny  
Ostrop, Britta Petersen, Ana Petrus, Stefan Reichmann,  
Najla Rettberg, Carmen Reverté, Nick Rochlin, Bregt  
Saenen, Birgit Schmidt, Jolien Scholten, Hugh Shanahan,  
Armin Straube, Veerle Van den Eynden, Justine  
Vandendorpe, Shanmugasundaram Venkataram, Cord  
Wiljes, Ulrike Wuttke, Joanne Yeomans, Biru Zhou

## How to be FAIR with your data

A teaching and training handbook  
for higher education institutions



Göttingen University Press  
2022

### Bibliographic information

The German National Library lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.dnb.de>

FAIRsFAIR has received funding from the European Commission's Horizon 2020 research and innovation programme under Grant Agreement no. 831558. The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.



### *Contact*

[www.fairsfair.eu](http://www.fairsfair.eu)

Dr. Birgit Schmidt, Email: [bschmidt@sub.uni-goettingen.de](mailto:bschmidt@sub.uni-goettingen.de)

This work is protected by German Intellectual Property Right Law.  
It is also available as an Open Access version through the publisher's homepage and the Göttingen University Catalogue (GUK) at the Göttingen State and University Library (<https://www.sub.uni-goettingen.de>).  
The license terms of the online version apply.

Setting and layout: Katja Töpfer  
Illustrations: Patrick Hochstenbach  
Language Editing: Andrew Rennison  
Coverdesign: Margo Bargheer  
Cover picture: Patrick Hochstenbach

© 2022 Universitätsverlag Göttingen  
<https://univerlag.uni-goettingen.de>  
ISBN: 978-3-86395-539-7  
DOI: <https://doi.org/10.17875/gup2022-1915>

## Contents

1 – Motivation .....	3
2 – About this book .....	7
2.1 How this book came about .....	7
2.2 What is FAIR? .....	8
2.3 Why make data FAIR? .....	11
2.4 Who will find this book useful and why? .....	12
3 – FAIR skills and competences .....	13
3.1 The FAIRsFAIR Competence Framework and Body of Knowledge for Higher Education .....	14
3.2 FAIR competence profiles for bachelor’s, master’s and doctoral degree levels .....	15
3.3 Learning outcomes .....	19
4 – Teaching and training designs for FAIR.....	31
4.1 Introduction .....	33
4.2 Elemental phases in course design.....	33
5 – FAIR lesson plans.....	49
6 – Implementing FAIR.....	53
6.1 Introduction .....	54
6.2 Arriving at FAIR institutional policies.....	54
6.3 Data management planning.....	58
6.4 Data processing and documentation .....	59
6.5 Support infrastructure.....	60
6.6 Data publication.....	63
6.7 Data reuse.....	66
7 – References .....	69
8 – About the authors & facilitators .....	75

Appendix A – Resources.....	79
Appendix B – Target audience personas.....	81
Appendix C – Data Stewardship Competence Groups (CF-DSP) and enumeration.....	93
Appendix D – Draft Body of Knowledge (supplement to FAIR Competence Framework).....	101
Appendix E – Knowledge units and corresponding learning outcomes for the bachelor’s, master’s and PhD degree levels.....	113
Appendix F – Lesson plans.....	130
Lesson plan 1: FAIR in a nutshell.....	132
Lesson plan 2: Data management plans (DMP).....	136
Lesson plan 3: Documentation.....	143
Lesson plan 4: Data creation.....	146
Lesson plan 5: File formats.....	149
Lesson plan 6: Metadata.....	152
Lesson plan 7: Data standardisation and ontologies.....	155
Lesson plan 8: Persistent identifiers (PIDs).....	159
Lesson plan 9: Licences, copyright and intellectual property rights (IPR) issues.....	162
Lesson plan 10: Finding and reusing data.....	168
Lesson plan 11: Repositories.....	173
Lesson plan 12: Dealing with confidential, personal, sensitive and private data and ethical aspects.....	178
Lesson plan 13: Data access.....	188
Lesson plan 13: Additional material – data availability statements.....	191
Lesson plan 14: FAIR software /citable code.....	192
Lesson plan 14: Additional material on software citation.....	196
Lesson plan 15: Research data management – overview and best practices..	200
Lesson plan 16: Data management and governance in industry and research ..	203



# 1 – Motivation

This handbook aims to support higher education institutions that are integrating research data management (RDM) skills and findable, accessible, interoperable, and reusable (FAIR) data principles (Wilkinson et al. 2016) into their educational programmes. Managing, curating, and preserving research data in line with the FAIR principles have undoubtedly acquired strategic importance in the institutional agendas of universities. Higher education institutions across Europe and other parts of the world recognise that practicing good RDM is key to staying on par with the digital transition in the production and dissemination of scientific knowledge and, at the same time, to driving the shift towards the mainstreaming of Open Research, commonly known as Open Science.

This handbook offers a practical tool to support universities in this endeavour, providing guidelines and model lesson plans for universities to integrate RDM and FAIR data-related content in bachelor's, master's and doctoral degree programmes. It will also be of interest to other stakeholders looking to deepen their knowledge of the FAIR data principles and searching for material to support them in the design and implementation of teaching or training in line with FAIR.

Survey data gathered from universities across 36 European countries indicated a gap between recognising the strategic importance of research data skills and securing their presence in university programmes (Morais et al. 2021). While 55% to 70% of 272 universities surveyed from 26 October 2020 to 15 January 2021 acknowledged the strategic importance of RDM and FAIR practices, the data showed a substantial gap with a view to their implementation. High levels of implementation were in fact reported by only 15% to 25% of the surveyed institutions. This gap was not limited to the level of institutional policies or infrastructure, but was also evident in the coverage of RDM and FAIR-related topics in current curricula and teaching, as shown by Stoy et al. (2020). This handbook responds to the need, expressed by responding uni-

versities in the same study, for practical guidance on the implementation of the FAIR principles and related skills and competences into curricula and research activities.

Universities are enhancing RDM skills and FAIR data principles education in response to changes within their communities and in the research and innovation landscape. From within their own academic communities, universities are faced with the need to tackle challenges concerning (open) research data. These are mainly determined by a general lack of awareness among their research communities of what the FAIR principles are, along with a widespread shortage of skills and competences as to how they can be put into practice (Morais et al. 2021).

New policies and frameworks are emerging at national, European, and international levels to promote the mainstreaming of Open Research. They present universities with opportunities to tackle the aforementioned challenges and receive financial and capacity assistance to develop their own initiatives in support of Open Research practices. However, efforts to effectively leverage these opportunities are unlikely to achieve their potential if research and support staff are not equipped with adequate skills and competences.

At the European level, the FAIR principles will form a cornerstone of the European Open Science Cloud (EOSC) implementation. The EOSC will collate existing research data infrastructures from EU Member States and Associated Countries to provide a new, shared virtual environment aimed at offering the scientific community seamless access to FAIR research data and services. By doing so, the EOSC aims to ‘help deliver Europe’s contribution to enabling the realisation of scientists’, and science’s, potential in the digital age’ (EOSC 2021, p. 11). Universities widely recognise the positive role that the EOSC can play in facilitating collaborative research and increasing the visibility of institutional research activities (Morais et al. 2021). Universities also have a key role to play, especially in providing more and better-targeted teaching and training activities to develop the next generation of researchers and data professionals. By upskilling and reskilling future graduates, researchers, and support staff, universities will increase their capacities to fully exploit the benefits of the EOSC, both present and in the future, and to contribute to its mission and implementation. At the same time, universities cannot, and indeed should not, set out to do this on their own, as building the EOSC and its skilled workforce is a responsibility that needs to be shared with European and national stakeholders. Practical use cases that can guide and enhance the engagement of universities in the new European infrastructure should be integrated into further implementation strategies of the EOSC (Stoy et al. 2020). Top-down support for the development of new policies and funding schemes, as well as the alignment of existing frameworks, are also key to boosting the capacity of universities to take on this role and be drivers for change.

At the European level, the Open Research transition is notably being promoted by the European Commission. A recent and prominent example is the require-

ment for Data Management Plans (DMPs) for all projects generating or reusing data introduced by the European Commission for Horizon Europe, the 9th European Framework Programme for Research & Innovation, and by a growing number of other funding organisations. Model Grant Agreements for EU-funded programmes between 2021 and 2027 will also require data gleaned from new projects to be compliant with the FAIR principles. However, this is not just a European endeavour. Funding organisations across the world, be they national or international, now require grant holders to deliver reusable and accessible data from their funded research projects. Whilst funders have previously tended to encourage data sharing, it is increasingly becoming a requirement for data created by publicly funded research projects to be made available with as few restrictions as possible where ethical and legal obligations permit, with secondary use of data being enabled wherever possible. This reflects funding organisations' efforts to secure public trust in scientific enquiries and to ensure accountability in public funding. In this landscape, national, funder, and institutional policies all play an important role and are constantly in flux (Sveinsdottir et al. 2021). Enhancing teaching and training provisions for RDM and FAIR data will be instrumental in addressing these new expectations on how research data should be managed and, hence, to ensuring the continued access of institutions to national, European, and international funding schemes.

At the national level, the landscape of policies addressing Open Research, while diverse, is becoming richer, with many European countries having already adopted such regulations or preparing to do so (EOSC 2020; Sveinsdottir et al. 2021). While the provisions related to FAIR data can still be improved in the context of these policies, universities should be aware of the opportunities they create. Having a sound framework of policies at the national or regional level can be instrumental not only as a driver for the development of top-down initiatives in institutions, but also to ensure that the impact of these efforts will be sustainable in the long term.

There are also significant economic benefits in making research data FAIR. A recent study commissioned by the European Commission (EC 2019) has shown that there are substantial additional costs when research data are not managed in compliance with the FAIR principles. These costs vary from storage and licence costs to more qualitative costs linked to the time spent by researchers on the creation, collection and management of data, and the risks of research duplication. In Europe, these are estimated to amount to at least EUR 10.2 billion per year (ibid.). Moreover, the same report highlights how, once the right infrastructures are in place, the benefits of having FAIR data are expected to increase in the long run. At the same time, making research data FAIR can offer different benefits to academic institutions and their researchers, particularly in terms of opportunities to manage time and storage costs in a more efficient way, while improving collaboration across scientific communities (ibid.). Despite the economic argument forming part of the discussion around FAIR

data, universities should strive to develop good RDM practices, and receive the support needed to do so, regardless of any potential returns on investment. Making FAIR data management an established practice across research performing organisations (RPOs) is in fact a key step in ensuring high-quality standards in terms of findability, accessibility, interoperability, and reusability of new scientific knowledge and in fostering the sharing of data in an ethical and responsible way.

To tackle the aforementioned challenges posed by the lack of awareness and skills, universities need to provide more and better-targeted teaching and training activities to their students and (early-stage) researchers. Students at the bachelor and master levels need to acquire general knowledge on how to sustainably manage data, document them accordingly, and make them FAIR. This will be instrumental for them not only if they choose to enter doctoral education, but also if they are interested in pursuing a career in other sectors where demand for data-skilled professionals is growing exponentially (OECD 2020). Researchers also need to be equipped with a basic level of data management skills that allow them to work efficiently within their research teams where the distribution of competences among their support staff is becoming increasingly variable (*ibid.*). However, at the doctoral level, general training will not be enough and must be accompanied by a discipline-specific approach.

In conclusion, a growing number of national, European and, international initiatives are emerging to establish Open Research practices as the standard way of conducting research. Investing in new and better training for RDM and FAIR data skills will be key to taking full advantage of the opportunities they have to offer. Top-down regulations will also act as a driver for the further uptake of a FAIR culture at universities, requiring higher education institutions to take the lead in advancing the implementation of (FAIR) research data management practices. At the same time, efforts will be needed at the institutional level to ensure that complying with RDM and FAIR is not seen as an extra burden on the shoulders of researchers, but rather as an integral and supported part of their research activities.

Fostering the integration of FAIR skills and competences in university programmes is a key step to furthering the transition towards FAIR and Open Research (EC 2018). This handbook supports universities in taking this step by providing ready-to-use material for teaching FAIR principles at different levels. It also presents didactic approaches on how to teach FAIR, equipping readers with the knowledge they need to get started with designing their own courses and training activities to be implemented at their institutions.

## 2 – About this book

### 2.1 How this book came about

This handbook was first written in a book sprint organised by the EU-funded FAIRsFAIR project. Led by the University of Göttingen, the project brought together a variety of RDM and teaching experts who wrote, edited and finalised the handbook. The aim of FAIRsFAIR, which ran from March 2019 to February 2022, is to develop and supply practical solutions to support the implementation and use of the FAIR principles throughout the research data lifecycle, including uptake of the principles in higher education.

Based on a survey and a number of focus groups (Stoy et al. 2020), an analysis of job advertisements as well as previous work by EDISON and other projects (Demchenko et al. 2021), FAIRsFAIR has developed a FAIR Competence Framework for Higher Education (ibid.). This handbook is a practical tool complementing the framework, supporting its application and implementation.

To extend the available pool of expertise beyond the project partners involved in this task (University of Göttingen, European University Association, University of Amsterdam and University of Minho), a book sprint was chosen as the method to prepare the handbook since this has proven successful in the past, as recently demonstrated by the *FOSTER Open Science Training Handbook* (Bezjak et al. 2019), *Engaging Researchers with Research Data Management: The Cookbook* (Clare et al. 2019), *The Turing Way* (The Turing Way n.d.), *FAIR Cookbook* for the Life Sciences (FAIR Cookbook n.d.), and the *Top 10 FAIR Data & Software Things* (Martinez et al. 2019).

The book sprint consisted of six three-hour sessions held between 1 and 10 June 2021 which involved a kick-off meeting, four dedicated sprint sessions, and a wrap-

up meeting. In view of the ongoing COVID-19 pandemic, the sprints were held virtually, using Google Docs for writing, Zoom for video conferencing, and Slack as an additional communication channel.

In a preceding application process, 38 experts from 14 European countries as well as the United States and Canada had been selected from a group of 53 applicants. Despite coming from diverse disciplinary backgrounds, they all possess ample relevant expertise in terms of RDM and the FAIR principles and, in most cases, experience in teaching and training and/or lesson, course, or curriculum design. Including the FAIRsFAIR colleagues, around 40 people contributed to the handbook by writing or reviewing and editing – or both.

The post-sprint editorial process was accompanied by an editorial team comprising book sprint participants and FAIRsFAIR project members. One step in this process was a public consultation on the first draft during summer 2021 to gather feedback and input from the wider community so as to further improve upon the first version. This was followed by a revision by the editorial team, a presentation of the revised draft in a workshop on 12 October 2021, and its subsequent finalisation. The handbook was first published as a FAIRsFAIR project deliverable in December 2021 (Engelhardt et al. 2021).

## 2.2 What is FAIR?

In 2016, the *‘FAIR Guiding Principles for scientific data management and stewardship’* were published in *Scientific Data* (Wilkinson et al. 2016). FAIR stands for findable, accessible, interoperable and reusable. The FAIR principles have become increasingly important, acting as guidelines to improve the entire lifecycle of research data management.

While FAIR and open data are overlapping yet distinct concepts, they both focus on data sharing to ensure that data are made available in ways that promote access and reuse (Higman et al. 2019). Open Research promotes a cultural shift towards sharing research outputs, whereas FAIR concentrates on how to prepare data so that they can be reused by others. However, FAIR does not require data to be open, and following FAIR can be beneficial for data that cannot be made open, e.g. for privacy reasons. FAIR provides a set of rules that are a robust standard to which curation of data should aspire. Consequently, it should be noted that FAIR-compliant data are not necessarily of high quality, and the issue of quality assurance of the data is a separate one extending beyond the scope of this book. Similarly, it should be noted that FAIR-compliant data may be necessary but not sufficient in some reuse scenarios, e.g. computational reproducibility (see Peer et al. 2021).

The term ‘FAIR’ was originally launched at a Lorentz workshop in the Netherlands in 2014 (Wilkinson et al. 2016; Data FAIRport n.d.), and in the following we will refer to the FAIR Guiding Principles as they were published in 2016 (see next page<sup>1</sup>).

The FAIR principles are typically translated into concrete complementary actions to be taken by researchers, infrastructure providers, research funders and other actors (European Commission 2018; Science Europe 2021). They are increasingly becoming a requirement by national and European funders, and institutional policies on good research practice, e.g. German Research Foundation 2019, UK Research and Innovation, National Institutes of Health, Dutch Research Council, all of which provide guidance on what they expect researchers to implement during the course of their projects, such as DMP templates or checklists to identify FAIR-compliant repositories (Davidson et al. 2019; Sveinsdottir et al. 2021).

---

1 On the next page, we quote the FAIR Guiding Principles as they appear in Wilkinson et al. (2016). Therefore, the spelling deviates in some places from the standard British English used in this document.



### To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

### To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
  - A1.1 the protocol is open, free, and universally implementable
  - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

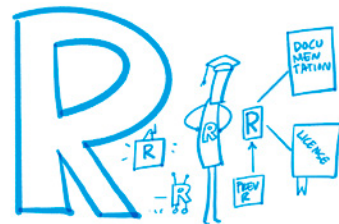


### To be Interoperable:

- I1. (meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta) data

### To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
  - R1.1. (meta)data are released with a clear and accessible data usage license
  - R1.2. (meta)data are associated with detailed provenance
  - R1.3. (meta)data meet domain-relevant community standards





## 2.3 Why make data FAIR?

Upholding integrity and reproducibility is key to any good research, and best practice in RDM is an essential part of efforts to accomplish this. Open Research and, in particular, the FAIR principles are a set of guidelines that could be viewed as a gold standard for RDM. When considering why adoption of the FAIR principles should be encouraged and embraced, there are many reasons extending beyond those of research integrity and reproducibility. Irrespective of whether they own or produce the data, or reuse data provided by others, researchers will find their lives much easier if they are able to find, retrieve and reuse data, while also increasing the value of the data due to their enhanced visibility. In addition, FAIR data enable easier data integration within and across disciplines, supporting worldwide, multi- and interdisciplinary research endeavours that address global challenges such as climate change, health emergencies or the realisation of sustainable development goals. When considering the financial implications, especially for publicly funded research, a reduction of double efforts and increasing reuse of existing data are key motivators, with studies underlining the implications of data management that is not FAIR-compliant, e.g. EC 2019. To this end, the FAIR principles go a considerable way in addressing this problem. Many funders and institutions, including the UN, WHO, OECD and others, have explicitly referenced the FAIR principles, providing a policy framework to support and sustain their growing importance. Funders' mandates mean that researchers will have to meet obligations to make their data FAIR-compliant. Meanwhile, data management plans (DMPs) are also becoming increasingly important and mandatory, with many templates explicitly providing guidance for the components of the FAIR principles, such as templates and guidelines provided until recently by Horizon 2020, and by Horizon Europe from 2021. Practical guidelines on how to comply with funding requirements and RDM policies were also developed by Science Europe (Science Europe 2021). Researchers can use these tools to identify the different considerations that need to be made for their project that correspond to each of the principles and which can be documented, such as file formats, standards and licences.

Although the FAIR principles do not necessitate data being open, the ambition is to increase alignment of the two concepts where possible, with the notable exception of data which cannot be made open for reasons such as their ethical sensitivity, copyright, cultural protocols, or commercial licensing. However, even in the case of such data, metadata should be made available for discoverability purposes which can then be requested and shared in a safe manner through access control mechanisms as and where appropriate. Not only does this aid data reuse, it also increases public trust and accountability, which is essential when considering publicly funded research.

The FAIR principles are complemented by other principles that focus on long-term governance, integrity and curation, such as the CARE Principles for Indigenous

Data Governance (Collective Benefit, Authority to Control, Responsibility, and Ethics; Carroll et al. 2020) which address ethical considerations, and the TRUST Principles for digital repositories (Transparency, Responsibility, User focus, Sustainability, Technology; Lin et al. 2020). Therefore, it is important to remember that applying the FAIR principles only covers part of best practice in RDM and Open Research, e.g. data curation practices, data services, and data visualisation.

## 2.4 Who will find this book useful and why?

This handbook aims to support higher education institutions in integrating content relating to the FAIR principles into their curricula and teaching. This involves a number of roles which contribute to this process at various levels.

To obtain a better grasp of the target audience(s) and their needs and expectations with regard to such a handbook, the book sprint started with an exercise dealing with personas representing different HEI staff groups for whom this work could be of relevance. The results of this exercise were used to develop the handbook's structure and content. For more information about the procedure and outcomes of the persona exercise, please refer to Appendix B.

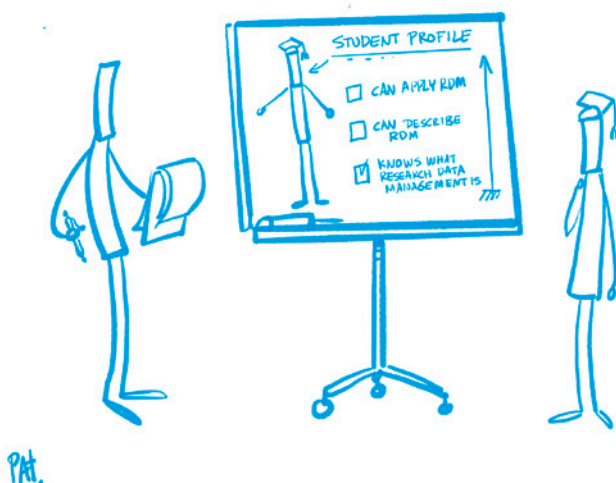
The following table summarises the main areas of activity with regard to the implementation of the FAIR principles in teaching and guides readers to the chapters in which each of these is addressed.

**Table 1:** Fields of activity and relevant chapters

Area of activity	Roles concerned (examples)	Most relevant chapters
lesson and course planning, creation, and teaching	lecturers, professors, trainers, support staff	3 - FAIR Skills and Competences 4 - Teaching and training designs 5 - FAIR lesson plans
design, implementation, and adaptation of curricula	doctoral programme managers, deans	3 - FAIR Skills and Competences
training of PhD students and early career researchers	support staff, trainers, lecturers, professors	3 - FAIR Skills and Competences 4 - Teaching and training designs 5 - FAIR lesson plans
consideration and implementation of FAIR in institutional strategies, policies, administration, and management	Vice presidents/Vice rectors/ Offices of research, data protection officer	6 - Implementing FAIR

### 3 – FAIR skills and competences

Before actually implementing topics surrounding FAIR in curricula and teaching, the first thing to do is to define the knowledge and competences which students at different educational levels should acquire. Here, we suggest a separate core set of Knowledge Units and associated learning outcomes for the bachelor's, master's and PhD degree levels.<sup>2</sup> These sets are discipline-agnostic and may need to be adapted slightly depending on the discipline in question. They can be used as a basis to develop a curriculum focused on the FAIR principles, or to map them to existing curricula and courses to identify which topics are already covered and which are not (and should therefore be added).



<sup>2</sup> Initially, we had considered six roles in total. In addition to bachelor's, master's and PhD degree students, we also looked at postdoc/researcher, PI and support staff. However, due to capacity limitations, the latter three were dropped in favour of the target audiences most relevant for HEI teaching.

The competence profiles suggested here were developed based on the FAIR Competence Framework for Higher Education – Data Stewardship Professional Framework (Demchenko et al. 2021) and the corresponding (draft) Body of Knowledge<sup>3</sup> (see Appendix D) created by the FAIRsFAIR project (which are both heavily based on the EDISON Data Science Framework, EDISONcommunity 2020). They are summarised below (section 3.1) and followed by a description of the approach used to create the competence profiles and the learning outcomes (sections 3.2 and 3.3).

### 3.1 The FAIRsFAIR Competence Framework and Body of Knowledge for Higher Education

The FAIRsFAIR Competence Framework for Higher Education (Demchenko et al. 2021) was designed to cover all knowledge, skills, and competences relevant to Data Stewardship. It defines Competence Groups for the following domains:

- Data Management (DSDM)
- Data Science Engineering (DSENG)
- Data Science Research Methods and Project Management (DSRMP); and
- Data Science Domain Knowledge (DSDK) as Business Process Management (DSBA).

The most relevant area in relation to the FAIR principles is Data Management which contains nine Competence Groups. For an overview of all Competence Groups, see Appendix C (taken from Demchenko et al. 2021, pp. 70 et sqq.).

The accompanying Body of Knowledge (BoK) breaks down the Competence Groups of the FAIR Competence Framework into a number of Knowledge Units (with each Knowledge Unit covering a specific aspect or topic), in turn making it easier to translate the framework into content and material for teaching and training. The Knowledge Units are grouped into Knowledge Area Groups (KAG) with a corresponding Knowledge Area Group (in the Body of Knowledge) for each Competence Group (of the Competence Framework).

As mentioned above, the FAIRsFAIR Competence Framework was developed based on the EDISON Data Science Framework, as was the (draft) Body of Knowl-

---

<sup>3</sup> In this draft version, one of four areas of the original version from the EDISON project (EDISONcommunity, 2020) – Research Data Management – has been updated and further developed. This is the domain most relevant to FAIR-related competences in university teaching. The other domains (Data Science Engineering, Data Science Research Methods and Project Management, as well as Data Science Domain Knowledge as Business Process Management) remain the same as in the original version.

edge. However, at the time the book sprint took place, the FAIRsFAIR BoK was still a work in progress with only one Competence Group having been updated compared to the original EDISON version: Data Management, the area cover most of the Knowledge Units that are of importance when teaching FAIR in Higher Education Institutions.

The Data Management KAG of the (draft) BoK comprises six Knowledge Areas (KA):

- General principles and concepts in Data Management and organisation
- Data Management Systems
- Data Management and Enterprise data infrastructure
- Data Governance
- Big Data storage
- Data archives and data libraries

For the full version of the KAG (draft) BoK, see Appendix D.

## 3.2 FAIR competence profiles for bachelor’s, master’s and doctoral degree levels

### Method

The scope of competences covered by the Competence Framework and BoK is geared towards the Data Steward role, encompassing a very wide range of Knowledge Units. Only a fraction of these is needed by students of other disciplines. To identify relevant competences and formulate corresponding learning outcomes, eight book sprint participants collaborated during (and after) the book sprint sessions in a multi-level process.

First, each of the Knowledge Units in the Data Management area of the BoK was assessed in terms of their relevance for bachelor’s, master’s and PhD degrees by assigning them one of five ranges (irrelevant, basic, intermediate, advanced or professional).<sup>4</sup> These ranges are based on the European Qualification Framework (EQF, European Union n.d.) which encompasses eight levels. The aim of creating the ranges was to reduce the complexity somewhat. The ‘basic’ range comprises levels 1-3 of the EQF, ‘intermediate’ levels 4-5, ‘advanced’ levels 6-7, and ‘professional’ level 8.

---

<sup>4</sup> The procedure in detail was as follows: First, for each of the six Knowledge Unit Areas, one (sometimes two) of the involved book sprint participants, based on their expertise and experience, estimated the required level for bachelor’s, master’s and PhD degree students. These were reviewed by the other participants before the next session. During the subsequent session, the group discussed each individual item, and then approved or amended the classification. Knowledge Units deemed irrelevant or redundant were removed to consolidate the table collaboratively.

This step also involved excluding or merging Knowledge Units considered irrelevant or redundant, e.g. a number of concepts closely related to the computer science and IT perspective on data management such as data warehouse architecture and processes, data models and query languages, or middleware for databases. On the other hand, a few topics seemed to be missing, e.g. ontologies and controlled vocabularies, or data discovery including data selection and use in research. Some of them are covered by Knowledge Units in other areas of the BoK. In such instances, the respective Knowledge Unit was added to the table. Topics not represented by any existing Knowledge Unit led to a new item being created and added.<sup>5</sup>

In a second step, the group discussed and agreed upon which of the selected Knowledge Units could be considered entry-level content, i.e. compulsory topics. This was then performed again for each of the bachelor's, master's and PhD degree levels. The competence profiles defined using this method are presented in the table below.

## Competence profiles

**Table 2:** Competence profiles for the bachelor's, master's and PhD degree levels

Topic	Bachelor (required level)	Master (required level)	PhD (required level)	Entry- level content?
General principles and concepts in data management – overview	basic	intermediate	advanced	yes
Overview of data types, data type registries and data formats	basic	basic	intermediate	yes
Metadata, metadata formats, standards and registries	basic	intermediate	advanced	yes
Open Research, Open Access, Open Data	basic	intermediate	advanced	yes
Metadata management, registries and publication	basic	basic	intermediate	no
Persistent Identifiers (PID), Open Researcher and Contributor ID (ORCID), Research Organization Registry (ROR)	basic	basic	intermediate	yes

<sup>5</sup> In addition to bachelor's, master's and PhD degree students, we considered three other roles in this first step: Postdoc/Researcher, PI and support staff. These were later dropped due to capacity reasons and to focus on the most relevant target audiences of HEI teaching.

<b>Topic</b>	<b>Bachelor (required level)</b>	<b>Master (required level)</b>	<b>PhD (required level)</b>	<b>Entry- level content?</b>
FAIR (Findable, Accessible, Interoperable, Reusable) principles in data management	basic	basic	intermediate	yes
FAIR metadata management and tools for FAIR metadata management	basic	intermediate	advanced	no
Databases and database management systems, data modelling	basic	basic	basic	no
Data structures	basic	basic	basic	no
Master data management, data dictionaries	basic	basic	intermediate	yes
FAIR data management requirements and compliance	irrelevant	basic	intermediate	no
Data management including reference and master data	irrelevant	basic	basic	no
Data storage and operations	basic	intermediate	advanced	no
Data infrastructure, data registries and data factories	basic	basic	intermediate	no
Data security and protection	basic	basic	intermediate	yes
Data backup	basic	intermediate	advanced	yes
Personal data protection, GDPR compliance	basic	basic	intermediate (depending on discipline)	yes
Data anonymisation/pseudonymisation	irrelevant	basic (depending on discipline)	intermediate (depending on discipline)	no
Data management planning, FAIR data management and compliance	basic	basic	intermediate	yes
Data integration and interoperability, data preparation and cleaning	basic	intermediate	advanced	no

<b>Topic</b>	<b>Bachelor</b> (required level)	<b>Master</b> (required level)	<b>PhD</b> (required level)	<b>Entry-level content?</b>
Data interoperability and meta-data management	basic	basic	intermediate	yes (basic concept)
Organisational roles in data governance, data stewardship	basic	basic	intermediate	no
Data provenance, data lineage	basic	basic	intermediate	yes
Responsible data use, data privacy, ethical principles, Intellectual Property Rights (IPR) and legal issues	basic	intermediate	advanced	yes
Data quality management, best practices and frameworks, data quality metrics	basic	intermediate	advanced	yes (basic concept)
Data protection policies (including personal data), data access policies, GDPR (General Data Protection Regulation) compliance	basic	basic	intermediate	no
Trusted data repositories and certification	basic	basic	intermediate	yes (basic concept)
Data discovery (published data), data selection and use in research	basic	intermediate	advanced	yes (basic concept)
Research data lifecycle	basic	basic	intermediate	yes
Ontologies and controlled vocabularies	basic	intermediate	advanced	yes



### 3.3 Learning outcomes

Finally, learning outcomes were formulated (using Bloom’s taxonomy). Via et al. (2020, p. 2) define learning outcomes as “the KSAs [i.e. knowledge, skills and abilities] that learners should be able to demonstrate after instruction, the tangible evidence that the teaching goals have been achieved”. They play an important role in the course design process (more information about this is provided in chapter 4).

The learning outcomes for the Knowledge Units deemed entry level content are presented in the tables below. For the full list of learning outcomes, please refer to Appendix E.

**Table 3:** Entry-level content including learning outcomes – bachelor level

Topic	Required level	Learning outcomes [b]=basic, [i]=intermediate, [a]=advanced]
General principles and concepts in data management – overview	basic	[b] Can define Research Data Management (RDM) and can describe its relevance and benefits.
Overview of data types, data type registries and data formats	basic	[b] Can describe what types of data exist (Knowledge). [b] Can explain what data type registries are (Knowledge). [b] Can identify data formats (Knowledge).
Metadata, metadata formats, standards and registries	basic	[b] Can describe types of metadata. [b] Can recognise metadata formats. [b] Can identify metadata standards. [b] Can use metadata standards to describe resources. [b] Can explain what metadata registries are. [b] Can search and find data and metadata standards registries.
Open Research, Open Access, Open Data	basic	[b] Can paraphrase the concept of Open Research. [b] Can describe the benefits of Open Research.

<b>Topic</b>	<b>Required level</b>	<b>Learning outcomes</b> [b]=basic, [i]=intermediate, [a]=advanced]
Persistent Identifiers (PID), Open Researcher and Contributor ID (ORCID), Research Organization Registry (ROR)	basic	[b] Can recognise PIDs and explain the different use cases for PIDs. [b] Can explain the importance of PIDs for FAIR data. [b] Can use PIDs to access data or other resources.
FAIR (Findable, Accessible, Interoperable, Reusable) principles in data management	basic	[b] Can paraphrase the FAIR principles. [b] Can explain why the FAIR principles were developed. [b] Can recognise the relationship between FAIR, Open and RDM.
Master data management, data dictionaries	basic	[b] Can develop a data management plan for their own work. [b] Can identify different types of data documentation. [b] Can explain the purpose of the documentation. [b] Can use existing documentation.
Data security and protection	basic	[b] Can define different levels of data security (user, folder, files). [b] Can explain different ways of data protection (physical, encryption etc.).
Data backup	basic	[b] Can describe what a backup is and tell reasons for backup creation. [b] Can explain the 3-2-1 rule and apply it to their own files. [b] Can identify institutional backup solutions.
Personal data protection, GDPR compliance	basic	[b] Can explain reasons for data protection. [b] Knows basic rules and legal regulations for sensitive data (e.g. GDPR). [b] Knows how to comply with these rules and laws.
Data management planning, FAIR data management and compliance	basic	[b] Can describe what a data management plan (DMP) is. [b] Can explain why data management planning is a step towards FAIR.
Data interoperability and metadata management	basic	[b] Can explain aspects of interoperability (Knowledge). [b] Can relate metadata management to interoperability (Understand).

<b>Topic</b>	<b>Required level</b>	<b>Learning outcomes</b> [b]=basic, [i]=intermediate, [a]=advanced]
Data provenance, data lineage	basic	[b] Can illustrate with an example what data provenance/data lineage means.
Responsible data use, data privacy, ethical principles, IPR and legal issues	basic	[b] Can summarise and explain ethical principles and responsible data use (e.g. CARE, indigenous data). [b] Can describe legal issues around data use and management (e.g. licences, patents, policies, contracts etc.).
Data quality management, best practices and frameworks, data quality metrics	basic	[b] Can summarise best practices ensuring data quality.
Trusted data repositories and certification	basic	[b] Can explain what a trusted data repository is and how to find it (re3data.org and FAIRsharing). [b] Can compare different certifications for data repositories (e.g. CoreTrustSeal, CLARIN certification).
Data discovery (published data), data selection and use in research	basic	[b] Can explain the importance of data discovery and reuse.
Research data lifecycle	basic	[b] Can explain the steps of the research data lifecycle. [b] Can compare different lifecycle models.
Ontologies, controlled vocabularies	basic	[b] Can explain the role of ontologies and vocabularies (Knowledge). [b] Can recognise the use of ontologies and vocabularies (Knowledge). [b] Can identify a few domain-relevant ontologies. (Knowledge). [b] Can search and find terminologies in registries.

**Table 4:** Entry-level content including learning outcomes – master level

<b>Topic</b>	<b>Required level</b>	<b>Learning outcomes</b> [b]=basic, [i]=intermediate, [a]=advanced]
General principles and concepts in data management – overview	intermediate	[b] Can define Research Data Management (RDM) and can describe its relevance and benefits. [i] Can describe RDM measures to be taken (including explaining why) at different stages of the research process.
Overview of data types, data type registries and data formats	basic	[b] Can describe what types of data exist (Knowledge). [b] Can explain what data type registries are (Knowledge). [b] Can identify data formats (Knowledge).
Metadata, metadata formats, standards and registries	intermediate	[b] Can describe types of metadata. [b] Can recognise metadata formats. [b] Can identify metadata standards. [b] Can use metadata standards to describe resources. [b] Can explain what metadata registries are. [b] Can search and find data and metadata standards registries [i] Can articulate metadata of different types to describe a resource. [i] Can write metadata in a relevant format. [i] Can appraise the usefulness of metadata standards to describe a resource. [i] Can search metadata registries to find resources.
Open Research, Open Access, Open Data	intermediate	[b] Can paraphrase the concept of Open Research. [b] Can describe the benefits of Open Research. [b] Can describe Open Access and Open Data as areas of Open Research. [i] Can recognise if a publication is open access. [i] Can discover platforms for Open Access/Open Data. [i] Can articulate what is required to make research outputs open. [i] Can contrast FAIR and open.

Topic	Required level	Learning outcomes [b]=basic, [i]=intermediate, [a]=advanced]
Persistent Identifiers (PID), Open Researcher and Contributor ID (ORCID), Research Organization Registry (ROR)	basic	<ul style="list-style-type: none"> <li>[b] Can recognise PIDs and explain the different use cases for PIDs.</li> <li>[b] Can explain the importance of PIDs for FAIR data.</li> <li>[b] Can use PIDs to access data or other resources.</li> </ul>
FAIR (Findable, Accessible, Interoperable, Reusable) principles in data management	basic	<ul style="list-style-type: none"> <li>[b] Can paraphrase the FAIR principles.</li> <li>[b] Can explain why the FAIR principles were developed.</li> <li>[b] Can recognise the relationship between FAIR, Open and RDM.</li> </ul>
Master data management, data dictionaries	basic	<ul style="list-style-type: none"> <li>[b] Can develop a data management plan for their own work.</li> <li>[b] Can identify different types of data documentation.</li> <li>[b] Can explain the purpose of the documentation.</li> <li>[b] Can use existing documentation.</li> </ul>
Data security and protection	basic	<ul style="list-style-type: none"> <li>[b] Can define different levels of data security (user, folder, files).</li> <li>[b] Can explain different ways of data protection (physical, encryption etc.).</li> </ul>
Data backup	intermediate	<ul style="list-style-type: none"> <li>[b] Can describe what a backup is and give reasons for backup creation.</li> <li>[b] Can explain the 3-2-1 rule and apply it to their own files.</li> <li>[b] Can identify institutional backup solutions.</li> <li>[i] Can explain institutional backup solutions and apply them to own files.</li> </ul>
Personal data protection, GDPR compliance	basic	<ul style="list-style-type: none"> <li>[b] Can explain reasons for data protection.</li> <li>[b] Knows basic rules and legal regulations for sensitive data (e.g. GDPR).</li> <li>[b] Knows how to comply with these rules and laws</li> </ul>
Data management planning, FAIR data management and compliance	basic	<ul style="list-style-type: none"> <li>[b] Can describe what a data management plan (DMP) is.</li> <li>[b] Can explain why data management planning is a step towards FAIR.</li> </ul>

Topic	Required level	Learning outcomes [b]=basic, [i]=intermediate, [a]=advanced
Data interoperability and metadata management	basic	[b] Can explain aspects of interoperability (Knowledge). [b] Can relate metadata management to interoperability (Understand).
Data provenance, data lineage	basic	[b] Can illustrate with an example what data provenance/data lineage means.
Responsible data use, data privacy, ethical principles, IPR and legal issues	intermediate	[b] Can summarise and explain ethical principles and responsible data use (e.g. CARE, indigenous data). [b] Can describe legal issues around data use and management (e.g. licences, patents, policies, contracts etc.). [i] Can analyse if ethical principles or legal issues play a role in their own work.
Data quality management, best practices and frameworks, data quality metrics	intermediate	[b] Can summarise best practices ensuring data quality. [i] Can describe how to recognise quality data.
Trusted data repositories and certification	basic	[b] Can explain what a trusted data repository is and how to find it (re3data.org and FAIRsharing). [b] Can compare different certifications for data repositories (e.g. CoreTrustSeal, CLARIN certification).
Data discovery (published data), data selection and use in research	intermediate	[b] Can explain the importance of data discovery and reuse. [i] Can discover published datasets in their discipline. [i] Can cite data.
Research data lifecycle	basic	[b] Can explain the steps of the research data lifecycle. [b] Can compare different lifecycle models.
Ontologies, controlled vocabularies	intermediate	[b] Can explain the role of ontologies and vocabularies (Knowledge). [b] Can recognise the use of ontologies and vocabularies (Knowledge). [b] Can identify a few domain-relevant ontologies (Knowledge). [b] Can search and find terminologies in registries. [i] Can use ontologies to describe resources (Apply).

**Table 5:** Entry-level content including learning outcomes – doctoral level

<b>Topic</b>	<b>Required level</b>	<b>Learning outcomes</b> [b]=basic, [i]=intermediate, [a]=advanced]
General principles and concepts in data management – overview	advanced	[b] Can define Research Data Management (RDM) and can describe its relevance and benefits. [i] Can describe RDM measures to be taken (including explaining why) at different stages of the research process. [a] Can practically apply theoretical knowledge about proper RDM measures to be taken at different stages to their own research process/project.
Overview of data types, data type registries and data formats	intermediate	[b] Can describe what types of data exist (Knowledge). [b] Can explain what data type registries are (Knowledge). [b] Can identify data formats (Knowledge). [i] Can determine proper data types for a resource (Analyse). [i] Can use a data type registry (Apply). [i] Can use proper data formats to express resources (Apply).
Metadata, meta-data formats, standards and registries	advanced	[b] Can describe types of metadata. [b] Can recognise metadata formats. [b] Can identify metadata standards. [b] Can use metadata standards to describe resources. [b] Can explain what metadata registries are. [b] Can search and find data and metadata standards registries. [i] Can articulate metadata of different types to describe a resource. [i] Can write metadata in a relevant format. [i] Can appraise the usefulness of metadata standards to describe a resource. [i] Can search metadata registries to find resources. [a] Can design rich metadata to describe a resource. [a] Can use proper metadata formats and models to express these metadata. [a] Can deposit metadata in a repository.

Topic	Required level	Learning outcomes [b]=basic, [i]=intermediate, [a]=advanced
Open Science/ Research, Open Access, Open Data	advanced	<ul style="list-style-type: none"> <li>[b] Can paraphrase the concept of Open Science.</li> <li>[b] Can describe the benefits of Open Science.</li> <li>[b] Can describe Open Access and Open Data as areas of Open Science.</li> <li>[i] Can recognise if a publication is open access.</li> <li>[i] Can discover platforms for Open Access/Open Data.</li> <li>[i] Can articulate what is required to make research outputs open.</li> <li>[i] Can contrast FAIR and open.</li> <li>[a] Can plan publication of Open Access publications and FAIR data.</li> </ul>
Persistent Identifiers (PID), Open Researcher and Contributor ID (ORCID), Research Organization Registry (ROR)	intermediate	<ul style="list-style-type: none"> <li>[b] Can recognise PIDs and explain the different use cases for PIDs.</li> <li>[b] Can explain the importance of PIDs for FAIR data.</li> <li>[b] Can use PIDs to access data or other resources.</li> <li>[i] Can apply PIDs to their own research outputs.</li> <li>[i] Can use PIDs to collaborate with others.</li> </ul>
FAIR (Findable, Accessible, Interoperable, Reusable) principles in data management	intermediate	<ul style="list-style-type: none"> <li>[b] Can paraphrase the FAIR principles.</li> <li>[b] Can explain why the FAIR principles were developed.</li> <li>[b] Can recognise the relationship between FAIR, Open and RDM.</li> <li>[i] Can plan for FAIR research outputs.</li> <li>[i] Can write and develop a research data management plan.</li> <li>[i] Can apply the principles to their own work.</li> <li>[i] Can evaluate the FAIRness of their own work or the work of others.</li> </ul>
Master data management, data dictionaries	intermediate	<ul style="list-style-type: none"> <li>[b] Can develop a data management plan for their own work.</li> <li>[b] Can identify different types of data documentation.</li> <li>[b] Can explain the purpose of the documentation.</li> <li>[b] Can use existing documentation.</li> <li>[i] Can modify existing documentation.</li> <li>[i] Can evaluate and prioritise data management activities.</li> </ul>



Topic	Required level	Learning outcomes [b]=basic, [i]=intermediate, [a]=advanced
Data security and protection	intermediate	<ul style="list-style-type: none"> <li>[b] Can define different levels of data security (user, folder, files).</li> <li>[b] Can explain different ways of data protection (physical, encryption etc.).</li> <li>[i] Can use different levels of security for their own work.</li> <li>[i] Can apply data protection methods like password protection and encoding.</li> <li>[i] Does share and collaborate in a secure way.</li> </ul>
Data backup	advanced	<ul style="list-style-type: none"> <li>[b] Can describe what a backup is and tell reasons for backup creation.</li> <li>[b] Can explain the 3-2-1 rule and apply it to their own files.</li> <li>[b] Can identify institutional backup solutions.</li> <li>[i] Can explain institutional backup solutions and apply them to own files.</li> <li>[a] Can analyse and evaluate backup.</li> <li>[a] Can solve backup problems independently or with further assistance from support staff.</li> </ul>
Personal data protection, GDPR compliance	intermediate (depending on discipl.)	<ul style="list-style-type: none"> <li>[b] Can explain reasons for data protection.</li> <li>[b] Knows basic rules and legal regulations for sensitive data (e.g. GDPR).</li> <li>[b] Knows how to comply with these rules and laws.</li> <li>[i] Can analyse compliance to legal regulations for sensitive data.</li> <li>[i] Can apply mechanisms to protect data appropriately.</li> </ul>
Data management planning, FAIR data management and compliance	intermediate	<ul style="list-style-type: none"> <li>[b] Can describe what a data management plan (DMP) is.</li> <li>[b] Can explain why data management planning is a step towards FAIR.</li> <li>[i] Can tell which areas should be covered in a DMP.</li> <li>[i] Can sketch a DMP for their own research project.</li> </ul>

Topic	Required level	Learning outcomes [b]=basic, [i]=intermediate, [a]=advanced
Data interoperability and metadata management	intermediate	<ul style="list-style-type: none"> <li>[b] Can explain aspects of interoperability (Knowledge).</li> <li>[b] Can relate metadata management to interoperability (Understand).</li> <li>[i] Use domain-relevant standards, models and formats for interoperable data (Apply).</li> <li>[i] Can relate metadata management to interoperability (Apply).</li> </ul>
Data provenance, data lineage	intermediate	<ul style="list-style-type: none"> <li>[b] Can illustrate with an example what data provenance/data lineage means.</li> <li>[i] Can transfer how data provenance/data lineage plays a role in their own research project.</li> <li>[i] Can apply data provenance good practices to their own data and ensure that an unbroken data lineage is established for their work.</li> </ul>
Responsible data use, data privacy, ethical principles, IPR and legal issues	advanced	<ul style="list-style-type: none"> <li>[b] Can summarise and explain ethical principles and responsible data use (e.g. CARE, indigenous data).</li> <li>[b] Can describe legal issues around data use and management (e.g. licences, patents, policies, contracts etc.).</li> <li>[i] Can analyse if ethical principles or legal issues play a role in their own work.</li> <li>[a] Can detect ethical or legal issues and solve them together with ethical and legal experts like e.g. ethics committee, data protection officers or lawyers from the institution.</li> </ul>
Data quality management, best practices and frameworks, data quality metrics	advanced	<ul style="list-style-type: none"> <li>[b] Can summarise best practices ensuring data quality.</li> <li>[i] Can describe how to recognise quality data.</li> <li>[a] Can use best practices and frameworks on their own data to ensure their quality.</li> </ul>

Topic	Required level	Learning outcomes [b]=basic, [i]=intermediate, [a]=advanced
Trusted data repositories and certification	intermediate	[b] Can explain what a trusted data repository is and how to find it (re3data.org and FAIRsharing). [b] Can compare different certifications for data repositories (e.g. CoreTrustSeal, CLARIN certification). [i] Can discover trusted repositories and identify those that are certified. [a] Can use a trusted repository to share research output.
Data discovery (published data), data selection and use in research	advanced	[b] Can explain the importance of data discovery and reuse. [i] Can discover published datasets in their discipline. [i] Can cite data. [a] Can develop a strategy to search for data. [a] Can articulate criteria for data selection. [a] Can extract datasets and build their own work on them.
Research data lifecycle	intermediate	[b] Can explain the steps of the research data lifecycle. [b] Can compare different lifecycle models. [i] Can apply the research data lifecycle on their own work.
Ontologies, controlled vocabularies	advanced	[b] Can explain the role of ontologies and vocabularies (Knowledge). [b] Can recognise the use of ontologies and vocabularies (Knowledge). [b] Can identify a few domain-relevant ontologies (Knowledge). [b] Can search and find terminologies in registries. [i] Can use ontologies to describe resources (Apply). [a] Can use ontologies for search and analysis (Apply).



## 4 – Teaching and training designs for FAIR

### 4.1 Introduction

FAIR has attracted considerable interest in higher education and research circles. Teaching FAIR can be positioned in the broader discussion about advancing data literacy (see figure 1 below; for more detail on information literacy for higher education, see ACRL 2015). Moreover, teaching FAIR is increasingly important since the FAIRsFAIR D7.1 survey (Stoy et al. 2020) has shown that courses on data handling (i.e. data analysis and/or scientific programming) rarely cover core FAIR topics like metadata standards, persistent identifiers and provenance.

This chapter introduces a structured approach to course design and does not serve to explain curriculum theory (for more information on course design, see Via et al. 2020). The various steps to help teachers and trainers design FAIR courses include articulating the importance of learning outcomes (see also chapter 3) for various audiences, taking into account the complexity of learning and its different levels, and comparing different forms of training delivery (also referred to as training experiences).

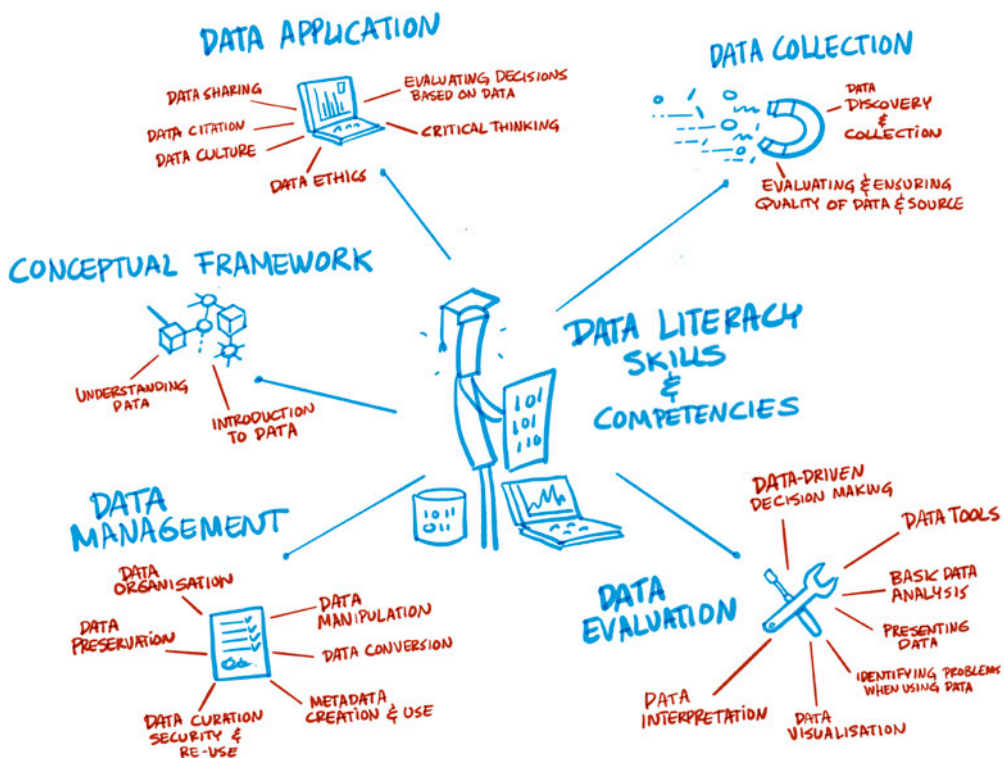
What these steps will help you with (based on FOSTER, n.d.):

- Integrating FAIR into your teaching: the lesson plans in chapter 5 and the didactical approaches in this chapter help you incorporate current FAIR data practices into your own teaching without having to organise a separate course for them (but they also allow you to offer a full course on FAIR data if you wish to).

- Stimulating FAIR data by design/practices: by using the good practices of this chapter and chapter 5, you can stimulate FAIR awareness and practices/workflows among your students, as well as staff members at your organisation who are involved in implementing the FAIR principles at the institutional level.
- Stimulating reuse: this chapter encourages the reuse of existing resources and learning activities, while allowing you to add your own examples.

After having read this chapter, as a teacher you should be able to:

- explain the benefits of learning FAIR;
- find new ideas for activities by learning from existing practices (see also chapter 5);
- encourage active learning using hands-on activities (see also chapter 5);
- help your students (or other persons whom it may concern) become aware of the FAIR principles and increase their FAIR data literacy; and
- help your students (or other persons whom it may concern) to use open resources combined with disciplinary theories and models.



**Figure 1:** Schematic representation of data literacy skills and competencies by Patrick Hochstenbach, based on Guler (2019, p. 15), originally adapted from Ridsdale et al. (2015, p. 38).

Before thinking about and working on the structure and content of a course or learning programme, it is important to take the target audience into consideration, e.g. researcher-facing vs. undergraduate student-facing. Identifying their needs, previous knowledge and existing skills with regard to RDM and the FAIR principles, as well as the gaps that need to be addressed is a crucial step for a successful course. Step 2 in chapter 4.2 suggests a number of measures that can be taken in this regard.

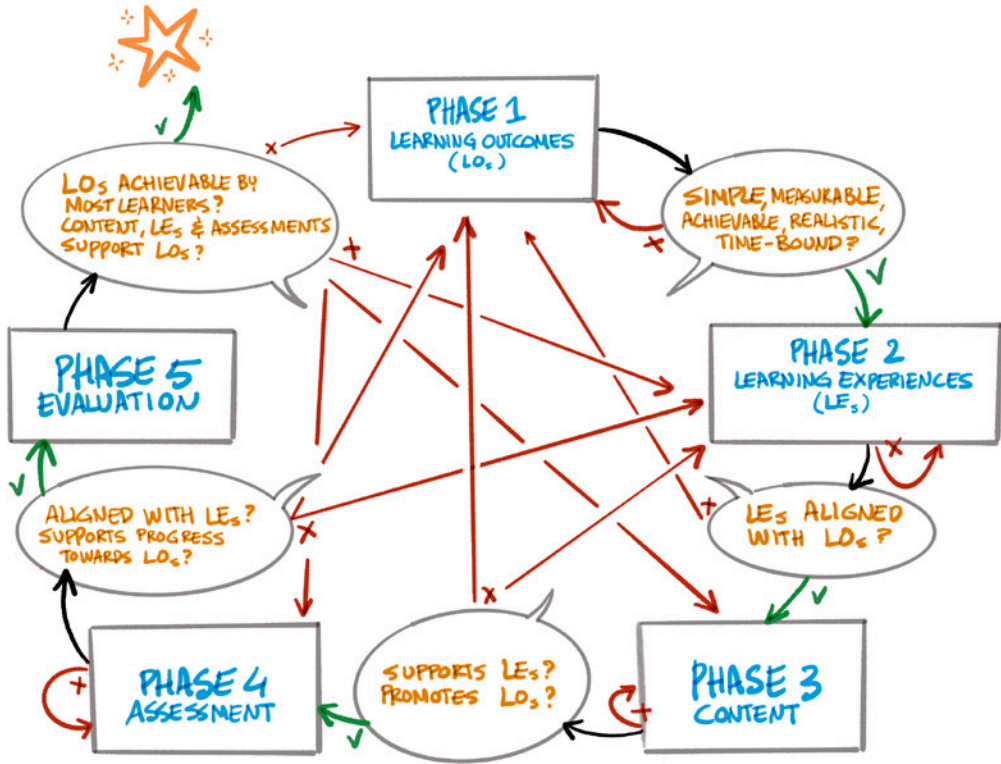
## 4.2 Elemental phases in course design

Once the needs and gaps of learners have been identified, the next steps of the course design can follow. To help teachers and trainers with this, we introduce Nicholls' paradigm for curriculum development, summarised by Via et al. (2020, adapted from Tractenberg et al. 2020) into five elemental phases (see also figure 2 below).

1. Select or identify learning outcomes (LOs):
  - “Learning Outcomes (LOs): the knowledge, skills and abilities (KSAs) that learners should be able to demonstrate after instruction, the tangible evidence that the teaching goals have been achieved; LOs are learner-centric” (Via et al. 2020, p. 2, emphasis omitted).
2. Select or develop learning experiences (LEs) that will help learners achieve the LOs:
  - “Learning Experience (LE): any setting or interaction in or via which learning takes place: e.g., a lecture, game, exercise, role-play” (ibid.).
3. Select or develop content relevant to the LOs.
4. Identify or develop assessments to ensure learners progress toward LOs:
  - “Assessment: the evaluation or estimation of the nature, quality or ability of someone or something” (ibid.).
5. Evaluate the course effectiveness.

Ideally, following these steps will help teachers to create an effective learning path for their intended learners. A learning path describes the chosen route, or a set of independent learning modules, taken by a learner through a range of courses or other training events. A learning path can also consist of independent training events by learners who only need to fill specific gaps. Practical implementation of this approach should include specification of the prerequisites or entry knowledge requirements

and may include an entry knowledge assessment to track the learners' progress and achievements at the end of the course.



**Figure 2:** Nicholls' phases of curriculum design & their dependencies by Patrick Hochstenbach, adapted from Via et al. (2020, p. 4). The rectangles show the key considerations in each phase. Red arrows represent revisions in the event that requirements resulting from the considerations have not been met yet, while green arrows depict a move to the next phase. If all requirements have been satisfied, the course or curriculum can be regarded as successful (represented by the star in the upper left).

These five steps are elaborated below, not so much to explain a curriculum development theory but to help integrate FAIR in teaching, stimulate FAIR data by teaching it, and enhance reuse of existing teaching materials on the topic of FAIR (for the latter, see particularly chapter 5).



## Step 1. Select or identify learning outcomes (LOs)

Learning outcomes are the starting point and driver of decision-making when developing training and teaching (see also Via et al. 2020). They are a reflection of the desired state and describe the overall purpose of participating in an educational activity. Via et al. (2020, p. 4) note a number of features that must be considered when developing measurable learning outcomes:

- be specific and well-defined;
- be realistic;
- rely on active verbs;
- focus on learning products, not the learning process<sup>6</sup>;
- be simple<sup>7</sup>;
- be appropriate in number; and
- support assessments that generate actionable evidence.

To summarise: Learning outcomes should be based on competences that learners gain or improve, and should be formulated from the learner's perspective. They describe a specific action (either practical or cognitive) on a specific level (knowing what vs. knowing how). In other words, they describe what learners can do after having attended the unit, course or module. When writing a FAIR module description or workshop announcement, it may make sense to include how learning will be achieved (this part is more about the content), and why (this part is more about the incentives).

A helpful tool when formulating learning outcomes are taxonomies like the taxonomy of educational objectives by Benjamin Bloom (known as Bloom's taxonomy or BT) which defines cognitive levels of learning outcomes (Bloom et al. 1956), along with its revised version by Andersen and Krathwohl (Andersen and Krathwohl 2001) which provides suggestions for using actionable verbs to describe learning outcomes. A common practice is to define learning outcomes on different levels and with different granularity, e.g. for a whole course, a specific session, part of a session, macro and micro-goals. As a general rule, one session might have around 3 to 5 individual learning outcomes (this can be discussed and adapted to the given context, but it is important not to aim for more than can be achieved in the time available).

On a more generic level, the following learning outcomes could, for instance, be formulated using the verbs of Bloom's taxonomy to make learning outcomes actionable:

- Students can recognise and define the FAIR principles.
- Students can explain and interpret the FAIR principles.

---

<sup>6</sup> This means to focus on what the learner will be able to do after the instruction (as opposed to what will be done during the instruction).

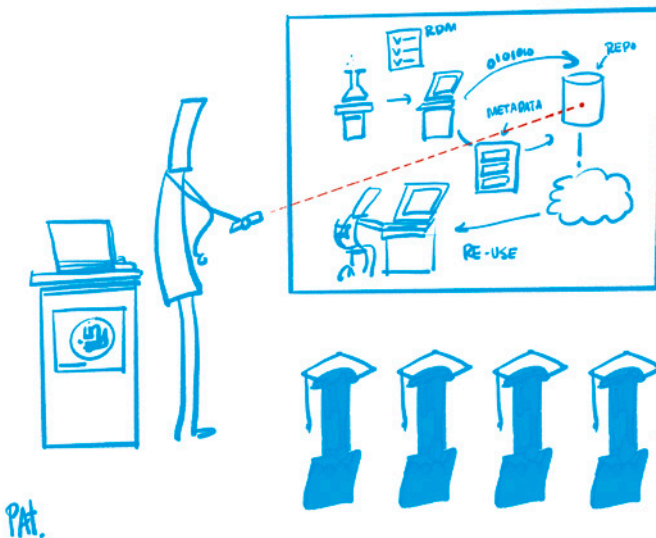
<sup>7</sup> This means not combining several pieces of knowledge, skills, or abilities in one learning outcome.

- Students can apply the FAIR principles.
- Students can analyse and critically discuss the FAIR principles.
- Students can evaluate commonly used data repositories in terms of their compliance with the FAIR principles so they can use them in their field of research.

Furthermore, learning outcomes may be formulated on a more granular level, e.g.:

- for each of the FAIR elements;
- how these FAIR elements relate to the different stages of the data or research lifecycle; and
- for different learning levels (beginner, intermediate, advanced).

For more detailed learning outcomes, see chapter 3.



## Step 2. Select or develop learning experiences (LEs)

Below is a list of learning experiences commonly used in teaching and training based on Via et al. (2020) and our own experiences.

Selecting the right learning experiences, i.e. the most suitable setting or environment for a specific learning activity or process, is not a straightforward thing to do. You need to tailor the methods used to the time available, along with the experience and skills of the target group and their expectations. If the course is part of a curriculum, students are unlikely to challenge the need for training. In this case, you can concentrate on thinking about how to get participants to learn. Informal training is often needed when looking to develop and enhance the skills of staff members. There

may be a whole host of reasons why people choose to attend informal training events. Therefore, tailoring relevant and directly applicable materials to meet participants' day-to-day research activities is a great way to motivate them. By offering different types of teaching or training, you as a teacher or trainer will learn what works best for different groups of learners over time.

FAIR training could be delivered as part of a formal course, part of a training or promotional event, or it can be embedded in managerial processes, e.g. grant application support, ethical review process, or basic training for new affiliate researchers. It could also be a lecture, a workshop, a series of events, an online course, self-learning materials, or training interventions.

It is easier to meet the expectations of students if you know what kind of understanding they already have about FAIR. If possible, try to get to know your course participants before or at the beginning of the training. This can be achieved by pre-tasks, a self-assessment survey, a poll or a discussion. If there are participants with pertinent prior knowledge, you can make use of that during the training.

No matter what type of teaching and training you choose when implementing FAIR in your institute, it is crucial to stay abreast of relevant local/regional resources that are available to your stakeholders to meet their day-to-day research needs and to be compliant with policies and regulations.

## Lectures

Lecturing as a traditional form of teaching/training is an effective way to provide basic information about the topic. Lectures can be recorded and used as flipped classroom<sup>8</sup> material combined with an interactive workshop. Starting a lecture with researchers describing their experiences, how they have implemented the elements of FAIR in their work, or a typical researcher's most urgent questions about data handling will help to engage the audience from the beginning. Basic concepts of a topic can be communicated effectively through brief lectures. Due to the far-reaching goal of FAIR, instructors should anticipate many questions from the audience, in turn making it good practice to include discussions, other activating methods and hands-on exercises after the lecture to consolidate the key points of learning. It is important to stress the role of FAIR in terms of good research practice, but it should be made clear that it is not always feasible to implement all aspects of FAIR to their fullest extent.

**Pros:** A lecture is a great delivery format for experienced and motivated learners where instructors can maximise content delivery in a dedicated time frame. Going beyond a dedicated lecture on FAIR, with a bit of planning, instructors may be able

---

<sup>8</sup> In a flipped classroom setting, students acquire basic knowledge about a new topic by self-study at home, e.g. by watching online lessons or reading textbooks, while in class, the focus is on the practical application of this knowledge (see [https://en.wikipedia.org/wiki/Flipped\\_classroom](https://en.wikipedia.org/wiki/Flipped_classroom)).

to fully incorporate FAIR teaching in any existing course, e.g. an introduction to research methods.

**Cons:** It can be time-consuming in the course design phase to incorporate relevant materials into a course without overloading information for learners. Learner engagement is key.

## Workshops

Workshops can be organised around a certain FAIR topic, or they can be more general in scope. By way of example, a ‘What should I know about FAIR’ workshop can allow participants to discuss what FAIR means to them. In a ‘Where should I deposit my data to be FAIR’ workshop, participants can choose a repository and deposit a dataset. In a ‘How to write a DMP’ workshop, participants can write their own DMP. Workshops can also focus on a research method where you can embed tasks involving FAIR, such as local institutional data storage options, documentation, and file naming conventions.

Arranging a workshop gives you an opportunity to find out and discuss the main questions or problems your target audience has concerning FAIR. Organisers can also provide standard offerings of FAIR workshops that will be repeated every year and plan for add-on workshops that would vary from year to year to meet the specific needs of the audience.

**Pros:** Workshops are ideal for delivering content on a single topic or to a specific target audience. They are short and easy to organise, with great flexibility in modifying materials to meet the different needs of different audiences, e.g. researchers vs. entry-level graduate students.

**Cons:** It is almost impossible to cover all FAIR topics in one single workshop. Therefore, teachers or training providers can design and conduct a workshop series covering various FAIR topics. Sometimes, learners might miss out on important topics covered in individual workshops due to self-selection biases, e.g. I only attend the workshops I deem interesting, or because of time constraints. Making connections from one workshop to another with brief recaps or highlighting key points of past and future workshops will be a useful strategy to promote full training in FAIR.

## Events

Your audience may not know about the FAIR principles. A good way to influence these types of audiences is to raise awareness with brief presentations at the events they already participate in, e.g. unit meetings, events of the faculty, newcomer events at the university, and all kinds of Open Research events. FAIR can also be a topic of coffee lectures or working lunches.

Take advantage of opportunities to reach your audience in a motivated state. For example, if a funder requires FAIR data, try to get a time slot at an event organised by the funder to explain what FAIR means. Funders are generally happy to accommodate this type of collaboration.

**Pros:** Outreach events are most suitable for promotional purposes. They are usually concise and provide a great opportunity to make allies of those willing to advance the FAIR agenda.

**Cons:** Time is often limited at outreach events. The messages about FAIR you want to convey must be clear and concise. They will be ideal ways to provide information about future training offerings, or to direct attendees to self-learning materials.

### Online courses

Online courses are a convenient way to organise training for a large number of participants or for participants from many locations. They can be taught fully online without any live interactions (i.e. asynchronous online learning), as a course where live training is given (i.e. synchronous online learning), or as a combination of the two.

**Pros:** Online courses, particularly in the form of asynchronous learning, might suit the needs of many busy learners who would appreciate a flexible format where they can take the course independently and at their own pace. Updates and adjustments to materials in common online course Learning Management Systems (LMS) are easy to manage with minimal impact on learner experiences.

**Cons:** The risk of losing learners is very high in online courses (i.e. high enrolment rate but low completion rate). While traditional courses usually retain about 80% of students (Atchley et al. 2013), the median completion rate for large-scale online courses (i.e. Massive Open Online Courses) is about 13% (Jordan 2015). This is partly due to the lack of live interactions and low engagement with the course materials (Muljana et al. 2019). An easy remedy for this could be to make part of your online course synchronous by providing weekly or fortnightly live office hours. Using interactive learning content (e.g. <https://h5p.org/>) embedded in the LMS will also facilitate retention of the learner's interest.

### Self-learning material

Self-learning material is an important part of any training format. This material is a reference for learners to consult as a recommended information source after a course or event. Self-learning material can also be used separately to acquire the basics of FAIR or to check a certain fact. This can include fact sheets, short instructional videos, quizzes to check the level of knowledge, and links to university guidelines and policies. It might be handy to have some instructional print materials, such as flyers

and fact sheets. You can use self-learning materials created by other parties, but each higher education institute should still have a clear starting point for its students and researchers on how to follow the FAIR principles at the organisation and where to get help.

When creating self-learning materials, extra attention is needed to organise the content to make it easy for users to browse and find the information they are looking for. The inventory of self-learning materials will grow over time, making it essential to provide users with a clear table of contents or a glossary.

**Pros:** Self-learning materials can be used and referenced in conjunction with other training formats, such as a workshop or an outreach event. They can be used as references not only by learners but also by teachers, trainers, and research support staff, e.g. grant officers who need to access DMPs for grant applications.

**Cons:** Self-learning materials are a rather passive learning experience, making it difficult to track learning progress and outcomes. Many learners will fall into a scenario where ‘I will look at it later’ means ‘Never’. Since it is relatively easy to produce and compile a large number of self-learning materials, self-learning materials could very quickly become a mess by not paying proper attention to the organisation and by not keeping information up to date, in turn creating difficulties for learners trying to find and access relevant information.

### Training interventions<sup>9</sup>

In higher education institutions, we may face situations where the level of our stakeholders’ knowledge about the FAIR principles does not meet their everyday research needs. For instance, when reviewing a data management plan, we may encounter a clear knowledge gap, and these situations might be the right entry points to provide specific/customised information about FAIR and start a discussion with the aim of promoting FAIRness in data management, not only to meet grant application and policy requirements, but also to improve the research workflow. Connecting local services, e.g. upcoming workshops or self-learning materials, to the researchers could be an effective way to address the knowledge gap.

**Pros:** Identifying knowledge gaps and providing locally available resources to address these gaps on a one-on-one basis is a great way to keep in touch with the research community and to effectively meet stakeholders’ needs.

---

<sup>9</sup> Definition: “Having perceived that the individual has short-fall in [their] output, and that it is expedient that [they] perform[...] at optimal level, training activity is undertaken by the individual in order to equip [them] with the wherewithal for performance at the required level. In other words, training is provided for the individual, to ‘salvage’ [them] from steady downward performance. This is referred to as ‘Training Intervention.’” (Abdul 2015, p. 108)

**Cons:** This service model operates on a case-by-case basis, which could prove to be a time-consuming task if you need to reach all the stakeholders in your institute. This could render the service model unscalable, especially if you are operating with a very small service provision team.

**Table 6:** Overview of advantages and disadvantages of different forms of teaching and training delivery

Type of learning experience	Pros	Cons
Lectures	<ul style="list-style-type: none"> <li>• Great delivery format for experienced and motivated learners;</li> <li>• Possibility to fully incorporate FAIR content in any existing course.</li> </ul>	<ul style="list-style-type: none"> <li>• Could be time-consuming;</li> <li>• Learner engagement is key.</li> </ul>
Workshops	<ul style="list-style-type: none"> <li>• Ideal for delivering single topic content or for a targeted group;</li> <li>• Flexible, short and easy to organise.</li> </ul>	<ul style="list-style-type: none"> <li>• To cover all topics of FAIR, a workshop series may be required;</li> <li>• Learners might miss out on important topics due to self-selection biases.</li> </ul>
Events	<ul style="list-style-type: none"> <li>• Most suitable for promotional purposes;</li> <li>• Great for networking;</li> <li>• Ideal to provide information for follow-up services and/or direct links to access existing materials.</li> </ul>	<ul style="list-style-type: none"> <li>• Limited time;</li> <li>• Messages need to be clear and concise.</li> </ul>
Online courses	<ul style="list-style-type: none"> <li>• Flexible, self-paced, independent.</li> <li>• Easy to manage and update at the back-end with minimal impact on learner experiences.</li> </ul>	<ul style="list-style-type: none"> <li>• Need to be aware of issues of student retention in asynchronous learning;</li> <li>• Lack of live interactions and low engagement;</li> <li>• Could include weekly or bi-weekly live office hours to include partial synchronous learning;</li> <li>• Use interactive course content to facilitate learner engagement.</li> </ul>

Type of learning experience	Pros	Cons
Self-learning material	<ul style="list-style-type: none"> <li>• Could be used and referenced in conjunction with other training formats;</li> <li>• Could be used by learners, teachers, trainers, and research support staff as references.</li> </ul>	<ul style="list-style-type: none"> <li>• Passive learning;</li> <li>• Not easy to track learning progress and outcomes;</li> <li>• Need to pay attention to the organisation and maintenance of the most up-to-date information in self-learning materials.</li> </ul>
Training interventions	<ul style="list-style-type: none"> <li>• Great way to keep in touch with the research community by directly addressing their knowledge gaps.</li> </ul>	<ul style="list-style-type: none"> <li>• Case-by-case consultation;</li> <li>• Might be time-consuming.</li> </ul>

### A hybrid model

When planning teaching and training strategies for FAIR, service providers might need to count on resources and collaborations from different units within the institution, while also making use of institutional, local, regional, national and/or international resources, and forging alliances with those willing to maximise the impact of the FAIR teaching and training. Below is a simplified hypothetical hybrid plan to implement FAIR teaching and training strategies using the different delivery formats mentioned above:

With the joint efforts of the *Office of Research and Innovation and the University Library*, University M implements an independent self-paced learning programme (**online courses**) using the existing university course management system (Moodle) to provide general training on FAIR principles along a typical research lifecycle. At the same time, the University Library complements this self-paced online learning programme with a series of hands-on **workshops**, spanning one academic year, to provide more tailored and focused training on domain/discipline-specific topics. All relevant training materials can be downloaded and used as **self-learning materials**. Both the learning programme and the library workshop materials are centralised in the institutional file repository and maintained jointly by the *Office of Research and Innovation and the University Library*.

Outreach/Awareness **events** are organised in conjunction with new faculty onboarding meetings as well as with student orientations. Representatives from the *Office of Research and Innovation and the University Library* are also present in certain monthly faculty meetings to promote various service offerings to researchers. Given that the independent self-paced learning programme capitalises on the convenience



of the university's course management system, materials in Moodle for the online learning programme can be easily transferred to other courses within University M for instructors and lecturers to use in their own **lectures** and curriculum in order to reach a much broader audience at the university. Course instructors and lecturers are all invited to contribute back to the online learning programme where appropriate. Representatives and liaison librarians from the University Library can also provide **short lecture services** for instructors, lecturers and research centres who would like to promote FAIR in their own courses or research units.

### **Step 3: Select content relevant to the learning outcomes**

The content of a course is the specific subject it covers. As FAIR encompasses a wide range of sub-topics, it needs to be broken down into individual content blocks, such as 'copyright law', 'metadata' or 'data repositories'. For a more comprehensive list, see chapter 5, which can be used as a source of inspiration you can blend with your own course formats. Your choice of content and teaching format will of course depend on your audience and the time available.

When teaching the FAIR principles – as with most other topics – there is a very real danger of cramming too much content into too little time. Consequently, you should drop all content not aligned with the learning outcomes. If you identify content you deem essential but does not support the learning outcomes, e.g. an existing institutional data policy, you should adapt the learning outcomes accordingly. This also ensures that the content is aligned with the learning assessment and course evaluation covered in the following two steps.

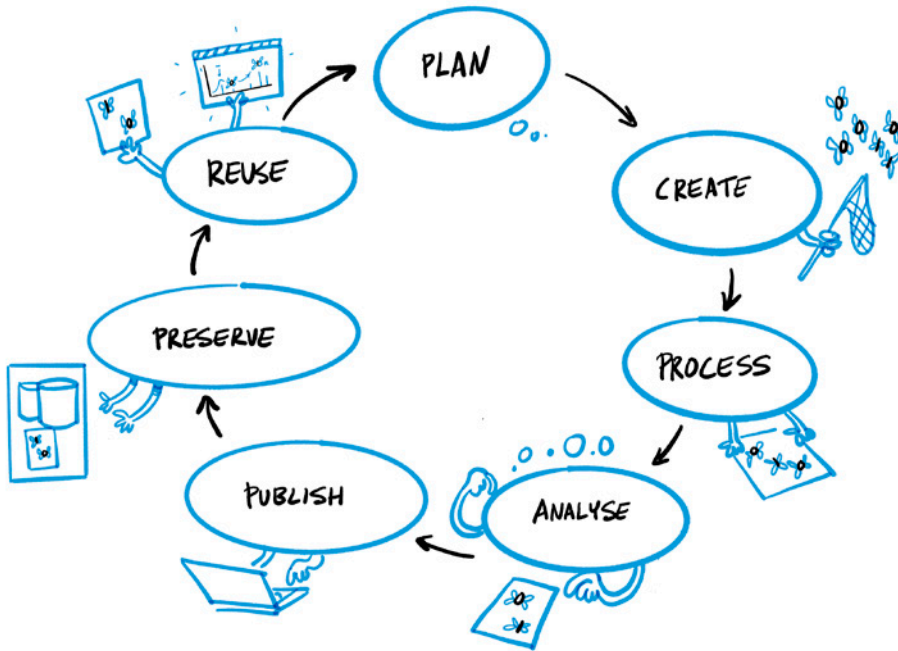
You will probably concentrate on a specific aspect of FAIR during a talk or workshop. If, however, your course covers all FAIR-relevant topics, there are several ways to organise and connect the individual content blocks:

#### **1. Follow the FAIR acronym**

Topics may be presented in the order in which they appear in the FAIR acronym: findable, accessible, interoperable, reusable. This approach makes most sense if the course's main topic is the FAIR principles from a generic or disciplinary perspective (e.g. Martinez et al. 2019). However, as several sub-topics, e.g. metadata, apply to more than one principle, and given that it is usually helpful to build on students' existing knowledge, you should also consider using one of the other three approaches instead. If you do in fact opt for one of the other approaches in your course, we recommend you include a special learning unit on FAIR in your overall curriculum to link topics with the four key FAIR principles.

## 2. Follow the research data lifecycle

The research data lifecycle<sup>10</sup> provides a generalised, structured look at the individual steps of how research projects handle research data. While this is clearly an idealised model, it has proven useful in teaching RDM, particularly when writing data management plans (DMPs).



**Figure 3:** Research Data Lifecycle by Patrick Hochstenbach, adapted from UK Data Service, n.d.

The process starts with a research question and selection of possible approaches. Ideally, this early stage involves an exploration of existing data to see what can be reused (in part) and encompasses every aspect of FAIR here. After drafting how data will be managed (ideally supported by a data management plan), data are collected, stored, described and analysed. Selection of the data to be preserved for the long-term depends upon a number of conditions (ethical and legal restrictions, plans for further use, hardware costs, etc.). After these steps, the data can be prepared for publication and possible reuse by others, or it can serve as input for a future project.

<sup>10</sup> Due to different disciplines and contexts, there is a large variety of such models (see, for example, Ball 2012).

### 3. Link FAIR practices to data management plans and planning

A data management plan (DMP) provides guidance throughout the whole research data management process and outlines how the data relevant for the research question will be retrieved, collected, described, stored, processed, analysed, preserved for the long term, and published.

DMPs cover all core aspects of the FAIR principles. As a result, following the topics of a DMP template, e.g. Science Europe 2021, is a sound approach, especially if the motivation for the course is a requirement to deliver a DMP, e.g. for a funder, or if the course requirement is to write an individual DMP. Furthermore, a DMP, when treated as a ‘living document’ which the researcher comes back to from time to time during a project, can serve as a powerful tool to stay organised during the research process.

### 4. Connect topics in a way that fulfils individual needs

Depending on learners’ existing knowledge and individual needs, content can also be ordered in other ways. This is especially relevant if the overall course has a specific topic, e.g. ‘metadata’, that you also want to present *in situ* and with relevance to the overall FAIR landscape.

#### **Step 4: Identify or develop assessments to ensure the learning is progressing towards learning outcomes**

Developing appropriate assessments for teaching and training strategies, e.g. a workshop or an online course, is an important step for any successful and sustainable training programme. It will not only help to improve learners’ experiences but will also aid instructors in improving and updating content (Via et al. 2020). As illustrated in table 7 below, different assessments can be conducted at different levels and serve different purposes.

**Table 7:** Approaches to assess progress towards learning outcomes

Assessment goals	Participants	Stakeholders	Example
<ul style="list-style-type: none"> <li>• To identify learning progress of learners.</li> <li>• To evaluate whether learners can apply their FAIR knowledge readily in a hands-on situation.</li> </ul>	Learners	Learners & Trainers	<p>Ask learners to draft a DMP for their projects. This could be for an ongoing project or an example project.</p> <p>Work with learners on the DMP to identify immediate knowledge gaps and provide relevant feedback and resources to learners via the drafted DMP.</p> <p>Instructors may use resources such as DMP evaluation guidelines to facilitate the evaluation of the learners' DMPs, e.g. Tuuli Working Group (2021), Donaldson et al. (2017).</p>
<ul style="list-style-type: none"> <li>• To collect information from learners on what they like or dislike about the content and format of delivery.</li> <li>• The ultimate goal is to improve educational experiences for future learners.</li> </ul>	Learners	Trainers	<p>Teaching and training evaluation surveys are often used in this context.</p> <p>Teaching and training programme developers might want to consult their institution's teaching and learning services (or equivalent) when preparing suitable evaluation survey questions.</p> <p>Shorter training usually requires shorter evaluation surveys or a simple 3-2-1 assessment at the end of the training to ask learners to identify 3 things they have learned, 2 things they want to know, and 1 question they want to ask (Via et al. 2020).</p> <p>Samples of course evaluation questions for a full course or a comprehensive learning programme (consultation with your local teaching and learning services is recommended):</p> <ol style="list-style-type: none"> <li>1. UC Berkeley: <a href="https://teaching.berkeley.edu/course-evaluations-question-bank">https://teaching.berkeley.edu/course-evaluations-question-bank</a></li> <li>2. McGill University: <ul style="list-style-type: none"> <li>• English: <a href="https://www.mcgill.ca/mercury/files/mercury/course-evaluation-questionnaires-en-final.pdf">https://www.mcgill.ca/mercury/files/mercury/course-evaluation-questionnaires-en-final.pdf</a></li> <li>• French: <a href="https://www.mcgill.ca/mercury/files/mercury/course-evaluation-questionnaires-fr-final.pdf">https://www.mcgill.ca/mercury/files/mercury/course-evaluation-questionnaires-fr-final.pdf</a></li> </ul> </li> </ol>

Assessment goals	Participants	Stakeholders	Example
<ul style="list-style-type: none"> <li>• Primarily for administrative purposes.</li> <li>• To perform programme evaluation at the institutional level.</li> <li>• To budget and plan for resources.</li> </ul>	Trainers	Trainers and Institutions	Sample key performance indicators (KPIs): <ul style="list-style-type: none"> <li>• Enrolment rates and completion rates of a specific programme</li> <li>• Learner satisfaction survey</li> <li>• Working hours used</li> </ul>

### Step 5: Evaluate course effectiveness

The final step is to evaluate whether the course guided learners to the learning outcomes defined initially. The results of this evaluation will help identify problems with the course design and allow for adjustments to improve course effectiveness in future iterations. If time and resources permit, it is good practice to pilot the first versions of your course and allow for incorporation of quick feedback and modifications shortly after launch.

Therefore, the evaluation needs to be *actionable*, i.e. it needs to be able to inform decisions.

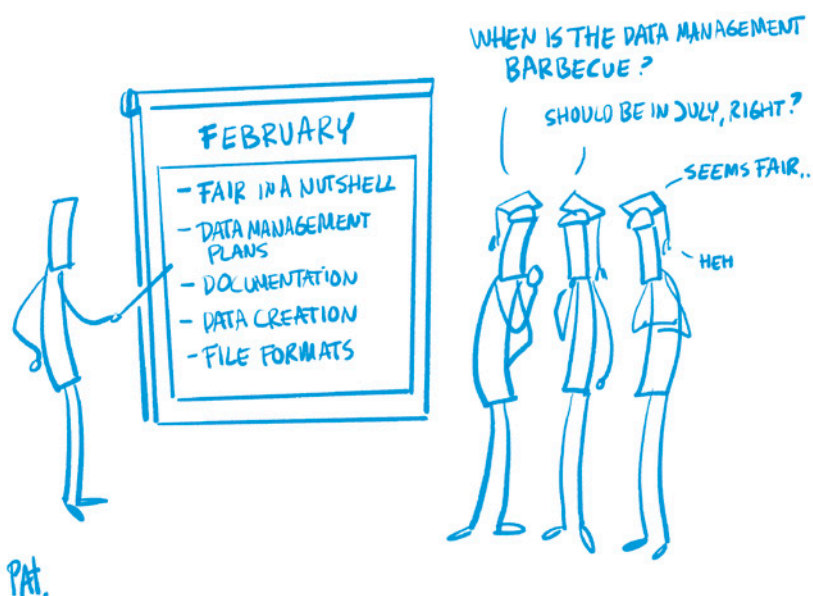
For longer courses with a full curriculum, it can be straightforward to define reliable metrics for course effectiveness. By way of example, course evaluation for a full semester of student seminars (Wiljes and Cimiano 2019) can be built on the study requirements that students must meet in order to receive credit points. Writing an individual DMP as a seminar paper provides a sound basis for evaluating whether students have acquired the knowledge, skills and abilities as defined by the learning outcomes. In addition, this allows you to identify problems with specific topics and narrow them down to the specific methods (i.e. learning experiences) that were used. To give an example, if ‘metadata’ is presented as the topic of a talk and the final evaluation of students’ DMPs reveals that they are not able to apply the content of the talk properly, another teaching method should be trialled instead. You could, for instance, provide students with a specific metadata standard and have them work out on their own how to apply it. Biernacka et al. (2020) provide examples of teaching methods for a wide variety of RDM/FAIR topics.

With shorter courses, e.g. a 4-hour workshop, evaluating course effectiveness is more challenging. We recommend leaving enough time for students to write down and ask questions. A lively discussion is generally a good sign that students are progressing.

To some extent, the metrics provided in step 4 to assess learners' progress can also be applied to evaluate overall course effectiveness. However, you should note that these metrics may also need to be improved upon iteratively.

Conducting an anonymous survey on student satisfaction can complement an evaluation of course effectiveness. However, this should be interpreted with care because student satisfaction may be influenced by factors other than successful learning (Denson et al. 2010). In addition, students are biased in evaluating how their own skills and knowledge improve (Dunning et al. 2004; Karpen 2018).

## 5 – FAIR lesson plans



While chapter 4 introduced an approach to developing FAIR courses and elaborated on a number of relevant considerations in this respect, this chapter provides examples of lesson plans for a number of topics related to RDM and the FAIR principles. The following list of lesson plans is not exhaustive and can be updated.

All lesson plans follow the same format<sup>11</sup> which includes the FAIR elements concerned, the learning outcomes, a summary of tasks/actions, material/equipment needed, references and take-home tasks. More details on the implementation of FAIR aspects, i.e. the practical application of the content taught through the lesson plans, are provided in chapter 6.

<sup>11</sup> The lesson plan template used here is based on this template: [https://www.class-templates.com/support-files/lpt\\_word\\_001-printable\\_lesson\\_plan\\_template.pdf](https://www.class-templates.com/support-files/lpt_word_001-printable_lesson_plan_template.pdf)

List of lesson plans on RDM- and FAIR-related topics:

- Lesson plan 1: FAIR in a nutshell
- Lesson plan 2: Data management plans (DMP)
- Lesson plan 3: Documentation
- Lesson plan 4: Data creation
- Lesson plan 5: File formats
- Lesson plan 6: Metadata
- Lesson plan 7: Data standardisation and ontologies
- Lesson plan 8: Persistent identifiers (PIDs)
- Lesson plan 9: Licences, copyright and intellectual property rights (IPR) issues
- Lesson plan 10: Finding and reusing data
- Lesson plan 11: Repositories
- Lesson plan 12: Dealing with confidential, personal, sensitive & private data and ethical aspects
- Lesson plan 13: Data access
- Lesson Plan 14: FAIR software/citable code
- Lesson plan 15: Research data management – overview and best practices
- Lesson plan 16: Data management and governance in industry and research







## 6 – Implementing FAIR



## 6.1 Introduction

Researchers cannot be left alone to do the heavy lifting in data management according to FAIR principles; they need to rely on support services provided by their institutions. In view of this, this chapter shifts the perspective from the individual researcher or research projects to the institution: How can they support their researchers with FAIR data management? What support services are necessary, what infrastructure needs to be put in place, and what policies need to be enacted? Each section in this chapter links back to the lesson plans to connect this institutional overview with the details provided there.

It should be noted that this chapter focuses on the requirements and measures to be taken within an institution. FAIRness is a global and institutional goal. A large amount of research is done in cooperation with external parties. This should be reflected by incorporating respective elements in, e.g. policies or data sharing agreements, but covering such points extends far beyond the scope of this handbook.

## 6.2 Arriving at FAIR institutional policies

Adopting an institutional research data policy that embraces the FAIR principles can result in recognition, energy and resources for the implementation of good practices since FAIR implementation requires reshaping and alignment of existing policies. This section looks at key stakeholders and ways to cultivate an institution-wide FAIR research data environment.

### Research data in the institutional policy framework

Institutional policies underpin staffing and resource allocation, approaches and workflows, and can enable and support (or hinder) new practices. Therefore, implementing the FAIR principles for research data at the institutional level needs a review of existing policies to remove potential stumbling blocks and adoption of research data policies with the aim of embracing FAIR.

An institutional policy commitment to the FAIR principles can strengthen policies and efforts in safeguarding **research integrity**, and should thus be included in policies related to institutional research data. Moreover, institutional commitments to **Open Access** or **Open Research** in general can also be bolstered by references to FAIR principles.

A great push for adopting FAIR principles at the institutional level stems from the fact that more and more funders are embracing FAIR as a requirement for their grants. Institutional policies can help to navigate conflicting interests in collaborative research projects. One such example would be to point out the benefits of FAIR data management to (potentially) sceptical industry partners in showing that the principles can be aligned with the need to protect commercially sensitive data. Some institutions may already have dedicated **research policies** in place for particular areas of research, e.g. for clinical research practices, either at the institutional or the departmental level. These existing policies should be checked for alignment with the FAIR principles as well.

Institutional policies regarding **data protection, research ethics**, commercialisation and **intellectual property rights (IP)** are sometimes seen to contradict or impede the implementation of FAIR for some research projects. Striving for FAIR data management can make the task of protecting personally identifiable data and any other sensitive data easier while maintaining the possibility to validate research results. Good (FAIR) data management enables greater control over data and supports a more targeted approach to achieve the aim of making research data ‘as open as possible, as closed as necessary’ (as outlined in the Programme Guidelines on FAIR Data Management in Horizon 2020). Institutional policies that need to restrict access to data for ethical, legal and commercial reasons can and should embrace the commitment to FAIR data management at the same time.

Research data might also be implicated in policies on **technical services**, e.g. cloud storage or repositories, **IT security (or cybersecurity)**, or in retention schedules of **record management**. It is important to engage with different policy owners from different units, e.g. IT or ethics, to develop a cohesive FAIR research data framework at the institutional level which also complies with applicable laws and regulations.

### **Influencing policymaking**

Writing and implementing institutional policies is a collaborative effort. Integrating FAIR principles into existing institutional policies, or developing a dedicated research data policy at an institution requires effective communication and networking with relevant stakeholders.

Understanding policy-making processes and workflows at the institution is the first step towards integrating FAIR in an institutional policy framework. Most institutions maintain a **central policy hub** and will have someone (an individual or a group of people) tasked with maintaining coherence between all institutional policies and ensuring currentness of all policies. Every individual policy will then have a primary owner tasked with maintaining the policy, supervising compliance and organising periodic reviews and a consultative approach for necessary updates. Ownership of

a policy is tied to a function. The owner of a Research Data Policy, for instance, could be the Data Steward, regardless of which individual currently holds the position. Each policy will also have a number of affected stakeholders whose interests need to be taken into account when proposing policy changes.

Typical steps to implement new or updated policies will involve:

1. identifying the relevant policy documents, their owners and relevant stakeholders;
2. understanding the interdependencies between policies and the procedures in place to implement or update them;
3. informal discussions with relevant stakeholders about the needs and benefits of new or updated policies. Understanding requirements and potential roadblocks;
4. proposing new policy statements (in new or updated policy documents);
5. consultations and discussions to reach a consensus with all stakeholders; and
6. policy owners forwarding the proposed changes (or new policies) for approval by senior management, such as the school council or senate.

Institutional setups vary widely and relevant stakeholders will go by various names. The following list therefore only provides a rough overview of potential stakeholders who might be involved in the policy implementation or update process:

**Research offices** monitor compliance with funder requirements and can be a key driver of institutional adoption of FAIR principles. Other involvements could include the provision of training and the enforcement of policies about research integrity.

**IT departments** offer a variety of support services relevant for research data that are governed by relevant policies and applicable laws and regulations. IT support services may include, but are not limited to, the provision of computers, servers and cloud storage, institutional repository hosting, and cyber and IT security maintenance.

**Libraries** often provide services supporting research data management. Sharing and publishing data are important aspects of Open Research. Other services libraries may provide include Open Access, repository support, DMP reviews, as well as RDM training and consultations.

**Ethics boards** need to approve a wide range of research proposals. Processes and procedures surrounding research data are key to gaining ethics clearance. Policies and procedures need to be aligned and integrated with the FAIR principles.

**Data protection offices** are concerned with implementing and safeguarding provisions laid down by applicable privacy laws and regulations, such as the EU General Data Protection Regulation (GDPR) and Canada's Personal Information Protection and Electronic Documents Act (PIPEDA). Data protection practices can and should be aligned with FAIR principles.

**Technology transfer offices** encourage and support researchers and their institutions with the commercialisation of research results by safeguarding intellectual property rights. Policies and procedures are in place to safeguard intellectual property rights. These policies can and should be aligned with FAIR principles to make data as open as possible and as closed as necessary.

**Departments, research centres and units, and individual researchers** are stakeholders in all research data-related policies. They might be the owners of some policies governing specific areas of research. It is a strategic advantage to have them as close allies for implementing or updating FAIR research data policies (Association of American Universities and Association of Public and Land-grant Universities 2021).

**Senior management** needs to formally put policies into effect and is ultimately responsible for maintaining alignment of all policies and organising review and update processes. In order to move towards institutional implementation of FAIR, senior management will need to recognise that research data are valuable assets of an institution, and that it is important to endorse FAIR principles to harness the ultimate value of research data.

### Resources:

Sample guides and perspectives on institutional approaches:

- Open Science and its role in universities: A roadmap for cultural change – Discussion and analyses on Open Research approaches at the university level, with recommendations on what universities can do to embrace Open Science principles, policies and practices.
- LEARN Toolkit of Best Practices for Research Data Management – 23 Best-Practice Case Studies from institutions around the world, drawn from issues in the original LERU<sup>12</sup> Roadmap.
- Association of American University’s Guide to Accelerate Public Access to Research Data – Discussions and recommendations on institutional strategies to advance public access to research data.
- Institutional policies are registered by FAIRsharing, making them discoverable and citable (see this example from the University of Oxford).

Learn more:

Lesson plan 9: Licences, copyright and intellectual property rights (IPR) issues

Lesson plan 12: Dealing with confidential, personal, sensitive & private data and ethical aspects

Lesson plan 16: Data management and governance in industry and research

---

<sup>12</sup> LERU: League of European Research Universities

## 6.3 Data management planning

Data management plans (DMPs) can be used to ensure the quality and consistency of data management throughout the data lifecycle and are required by many funders. Responsibilities for data management lie with researchers or research teams, but institutions need to offer support with many of the issues raised in DMPs.

DMPs provide a list of topics that need to be considered to achieve FAIR data management. Researchers rely on a wide range of institutional support services to meet these requirements. DMPs usually include the following topics:

- **Data description and collection or reuse of existing data**  
Existing data from institutional repositories or digital data collections at the library can be made available for reuse. Support and guidance can be provided for data creation, collection, and description.
- **Documentation and data quality**  
Having local expertise and fostering good practice at departmental level is a good way to provide guidance on the multitude of standards and approaches.
- **Storage and backup** (during the research project) and **data sharing and long-term preservation** (at the end of a research project) Researchers will require support from IT departments.
- **Legal and ethical requirements and data management responsibilities and resources**  
Ethics boards, data protection offices, IP offices, legal and financial departments need to guide researchers in safeguarding these aspects.

Coordinating this support and aiding researchers with the planning process via training and consultancy are key tasks of institutional Data Stewards. Services can also include institutional participation in tools like DMP Online or DMP OPIDoR. These web-based services provide guidance for all criteria, offer sample plans, include DMP templates from multiple funding bodies, and allow researchers to work collaboratively on their plans.

DMPs are often described as living documents and should be updated according to changing circumstances.

Learn more:

Lesson plan 2: DMPs



## 6.4 Data processing and documentation

Data processing constitutes a key step in the data lifecycle and one that researchers must undertake to make data useful for analyses (Paine et al. 2015). Many scholars in information science and other fields point out that knowledge and understanding of the context of data creation are necessary to be able to analyse, share, and reuse data, e.g. Faniel and Jacobsen 2010. Initially, research data are often referred to as ‘raw’, meaning they are yet to undergo processing following their creation. However, Gitelman’s (2013) impressive edited volume ‘Raw data is an oxymoron’ emphasised that data are never raw and always already embody decisions. Embracing the FAIR principles helps to ensure that data processing decisions remain explicit and are documented.

There is a bewildering diversity of processes and practices that fall under ‘data processing’. Among other things, ‘processing’ can mean entering data into lists, transcribing recorded conversations, checking data, validating data, cleaning data, anonymising data, describing data using metadata, choosing appropriate data formats, and choosing appropriate repositories. Research fields (sometimes) differ markedly in all these parameters, e.g. by the extent to which data need to be cleaned before further analysis can happen (Paine et al. 2015), the extent to which data from different sources need to be integrated into new data products to answer research questions, and in terms of finding common data formats. On the one hand, appropriate standards need to be followed in order to make your research data as FAIR as possible; on the other hand, the variability of disciplinary or domain-specific research processes is considerable. Therefore, this may require specific sets of knowledge and skills from researchers and/or research support staff to meet these disciplinary or domain-specific standards.

Support services at the institutional level can usually only provide general guidelines. The minutiae of discipline and method-specific practices need to be provided and supported at the departmental and research group level. In order to make data reusable and interoperable, there should be clear expectations and support at each level to help researchers to:

- think about how the data generated might be used by others, and under what conditions;
- think about the information others will need to be able to reuse data and translate that information into documentation complying with appropriate metadata standards (if applicable);
- make sure they appropriately document every step, ranging from raw to processed data and on to research-ready data;

- save and back up documentation alongside important iterations of both primary and processed data; and
- think about the appropriateness of file formats in use and for sharing/publication.

Learn more:

Lesson plan 3: Documentation

Lesson plan 4: Data creation

Lesson plan 6: Metadata

Lesson plan 7: Data standardisation and ontologies

## 6.5 Support infrastructure

Resources and infrastructure are required to enable FAIR data practices within a higher education institution. However, some of the requirements in FAIR can be met through open-source solutions so as to maximise potential reuse of data. Investment in both staffing and platforms is recommended to enable academics to optimise FAIR data.

### Systems for storage, backup and collaboration

Researchers increasingly depend upon technological platforms and tools throughout the data lifecycle. Data, metadata and other artefacts of the research process, including ontologies, software, documentation and papers, all need to be stored in environments where they are backed up and made available for collaboration with partners while being appropriately protected. A common backup method used for research data is the ‘3-2-1 rule’ which involves three copies of the research data being saved: two locally and one off-site. These technical environments may be locally available in a higher education institution or delivered through other services such as cloud computing, including hosting by third parties using a variety of open source or proprietary technologies. Whatever the selected technical infrastructure, investment is required to enable academics to optimise FAIR data to maximise their potential for reuse.

In addition to backup and restoration services that safeguard researchers against data loss, theft, or failure of computers or storage media, and against accidental deletion or unintentional changes to the data, appropriate access management is vital. Authentication (identification and login) and authorisation (permissions control) mechanisms can facilitate collaboration and data sharing within teams. Access control

is also critical to protect sensitive data, e.g. personal information about data subjects, as this has both legal and ethical implications.

Questions about storage, backup and data security are included in many DMP templates issued by funders. Institutions should therefore supply their researchers with information on how their services meet the requirements to help them write their DMPs.

### **Repository services**

Research data repositories are key pieces of infrastructure needed in the research data lifecycle to enable FAIR. They provide a persistent identifier, make the descriptive metadata available, and give access to the data (if applicable). Some repositories offer preservation, which is covered in the following section.

Repositories offer supporting services for the deposit of and access to information vital to research data. A repository or archive may focus on research datasets, but also provide services covering metadata, ontologies, software, etc. These repositories can provide technical infrastructure for ongoing storage, resource discovery and access to (meta)data with persistent identifiers assigned to support citations, credit and vital links between ‘digital objects’ that support interoperability.

For researchers, these repositories provide both a source of data for reuse, and a reliable location for the results of their work. The underlying principles of FAIR are central to the role of repositories, but different repositories offer varying services and degrees of compliance with the expectations of FAIR data. It is important to note that data which are FAIR at the point of deposit in a repository may not remain so without active curation and preservation. Repositories can enable FAIR data by continuing to manage the data formats, e.g. through emulation of ongoing migration to long-term formats. They should also manage the supporting technologies and associated metadata and ontologies as they change over time. The FAIRsFAIR project has developed a capability/maturity approach that aligns repository capability with the requirements for enabling FAIR data over time. Repositories that support more specialised metadata, e.g. disciplinary or domain-specific, will be able to support more sophisticated resource discovery.

Many institutions run their own institutional repositories, but there are a large number of repositories available elsewhere. Some disciplines or domains have dedicated national or multinational repositories.

From a researcher’s point of view, the choice of repository should depend upon the level of support required by their data types and that offered by the repository. These can range from basic storage to resource discovery and on to managing access and use of sensitive data, supporting the peer review of data associated with publications or services involving digital preservation.

The OpenAIRE repository guide advises users to check the availability of a suitable repository in this order:

1. The best choice (if available) is to use an established, dedicated (external) data archive or repository that caters specifically to the research domain to preserve the data according to recognised discipline-specific standards.
2. The second-best choice is to use institutional data repositories.
3. If the above are not viable options, a cost-free data repository should be used.

Dedicated disciplinary repositories are more likely to support community (meta)data standards that will make data more interoperable and more FAIR in general. Institutional repositories may offer integration with local support services, but this may be more generic and therefore less likely to use community standards and rich/specific metadata. Up-to-date lists of available registered data repositories can be found at [re3data](#) and [FAIRsharing](#).

Free-of-charge public repositories are good alternatives for small institutions with limited resources. Three of the more widely known and free to use data repositories are:

- Zenodo – An open access data, software and publication repository for researchers who want to share multidisciplinary research results. It is suitable for all types of research data. It is free to use and has guaranteed funding from the EU for the foreseeable future. Runs on open-source software.
- Harvard Dataverse – An open access repository for research data, code and related material. Open to data from all disciplines worldwide. Runs on open-source software.
- Figshare – An open access repository that provides DOIs and Creative Commons Licences for all datasets. Runs on proprietary software.

Relying purely on external services does not help when developing institutional capacities in this dynamic field, for example in regard to digital preservation. As a result, higher education institutes should invest in their own data repositories if funding/resources are available. A detailed guide on considerations in relation to setting up a data repository was developed by the DCC: [Where to keep research data](#). Integrating institutions in the emerging global research support infrastructure requires awareness and engagement with initiatives like the European Open Science Cloud.

Institutions can provide their researchers with guidance in navigating the world of repositories. Combining a multi-purpose institutional repository with advice on the selection of suitable special purpose repositories elsewhere for suitable datasets is a way forward for many institutions.

## Digital preservation

While most repositories have at least some features that can be used to ensure long-term FAIRness of datasets, thorough digital preservation requires a specific set of organisational, technical and digital object management abilities based on mature standards and assessment processes. Repositories that reach these standards may be certified as ‘trustworthy digital repositories’ (TDR) to signify that they offer active preservation of data and metadata to maintain their value to their community of users over time. Initiatives include the CoreTrustSeal and the nestor Seal. The FAIRsFAIR project has developed a capability/maturity approach that aligns TDR capability with the requirements for enabling FAIR data over time

Digital preservation as defined by the *OAIS reference model* (CCSDS 2012) ensures that data are secure, findable and usable for as long as needed. Not only do many research funders require datasets to be made available for up to ten years, or in perpetuity, it is also best practice and in the interest of both the institute and individual researchers to ensure that generated research data remain accessible after a period of time, even if the software and technology in use at that time are now outdated.

The DPC Rapid Assessment Model (DPC 2021) has been designed to perform rapid benchmarking of an organisation’s digital preservation capacity. This includes tools and considerations when making a business case to implement digital preservation as well as procurement and training. The DPC also hosts the Digital Preservation Handbook (DPC 2015) which offers plenty of advice on how institutions can develop their capacities in this area.

Learn more:

Lesson plan 8: Persistent identifiers (PIDs)

Lesson plan 11: Repositories

## 6.6 Data publication

Proper recognition of researchers’ contributions is fundamental to ensuring widespread adoption of FAIR principles. Once the data have been created, processed, analysed, and their preservation ensured, a clear pathway to crediting the authors in all data-related publications needs to be established. As a minimum standard, datasets need to be cited like other references so as to credit the researchers involved.

Most datasets are published in repositories, often to support and underpin article publication. Linking academic articles and associated data is important for the findability of data and reproducibility of research. The last ten years have also seen the

emergence of dedicated data papers and data journals where peer-reviewed datasets are taking centre stage.

Alongside traditional publications and datasets, there are numerous items of research support information that should be published to make research reproducible and data reusable. These include documentation of methods and protocols, or software and code.

All these research outputs are essential, and researchers can get credit for these parts of their research by publishing them, in turn making the work more shareable, discoverable, comprehensible, reusable, and reproducible.

Authors need to provide contextual information on the relevant dataset, method, software code or other element to be published, and institutions can support their researchers in navigating the emerging publication landscape.

### **Data availability statements**

Data availability statements or statements of availability of supporting data provide information about where the data supporting the results described in a research article can be found and how they can be accessed. These statements can link to a data repository location where the data have been publicly deposited, or can refer to the supplementary information published as part of the article; data availability statements can also clarify when the data are not available or only available privately upon request to the authors. Since these statements are often in free-text form, it is often difficult to identify the level of data access and availability expressed in them. However, a study on 531,889 research articles from PLOS<sup>13</sup> and BMC<sup>14</sup> (Colavizza et al. 2020) has shown that only 12% to 21% of all analysed articles published in 2017 and 2018 included a data availability statement containing a link to a repository, but there is an association between those articles and up to 25% higher citation counts. This has contributed to encouraging the adoption of such statements in the research community as it shows a clear benefit for researchers in terms of the academic impact of their work.

### **Data papers, data journals and peer review for datasets**

Alongside publication of the data in a repository and referencing it in research papers, dedicated data papers can also contribute to the increased visibility of the data and recognition of the researchers' work.

---

13 Public Library of Science (PLOS)

14 BioMed Central (BMC)

Data papers provide an easy channel for researchers to publish their datasets and receive proper credit and recognition for the work they have done. This is particularly true for replication data, negative datasets or data from intermediate experiments, which often go unpublished. Data papers enable researchers to easily share a brief, thorough description of their data, and contain or link to relevant raw data in a repository, in turn helping others discover, understand and reuse the data and reproduce results (Walters 2020).

Data journals have been around for a decade and were established to ensure that researchers creating datasets were appropriately credited with citable outputs. Examples of such journals include *Scientific Data*, *GigaScience*, *F1000Research* for scientific disciplines, and the *Journal of Open Humanities Data* and *Research Data Journal* for the Humanities and Social Sciences for humanities and social sciences.

Recognised pathways to data publication raise the important topic of peer review of data, which needs to become a fundamental part of the publication process. From a researcher's perspective, the considerable time and resource commitment involved in data management and publication need to be supported by appropriate incentives.

## **Methods and protocols**

Method and protocol articles provide details of the methods and/or protocols developed and the materials used during a research cycle. They recognise the time researchers spend customising methods and creating original laboratory resources. Not every method is novel enough to warrant a full research article. However, the customisations that researchers make to methods and the new materials they use can be useful for others, saving them valuable time in developing their own approaches. A platform for developing and sharing reproducible methods is provided by *Protocols.io*.

## **Software**

Making software and code generated in the course of research available via platforms like *GitHub* is part of an Open Research workflow. Software research articles go a step further and may describe significant software and/or code, including relevant post-publication version updates, and/or capture metadata needed to help others apply the software in their own research. They also may describe the impact the software has had on scientific research. Software may also be published as a standalone output, using, e.g. the integration between *GitHub* and *Figshare/Zenodo*. The *Software Sustainability Institute* offers advice on this.

## Other forms of articles relating to specific elements of the research process

Other forms of articles covering a specific aspect of research or the research process focus on hardware and lab resources as well as microarticles and visual case discussions (see Elsevier Research Elements).

Learn more:

Lesson plan 13: Data access

Lesson Plan 14: FAIR software/citable code

## 6.7 Data reuse

Enabling and supporting the reuse of data is one of the core aims of the FAIR principles, and the preceding chapters looked at the reusability of data from many angles, mostly in regard to workflows and practices from a researcher's point of view. This chapter looks at measures that institutions can implement to support and promote the reuse of data.

### Facilitate data sharing agreements

When multiple parties are involved in a research project, it is good practice to have a data sharing agreement in place. Data sharing agreements define the purpose of data sharing, govern what happens to data at each stage of the research process, specify the standards used, and help all parties involved to be clear about their roles and responsibilities. A data sharing agreement can either be set up as a separate document, or data sharing clauses can form part of a broader contract or collaboration agreement.

Before data are shared, involved parties should talk to each other to discuss data sharing issues and come to a joint agreement, which is then documented in a data sharing agreement. The process for creating data sharing agreements may vary from country to country and from institution to institution. It is also possible that other terminology is in use, such as 'information sharing agreement' rather than data sharing agreement.

A data sharing agreement

- establishes roles and responsibilities;
- specifies the purpose of the data sharing;
- governs what happens to the data at each stage of the research process; and
- establishes common standards.



Data sharing agreements are designed to help justify data sharing and demonstrate that all relevant compliance aspects have been considered and documented. A data sharing agreement provides a common framework that also helps meet legal requirements, e.g. for data protection principles.<sup>15</sup>

### **Enhancing discoverability**

Researchers following the FAIR data principles will have documented their data with rich metadata. To make datasets findable, these metadata need to be as widely available as possible. While the original repository in which the data are hosted will provide search functions, metadata should also be indexed in other discovery portals. Descriptive metadata can be indexed (made findable) by general search engines. A more targeted search across multiple repositories is made possible by dedicated dataset search engines like Google DataSet Search, the Data Citation Index of Web of Science, or the Open Source-based service, BASE. This does not happen automatically, but requires conscious effort by the repository. Search engines rely on the mapping of metadata into their underlying metadata schemas, which are schema.org for Google and a custom Data Citation Index schema for Web of Science. BASE curates sources that provide information via OAI-PMH.

Another way of enhancing the discoverability of datasets is by linking the dataset as widely as possible to other information resources. Examples include keywords, links to research articles via DOIs, and authors via ORCIDs. A more advanced level involves the interlinking of datasets or into the linked data world of the semantic web.

### **Promotion of data reuse**

An institutional aim should be to create a virtuous cycle in which researchers become part of communities of practice who consider data reuse and interlinking of various datasets to be an integral part of their research process. Activities supporting this aim include:

- showcasing examples of successful reuse of datasets in blogs and social media;
- organising events like hackathons focused on existing datasets;
- promoting hands-on teaching with existing datasets on all levels and in all disciplines; and
- creating experimental and collaborative spaces like Data Labs (Open a Glam Lab offers advice how to approach this task).

---

15 For more detail, see, for example, the data sharing agreement framework template of the University of Wageningen: <https://www.wur.nl/web/file?uuid=b8299644-97b7-4d8f-959e-25f8fce9fb77&owner=497277b7-cdf0-4852-b124-6b45db364d72&contentid=546669>

Learn more:

Lesson plan 9: Licences, copyright and intellectual property rights (IPR) issues

Lesson plan 10: Finding and reusing data

## 7 – References

Zotero group: [https://www.zotero.org/groups/4273178/fairsfair\\_booksprint](https://www.zotero.org/groups/4273178/fairsfair_booksprint)  
All online references accessed 21 January 2022.

- Abdul, H., 2015. Training Intervention Strategies for Positive Learning Transfer. *Journal of Resources Development and Management* 11, 107–15.
- ACRL, 2015. *Framework for Information Literacy for Higher Education* [online]. Available from: <http://www.ala.org/acrl/standards/ilframework>.
- Anderson, L. W. and Krathwohl, D. R. (eds.), 2001. *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Association of American Universities and Association of Public and Land-grant Universities, 2021. *AAU APLU Guide to Accelerate Public Access to Research Data* [online]. Washington, DC. <https://doi.org/10.31219/osf.io/tjybn>.
- Atchley, T. W., Wingenbach, G. and Akers, C., 2013. Comparison of course completion and student performance through online and traditional courses. *The International Review of Research in Open and Distributed Learning* [online]. 14 (4). <https://doi.org/10.19173/irrodl.v14i4.1461>.
- Australian Government. *Federal Register of Legislation* [online]. Available from: <https://www.legislation.gov.au/Series/C2004A03712>.
- Ball, A., 2012. *Review of Data Management Lifecycle Models* [online]. Bath, UK: University of Bath. <https://researchportal.bath.ac.uk/en/publications/review-of-data-management-lifecycle-models>.
- Bezjak, S., Conzett, P., Fernandes, P. L., Görögh, E., Helbig, K., Kramer, B., Labastida, I. et al., 2019 *The Open Science Training Handbook*. Zenodo. <https://doi.org/10.5281/zenodo.2587951>.

- Biernacka, K., Bierwirth, M., Buchholz, P., Dolzycka, D., Helbig, K., Neumann, J., Odebrecht, C., Wiljes, C. and Wuttke, U., 2020. *Train-the-Trainer Concept on Research Data Management* [online]. (Version 3.0). Zenodo. <https://doi.org/10.5281/zenodo.4071471>.
- Bloom, B. S., Krathwohl, D. R. and Masia, B. B., 1956. *Taxonomy of educational objectives. 1: Cognitive domain*. New York: McKay.
- Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materchera, S., Parsons, M., et al. 2020. The CARE Principles for Indigenous Data Governance. *Data Science Journal* [online], 19 (1), 43. <https://doi.org/10.5334/dsj-2020-043>.
- CASRAI. *Research Data Management Glossary*. <https://codata.org/rdm-terminology/>.
- CCSDS, 2012. *Reference Model for an Open Archival Information System (OAIS). Recommended Practice* [online]. CCSDS 650.0-M-2. Magenta Book. <https://public.ccsds.org/pubs/650x0m2.pdf>.
- CESSDA Training Team, 2020. *CESSDA Data Management Expert Guide* [online]. <https://doi.org/10/gkc9v4>.
- Clare, C., Cruz, M., Papadopoulou, E., Savage, J., Teperek, M., Wang, Y., Witkowska, I. and Yeomans, J., eds, 2019. *Engaging Researchers with Data Management: The Cookbook*. Open Reports Series. Open Book Publishers. <https://doi.org/10.11647/OBP.0185>.
- Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K. and Barbara McGillivray, 2020. The Citation Advantage of Linking Publications to Research Data. *PLoS ONE* [online], 15 (4). <https://doi.org/10/ggtcrb>.
- Data FAIRport. *Data FAIRport Conference. Jointly Designing a Data FAIRport* [online]. Available from: [https://www.datafairport.org/component/content/article/8\\_news/9\\_item1/index.html](https://www.datafairport.org/component/content/article/8_news/9_item1/index.html).
- Davidson, J., Engelhardt, C., Proudman, V., Stoy, L., and Whyte, A., 2019. *D3.1 FAIR Policy Landscape Analysis* [online] Zenodo. <https://doi.org/10.5281/zenodo.5537032>.
- Demchenko, Y., Stoy, L., Engelhardt, C. and Gaillard, V., 2021. *D7.3 FAIR Competence Framework for Higher Education (Data Stewardship Professional Competence Framework)* [online]. Zenodo. <https://doi.org/10.5281/zenodo.4562088>.
- Denson, N., Loveday, T. and Dalton, H., 2010. Student evaluation of courses: what predicts satisfaction? *Higher Education Research & Development* [online] 29 (4), 339-352. <https://doi.org/10.1080/07294360903394466>.
- Donaldson, M., Schwamm, H. and Campbell, Fay, 2017. EPSRC DMP Assessment Rubric v2.0 [online]. Zenodo. <https://doi.org/10.5281/zenodo.247087>.

- DPC, 2021. *Digital Preservation Coalition Rapid Assessment Model* [online]. Version 2, March 2021). <https://doi.org/10.7207/dpcram21-02>.
- DPC, 2015. *Digital Preservation Handbook* [online]. 2nd edition. <https://www.dp-conline.org/handbook>.
- Dunning, D., Heath, C. and Suls, J. M., 2004. Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest* [online], 5 (3), 69-106. <https://doi.org/10.1111/j.1529-1006.2004.00018.x>.
- EDISONcommunity, 2020. *EDSF*. GitHub Repository. Available from: <https://github.com/EDISONcommunity/EDSF>.
- Engelhardt, C. et al., 2021. *D7.4 How to be FAIR with your data. A teaching and training handbook for higher education institutions* [online]. Zenodo. <https://doi.org/10.5281/zenodo.5665492>.
- European Open Science Cloud, 2021. *Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC)* [online]. Version 1.0. Available from: [https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0\\_15Feb2021.pdf](https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf).
- EOSC Executive Board, Landscape Working Group (WG), 2020. *Landscape of EOSC-Related Infrastructures and Initiatives* [online]. Available from: <https://data.europa.eu/doi/10.2777/132181>.
- European Commission, 2019. *Cost-Benefit Analysis for FAIR Research Data* [online]. Available from: <https://data.europa.eu/doi/10.2777/02999>.
- European Commission, 2018. *Turning FAIR into reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data* [online]. <https://doi.org/10.2777/1524>.
- European Union. *Description of the Eight EQF Levels* [online]. Available from: <https://europa.eu/europass/en/description-eight-eqf-levels>.
- European Union General Data Protection Regulation (GDPR), 2016. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)* [online]. OJ 2016 L 119 (1). Available from: <http://data.europa.eu/eli/reg/2016/679/oj>.
- FAIR Cookbook*, [online]. Available from: <https://w3id.org/faircookbook>.
- FAIRsharing, interlinking standards, repositories and policies* [online]. Available from: <https://fairsharing.org> and <https://doi.org/10.25504/FAIRsharing.2abjs5>.
- Faniel, I. M. and Jacobsen, T. E., 2010. Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. *Comput-*

- er Supported Coop Work* [online] 19 (3-4), 355-375. <https://doi.org/10.1007/s10606-010-9117-8>
- FOSTER. *Use Open Data in Teaching* [online]. Online Course. Available from: <https://www.fosteropenscience.eu/node/2656>.
- German Research Foundation, 2019. *Guidelines for Safeguarding Good Research Practice. Code of Conduct* [online]. <https://doi.org/10/gg927v>.
- Gitelman, L., 2013. *“Raw Data” is an oxymoron*. Cambridge, MA: The MIT Press.
- Guler, G., 2019. *Data Literacy from Theory to Reality: How Does It Look?* [online]. Thesis (Master), Vrije Universiteit Brussel. Available from: [https://www.researchgate.net/profile/Gulsen-Guler/publication/335620777\\_Data\\_literacy\\_from\\_theory\\_to\\_reality\\_How\\_does\\_it\\_look/links/5ddfbfb74585159aa4503cde/Data-literacy-from-theory-to-reality-How-does-it-look.pdf](https://www.researchgate.net/profile/Gulsen-Guler/publication/335620777_Data_literacy_from_theory_to_reality_How_does_it_look/links/5ddfbfb74585159aa4503cde/Data-literacy-from-theory-to-reality-How-does-it-look.pdf).
- Higman, R., Bangert, D. and Jones, S., 2019. Three Camps, One Destination: The Intersections of Research Data Management, FAIR and Open. *Insights* [online] 32 (18), <https://doi.org/10/gf4jhr>.
- ICPSR - *Preserving Respondent Confidentiality* [online]. Available from: <https://www.icpsr.umich.edu/web/pages/deposit/confidentiality.html>.
- Jordan, K. (2015). Massive open online course completion rates revisited: Assessment, length and attrition. *The International Review of Research in Open and Distributed Learning* [online], 16 (3), <https://doi.org/10.19173/irrodl.v16i3.2112>.
- Karpen, S. C., 2018. The Social Psychology of Biased Self-Assessment. *American Journal of Pharmaceutical Education* [online] 82 (5) 6299. <https://doi.org/10.5688/ajpe6299>.
- *Wet op de dierproeven* [online]. Wet. Available from: <https://wetten.overheid.nl/BWBR0003081/2019-01-01>.
- Learn, 2017. *LEARN Toolkit of Best Practice for Research Data Management* [online]. LEARN project. <https://doi.org/10.14324/000.learn.00>.
- Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giarretta, D., De Giusti, M. et al, 2020. The TRUST Principles for Digital Repositories. *Scientific Data* [online] 7 (1), 144. <https://doi.org/10.1038/s41597-020-0486-7>.
- Mahey et al., 2019. *Open a GLAM Lab* [online]. Doha <http://doi.org/10.21428/16ac48ec.f54af6ae>.
- Martinez, P. A., Erdmann, C., Simons, N., Otsuji, R., Labou, S., Johnson, R., Castellao, G. et al., 2019. Top 10 *FAIR Data & Software Things* [online]. <https://doi.org/10/gkbnxv>.
- Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 2021. *Wet medisch-wetenschappelijk onderzoek met mensen* [online]. Wet. Available from: <https://wetten.overheid.nl/BWBR0009408/2021-05-26>.

- Morais, R., Saenen, B., Garbuglia, F., Berghmans, S. and Gaillard, V., 2021. *From Principles to Practices: Open Science at Europe's Universities. 2020-2021 EUA Open Science Survey Results* [online]. Zenodo. <http://doi.org/10.5281/zenodo.5062982>.
- Muljana, P. S. and Luo, T., 2019. Factors contributing to student retention in online learning and recommended strategies for improvement: A systematic literature review. *Journal of Information Technology Education: Research* [online], 18, 19-57. <https://doi.org/10.28945/4182>
- Nicholls, G., 2001. *Developing Teaching and Learning in Higher Education*. 1st edition. Routledge. <https://doi.org/10.4324/9780203469231>.
- OECD, 2020. *Building Digital Workforce Capacity and Skills for Data-Intensive Science* [online]. <https://doi.org/10.1787/e08aa3bb-en>.
- Paine, D. and Lee, C., 2015. *Examining Data Processing Work as Part of the Scientific Data Lifecycle Comparing Practices Across Four Scientific Research Groups* [online] Figshare. <https://doi.org/10.6084/M9.FIGSHARE.1354039>.
- Peer, L., Arguillas, F., Honeyman, T., Miljković, N., Peters-von Gehlen, K. and CURE-FAIR subgroup 3, 2021. *Challenges of Curating for Reproducible and FAIR Research Output* [online]. (2.1). <https://doi.org/10.15497/RDA00063>.
- PsychData - Terms of Use* [online]. Available from: <https://www.psychdata.de/index.php?main=take&sub=empfang&lang=eng>.
- RDMkit* [online]. Available from: <https://rdmkit.elixir-europe.org/>.
- Registry of Research Data Repositories* [online]. Available from: <https://www.re3data.org/>.
- Ridsdale, C., Rothwell, J., Smit, M., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., Wuetherick, B. and Ali-Hassan, H., 2015. *Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report* [online]. <https://doi.org/10.13140/RG.2.1.1922.5044>.
- Science Europe, 2021. *Practical Guide to the International Alignment of Research Data Management - Extended Edition* [online]. <https://doi.org/10.5281/ZENODO.4915861>.
- Stoy, L., Saenen, B., Davidson, J., Engelhardt, C. and Gaillard, V., 2020. *D7.1 FAIR in European Higher Education* [online]. Zenodo. <https://doi.org/10.5281/zenodo.3629682>.
- Sveinsdottir, T., Davidson, J. and Proudman, V., 2021. *An Analysis of Open Science Policies in Europe, v7* [online]. Zenodo. <https://doi.org/10.5281/ZENODO.4725817>.
- The Turing Way. *The Turing Way Book Dashes* [online]. Available from: <https://the-turing-way.netlify.app/community-handbook/bookdash.html>.

- Tractenberg, R. E., Lindvall, J. M., Attwood, T. and Via, A., 2020. *Guidelines for Curriculum and Course Development in Higher Education and Training* [online]. Preprint. SocArXiv. <https://doi.org/10.31235/osf.io/7qeht>.
- Tuuli Working Group, 2021. *Finnish DMP Evaluation Guidance* [online]. Zenodo. <https://doi.org/10.5281/ZENODO.4729831>.
- UK Data Service. *Research Data Lifecycle* [online]. UK Data Service. Available from: <https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx>.
- Via, A., Palagi, P. M., Lindvall, J. M., Tractenberg, R. E., Attwood, T. K. and The GOBLET Foundation, 2020. Course Design: Considerations for Trainers – a Professional Guide. *F1000Research* [online], 9. <https://doi.org/10/gkc9v6>.
- Walters, W. H., 2020. Data Journals: Incentivizing Data Access and Documentation Within the Scholarly Communication System. *Insights* [online] 33 (1). <https://insights.uksg.org/articles/10.1629/uksg.510/>
- Wiljes, C. and Cimiano, P., 2019. Teaching Research Data Management for Students. *Data Science Journal* [online] 18 (1), 38. <https://doi.org/10.5334/dsj-2019-038>.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N. et al., 2016. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* [online], 3 (160018). <https://doi.org/10/bdd4>.



## 8 – About the authors & facilitators

<b>Name, Affiliation &amp; ORCID</b>	<b>CRedit Roles</b> ( <a href="https://credit.niso.org/">https://credit.niso.org/</a> )
Raisa Barthauer, University of Göttingen, 0000-0003-0938-8104	Writing – reviewing and editing
Katarzyna Biernacka, Humboldt-Universität zu Berlin, 0000-0002-6363-0064	Writing – original draft Writing – reviewing and editing
Aoife Coffey, University College Cork, 0000-0001-5008-3711	Writing – original draft
Ronald Cornet, Amsterdam University Medical Centre, 0000-0002-1704-5980	Writing – original draft Writing – reviewing and editing
Alina Danciu, Sciences Po Paris, 0000-0002-5126-0078	Writing – original draft Writing – reviewing and editing
Yuri Demchenko, University of Amsterdam, 0000-0001-7474-9506	Writing – original draft Writing – reviewing and editing
Stephen Downes, National Research Council of Canada, 0000-0001-6797-9012	Writing – original draft Writing – reviewing and editing
Claudia Engelhardt, University of Göttingen, 0000-0002-3391-7638	Writing – original draft Writing – reviewing and editing Project administration

<b>Name, Affiliation &amp; ORCID</b>	<b>CRedit Roles</b> ( <a href="https://credit.niso.org/">https://credit.niso.org/</a> )
Christopher Erdmann, American Geophysical Union, 0000-0003-2554-180X	Writing – original draft
Federica Garbuglia, European University Association, 0000-0002-9162-4446	Writing – original draft Writing – reviewing and editing
Kerstin Germer, Humboldt-Universität zu Berlin, 0000-0002-4707-5793	Writing – original draft Writing – reviewing and editing
Kerstin Helbig, Humboldt-Universität zu Berlin, 0000-0002-2775-6751	Writing – original draft
Margareta Hellström, Lund University and ICOS Carbon Portal, 0000-0002-4154-2610	Writing – original draft Writing – reviewing and editing
Kristina Hettne, Leiden University Libraries, 0000-0002-4182-7560	Writing – original draft
Dawn Hibbert, University of Northampton, 0000-0001-9741-5840	Writing – original draft Writing – reviewing and editing
Mijke Jetten, Dutch Techcentre for Life Sciences and Health-RI, 0000-0001-9114-2896	Writing – original draft Writing – reviewing and editing
Yulia Karimova, Institute for Systems and Computer Engineering, Technology and Science, 0000-0002-1015-6709	Writing – original draft Writing – reviewing and editing
Karsten Kryger Hansen, Aalborg University, 0000-0002-2407-8764	Writing – original draft
Mari Elisa Kuusniemi, University of Helsinki, 0000-0002-7675-287X	Writing – original draft Writing – reviewing and editing
Viviana Letizia, Elsevier, 0000-0003-1088-5255	Writing – original draft

<b>Name, Affiliation &amp; ORCID</b>	<b>CRedit Roles</b> ( <a href="https://credit.niso.org/">https://credit.niso.org/</a> )
Valerie McCutcheon, University of Glasgow, 0000-0002-0705-4832	Writing – reviewing and editing
Barbara McGillivray, King’s College London and The Alan Turing Institute, 0000-0003-3426-8200	Writing – original draft
Jenny Ostrop, University of Bergen, 0000-0003-2752-8377	Writing – original draft Writing – reviewing and editing
Britta Petersen, Christian-Albrechts-Universität zu Kiel, 0000-0002-0355-2594	Writing – original draft Writing – reviewing and editing
Ana Petrus, University of Applied Sciences of the Grisons, 0000- 0002-0928-8894	Writing – original draft Writing – reviewing and editing
Stefan Reichmann, TU Graz, 0000-0003-1544-5064	Writing – original draft
Najla Rettberg, University of Göttingen, 0000-0003-1888-2294	Writing – original draft Project administration
Carmen Reverté, Institute of Agrifood Research and Technology, 0000-0003-4768-7180	Writing – original draft Writing – reviewing and editing
Nick Rochlin, University of British Columbia, 0000-0002-0772-9342	Writing – original draft
Bregt Saenen, European University Association, 0000-0002-2827-9504	Writing – reviewing and editing
Birgit Schmidt, University of Göttingen, 0000-0001-8036-5859	Writing – original draft Writing – reviewing and editing
Jolien Scholten, Vrije Universiteit Amsterdam, 0000-0002-0839-3916	Writing – original draft

<b>Name, Affiliation &amp; ORCID</b>	<b>CRedit Roles</b> ( <a href="https://credit.niso.org/">https://credit.niso.org/</a> )
Hugh Shanahan, Royal Holloway, University of London, 0000-0003-1374-6015	Writing – reviewing and editing
Armin Straube, University of Limerick, 0000-0002-5229-1968	Writing – original draft Writing – reviewing and editing
Veerle Van den Eynden, KU Leuven, 0000-0003-2542-2747	Writing – original draft Writing – reviewing and editing
Justine Vandendorpe, ZB Med – Information Centre for Life Sciences, 0000-0002-9421-8582	Writing – original draft
Shanmugasundaram Venkataram, DCC and OpenAIRE, 0000-0002-3200-2698	Writing – original draft Writing – reviewing and editing
André Vieira, University of Minho, 0000-0002-4302-645X	Project administration
Cord Wiljes, Universität Bielefeld, 0000-0003-2528-5391	Writing – original draft
Ulrike Wuttke, University of Applied Sciences Potsdam, 0000-0002-8217-4025	Writing – original draft Writing – reviewing and editing
Joanne Yeomans, Leiden University, 0000-0002-0738-7661	Writing – original draft
Biru Zhou, McGill University, 0000-0001-6914-1432	Writing – original draft Writing – reviewing and editing

## **Appendix A – Resources**

This is not intended to be an exhaustive list but a starting point.

### **Glossaries**

- CASRAI RDM Terminology
- Terms4FAIRSkills

### **DMP (and other) tools**

- DMP Online
- Research Data Management Toolkit
- Argos DMP
- Data Stewardship Wizard
- Research Data Management Organiser (RDMO)
- MapleDocs
- DLCM
- Science Europe Practical Guide to International Alignment of Research Data Management
- GFBio tool for DMP
- FAIR Data Tools (DTL Data FAIRport)
- FAIR Aware Tool

### **Guides/Practices**

- ELIXIR RDMkit
- CESSDA Data Management Expert Guide  
(offline PDF version: <https://zenodo.org/record/3820473#.YDzA7twxk-U>)
- CESSDA DMP Questions Qualitative data

- CESSDA DMP Questions Quantitative data
- FAIR Data Management in Horizon 2020 Guidelines
- ICPSR Framework for Creating a Data Management Plan
- FAIR Cookbook, practical recipes with applied examples
- The Turing Way (also: <https://doi.org/10.5281/zenodo.3233853>)
- Assessing capability maturity and engagement with FAIR-enabling practice (ACME-FAIR)
  - Defining the Policy Environment: ACME-FAIR Issue #1
  - Professionalising Roles through Training, Mentoring, and Recognition: ACME-FAIR Issue#3
  - Supporting Data Management Planning: ACME-FAIR Issue#4
  - Defining Data Interoperability Frameworks: ACME-FAIR Issue #5
  - Ensuring Trustworthy Curation: ACME-FAIR Issue #7

### **Support for licence selection**

- EUDAT License Selector
- DCC “How to License Research Data”
- <https://creativecommons.org/choose/>
- <https://opendefinition.org/licenses/>
- <https://choosealicense.com/>
- Data Licencing: Choose the right right, use the data right:
  - <http://eprints.gla.ac.uk/171314/>
  - <http://eprints.gla.ac.uk/171315/>
  - <http://eprints.gla.ac.uk/171316/>
  - <http://eprints.gla.ac.uk/171317/>

### **Metadata**

- Metadata Standards Catalogue
- Linked Open Vocabularies
- FAIRsharing interlinked (meta)data standards

### **Repositories**

- Registry of research data repositories (re3data.org)
- FAIRsharing interlinked repositories to (meta) data standards
- Zenodo
- B2Share
- Dryad
- Atmospheric Radiation Measurement repository
- List of Core Certified Repositories

## Appendix B – Target audience personas

This appendix documents an exercise on target audience personas that was conducted in breakout groups during the kick-off meeting on 1 June 2021. The aim was to have the book sprinters put themselves in the place of a reader of the book they were going to collaboratively write, and then think about the needs and requirements regarding the handbook from the recipient's perspective. The outcomes informed discussions about the structure and content of the handbook.

The participants split up into five groups, each discussing a persona with one of the following roles: junior lecturer, professor, doctoral programme manager, support staff member, management. Working with sticky notes on digital whiteboards, they collected their thoughts and ideas about the role, subject/discipline, employment situation, familiarity with technology, as well as with the FAIR principles of each of the above mentioned. Most importantly, they also stepped into the persona's shoes and tried to answer the following questions:

- In what way / for what purpose would this person use the handbook?
- Which needs and expectations does this person have with regard to the handbook?

Below is a copy of each whiteboard with all the information gathered during this exercise.

For a summary of the breakout session outcomes, see page 92.



## Junior Lecturer

### Age

late 20s  
early 30s  
(UK)

late 20s  
mid 30s

ACADEMIC  
AGE?

### Role

Lecture (UK)

### Subject / discipline

all disciplines

specific subjects  
e.g. on  
methodology?

### Employment situation (full-time, part-time, etc.)

part-time  
(in most of  
the rest of  
Europe)

can be „external“ e.g. in  
Austria, i.e. no affiliation  
beyond one or two teaching  
appointments

fulltime and  
permanent  
(in the UK)

### Technology / familiarity

It depends on the  
discipline, but generally  
average

Have  
technological  
skills

need to know technology  
used in University/Institution  
(data repository, for example,  
DMP tools, etc.)

### Familiarity with the FAIR principles

very diverse  
but often not  
high

Should have  
heard of it at  
least

Basic knowledge about RDM  
in general, FAIR principles  
and open science



**In what way / for which purpose would this person use the Handbook?**

to assist them in preparing the lectures	to familiarize with FAIR principles	Additional material for lecture	to find structure, examples, ideas, tools
adding to course on research methods or research ethics	Improve knowledge related to FAIR principles, RDM issues in general	Junior lecturers would need to educate both themselves and their students, with the help of the handbook	disseminate FAIRness

**Which needs and expectations does this person have with regard to the Handbook?**

Helping researchers to recognise their own research processes in the FAIR ecosystem	Important to define data and FAIR data and meta-data. Also, (meta-data) standards, tools, platforms.	Contents will need to be relevant to research project workflow. It needs to have direct connections and concrete examples (both generic and disciplinary specific) to be readily incorporated into teaching and training activities.
---	--	--

**Other**



## Professors

### Age

45

### Role

Recently tenured full professor

### Subject / discipline

Physical geography  
(Alpine region specialist)

### Employment situation (full-time, part-time, etc.)

full-time, tenur,  
10 hrs per week  
teaching  
obligation

Research  
Centre / Institute  
Lead

Part time

PGR  
Super-  
vision

Lead on Re-  
search Projects  
(PI) Responsibility  
for bringin in  
funding

### Technology / familiarity

GIS systems only,  
basic Internet

statistical  
programs  
(R, etc.)

E-learning

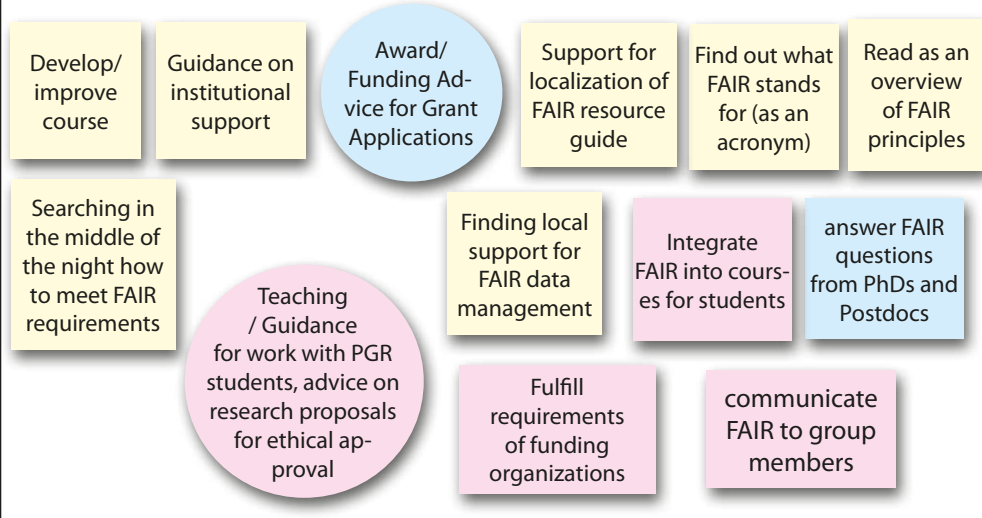
### Familiarity with the FAIR principles

just some  
basic  
knowledge

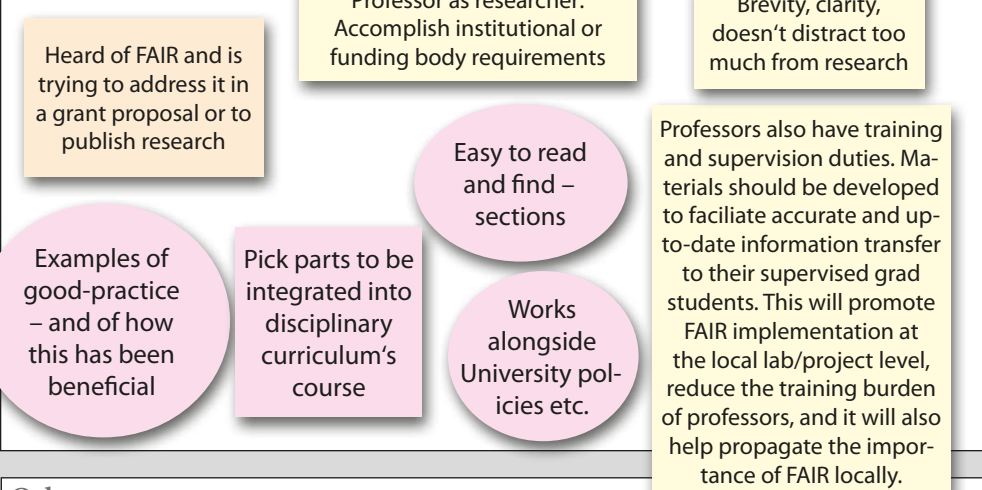
FAIR what  
now?

depending  
on the  
discipline

**In what way / for which purpose would this person use the Handbook?**



**Which needs and expectations does this person have with regard to the Handbook?**



**Other**





## Doctoral programme manager

### Age

Mid-late  
40s?

### Role

organise  
trainings

Coordinate to make a training happen:  
find the right people, decide when and  
where, how many credits, send out  
announcements and information

### Subject / discipline

Discipline  
Specific

### Employment situation (full-time, part-time, etc.)

Full-time,  
middle  
administration

### Technology / familiarity

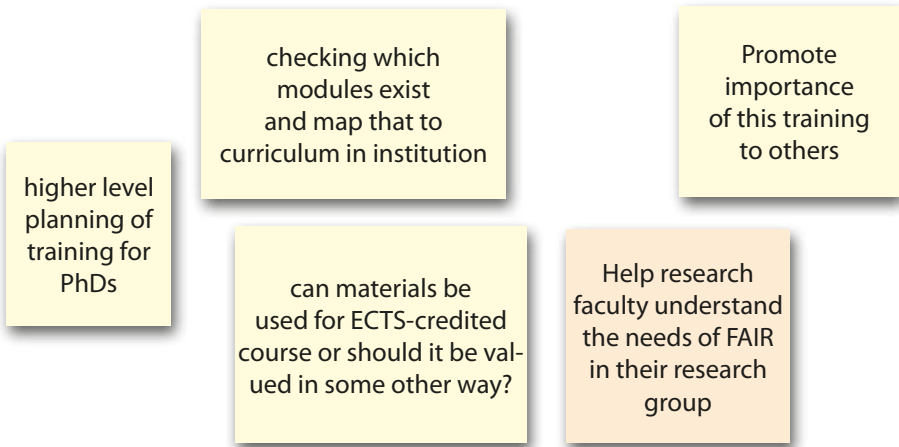
Knows/uses some  
of the tools but not  
very technical

### Familiarity with the FAIR principles

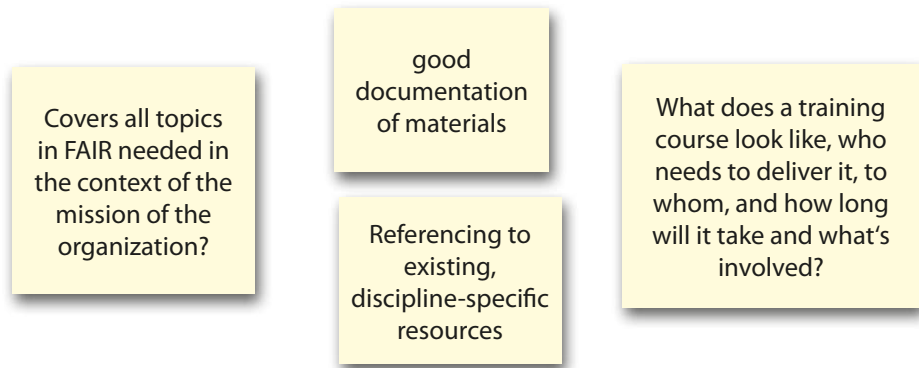
not very  
much

Knows they  
are important  
but not the  
details

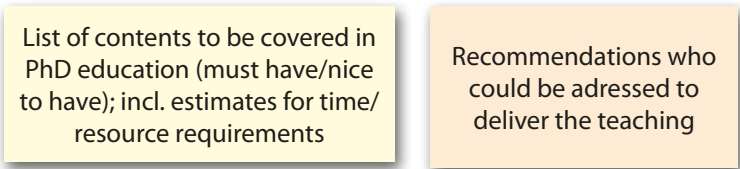
**In what way / for which purpose would this person use the Handbook?**



**Which needs and expectations does this person have with regard to the Handbook?**



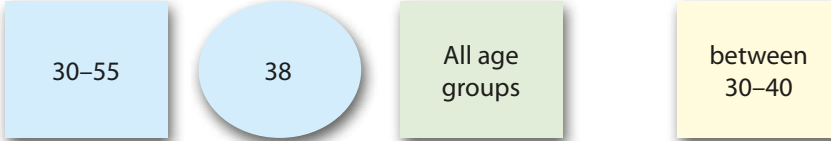
**Other**



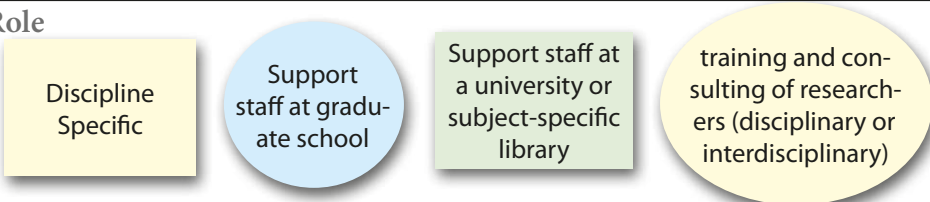


## Support staff member

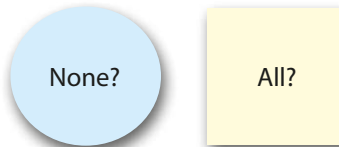
### Age



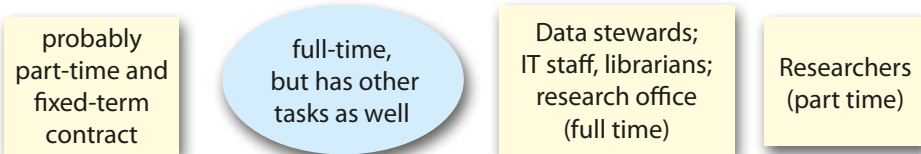
### Role



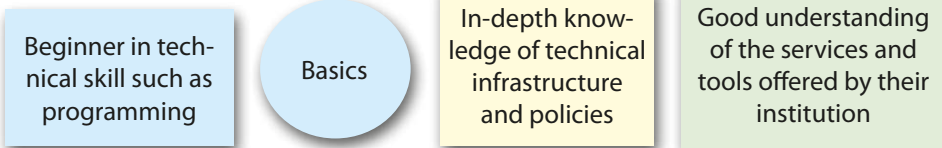
### Subject / discipline



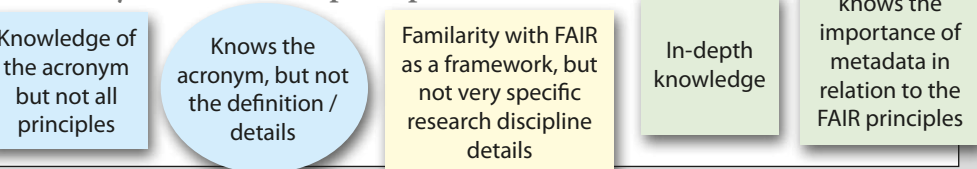
### Employment situation (full-time, part-time, etc.)



### Technology / familiarity



### Familiarity with the FAIR principles



**In what way / for which purpose would this person use the Handbook?**

As a reference for getting around “all” aspects of good data practice, to determine next available steps.

Find suitable resources for planning teaching

A living document

Checking compliance with FAIR; training; reference manual

Learn about FAIR principles in detail and teach it to others, use materials from the book

**Which needs and expectations does this person have with regard to the Handbook?**

Identifiable “situations” from research practice

Practical exercises

Can be easily understood by lay persons

Step-by-step guidelines

Discipline specific examples

Definition of FAIR, details on FAIR, materials, resources, literature references

Both fish (actual things you can do) and some learning how to fish yourself in the future

**Other**

Familiarise with the contract of the researchers. Find out any other applications towards data handling. How to maximise the potential from the researchers’ point of view. Make a framework to identify the situation and the good mix that it’s for the researchers to take.  
Example: <https://twitter.com/karstenkryger/status/1385501114066477061/photo/1>

Might need icons or images for training slides and marketing of FAIR



Management (vice president for research, director of doctoral schools. etc.)

Age

50–65

Fine with any age, just not too young, I would say

40-up

Role

Organising the engagement (timeline, resource allocation) of the institution’s activities

Responsible for overarching issues

Needs to be able to make the “business case” for FAIR data

Person that makes the KPIs for FAIR in their institution policies

Subject / discipline

Both domain agnostic and subject specific – we also have managers in the various domain specific institutes and in national domain infrastructures / thematics organisations

Interdisciplinary: from computer science to humanities

Employment situation (full-time, part-time, etc.)

Full-time

Fine with part time and full time

Technology / familiarity

Probably rather low when it comes to details

When in a faculty probably a little more detailed compared to when in a overarching part of the organisation (like executive boards)

concentrate on faculties they have familiarities with

Familiarity with the FAIR principles

Has heard of the principles, but probably has quite vague understanding of what the principles mean and imply, especially concerning practical implementations and / or required

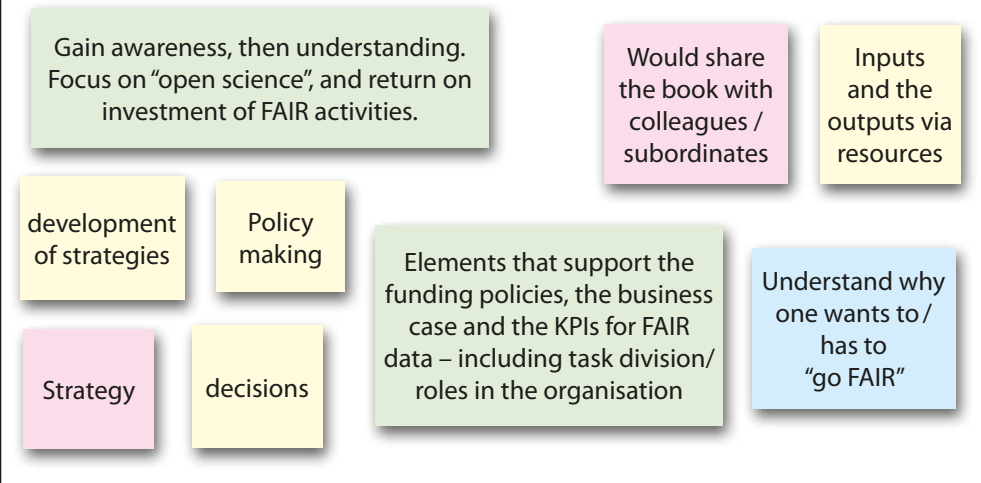
Initially low, might have heard of FAIR, wants to have helicopter view on it

uses the term for proposals

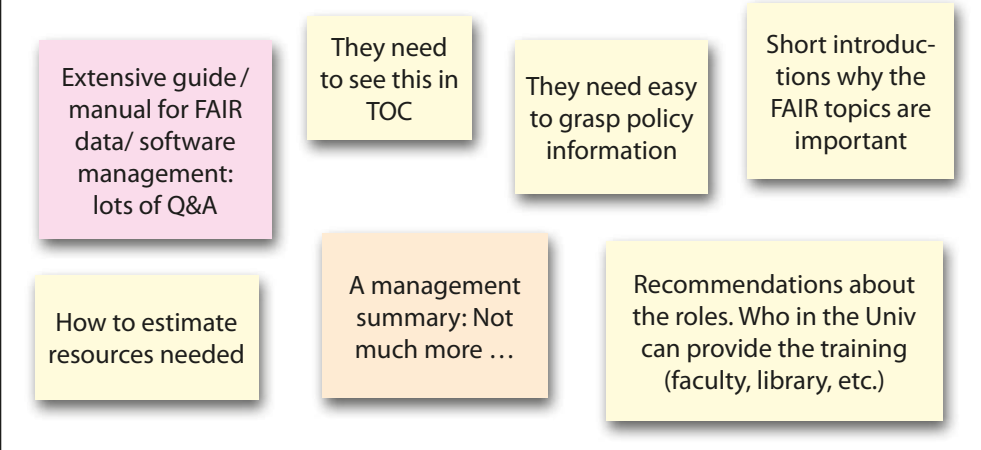
They should have as they make the policy, right? Or at least make the KPIs?



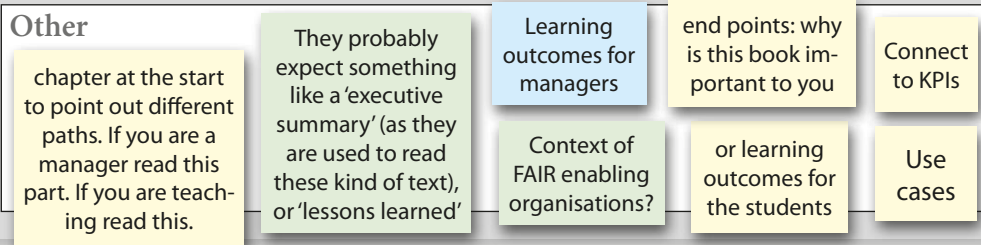
**In what way / for which purpose would this person use the Handbook?**



**Which needs and expectations does this person have with regard to the Handbook?**



**Other**



Here is a brief summary of the outcome for each breakout session.

In terms of purposes, there are strong similarities between the junior lecturer, the professor and the support staff member. All three are envisioned as using this handbook to prepare lectures, courses or training in which they teach others (students, researchers) about the FAIR principles. At the same time, participants thought they would also use it as a tool to find out about and teach themselves the FAIR principles (all three groups), to get advice for grant applications (professor), or as a reference for good FAIR practice and to check FAIR compliance (support staff). The doctoral programme manager is seen as using the handbook for higher-level planning of training for PhD students which also involves the mapping of relevant content to the existing curriculum and thinking about assessment and accreditation. Furthermore, they would use it as a resource when supporting or advising colleagues on FAIR matters. For someone working at management level, such as a vice-rector for research, it is crucial to know why the FAIR principles are important, what their implications for strategic planning and policy-making are, and how to make the case for FAIR.

As for expectations, the handbook should enable the user to fulfil the task they are using it as a tool for in the best way possible. It should therefore be easy to navigate and understand, with content accurate, up to date and easy to integrate into courses. Practical exercises and materials help with the latter. Concrete examples of good practice and use cases illustrate the relevance to research. References to existing resources, especially discipline-specific ones, can serve as a starting point for finding additional information to tailor courses to a specific audience.

**Appendix C – Data Stewardship Competence Groups  
(CF-DSP) and enumeration  
(according to FAIRsFAIR Deliverable D7.3)**

This table from Demchenko et al. (2021, pp. 70 et sqq.) is a reference for the work done in chapter 3. It was used as the basis for developing the competence profiles and learning outcomes described there.

Data Management (DSDM)	Data Science Engineering (DSENG)
<p><b>DSDM – extended, relevant</b> Develop and implement data management strategy for data collection, storage, preservation, and availability for further processing, Ensure compliance with FAIR data principles.</p>	<p><b>DSENG – no changes, generally relevant</b> Use engineering principles and modern computer technologies to research, design, implement new data analytics applications; develop experiments, processes, instruments, systems, infrastructures to support data handling during the whole data lifecycle.</p>
<p><b>DSDM01 – extended, essential</b> Develop and implement data management and governance strategy, in particular, in the form of Data Governance Policy and Data Management Plan (DMP). Ensure compliance with standards and best practices in Data Governance and Data Management.</p>	<p><b>DSENG01 – no changes, low relevance</b> Use engineering principles (general and software) to research, design, develop and implement new instruments and applications for data collection, storage, analysis and visualisation</p>
<p><b>DSDM02 – extended, essential</b> Develop and implement relevant data models, define metadata using common standards and practices for different data sources in a variety of scientific and industry domains.</p> <ul style="list-style-type: none"> <li>• Ensure metadata compliance with FAIR requirements.</li> <li>• Be familiar with the metadata management tools.</li> </ul>	<p><b>DSENG02 – no changes, low relevance</b> Develop and apply computational and data driven solutions to domain related problems using wide range of data analytics platforms, with a special focus on Big Data technologies for large datasets and cloud based data analytics platforms</p>

<p><b>Data Science Research Methods and Project Management (DSRMP)</b></p>	<p><b>Data Science Domain Knowledge (DSDK) as Business Process Management (DSBA)</b></p>
<p><b>DSRMP – revised, generally relevant</b>                      Create new understandings and capabilities by using the scientific method (hypothesis, test/artefact, evaluation) or similar engineering methods to discover new approaches to create new knowledge and achieve research or organisational goals</p> <ul style="list-style-type: none"> <li>• Base research on collected scientific facts and collected data</li> </ul>	<p><b>DSDK – generally relevant</b>                      Use domain knowledge (scientific or business) to develop relevant data analytics applications; adopt general Data Science methods to domain specific data types and presentations, data and process models, organisational roles and relations</p>
<p><b>DSRMP01 – generally relevant</b>                      Create new understandings, discover new relations by using the research methods (including hypothesis, artefact/experiment, evaluation) or similar engineering research and development methods</p>	<p><b>DSBA01 – relevant for organisation processes and data</b>                      Analyse information needs, assess existing data and suggest/identify new data required for specific business context to achieve organizational goal, including using social network and open data sources</p> <ul style="list-style-type: none"> <li>• Data management and Quality Assurance of organisational data assets</li> </ul>
<p><b>DSRMP02 – generally relevant</b>                      Direct systematic study toward the understanding of the observable facts, and discovers new approaches to achieve research or organisational goals</p>	<p><b>DSBA02 – relevant for organisation processes and data</b>                      Operationalise fuzzy concepts to enable key performance indicators measurement to validate the business analysis, identify and assess potential challenges</p> <ul style="list-style-type: none"> <li>• Specify requirements/develop data models for organisational data</li> </ul>

Data Management (DSDM)	Data Science Engineering (DSENG)
<p><b>DSDM03 – extended, essential</b></p> <p>Integrate heterogeneous data from multiple sources and provide them for further analysis and use</p> <p>Perform data preparation and cleaning</p> <p>Match/transfer data models of individual datasets</p>	<p><b>DSENG03 – extended, relevant</b></p> <p>Develop and prototype specialised data analysis applications, tools and supporting infrastructures for data driven scientific, business or organisational workflow; use distributed, parallel, batch and streaming processing platforms, including online and cloud based solutions for on-demand provisioned and scalable services</p> <ul style="list-style-type: none"> <li>• Develop new tools and applications, ensure support of the data FAIRness requirements by existing and new tools and applications</li> </ul>
<p><b>DSDM04 – extended, highly essential</b></p> <p>Maintain historical information on data handling, including reference to published data and corresponding data sources</p> <ul style="list-style-type: none"> <li>• Publish data, metadata and related metrics</li> <li>• Perform and maintain data archiving</li> <li>• Develop necessary archiving policy, comply with Open Science and Open Access policies if applicable</li> <li>• Perform data provenance and ensure continuity through the whole data lifecycle, ensure data provenance</li> </ul>	<p><b>DSENG04– extended, essential</b></p> <p>Develop, deploy and operate data infrastructure, including data storage and processing facilities, using different distributed and cloud based platforms.</p> <ul style="list-style-type: none"> <li>• Implement requirements for data storage facilities to comply with the data management policies and FAIR data principles in particular.</li> </ul>
<p><b>DSDM05 – extended, essential</b></p> <p>Develop policy and metrics for data quality management (e.g. Altmetrix), maintain data quality and compliance to standards, perform data curation</p> <p>Interact/Collaborate with data providers and data owners to ensure data quality</p>	<p><b>DSENG05– extended, relevant</b></p> <p>Consistently apply data security mechanisms and controls at each stage of the data processing, including data anonymisation, privacy and IPR protection, ensure standards and corresponding data protection regulation compliance, in particular GDPR.</p> <ul style="list-style-type: none"> <li>• Define and implement (coordinate) data access policies for different stakeholders and organisational roles</li> </ul>

<b>Data Science Research Methods and Project Management (DSRMP)</b>	<b>Data Science Domain Knowledge (DSDK) as Business Process Management (DSBA)</b>
<p><b>DSRMP03- extended, essential</b>            Analyse domain related research process model, identify and analyse available data to identify research questions and/ or organisational objectives and formulate sound hypothesis            Link domain related concepts and models to general/abstract Data Science concepts and models,</p>	<p><b>DSBA03 – generally relevant</b>            Deliver business focused analysis using appropriate BA/BI methods and tools, identify business impact from trends; make business case as a result of organisational data analysis and identified trends            Ensure data availability and quality for BA/BI needs</p>
<p><b>DSRMP04 – generally relevant</b>            Undertake creative work, making systematic use of investigation or experimentation, to discover or revise knowledge of reality, and use this knowledge to devise new applications (data driven), contribute to the development of organisational or project objectives</p>	<p><b>DSBA04 – relevant for organisation processes and data</b>            Analyse opportunity and suggest the use of historical data available at organisation for organisational processes optimisation</p> <ul style="list-style-type: none"> <li>• Coordinate implementation of FAIR data principles for collected data, ensure proper lineage and provenance of collected data</li> </ul>
<p><b>DSRMP05 – extended, essential</b>            Design experiments which include data collection (passive and active) for hypothesis testing and problem solving</p> <ul style="list-style-type: none"> <li>• Work with Data Science, Data Stewardship and data infrastructure teams to develop project/research goals.</li> </ul>	<p><b>DSBA05 – relevant for organisation processes and data</b>            Analyse customer relations data to optimise/ improve interaction with the specific user groups or in the specific business sectors</p>

Data Management (DSDM)	Data Science Engineering (DSENG)
<p><b>DSDM06 – extended, essential</b> Develop and manage/supervise policies on data protection, privacy, IPR and ethical issues in data management, address legal issues if necessary.</p> <ul style="list-style-type: none"> <li>• Ensure GDPR compliance in data management and access</li> <li>• Develop data access policies and coordinate their implementation and monitoring, including security breaches handling</li> </ul>	<p><b>DSENG06– extended, essential</b> Design, build, operate relational and non-relational databases (SQL and NoSQL), integrate them with the modern Data Warehouse solutions, ensure effective ETL (Extract, Transform, Load) and ELT (Extract, Load, Transform), OLTP, OLAP processes for large datasets</p> <ul style="list-style-type: none"> <li>• Define, implement and maintain data model, reference data, master data definitions, implement consistent metadata</li> </ul>
<p><b>DSDM07* - added new, essential</b> Manage Data Management/Data Stewards team, coordinate related activity between organisational departments, external stakeholder to fulfil Data Governance policy requirements, provide advice and training to staff. Define domain/organisation specific data management requirements, communicate to all departments and supervise/coordinate their implementation. Coordinate/supervise data acquisition.</p>	
<p><b>DSDM08* - added new, essential</b> Develop organisational policy and coordinate activities for sustainable implementation of the FAIR data principles and Open Science, define corresponding requirements to data infrastructure and tools, ensure organisational awareness.</p>	
<p><b>DSDM09* - added new, essential</b> Specify requirements to and supervise the organisational infrastructure for data management and (and archiving), maintain the park for data management tools, provide support to staff (researchers or business developers), coordinate solving problems.</p>	



<b>Data Science Research Methods and Project Management (DSRMP)</b>	<b>Data Science Domain Knowledge (DSDK) as Business Process Management (DSBA)</b>
<p><b>DSRMP06 – extended, essential</b>                      Develop and guide data driven projects, including project planning, experiment design, data collection and handling</p>	<p><b>DSBA06 – relevant for organisation processes and data</b>                      Analyse multiple data sources for marketing purposes; identify effective marketing actions</p>
	<p><b>DSBA07 – added, essential</b>                      Coordinate intra organisational activities related to data analytics, data management and data provenance/lineage along all data flow stages, ensure data FAIRness</p>



## Appendix D – Draft Body of Knowledge (supplement to FAIR Competence Framework)

as of 27 May 2021

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG1-DSDA: Data Science Analytics	DSDA.01/ SMDA	KA01.01	Statistical methods for data analysis

Suggested Knowledge Units (KU)	
KU1.01.00	Statistical methods overview
KU1.01.01	Probability & Statistics
KU1.01.02	Statistical paradigms (regression, time series, dimensionality, clusters)
KU1.01.03	Probabilistic representations (causal networks, Bayesian analysis, Markov nets)
KU1.01.04	Frequentist and Bayesian statistics
KU1.01.05	Probabilistic reasoning
KU1.01.06	Exploratory and confirmatory data analysis
KU1.01.07	Quantitative analytics
KU1.01.08	Qualitative Analytics
KU1.01.09	Data preparation and preprocessing
KU1.01.10	Performance analysis
KU1.01.11	Markov models, markov networks
KU1.01.12	Operations research
KU1.01.13	Information theory
KU1.01.14	Discrete Mathematics and Graph Theory
KU1.01.15	Mathematical analysis
KU1.01.16	Mathematical software and tools

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG1-DSDA: Data Science Analytics	DSDA.02/ ML	KA01.02	Machine Learning

Suggested Knowledge Units (KU)	
KU1.02.00	Machine Learning methods overview and use cases
KU1.02.01	Machine Learning theory and algorithms
KU1.02.02	Supervised Machine Learning
KU1.02.03	Unsupervised Machine Learning
KU1.02.04	Reinforced learning
KU1.02.05	Classification methods
KU1.02.06	Design and Analysis of Algorithms
KU1.02.07	Game Theory & Mechanism design
KU1.02.08	Artificial Intelligence
KU1.01.02	Statistical paradigms (regression, time series, dimensionality, clusters)
KU1.01.03	Probabilistic representations (causal networks, Bayesian analysis, Markov nets)
KU1.01.04	Frequentist and Bayesian statistics
KU1.01.05	Probabilistic reasoning
KU1.01.08	Performance analysis

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG1-DSDA: Data Science Analytics	DSDA.03/ DM	KA01.03	Data Mining

Suggested Knowledge Units (KU)	
KU1.03.00	Data Mining methods and technologies overview
KU1.01.08	Performance analysis
KU1.02.01	Machine Learning theory and algorithms
KU1.02.02	Supervised Machine Learning
KU1.02.03	Unsupervised Machine Learning
KU1.02.04	Reinforced learning
KU1.02.05	Classification methods
KU1.03.01	Data mining and knowledge discovery
KU1.03.02	Knowledge Representation and Reasoning
KU1.03.03	CRISP-DM and data mining stages
KU1.03.04	Anomaly Detection

KU1.03.05	Time series analysis
KU1.03.06	Feature selection, Apriori algorithm
KU1.03.07	Graph data analytics

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG1-DSDA: Data Science Analytics	DSDA.04/ TDM	KA01.04	Text Data Mining

Suggested Knowledge Units (KU)	
KU1.04.00	Text Data Mining overview
KU1.04.01	Text analytics including statistical, linguistic, and structural techniques to analyse structured and unstructured data
KU1.04.02	Data mining and text analytics
KU1.04.03	Natural Language Processing
KU1.04.04	Predictive Models for Text
KU1.04.05	Retrieval and Clustering of Documents
KU1.04.06	Information Extraction
KU1.04.07	Sentiments analysis

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG1-DSDA: Data Science Analytics	DSDA.05/ PA	KA01.05	Predictive Analytics

Suggested Knowledge Units (KU)	
KU1.05.00	Predictive analytics methods overview
KU1.05.01	Predictive modeling and analytics
KU1.05.02	Inferential and predictive statistics
KU1.05.03	Machine Learning for predictive analytics
KU1.05.04	Prescriptive Analytics
KU1.05.05	Regression and Multi Analysis
KU1.05.06	Generalised linear models
KU1.05.07	Time series analysis and forecasting
KU1.05.08	Deploying and refining predictive models

<b>Knowledge Area Groups (KAG)</b>			<b>Knowledge Areas (KA)</b>
KAG1-DSDA: Data Science Analytics	DSDA.06/ MODSIM	KA01.06	Computational modelling, simulation and optimisation

<b>Suggested Knowledge Units (KU)</b>	
KU1.06.01	Modelling and simulation theory and techniques (general and domain oriented)
KU1.06.02	Operations research and optimisation
KU1.06.03	Large scale modelling and simulation systems
KU1.06.04	Network optimisation
KU1.06.05	Risk simulation and queueing

<b>Knowledge Area Groups (KAG)</b>			<b>Knowledge Areas (KA)</b>
KAG1-DSDA: Data Science Analytics	DSDA.07/ DAVIZ	KA01.07	Data Analytics Visualisation and Story Telling

<b>Suggested Knowledge Units (KU)</b>	
KU1.07.01	Data Analytics Visualisation Methods
KU1.07.02	Data Analytics Visualisation Tools and Software (desktop and cloud based)
KU1.07.03	Storytelling best practices, dashboards and reports design

<b>Knowledge Area Groups (KAG)</b>			<b>Knowledge Areas (KA)</b>
KAG2-DSENG: Data Science Engineering	DSENG.01/ BDI	KA02.01	Big Data Infrastructure and Technologies

<b>Suggested Knowledge Units (KU)</b>	
KU2.01.00	Big Data Infrastructure Technologies and tools overview
KU2.01.01	Computer systems organisation for Big Data applications, CAP, BASE and ACID theorems
KU2.01.02	Parallel and Distributed Computer Architecture
KU2.01.03	High Performance and Cloud Computing
KU2.01.04	Clouds and scalable computing
KU2.01.05	Cloud based Big Data platforms and services
KU2.01.06	Big Data (large scale) storage and filesystems (HDFS, Ceph, etc)
KU2.01.07	NoSQL databases
KU2.01.08	Computer networks for high-performance computing and Big Data infrastructure
KU2.01.09	Computer networks: architectures and protocols
KU2.01.10	Big Data Infrastructure management and operation

<b>Knowledge Area Groups (KAG)</b>			<b>Knowledge Areas (KA)</b>
KAG2-DSENG: Data Science Engineering	DSENG.02/ DSIAPP	KA02.02	Infrastructure and platforms for Data Science applications

<b>Suggested Knowledge Units (KU)</b>	
KU2.02.00	Overview Infrastructure and platforms for Data Science applications
KU2.02.01	Big Data Infrastructure: services and components, including data storage infrastructure
KU2.02.02	Big Data analytics platforms and tools (including Hadoop, Spark, and cloud based Big Data services)
KU2.02.03	Large scale cloud based storage and data management
KU2.02.04	Cloud based applications and services operation and management
KU2.02.05	Big Data and cloud based systems design and development, including tools
KU2.02.06	Data processing models (batch, streaming, parallel)
KU2.02.07	Enterprise information systems
KU2.02.08	Data security and protection

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG2-DSENG: Data Science Engineering	DSENG.03/ CCT	KA02.03	ICloud Computing technologies for Big Data and Data Analytics

Suggested Knowledge Units (KU)	
KU2.03.01	Cloud Computing architecture and services
KU2.03.02	Cloud Computing Engineering (infrastructure and services design, management and operation)
KU2.03.03	Cloud based applications and services operation and management

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG2-DSENG: Data Science Engineering	DSENG.04/ SEC	KA02.04	Data and Applications security

Suggested Knowledge Units (KU)	
KU2.04.01	Infrastructure, applications and data security
KU2.04.02	Data encryption and key management, blockchain based technologies
KU2.04.03	Access Control and Identity Management
KU2.04.04	Security services management, including compliance and certification
KU2.04.05	Data anonymisation
KU2.04.06	Personal data protection, GDPR compliance control

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG2-DSENG: Data Science Engineering	DSENG.05/ BDSE	KA02.05	Big Data systems organisation and engineering

Suggested Knowledge Units (KU)	
KU2.05.00	Big Data systems organisation and design Overview
KU2.05.01	Big Data systems organisation and design
KU2.05.02	Big Data algorithms for large scale data processing
KU2.05.03	Big Data Analytics
KU2.05.04	Big Data analytics platforms and tools (including Hadoop, Spark, and cloud based Big Data services)
KU2.05.05	Big Data algorithms for data ingest, pre-processing, and visualisation
KU2.05.06	Big Data systems for application domains
KU2.05.07	Big Data software (systems) architectures
KU2.05.08	Requirements engineering and software systems development



KU2.05.09	Large and ultra-large scale software systems organisation
KU2.05.10	DevOps and cloud enabled applications development
KU2.05.11	Big Data Infrastructure management and operation

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG2-DSENG: Data Science Engineering	DSENG.06/ DSAPPD	KA02.06	Data Science (Big Data) applications design

Suggested Knowledge Units (KU)	
KU2.06.01	Data analytics, data handling software requirements and design
KU2.06.02	Applications engineering management
KU2.06.03	Software engineering models and methods
KU2.06.04	Software quality assurance
KU2.06.05	Programming languages for Big Data analytics: R, python, Pig, Hive, others
KU2.06.06	Models and languages for complex interlinked data presentation and visualisation
KU2.06.07	Agile development methods, platforms and tools
KU2.06.08	DevOps and continuous deployment and improvement paradigm

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG2-DSENG: Data Science Engineering	DSENG.07/ IS	KA02.07	Information systems (to support data-driven decision making)

Suggested Knowledge Units (KU)	
KU2.07.01	Decision Analysis and Decision Support Systems
KU2.07.02	Predictive analytics and predictive forecasting
KU2.07.03	Data Analysis and statistics
KU2.07.04	Data warehousing and Data Mining
KU2.07.05	Data Mining
KU2.07.06	Multimedia information systems
KU2.07.07	Enterprise information systems
KU2.07.08	Collaborative and social computing systems and tools

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG3-DSDM: Data Management	DSDM.01/ DMORG	KA03.01	General principles and concepts in Data Management and organisation

Suggested Knowledge Units (KU)	
KU3.01.00	General principles and concepts in Data Management - Overview
KU3.01.01	Overview Data type, data type registries, data formats
KU3.01.02	Metadata, metadata formats, metadata standards, metadata registries
KU3.01.03	Data Lifecycle Management
KU3.01.04	Data Factories and data infrastructure
KU3.01.05	Open Science, Open Access, Open Data
KU3.01.06	Metadata registries, publishing metadata
KU3.01.07	Persistent Identifiers (PID), Open Researcher and Contributor ID (ORCID), Research Organization Registry (ROR)
KU3.01.08	Ethical principle and data privacy
	FAIR (Findable, Accessible, Interoperable, Reusable) principles in Data Management
KU3.01.09	FAIR metadata management, tools for FAIR metadata management

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG3-DSDM: Data Management	DSDM.02/ DMS	KA03.02	Data management systems

Suggested Knowledge Units (KU)	
KU3.02.00	Overview data management systems
KU3.02.01	Data Warehouse architecture and processes (OLAP, OLTP, ETL, ELT)
KU3.02.02	Databases and Database Management Systems, Data Modelling
KU3.02.03	Data structures
KU3.02.04	Data Models and Query Languages, sql
KU3.02.05	Database design and database models
KU3.02.06	Database administration
KU3.02.07	Enterprise Data Warehouses, architectural components and popular platforms
KU3.02.08	Middleware for databases
KU3.02.09	Master Data Management, Data Dictionaries
KU3.02.10	FAIR data management requirements and compliance
KU3.02.11	User data management tools and user support

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG3-DSDM: Data Management	DSDM.03/ EDMI	KA03.03	Data Management and Enterprise data infrastructure

Suggested Knowledge Units (KU)	
KU3.03.00	Overview Data Management and Enterprise Data Infrastructure
KU3.03.01	Data management, including Reference and Master Data
KU3.03.02	Data Warehousing and Business Intelligence
KU3.03.03	Data storage and operations
KU3.03.04	Data archives/storage compliance and certification
KU3.03.05	Metadata, linked data, provenance
KU3.03.06	Data infrastructure, data registries and data factories
KU3.03.07	Data security and protection
KU3.03.08	Data backup
KU3.03.09	Data anonymisation
KU3.03.10	Personal data protection, GDPR compliance

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG3-DSDM: Data Management	DSDM.04/ DGOV	KA03.04	Data Governance

Suggested Knowledge Units (KU)	
KU3.04.00	Overview Data governance principles and organisation
KU3.04.01	Data Governance Policy, KPI (Key Performance Indicators), best practices
KU3.04.02	Data Management Planning, FAIR data management and compliance
KU3.04.03	Data Integration and Interoperability, Data preparation and data cleaning
KU3.04.04	Data interoperability and metadata management
KU3.04.05	Organisational roles in data governance, data stewardship
KU3.04.06	Data provenance, data lineage
KU3.04.07	Responsible data use, data privacy, ethical principles, IPR, legal issues
KU3.04.08	Data quality management, best practices and frameworks, data quality metrics
KU3.04.09	Data infrastructure compliance and certification, compliance standards
KU3.04.10	Data protection policies (including personal data), data access policies, GDPR compliance
KU3.04.11	User needs analysis and definition of requirements to supporting infrastructure and tools
KU3.04.12	Data management costs, funding models, budgeting

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG3-DSDM: Data Management	DSDM.05/ BDSTOR	KA03.05	Big Data storage (large scale)

Suggested Knowledge Units (KU)	
KU3.05.00	Big Data storage Overview
KU3.05.01	Cloud-based Big Data storage, Data Lakes, Data Fabrics
KU3.05.02	Storage architectures, distributed files systems (HDFS, RAID, Ceph, Lustre, Gluster, etc.)
KU3.05.03	Data storage redundancy, replication and backup

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG3-DSDM: Data Management	DSDM.06/ DLIB	KA03.06	Data archives and data libraries

Suggested Knowledge Units (KU)	
KU3.06.01	Data archives and data libraries organisation
KU3.06.02	Information retrieval
KU3.06.03	Data curation and provenance
KU3.06.04	Search Engines technologies
KU3.06.05	Metadata management and publication
KU3.06.06	Trusted data repositories and certification

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG4-DSRMP: Research Methods and Project Management	DSRMP.01/ RM	KA04.01	Research Methods and Research data

Suggested Knowledge Units (KU)	
KU4.01.01	Research methods and research cycle, research questions and hypothesis evaluation
KU4.01.02	Research types and research process models
KU4.01.03	Modelling and experiment planning
KU4.01.04	Research data collection and quality assessment
KU4.01.05	Data discovery (published data), data selection and use in research
KU4.01.06	Data lifecycle management and data provenance
KU4.01.07	Research data management plan and ethical issues
KU4.01.08	Use cases analysis: research infrastructures and projects

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG4-DSRMP: Research Methods and Project Management	DSRMP.02/ PM	KA04.02	Research Project Management

Suggested Knowledge Units (KU)	
KU4.02.01	Research Project Management based on general Project Management practices
KU4.02.02	Project Scope Management
KU4.02.03	Project Quality and KPI (Key Performance Indicators)
KU4.02.04	Project Risk Management
KU4.02.05	Grant application and management
KU4.02.06	European Research Area. Open Science, Open data and FAIR data sharing
KU4.01.07	Research data management plan and ethical issues
KU4.01.08	Use cases analysis: research infrastructures and projects

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG5-DSBPM: Business Analytics	DSBA.01/ BAF	KA05.01	Business Analytics Founda- tion

Suggested Knowledge Units (KU)	
KU5.01.00	Business Analytics and Business Intelligence: Overview
KU5.01.01	Business Analytics and Business Intelligence: Data, Models (statistical) and Decisions
KU5.01.02	Data-driven Customer Relations Management (CRP), User Experience (UX) requirements and design
KU5.01.03	Operations Analytics
KU5.01.04	Business Process Optimisation and effective data management
KU5.01.05	Data Warehouses technologies, data modelling, data integration (from multiple sources, including historical data) and analytics
KU5.01.06	Data-driven marketing technologies
KU5.01.07	Business Analytics Capstone
KU5.01.08	Econometrics methods and application for Business Analytics
KU5.01.09	Cognitive technologies for Business Analytics

Knowledge Area Groups (KAG)			Knowledge Areas (KA)
KAG6-DSBA: Business Analytics	DSBA.02/ BAEM	KA05.02	Business Analytics organisati- on and enterprise management

Suggested Knowledge Units (KU)	
KU5.02.01	Business processes and operations
KU5.02.02	Project scope and risk management
KU5.02.03	Business Analysis Planning and Monitoring
KU5.02.04	Requirements Analysis and Design Definition
KU5.02.05	Requirements Life Cycle Management (from inception to retirement)
KU5.02.06	Solution Evaluation and improvements recommendation
KU5.02.07	Agile Data-Driven methodologies, processes and enterprises
KU5.02.08	Use cases analysis: business and industry
KU5.02.09	Data management for BA/BI (Business Analytics, Business Intelli- gence), organisational models and requirements
KU5.02.10	Data quality management, FAIR data principles for organisational data

**Appendix E – Knowledge units and corresponding  
learning outcomes for the bachelor’s, master’s and  
PhD degree levels**

Content/topic from [FAIR Competences BOK], based on: EDISON Data Science Framework	<b>basic learning outcomes</b>	<b>intermediate learning outcomes</b> (include and build on basic learning outcomes)	
General principles and concepts in data management – overview	<ul style="list-style-type: none"> <li>• Can define Research Data Management (RDM) and can describe its relevance and benefits.</li> </ul>	<ul style="list-style-type: none"> <li>• Can describe RDM measures to be taken (including explaining why) at different stages of the research process.</li> </ul>	
Overview of data types, data type registries and data formats	<ul style="list-style-type: none"> <li>• Can describe what types of data exist (Knowledge).</li> <li>• Can explain what data type registries are (Knowledge).</li> <li>• Can identify data formats (Knowledge).</li> <li>• Can search and find data formats in registries.</li> </ul>	<ul style="list-style-type: none"> <li>• Can determine proper data types for a resource (Analyse).</li> <li>• Can use a data type registry (Apply).</li> <li>• Can use proper data formats to express resources (Apply).</li> </ul>	
Metadata, metadata formats, standards and registries	<ul style="list-style-type: none"> <li>• Can describe types of metadata.</li> <li>• Can recognise metadata formats.</li> <li>• Can identify metadata standards.</li> <li>• Can use metadata standards to describe resources.</li> <li>• Can explain what metadata registries are.</li> <li>• Can search and find data and metadata standards in registries.</li> </ul>	<ul style="list-style-type: none"> <li>• Can articulate metadata of different types to describe a resource.</li> <li>• Can write metadata in a relevant format.</li> <li>• Can appraise the usefulness of metadata standards to describe a resource.</li> <li>• Can search metadata registries to find resources.</li> </ul>	
Open Research, Open Access, Open Data	<ul style="list-style-type: none"> <li>• Can paraphrase the concept of Open Research.</li> <li>• Can describe the benefits of Open Research.</li> <li>• Can describe Open Access and Open Data as areas of Open Research.</li> </ul>	<ul style="list-style-type: none"> <li>• Can recognise if a publication is open access.</li> <li>• Can discover platforms for Open Access/Open Data.</li> <li>• Can articulate what is required to make research outputs open.</li> <li>• Can contrast FAIR and Open.</li> </ul>	



advanced learning outcomes (include and build on intermediate learning outcomes)	Bachelor	Master	PhD	Entry- Level Content?
<ul style="list-style-type: none"> <li>• Can practically apply theoretical knowledge about proper RDM measures to be taken at different stages to their own research process/project.</li> </ul>	basic	inter- mediate	advanced	Yes
None.	basic	basic	inter- mediate	Yes
<ul style="list-style-type: none"> <li>• Can design rich metadata to describe a resource.</li> <li>• Can use proper metadata formats and models to express these metadata.</li> <li>• Can deposit metadata in a repository.</li> </ul>	basic	inter- mediate	advanced	Yes
<ul style="list-style-type: none"> <li>• Can plan publication of Open Access publications and FAIR data.</li> </ul>	basic	inter- mediate	advanced	Yes

Content/topic from [FAIR Competences BOK], based on: EDISON Data Science Framework	<b>basic learning outcomes</b>	<b>intermediate learning outcomes</b> (include and build on basic learning outcomes)
Metadata management, registries and publication	<ul style="list-style-type: none"> <li>• Can explain aspects of metadata management and the publication process in metadata registries.</li> </ul>	<ul style="list-style-type: none"> <li>• Can perform basic steps related to metadata management.</li> <li>• Can execute steps in metadata publication.</li> </ul>
Persistent Identifiers (PID), Open Researcher and Contributor ID (ORCID), Research Organization Registry (ROR)	<ul style="list-style-type: none"> <li>• Can recognise PIDs and explain the different use cases for PIDs.</li> <li>• Can explain the importance of PIDs for FAIR data.</li> <li>• Can use PIDs to access data or other resources.</li> </ul>	<ul style="list-style-type: none"> <li>• Can apply PIDs to their own research outputs.</li> <li>• Can use PIDs to collaborate with others.</li> </ul>
FAIR (Findable, Accessible, Interoperable, Reusable) principles in data management	<ul style="list-style-type: none"> <li>• Can paraphrase the FAIR principles.</li> <li>• Can explain why the FAIR principles were developed.</li> <li>• Can recognise the relationship between FAIR, RDM and Open.</li> </ul>	<ul style="list-style-type: none"> <li>• Can plan for FAIR research outputs.</li> <li>• Can write and develop a research data management plan.</li> <li>• Can apply the principles to their own work.</li> <li>• Can evaluate the FAIRness of their own work or the work of others.</li> </ul>
FAIR metadata management and tools for FAIR metadata management	<ul style="list-style-type: none"> <li>• Can name aspects related to FAIR metadata management.</li> <li>• Can give an example of tools for FAIR metadata management.</li> </ul>	<ul style="list-style-type: none"> <li>• Can describe aspects of metadata management to comply with FAIR.</li> <li>• Can work with one of the FAIR metadata management tools.</li> </ul>

advanced learning outcomes (include and build on intermediate learning outcomes)	Bachelor	Master	PhD	Entry- Level Content?
<ul style="list-style-type: none"> <li>• Can select appropriate metadata formats and a metadata registry appropriate for the subject domain of a research project.</li> </ul>	basic	basic	inter- mediate	No
None.	basic	basic	inter- mediate	Yes
None.	basic	basic	inter- mediate	Yes
<ul style="list-style-type: none"> <li>• Is able to support FAIR metadata management for the selected subject domain.</li> <li>• Can assess and select tools for FAIR metadata management.</li> </ul>	basic	inter- mediate	advanced	No

Content/topic from [FAIR Competences BOK], based on: EDISON Data Science Framework	<b>basic learning outcomes</b>	<b>intermediate learning outcomes</b> (include and build on basic learning outcomes)
Databases and data-base management systems, data modelling	<ul style="list-style-type: none"> <li>• Can explain what a database is, including common database terminology.</li> <li>• Can explain and list some of the advantages and disadvantages of using databases.</li> <li>• Can distinguish between databases and spreadsheets.</li> <li>• Can recall basic concept of data modelling.</li> </ul>	<ul style="list-style-type: none"> <li>• Can identify basic database classifications and discuss their differences.</li> <li>• Can recall the most common database models and discuss their usage.</li> <li>• Understands how a relational database is designed, created, used, and maintained.</li> <li>• Is able to build and assess data-based models.</li> </ul>
Data structures	<ul style="list-style-type: none"> <li>• Understands and can restate the fundamentals of basic data structures.</li> <li>• Is able to implement and apply data structures.</li> </ul>	<ul style="list-style-type: none"> <li>• Is able to describe the usage of various data structures algorithms.</li> <li>• Is able to explain and summarise the advantages and disadvantages of various data structures implementations.</li> </ul>
Master data management, data dictionaries	<ul style="list-style-type: none"> <li>• Can develop a data management plan for their own work.</li> <li>• Can identify different types of data documentation.</li> <li>• Can explain the purpose of the documentation.</li> <li>• Can use existing documentation.</li> </ul>	<ul style="list-style-type: none"> <li>• Can modify existing documentation.</li> <li>• Can evaluate and prioritise data management activities.</li> </ul>

advanced learning outcomes (include and build on intermediate learning outcomes)	Bachelor	Master	PhD	Entry- Level Content?
<ul style="list-style-type: none"> <li>• Is able to design and implement own databases.</li> </ul>	basic	basic	basic	No
<ul style="list-style-type: none"> <li>• Is able to analyse the performance characteristics of algorithms using mathematical and measurement techniques.</li> <li>• Is able to design and apply appropriate data structures for solving computing problems.</li> </ul>	basic	basic	basic	No
None.	basic	basic	inter- mediate	Yes

Content/topic from [FAIR Competences BOK], based on: EDISON Data Science Framework	<b>basic learning outcomes</b>	<b>intermediate learning outcomes</b> (include and build on basic learning outcomes)
FAIR data management requirements and compliance	<ul style="list-style-type: none"> <li>• Can name the main stakeholders or parties that potentially mandate FAIR compliance and data management measures.</li> <li>• Can list FAIR data management requirements.</li> </ul>	<ul style="list-style-type: none"> <li>• Can identify the FAIR and RDM requirements that are relevant for the own research context.</li> <li>• Can explain where to get support with regard to RDM and the FAIR principles.</li> <li>• Can plan proper measures for RDM and making data FAIR (with support if necessary).</li> <li>• Can apply proper measures for RDM and making data FAIR (with support if necessary).</li> </ul>
Data management, including reference and master data	<ul style="list-style-type: none"> <li>• Can define reference and master data.</li> <li>• Understands the critical roles reference and master data play in data management.</li> </ul>	<ul style="list-style-type: none"> <li>• Can describe different Master Data Management (MDM) architectures and their suitability for different needs.</li> </ul>
Data storage and operations	<ul style="list-style-type: none"> <li>• Can identify different options for data storage and their operational aspects .</li> <li>• Can state different types and functions of storage systems.</li> </ul>	<ul style="list-style-type: none"> <li>• Can specify and explain requirements regarding data storage for specific data or organisational processes.</li> </ul>
Data infrastructure, data registries and data factories	<ul style="list-style-type: none"> <li>• Can list existing infrastructure elements and services required to support consistent data management and handling.</li> </ul>	<ul style="list-style-type: none"> <li>• Can specify and explain requirements with regard to the data infrastructure and its components for specific data or organisational data.</li> </ul>

advanced learning outcomes (include and build on intermediate learning outcomes)	Bachelor	Master	PhD	Entry- Level Content?
<ul style="list-style-type: none"> <li>• Can plan proper measures for RDM and making data FAIR (without support).</li> <li>• Can apply proper measures for RDM and making data FAIR (without support).</li> </ul>	irrelevant	basic	inter- mediate	No
<ul style="list-style-type: none"> <li>• Is able to design a Data Governance Framework and to manage master and reference data.</li> </ul>	irrelevant	basic	basic	No
<ul style="list-style-type: none"> <li>• Can compare different storage options.</li> <li>• Can select and justify a data storage solution for a project or organisation.</li> </ul>	basic	inter- mediate	advanced	No
<ul style="list-style-type: none"> <li>• Can compare different infrastructure solutions.</li> <li>• Can select and justify a data storage solutions for a project or organisation.</li> <li>• Understands the role and functions of the data factories.</li> </ul>	basic	basic	inter- mediate	No

Content/topic from [FAIR Competences BOK], based on: EDISON Data Science Framework	<b>basic learning outcomes</b>	<b>intermediate learning outcomes</b> (include and build on basic learning outcomes)	
Data security and protection	<ul style="list-style-type: none"> <li>• Can define different levels of data security (user, folder, files).</li> <li>• Can explain different ways of data protection (physical, encryption etc.).</li> </ul>	<ul style="list-style-type: none"> <li>• Can use different levels of security for their own work.</li> <li>• Can apply data protection methods like password protection and encoding.</li> <li>• Does share and collaborate in a secure way.</li> </ul>	
Data backup	<ul style="list-style-type: none"> <li>• Can describe what a backup is and tell reasons for backup creation.</li> <li>• Can explain the 3-2-1 rule and apply it to their own files.</li> <li>• Can identify institutional backup solutions.</li> </ul>	<ul style="list-style-type: none"> <li>• Can explain institutional backup solutions and apply them to own files.</li> </ul>	
Personal data protection, GDPR compliance	<ul style="list-style-type: none"> <li>• Can explain reasons for data protection.</li> <li>• Knows basic rules and legal regulations for sensitive data (e.g. GDPR).</li> <li>• Knows how to comply with these rules and laws.</li> </ul>	<ul style="list-style-type: none"> <li>• Can analyse compliance to legal regulations for sensitive data.</li> <li>• Can apply mechanisms to protect data appropriately.</li> </ul>	
Data anonymisation/pseudonymisation	<ul style="list-style-type: none"> <li>• Can describe directly identifying attributes and detect them in data.</li> <li>• Can explain the difference between anonymisation and pseudonymisation.</li> </ul>	<ul style="list-style-type: none"> <li>• Can anonymise/pseudonymise data by stripping identifying attributes.</li> </ul>	



advanced learning outcomes (include and build on intermediate learning outcomes)	Bachelor	Master	PhD	Entry- Level Content?
None.	basic	basic	inter- mediate	Yes
<ul style="list-style-type: none"> <li>• Can analyse and evaluate backup.</li> <li>• Can solve backup problems independently or with further assistance from support personnel.</li> </ul>	basic	inter- mediate	advanced	Yes
None.	basic	basic	inter- mediate (depending on disci- pline)	Yes
None.	irrelevant	basic (de- pending on disci- pline)	inter- mediate (depending on disci- pline)	No

Content/topic from [FAIR Competences BOK], based on: EDISON Data Science Framework	<b>basic learning outcomes</b>	<b>intermediate learning outcomes</b> (include and build on basic learning outcomes)	
Data management planning, FAIR data management and compliance	<ul style="list-style-type: none"> <li>• Can describe what a data management plan (DMP) is.</li> <li>• Can explain why data management planning is a step towards FAIR.</li> </ul>	<ul style="list-style-type: none"> <li>• Can tell which areas should be covered in a DMP.</li> <li>• Can sketch a DMP for their own research project.</li> </ul>	
Data integration and interoperability, data preparation and data cleaning	<ul style="list-style-type: none"> <li>• Can explain aspects related to data interoperability and integration.</li> <li>• Can explain aspects of data preparation and cleaning.</li> </ul>	<ul style="list-style-type: none"> <li>• Can perform basic tasks in data interoperability and integration.</li> <li>• Can perform basic tasks in data preparation and cleaning.</li> </ul>	
Data interoperability and metadata management	<ul style="list-style-type: none"> <li>• Can explain aspects of interoperability (Knowledge).</li> <li>• Can relate metadata management to interoperability (Understand).</li> </ul>	<ul style="list-style-type: none"> <li>• Use domain-relevant standards, models and formats for interoperable data (Apply).</li> <li>• Can relate metadata management to interoperability (Apply).</li> </ul>	
Organisational roles in data governance, data stewardship	<ul style="list-style-type: none"> <li>• Can define data governance and name its components.</li> <li>• Can name different roles involved in data governance.</li> </ul>	<ul style="list-style-type: none"> <li>• Can name roles and structures in data governance and knows how they work together.</li> <li>• Can recall goals and added value of data governance.</li> </ul>	

advanced learning outcomes (include and build on intermediate learning outcomes)	Bachelor	Master	PhD	Entry- Level Content?
<ul style="list-style-type: none"> <li>• Can develop a detailed DMP according to funder requirements and engage with relevant university instances/authorities.</li> <li>• Can collaborate on a DMP and modify the plan during the project progress (“living document”).</li> <li>• Can apply principles to protect personal sensitive data and develop Data Protection Impact Assessment, if required. (depending on discipline).</li> </ul>	basic	basic	inter- mediate	Yes
<ul style="list-style-type: none"> <li>• Can select best solutions/standards for data interoperability.</li> <li>• Can select appropriate tools and methods for data integration.</li> <li>• Can select appropriate methods and tools for data preparation and cleaning.</li> </ul>	basic	inter- mediate	advanced	No
None.	basic	basic	inter- mediate	Yes (very basic)
<ul style="list-style-type: none"> <li>• Can develop strategies to successfully embed data governance in an organisation.</li> </ul>	basic	basic	inter- mediate	No

Content/topic from [FAIR Competences BOK], based on: EDISON Data Science Framework	<b>basic learning outcomes</b>	<b>intermediate learning outcomes</b> (include and build on basic learning outcomes)
Data provenance, data lineage	<ul style="list-style-type: none"> <li>• Can illustrate with an example what data provenance/data lineage means.</li> </ul>	<ul style="list-style-type: none"> <li>• Can transfer how data provenance/data lineage plays a role in their own research project.</li> <li>• Can apply data provenance good practices to their own data and ensure that an unbroken data lineage is established for their work.</li> </ul>
Responsible data use, data privacy, ethical principles, IPR and legal issues	<ul style="list-style-type: none"> <li>• Can summarise and explain ethical principles and responsible data use (e.g. CARE, indigenous data).</li> <li>• Can describe legal issues around data use and management (e.g. licences, patents, policies, contracts etc.).</li> </ul>	<ul style="list-style-type: none"> <li>• Can analyse if ethical principles or legal issues play a role in their own work.</li> </ul>
Data quality management, best practices and frameworks, data quality metrics	<ul style="list-style-type: none"> <li>• Can summarise best practices ensuring data quality.</li> </ul>	<ul style="list-style-type: none"> <li>• Can describe how to recognise quality data.</li> </ul>
Data protection policies (including personal data), data access policies, GDPR compliance	<ul style="list-style-type: none"> <li>• Can state general requirements on data protection and access control.</li> <li>• Can give examples of policies related to data protection and access control.</li> <li>• Can list the main aspects related to the GDPR.</li> </ul>	<ul style="list-style-type: none"> <li>• Can explain general requirements on data protection and access control.</li> <li>• Can explain content and use of policies related to data protection and access control.</li> <li>• Can explain what are the main aspects related to GDPR in organisational data management.</li> </ul>

advanced learning outcomes (include and build on intermediate learning outcomes)	Bachelor	Master	PhD	Entry- Level Content?
<ul style="list-style-type: none"> <li>• Can use tools for data provenance management.</li> </ul>	basic	basic	inter- mediate	Yes
<ul style="list-style-type: none"> <li>• Can detect ethical or legal issues and solve them together with ethical and legal experts like e.g. ethics committee, data protection officers or lawyers from the institution.</li> </ul>	basic	inter- mediate	advanced	Yes
<ul style="list-style-type: none"> <li>• Can use best practices and frameworks on their own data to ensure their quality.</li> </ul>	basic	inter- mediate	advanced	Yes (basic concept)
<ul style="list-style-type: none"> <li>• Can write specific policies related to data protection and access.</li> <li>• Can analyse compliance to GDPR in organisational data management.</li> </ul>	basic	basic	inter- mediate	No

Content/topic from [FAIR Competences BOK], based on: EDISON Data Science Framework	<b>basic learning outcomes</b>	<b>intermediate learning outcomes</b> (include and build on basic learning outcomes)	
Trusted data repositories and certification	<ul style="list-style-type: none"> <li>• Can explain what a trusted data repository is and how to find it (re3data.org and FAIRsharing).</li> <li>• Can compare different certifications for data repositories (e.g. CoreTrustSeal, CLARIN certification).</li> </ul>	<ul style="list-style-type: none"> <li>• Can discover trusted repositories and identify those that are certified.</li> </ul>	
Data discovery (published data), data selection and use in research (added from DRSMPM Knowledge Area Group)	<ul style="list-style-type: none"> <li>• Can explain the importance of data discovery and reuse.</li> </ul>	<ul style="list-style-type: none"> <li>• Can discover published datasets in their discipline.</li> <li>• Can cite data.</li> </ul>	
Research data lifecycle (added)	<ul style="list-style-type: none"> <li>• Can explain the steps of the research data lifecycle.</li> <li>• Can compare different lifecycle models.</li> </ul>	<ul style="list-style-type: none"> <li>• Can apply the research data lifecycle to their own work.</li> </ul>	
Ontologies, controlled vocabularies (added)	<ul style="list-style-type: none"> <li>• Can explain the role of ontologies and vocabularies (Knowledge).</li> <li>• Can recognise the use of ontologies and vocabularies (Knowledge)</li> <li>• Can identify a few domain-relevant ontologies (Knowledge).</li> <li>• Can search and find terminologies in registries.</li> </ul>	<ul style="list-style-type: none"> <li>• Can use ontologies to describe resources (Apply).</li> </ul>	

advanced learning outcomes (include and build on intermediate learning outcomes)	Bachelor	Master	PhD	Entry- Level Content?
<ul style="list-style-type: none"> <li>• Can use a trusted repository to share research output.</li> </ul>	basic	basic	inter- mediate	Yes (basic concept)
<ul style="list-style-type: none"> <li>• Can develop a strategy to search for data.</li> <li>• Can articulate criteria for data selection.</li> <li>• Can extract datasets and build their own work on them.</li> </ul>	basic	inter- mediate	advanced	Yes (basic concept)
None.	basic	basic	inter- mediate	Yes
<ul style="list-style-type: none"> <li>• Can use ontologies for search and analysis (Apply).</li> </ul>	basic	inter- mediate	advanced	Yes

## Appendix F – Lesson plans

*Mapping of the competence profile topics to the lesson plans*

<b>Topic</b>	<b>Number of relevant lesson plan</b> (number in parantheses: the plan partly addresses the topic)
General principles and concepts in data management – overview	15, (4)
Overview of data types, data type registries and data formats	(5)
Metadata, metadata formats, standards and registries	6
Open Research, Open Access, Open Data	15, (9)
Metadata management, registries and publication	(6)
Persistent Identifiers (PID), Open Researcher and Contributor ID (ORCID), Research Organization Registry (ROR)	8
FAIR (Findable, Accessible, Interoperable, Reusable) principles in data management	1, (16)
FAIR metadata management and tools for FAIR metadata management	(6)
Databases and database management systems, data modelling	(16)
Data structures	
Master data management, data dictionaries	3, (16)



<b>Topic</b>	<b>Number of relevant lesson plan</b> (number in parantheses: the plan partly addresses the topic)
FAIR data management requirements and compliance	(1)
Data management including reference and master data	(16)
Data storage and operations	
Data infrastructure, data registries and data factories	(16)
Data security and protection	(12)
Data backup	
Personal data protection, GDPR compliance	12, (15), (16)
Data anonymisation/pseudonymisation	12
Data management planning, FAIR data management and compliance	2, (1), (15), (16)
Data integration and interoperability, data preparation and cleaning	7
Data interoperability and metadata management	(7)
Organisational roles in data governance, data stewardship	(15),(16)
Data provenance, data lineage	(8)
Responsible data use, data privacy, ethical principles, Intellectual Property Rights (IPR) and legal issues	12, (15)
Data quality management, best practices and frameworks, data quality metrics	(2)
Data protection policies (including personal data), data access policies, GDPR (General Data Protection Regulation) compliance	12, 13
Trusted data repositories and certification	11
Data discovery (published data), data selection and use in research	10
Research data lifecycle	(4)
Ontologies and controlled vocabularies	7

## Lesson plan 1: FAIR in a nutshell

### FAIR elements:

#### Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

#### Accessible

Once the user finds the required data, they need to know how can they be accessed, possibly including authentication and authorisation.

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
  - A1.1 The protocol is open, free, and universally implementable
  - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

#### Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

#### Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
- R1.1. (Meta)data are released with a clear and accessible data usage license
- R1.2. (Meta)data are associated with detailed provenance
- R1.3. (Meta)data meet domain-relevant community standards

**Primary audience(s):** Bachelor's, master's, PhD degree students

**Learning outcomes:**

- Can paraphrase the FAIR principles
- Can explain why the FAIR principles were developed
- Can recognise the relationship between FAIR, RDM and Open
- Can plan for FAIR research outputs
- Can write and develop a research data management plan
- Can apply the principles to their own work
- Can evaluate the FAIRness of their own work or the work of others

**Summary of tasks/actions:**

1. Introduction to the FAIR principles
  - a. What is FORCE11 and where did the need to define the FAIR principles come from?
  - b. What do the FAIR principles stand for [Wilkinson et al. 2016]?
    - Findable
    - Accessible
    - Interoperable
    - Reusable
2. Explain the difference and overlap between FAIR, Open Data and research data management
  - a. Define Open Data
  - b. Define research data management
  - c. Show the relationship between FAIR, Open Data and RDM [Higman et al. 2019]
    - Intersections between the terms
    - Distinctions between the terms
3. How to make data FAIR? [The Top 10 FAIR data and software things; Knight 2015; PARTHENOS 2019]

- a. F is for making data findable
  - Look for existing data in repositories (see Lesson plan 10)
  - Upload to and share your data via a repository (see Lesson plan 11)
  - Describe your data with as much detail as possible (see Lesson plan 6)
  - Apply a persistent identifier (see Lesson plan 8)
- b. A is for making data accessible
  - Consider what can and will be shared under which conditions (see Lesson plan 13)
  - Obtain participant consent and perform risk management (see Lesson plan 12)
- c. I is for making data interoperable
  - Use open, standardised and common formats (see Lesson plan 5)
  - Consistent vocabulary
  - Apply common metadata standards (see Lesson plan 6)
  - Linked data
- d. R is for making data reusable
  - Consider permitted use
  - Apply appropriate license (see Lesson plan 9)
  - Add sufficient documentation and provenance information (see Lesson plan 3)
  - When using data of others, give credit by data citation (see Lesson plan 10)

**Materials/Equipment:**

- Computer/laptop
- Internet/browser

**References:**

- DeiC. Myths about FAIR. (Part of FAIR for Beginners). <https://www.deic.dk/en/data-management/instructions-and-guides/FAIR-for-Beginners>
- Higman, R., et al., 2019. Three Camps, One Destination: The Intersections of Research Data Management, FAIR and Open. *Insights* [online] 32 (1). <https://doi.org/10/gf4jhr>
- Jones, S. and Grootveld, M., 2017. *How FAIR are your data?* [online]. Zenodo. <http://doi.org/10.5281/zenodo.3405141>
- Knight, G., 2015. Preparing Data for Sharing: *The FAIR Principles* [online]. Presentation, 1 December 2015. Available at: <https://www.slideshare.net/lshmt/preparing-data-for-sharing-the-fair-principles>

- Martinez, P. A., Erdmann, C., Simons, N., Otsuji, R., Labou, S., Johnson, R., Castella, G., Boas, B. V., Lamprecht, A.-L., Ortiz, C. M., Garcia, L., Kuzak, M., Stokes, L., Honeyman, T., Wise, S., Quan, J., Peterson, S., Neeser, A., Karvovskaya, L., ... Fankhauser, E., 2019. *The Top 10 FAIR Data and Software things*. Things [online]. Zenodo. <https://doi.org/10.5281/ZENODO.3409968>
- PARTHENOS, Hollander, H., Morselli, F., Uiterwaal, F., Admiraal, F., Trippel, T., & Di Giorgio, S. (2019). *PARTHENOS Guidelines to FAIRify data management and make data reusable* [online]. Zenodo. <https://doi.org/10.5281/zenodo.3368858>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ..., and Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* [online] 3(9). <https://doi.org/10.1038/sdata.2016.18>

## Lesson plan 2: Data management plans (DMP)

**FAIR elements:** All (see Summary of Tasks/Actions 1. a) for more detail)

**Primary audience(s):** Bachelor's, Master's, PhD degree students

### Learning outcomes:

- Can describe what a data management plan is
- Can explain why data management planning is a step towards FAIR
- Can tell which areas should be covered in a DMP
- Can sketch a DMP for their own research project
- or (depending on scope and intensity of the lesson): Can develop detailed DMP according to funder requirements and engage with relevant university instances/ authorities
- Can collaborate on a DMP and modify the plan during the project ('living document')
- Can apply principles to protect personal sensitive data and develop a data protection impact assessment, if required (depending on discipline)
- Can summarise best practices in data quality (principles, benefits, standards and tools)
- Understands when it is appropriate to create plans and knows the difference between DMP and other types of documents for the project, e.g. project management plan
- Knows tools, guides, templates and other types of support for DMP creation
- Knows the common difficulties during DMP creation
- Understands the concept of the machine-actionable DMP

### Summary of tasks/actions:

1. Introduction to data management plan (DMP)
  - a. DMP with reference to FAIRness

A good data management plan covers all **FAIR principles (Findable, Accessible, Interoperable, Reusable)**<sup>16</sup>.

A DMP helps to make the data *Findable (F principle)* because it includes all information about where data is stored and preserved, during and after the project.

---

<sup>16</sup> <https://www.go-fair.org/fair-principles/>

Moreover, a DMP also contains information about persistent identifiers, e.g. DOI, along with a description of the data and metadata standards used.

A DMP helps to make the data **Accessible (A principle)** because it also includes information about how data can be accessed, what is required to access the data (authentication or authorisation) and by what (standardised and universal) communications protocol, e.g. HTTP, HTTPS.

A DMP helps to make the data **Interoperable (I principle)**, indicating which metadata standards, vocabularies, methodologies, and tools were used to facilitate interoperability. Moreover, a machine-actionable DMP also helps to address the ability of different systems and services to exchange both metadata and data produced during the project.

A DMP helps to make the data **Reusable (R principle)** because it allows data to be described with more detail and accuracy, making it easier for others to understand. Moreover, during DMP creation, it is necessary to indicate the information that is needed to prepare the data for sharing and reuse with appropriate licences and rules, namely, how the data can be reused and for whom the data may be valuable.

- b. Benefits, advantages and importance of DMP creation for researchers, their host institutions and funders:



Source: CESSDA Data Management Expert Guide<sup>17</sup>, CC BY-SA 4.0.

- c. When is a DMP needed, at what stage of the project?

<sup>17</sup> [https://www.cessda.eu/var/cessda/storage/images/cessda-training/expert-tour-guide/a-training/20171119\\_benefitsdmp\\_tekengebied-12/33308-1-eng-GB/20171119\\_BenefitsDMP\\_Tekengebied-1\\_large.png](https://www.cessda.eu/var/cessda/storage/images/cessda-training/expert-tour-guide/a-training/20171119_benefitsdmp_tekengebied-12/33308-1-eng-GB/20171119_BenefitsDMP_Tekengebied-1_large.png)

2. Content of a good DMP
  - a. Context of the project (brief description and examples)
  - b. Data and resources produced/collected during the project (brief description of the type and formats of the data; examples)
  - c. Methodologies used for data collection (brief description and examples)
  - d. Organisation of the data during the project and in datasets (brief description of the structure and names of the folders and files; examples)
  - e. Metadata and metadata standards (brief description and examples)
  - f. Documentation (brief description of the additional documentation, such as confidentiality agreements, agreements between partners, informed consent, authorisation by Ethics Committee, Data Protection Impact Assessment (DPIA) or Data Protection agreement that can substitute DPIA; examples)
  - g. Data quality procedures during data collection, data processing, data sharing and reuse
    - What does data quality mean in research data management?
    - Quality assurance guidelines (data description, metadata standards, documentation, data checking, etc.)
    - Ensure quality control (curation processes, data entry programs, use of standardised data formats, etc.)
      - documenting the calibration of instruments
      - taking duplicate samples or measurements
      - standardised data capture, data entry or recording methods
      - data entry validation techniques
      - methods of transcription
      - peer review of data
    - Data quality for publishing in repositories (completeness, uniqueness, timeliness, validity, accuracy, consistency)
    - Data quality assessment (data quality checklist)
  - h. Ethics and intellectual property (brief description and examples)
  - i. Data sharing (data access and reuse) (brief description and examples)
  - j. Data storage and backup (brief description and examples)
  - k. Selection and preservation of data (brief description and examples)
  - l. Responsibilities for managing data and resources (brief description and examples)
  - m. Additional information (such as the DMP monitoring and update process, and its importance) (brief description and examples)
3. Tools for DMP creation
  - a. DMPOnline (brief description and demonstration of the tool)
  - b. Data Steward Wizard (brief description and demonstration of the tool)
  - c. Argos DMP (brief description and demonstration of the tool)



4. Guides and templates that help create a DMP
  - a. Guides developed by government institutions and funders (e.g. Guidelines on FAIR Data Management in Horizon 2020) (brief description and examples)
  - b. Guides for specific domains, e.g. cancer research, clinical research, biological research (brief description and examples)
  - c. Checklists, frameworks, e.g. Digital Curation Centre (DCC), Inter-university Consortium for Political and Social Research (ICPSR), Framework for Creating a Data Management Plan (brief description and examples)
5. Support for DMP at the institution
  - a. Data Steward (brief description and responsibilities)
  - b. Data Protection Officer (brief description and responsibilities)
  - c. Research data support in library (brief description and responsibilities)
  - d. Other types of support, e.g. IT staff, grant administrator, funder officer, project managers (brief description and responsibilities)
6. A different approach to DMP creation for sensitive, personal and private data
  - a. Difference between these types of data (brief description and examples)
  - b. Additional documents and procedures, GDPR, connection with ethics committee, DPO, DPIA (brief description and examples)
7. Common difficulties in DMP creation (brief description of each point and examples)
8. Creation of the DMP for a project relevant for learners (practice session with a presentation and defence)

**Materials/Equipment:**

- Computer/laptop
- Internet
- DMPOnline or other tool that helps to create a DMP

**References:*****Definitions***

Clare, C., Cruz, M., Papadopoulou, E., Savage, J., Teperek, M., Wang, Y., Witkowska, I. and Yeomans, J. (Eds.), 2019. *Engaging Researchers with Data Management: The Cookbook*. Open Book Publishers. <https://doi.org/10.11647/OBP.0185>

- Michener, W. K., 2015. Ten Simple Rules for Creating a Good Data Management Plan. *PLOS Computational Biology* [online] 11(10), e1004525. <https://doi.org/10.1371/journal.pcbi.1004525>
- Hausen, D. A., Schmitz, D. and Trautwein-Bruns, U., 2020. *Content of a Data Management Plan* [online]. Video. <http://doi.org/10.18154/RWTH-2019-10064>, <https://youtu.be/fcCj6sNvoOw>
- Research Data Netherlands, 2014. *The what, why and how of data management planning* [online]. Video. Available at: <https://youtu.be/gYDb-GP1CA4>
- Juran, J. M. and Godfrey, A. B., 1998. *Juran's quality handbook: Fifth Edition*. McGraw-Hill Education. Available at: <https://gmpua.com/QM/Book/quality%20handbook.pdf>
- Chapman, A. D., 2005. *Principles of data quality, version 1.0. Report for the Global Biodiversity Information Facility* [online]. Copenhagen. Available at: <https://docs.niwa.co.nz/library/public/ChaArPrindq.pdf>
- OpenAIRE. *How to create a Data Management Plan*. Available from: <https://www.openaire.eu/when-do-i-have-to-create-a-data-management-plan>
- Miksa, T., Simms, S., Mietchen, D. and Jones, S., 2019. Ten principles for machine-actionable data management plans. *PLOS Computational Biology* [online] 15(3), e1006750. <https://doi.org/10.1371/journal.pcbi.1006750>
- Science Europe, 2021. *Practical Guide to the International Alignment of Research Data Management—Extended Edition*. Zenodo [online]. <https://doi.org/10.5281/zenodo.4915861>

### **Tools**

- DMP Online
- Argos DMP
- Data Steward Wizard
- GFBio tool for DMP
- An inventory of tools for converting your data to RDF
- Software quality checklist
- QAMyData

### **Useful links**

- The Turing Way: Data Management Plan
- Metadata Standards Catalog
- FAIRsharing - data and metadata standards
- Data Management Plans Stanford Libraries
- Horizon 2020 DMP Template
- DCC Data Management Plan

- OpenAire DMP creation
- DMP Templates
- CC Licenses
- Personal Data
- GDPR
- DPIA
- CESSDA DMP Checklist
- CESSDA Data Management Expert Guide
- DCC Checklist for DMP
- ICPSR Framework for DMP creation
- The MIT Total Data Quality Management Program (TDQM)
- Data Quality Review
- DLCM DMP tools

### ***Use cases / Examples of DMP***

- CESSDA DMP Questions Qualitative Data
- CESSDA DMP Questions Quantitative Data
- Cancer research (CRUK)
- Clinical research (CRUK)
- Population research (CRUK)
- Biological research (NSF)
- Karimova, Y., Ribeiro, C. and David, G., 2021. Institutional Support for Data Management Plans: Five Case Studies. In E. Garoufallou & M.-A. Ovalle-Perandones (Eds.), *Metadata and Semantic Research* (Vol. 1355, pp. 308–319). Springer International Publishing. [https://doi.org/10.1007/978-3-030-71903-6\\_29](https://doi.org/10.1007/978-3-030-71903-6_29)
- Barbosa, S. and Karimova, Y., 2020. *SAIL Data Management Plan (Version 1.0.0)* [online]. Zenodo. <https://doi.org/10.5281/zenodo.4286210>
- Diepenbroek, M., et al., 2014. Biodiversity and Ecological Research Data: Towards an integrated biodiversity and ecological research data management and archiving platform: the German federation for the curation of biological data (GFBio). In: Plödereder, E., Grunske, L., Schneider, E. & Ull, D. (Eds.). *Informatik 2014*. Bonn: Gesellschaft für Informatik e.V. (pp. 1711-1721). <https://dl.gi.de/handle/20.500.12116/2782>
- Best Practices for Biomedical Research Data Management
- Harvard Longwood Medical Area Research Data Management Working Group

### **Use cases/Examples of data quality processes**

- Biodiversity:
  - OECD, 2017. Data quality. In *OECD Handbook for Internationally Comparative Education Statistics: Concepts, Standards, Definitions and Classifications* (pp. 77–83). OECD Publishing, Paris, <https://doi.org/10.1787/9789264279889-9-en>.
  - Chapman, A. D., 2005. *Principles of Data Quality, version 1.0. Report for the Global Biodiversity Information Facility* [online]. Copenhagen. Available at: <https://docs.niwa.co.nz/library/public/ChaArPrindq.pdf>
  - Chapman, A., Belbin, L., Zermoglio, P., Wiczorek, J., Morris, P. and Nicholls, M. et al., 2020. Developing Standards for Improved Data Quality and for Selecting Fit for Use Biodiversity Data. *Biodiversity Information Science And Standards* [online], 4. <https://doi.org/10.3897/biss.4.50889>
  - Biodiversity Data Quality Interest Group (TDWG)
- Agriculture:
  - Agriculture Statistics Data Quality
  - Agriculture Data Quality
- Medicine and Biomedicine:
  - Medical Data Quality
- Geospatial:
  - Geospatial databases
- Sensoring:
  - SAIL and Sensor data quality control procedures:
    - Documentation of Sensor Data Correction Script
    - Geo-referencing Data. GNSS Post-processing

### **Take-home tasks:**

- Analysis of existing metadata standards: <https://rdamsc.bath.ac.uk/scheme-index> and <https://fairsharing.org/standards>
- Choosing the right licence for data, e.g. <https://ufal.github.io/public-license-selector/>, more information on this can also be found in Lesson plan 9
- Analysis of the DMP examples for scientific domain relevant to learners
- Analysis of the examples of the data quality procedures
- Datasets validation from data quality perspective
- Creation of a data quality policy for an specific use case
- Creation of the DMP for a project relevant for learners
- Preparation of a presentation for defence

## Lesson plan 3: Documentation

### FAIR elements:

#### Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
- R1.2. (Meta)data are associated with detailed provenance

**Primary audience(s):** Bachelor's, master's, PhD degree students

### Learning outcomes:

- Can explain the purpose (benefits) of the documentation, and its relation to FAIRness
- Can identify different types of data documentation, and which are suitable to a specific discipline/domain
- Can use existing documentation
- Can modify existing documentation
- Can identify considerations and strategies for documentation

### Summary of tasks/actions:

1. Introduce concept of documenting research data
  - a. Outline that a key aspect of data reusability is that it is easily interpreted by people outside of the study, and that this can be achieved by proper documentation
2. Link to relevant section/question of DMP tool used in your country/region (The examples used below are from the Canadian DMP Assistant, <https://assistant.portagenetwork.ca/>).
  - a. What documentation will be needed for the data to be read and interpreted correctly in the future?
    - Project-level
    - File-level
    - Item-level
    - Any other contextual information necessary for others to interpret

- b. How will you make sure the documentation is created or captured consistently throughout the project?
  - Clear articulation of how this will be done and by whom
  - Standardised process for accurate, consistent, and complete documentation
3. Depending on the discipline/domain of the group, introduce relevant documentation formats
  - a. Readme file
  - b. Data dictionary
  - c. Codebook
  - d. Commented code
  - e. Lab/field notebook (including Jupyter Notebooks, R markdown, electronic lab notebooks, etc.)
    - If introducing multiple formats, outline similarities/differences and use cases
    - For each format that is showcased, articulate considerations and other important aspects by using exemplars and other material from the “References” section
4. Conduct an exercise in which learners complete one or more of the documentation formats, based on course/project work that is relevant to learners. Blank templates can be found/created using material from the “References” section. Review and discuss challenges, as well as strategies to mitigate challenges.

**References:*****READMEs***

- Guide to writing “readme” style metadata
- README template

***Data dictionaries***

- How to Make a Data Dictionary
- Data Dictionary Template
- Community defined models and formats in FAIRsharing

***Codebooks***

- Codebook Cookbook
- Sample Questionnaire with Coding

***Commented code***

- Coding and Comment Style

***Lab/field notebook***

- Examples of notebook pages and entries
- Guide for Taking Field Notes
- Electronic Lab Notebooks
- Jupyter
- R Markdown

**Exercises:**

- LEGO® Metadata for Reproducibility game pack - Enlighten: Publications

## Lesson plan 4: Data creation

### FAIR elements:

#### Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, making this is an essential component of the FAIRification process.

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

#### Accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
  - A1.1 The protocol is open, free, and universally implementable
  - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

#### Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

#### Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes



- R1.1. (Meta)data are released with a clear and accessible data usage license
- R1.2. (Meta)data are associated with detailed provenance
- R1.3. (Meta)data meet domain-relevant community standards

**Primary audience(s):** Bachelor's, master's, PhD degree students

**Learning outcomes:**

- Can define research data
- Can explain the steps of the research data lifecycle
- Can practically apply theoretical knowledge about proper RDM measures to be taken at the stage of data creation

**Summary of tasks/actions:**

1. Introduce the definition of research data and research data lifecycle
  - a. Learners create the research data lifecycle: Learners receive cards with key terms of the lifecycle. In groups, they should arrange the cards into a research data lifecycle, discussing what the terms might mean. At the end of the session, they should present their results to the other groups [Biernacka et al. 2020].
2. How can data be created?
  - a. New data collection
  - b. Reuse of existing data (see also Lesson plan 9)
    - Learners go to a repository (at best, a discipline-specific one suitable for their research field) and find data that they could use for their research.
3. First steps while creating data
  - a. Selection of research design
    - Quantitative
    - Qualitative
  - b. Research instruments
    - Questionnaires/surveys
    - Interviews
    - Field observations
    - Other
  - c. Data planning (see also Lesson plan 2)
    - Learners write a short data management plan based on a template. It does not have to be very detailed. It is important for participants to think about the data and write down their initial thoughts in bullet points.

- d. Locate existing research data (see also Lesson plan 10)
  - See task Reuse of existing data (2. b)
- e. Collect new research data
- f. Capture and create metadata (see also Lesson plan 6)
  - Create a board, e.g. Padlet, Miro or a flipchart, and let the learners write down which metadata they think would be useful for their data/in their discipline. Discuss.

**Materials/Equipment:**

- Computer
- Internet
- For 1a: cards with key terms or virtual tool, e.g. Padlet
- For 3f: a virtual board or flipchart

**References:**

Biernacka, K., Bierwirth, M., Dolzycka, D., Helbig, K., Neumann, J., Odebrecht, C., Wilkes, C. and Wuttke, U., 2020. *Train-the-Trainer Concept on Research Data Management (Version 3.0)* [online]. Zenodo. <http://doi.org/10.5281/zenodo.4071471>

## Lesson plan 5: File formats

### FAIR elements:

#### Accessible

Once the user finds the required data, they need to know how they can be accessed, possibly including authentication and authorisation.

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
  - A1.1 The protocol is open, free, and universally implementable
  - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

#### Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

#### Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
  - R1.1. (Meta)data are released with a clear and accessible data usage license
  - R1.2. (Meta)data are associated with detailed provenance
  - R1.3. (Meta)data meet domain-relevant community standards

**Primary audience(s):** Bachelor's, master's, PhD degree students

**Learning outcomes:**

- Knows which formats support FAIR data
- Understands what the differences are between open and proprietary formats
- Knows about open formats and how/where to check their openness
- Is able to apply knowledge by exporting/converting files into different formats

**Summary of tasks/actions:**

1. Raise awareness about file formats and their standards:
  - a. obsolescence
  - b. proliferation
  - c. lossless vs. lossy formats
  - d. significant properties
2. Show the differences between open and proprietary formats, and explain their role in making data FAIR (documentation, standards):
  - a. What are the advantages of open formats?
  - b. What are the disadvantages of proprietary formats?
  - c. What should you do if you still (need to) use proprietary formats?
    - How to convert file formats?
    - How to export files into a different format?
    - How to save the files in containers to preserve the original (proprietary) format along with a more open option?
3. Show tools for file format identification, e.g. PRONOM, and validation, e.g. JHOVE
4. Application of knowledge in practice (quiz, exercises)
  - a. Questionnaire: Open or not? Which of these file formats support FAIR data?
    - Which of these text formats are suitable for long-term archiving? (Multiple choice)
      - txt
      - docx
      - odt
      - html
    - Which of these tabular formats are suitable for long-term archiving? (Multiple choice)
      - xlsx
      - csv
      - tsv
      - spss portable

- Which of these image formats are suitable for long-term archiving? (Multiple choice)
  - jpg
  - png
  - tiff
  - gif
- b. Learners choose a random folder from their directory. They should check the stored file formats in terms of the FAIR principles and try to export or convert the file into a more open file format, if necessary.

**Materials/Equipment:**

- Computer/laptop
- Internet/browser

**References:**

- Digital Preservation Handbook
- PRONOM
- JHOVE
- FAIRsharing list of file formats in all disciplines

## Lesson plan 6: Metadata

### FAIR elements:

#### Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

#### Accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
  - A1.1 The protocol is open, free, and universally implementable
  - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

#### Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

#### Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

- R1.1. (Meta)data are released with a clear and accessible data usage license
- R1.2. (Meta)data are associated with detailed provenance
- R1.3. (Meta)data meet domain-relevant community standards

**Primary audience(s):** Bachelor's, master's, PhD degree students

**Learning outcomes:**

- Can describe types of metadata
- Can recognise metadata formats
- Can identify metadata standards
- Can use metadata standards to describe resources
- Can explain what metadata registries are
- Can search and find data and metadata standards in registries
- Can articulate metadata of different types to describe a resource
- Can write metadata in a relevant format
- Can appraise the usefulness of metadata standards to describe a resource

**Summary of tasks/actions:**

1. Metadata are 'data about data'
  - a. Present and describe the different types of metadata (can present the whole list, or pick specific elements relevant to your audience).
    - Metadata are:
      - standardised
      - structured
      - machine- and human-readable
      - a subset of documentation
  - b. Documentation (descriptive and/or technical info)
  - c. Controlled vocabularies and ontologies
  - d. Persistent identifiers (PIDs)
  - e. Licences
2. Learn syntax of example metadata standards:
  - a. Dublin Core is general and applicable to all datasets on a project level; on a data level there are discipline-specific standards to branch into such as:
    - Data Documentation Initiative (DDI) – social science
    - Ecological Metadata Language (EML) – ecology
    - Flexible Image Transport System (FITS) – astronomy
  - b. Minimum information standards

### 3. Use metadata catalogues/registries and search for suitable standards

Metadata form the core of machine- and human-readable descriptions of data, be they technical information or annotations, and cover all aspects of the FAIR principles. Metadata is an umbrella term that includes file formats, ontologies and licences, and documentation in general. For each of the principles, metadata can be used at different granularities and domain specificity, with more general metadata not providing as much usefulness and value to the underlying data than domain-specific metadata.

#### References:

Metadata for Machines workshops

- General information: <https://www.go-fair.org/how-to-go-fair/metadata-for-machines/>
- Example: Metadata for Machines workshops, including material. These were funded by the Dutch research foundation ZonMw in support of their COVID-19 research programme: <https://osf.io/bhzf8/>
- Handbook of Metadata, Semantics and Ontologies
- FAIR Cookbook, recipes for hands-on FAIRifications in the Life Sciences.
- FAIRsharing resource to discover (meta)data standards (and which repositories implement them)

#### Take-home tasks:

1. Create the metadata for a dataset
  - a. Search for standards in catalogues like:
    - <https://rdamsc.bath.ac.uk/>
    - <http://rd-alliance.github.io/metadata-directory/>
    - FAIRsharing data and metadata standards
    - <https://lov.linkeddata.es/dataset/lov/>
  - b. How to create a metadata profile or template
    - FAIRplus example
2. Encode the data in a dataset using controlled vocabularies/ontologies
  - a. FAIRsharing terminology artifacts
  - b. Jacob et al. Making experimental data tables in the life sciences more FAIR: a pragmatic approach GigaScience, Volume 9, Issue 12 2020

#### Exercises:

- LEGO® Metadata for Reproducibility game pack - Enlighten: Publications



## Lesson plan 7: Data standardisation and ontologies

### FAIR elements:

#### Findable

Standardisation of data identifiers makes data easier to find.

- F1. (Meta)data are assigned a globally unique and persistent identifier

#### Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

Interoperability is made easier through standardised representations of knowledge and by using standard variables that allow linking of data files, e.g. using standardised date and time stamps.

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

#### Reusable

Domain-relevant community standards make data easier to understand and reuse.

- R1.3. (Meta)data meet domain-relevant community standards

**Primary audience(s):** Bachelor's, master's, PhD degree students (without a knowledge management background)

### Learning outcomes:

- Can explain aspects related to data interoperability and integration (standardisation in data and how data standards are used)
- Can explain aspects of data preparation and cleaning
- Can explain the roles of ontologies and vocabularies
- Can recognise the use of ontologies and vocabularies
- Can recognise the role of data standards in making data FAIR.
- Understands that different communities use different data standards and ontologies to improve the understanding and interoperability of their research data
- Can identify a few domain-relevant ontologies

- Understands usage scenarios of ontologies during data collection, data analysis and when making data available through repositories, APIs, etc.
- Knows how to act when an ontology does not exist or elements are missing in an existing ontology

### Summary of tasks/actions:

1. Explain with easy and practical examples (from your discipline) how standardisation of data can be applied in research. Standardisation enables interoperability of data, common understanding of data and facilitates (interdisciplinary) reuse of data. Some simple examples are:
  - a. Standard coding structures (e.g. use 1=male, 2=female systematically, and not sometimes 1=female, 2=male, or 0=male, 1=female)
  - b. Standard units: degrees Celsius vs. degrees Fahrenheit; wind speed measured in m/s vs. knots/s, universal date and time stamps<sup>18</sup>
  - c. Standard geospatial representations, e.g. WGD84
  - d. Statistical Classification of Economic Activities in the European Community: NACE code
  - e. Universal system of (binomial) nomenclature and taxonomy to name and classify biodiversity, now also including DNA barcoding<sup>19</sup>
  - f. Standards for dates and times (ISO 8601), for countries (ISO 3166), for geographical names (Getty Thesaurus)
2. You could also show an example of how not using standards makes things more difficult, or involves more work to clean and translate data, for example:
  - a. Survey data where standardised responses are still captured as 'text' rather than numerical codes (dataset with 'male', 'female' rather than numeric codes)
  - b. Datasets where units of variables are not defined, so it is not possible to say whether the temperature is in Celsius or Fahrenheit
  - c. Any other example listed above where no standard was used in the dataset
3. Use examples of data standards in different disciplinary communities (see references)
  - a. Help define data procedures, standards and guidelines by discipline. For example, are there guidelines for data processing, are there metadata standards, are there controlled vocabularies, ontologies and taxonomies, are there specialised data repositories used by the scientific community?

18 Good example on standardising date time stamp in: Data Tree, module 2, topic 4, Data Handling and Formats: Practicalities: Presentation: Data Handling and Formats (datatree.org.uk)

19 Global Taxonomy Initiative (cbd.int)

4. Describe what ontologies are and their function in the semantic web. Learn the various types of ontologies.

Interoperability is also part of teaching and adheres to the following principles:

- F2. Data are described with rich metadata – using ontologies is part of good metadata practice
- R1. Meta(data) are richly described with a plurality of accurate and relevant attributes – using ontologies is part of good practice for rich and precise descriptions
- R1.3. (Meta)data meet domain-relevant community standards – the same as the previous two bullet points

## References:

### *Use cases*

- BioSharing registry: <https://biosharing.org/>
- Specifications of Standards in Systems and Synthetic Biology: <https://doi.org/10.1515/jib-2018-0013>
- Biodiversity standards:
  - Audubon Core: <https://www.tdwg.org/standards/ac/>
  - Darwin Core: <https://www.tdwg.org/standards/dwc/>
  - Natural Collections Descriptions (NDC): <https://www.tdwg.org/standards/ncd/>
  - GUID applicability statements: <https://github.com/tdwg/guid-as>
  - TDWG Access Protocol for information Retrieval (TAPIR): <https://www.tdwg.org/standards/tapir/>
  - TDWG Standards Documentation Standard (SDS): <https://www.tdwg.org/standards/sds/>
  - Vocabulary Maintenance Standard (VMS): <https://www.tdwg.org/standards/vms/>
  - Global Genome Biodiversity Network (GGBN Data Standard): <https://www.tdwg.org/standards/ggbn/>
  - Access to Biological Collection Data (ABDC): <https://www.tdwg.org/standards/abcd/>
  - Description Language for Taxonomy (DELTA): <https://www.tdwg.org/standards/delta/>
  - Structured Descriptive Data (SDD): <https://www.tdwg.org/standards/sdd/>
  - Taxonomic Schema (TCS): <https://www.tdwg.org/standards/tcs/>

### ***Agriculture data***

- Dzale Yeumo, E., Alaux, M., Arnaud, E. et al., 2017. Developing data interoperability using standards: A wheat community use case [version 2; peer review: 2 approved]. *F1000Research* [online] 6:1843. <https://doi.org/10.12688/f1000research.12234.2>
- Wheat Data Interoperability Guidelines and Recommendations: <https://www.rd-alliance.org/groups/wheat-data-interoperability-wg.html> and <https://www.rd-alliance.org/group/working-and-interest-group-chairs-wheat-data-interoperability-wg/outcomes/wheat-data>
- Agrisemantics Working Group Recommendations: <https://www.rd-alliance.org/groups/agrisemantics-wg.html>
- The eROSA Roadmap for a pan-European e-Infrastructure for Open Science in Agricultural and Food Sciences (led by INRA) significantly reflects outputs of several RDA groups, including Data Fabric’s “Recommendations for Implementing a Virtual Layer for Management of the Complete Life Cycle of Scientific Data”.
- The FAIRsharing Registry and Recommendations: Interlinking Standards, Databases and Data Policies: <https://www.rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases-wg/outcomes/fairsharing>

### ***Ocean data***

- Ocean Data Standards and Best Practices: <https://www.oceandatastandards.org/>
- SeaDataNet Metadata Profile ISO 19115: <https://www.seadatanet.org/content/download/1855/file/CDI-profile-V10.0.1.pdf>
- Basic Register of Thesauri, Ontologies & Classifications (BARTOC), <https://bartoc.org/>
- Zaharee, M., 2013. Building controlled vocabularies for metadata harmonization. *Bulletin of the American Society for Information Science and Technology* [online] 39(2), 39–42. <https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/bult.2013.1720390211>

### **Take-home tasks:**

- Analyse the existing standards (general and/or by discipline) required in FAIR principles
- Study/Analyse the standards that apply in a particular discipline
- Standardise a dataset: choose a discipline, create or download a dataset and standardise it according to the scientific community
- Activities related to data standardisation tools:
  - OpenRefine tool (data clean, data transformation, data normalisation, etc.)
  - Data FAIRification tools: <https://fairplus.github.io/the-fair-cookbook/content/recipes/interoperability/rdf-conversion.html>

## Lesson plan 8: Persistent identifiers (PIDs)

### FAIR elements:

#### Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, making this an essential component of the FAIRification process.

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

#### Accessible

Once the user finds the required data, they need to know how they can be accessed, possibly including authentication and authorisation.

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
  - A1.1 The protocol is open, free, and universally implementable
  - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

**Primary audience(s):** Bachelor's, master's, PhD degree students

### Learning outcomes:

- Can recognise PIDs and explain the different use cases for PIDs
- Can explain the importance of PIDs for FAIR data
- Understands the PID syntax
- Can use PIDs to access data or other resources
- Can apply PIDs to their own research outputs
- Can use PIDs to collaborate with others
- Knows about provenance and versioning of data
- Optional: knows PID graphs

**Summary of tasks/actions:**

1. Provide a use case to show the importance of persistent identifiers (PIDs). Define the problem, e.g. different scenarios where digital objects may have the same or similar names, such as different versions or authors – disambiguate; also for findability and accessibility of data – can be resolved by web browsers, etc. and are actionable
  - a. Identify different entities that can be assigned a PID, e.g. people, data, and institutions
  - b. Define together what persistent identifiers are
  - c. Explain the difference between persistent identifiers and authority files
2. Show the different types of PIDs and how their syntax can look:
  - a. DOI
  - b. Crossref
  - c. ORCID
  - d. ROR
  - e. RAID
  - f. other
3. Explain how to receive a PID
  - a. Repositories
  - b. PID minting
4. Show provenance as an important aspect of FAIR data
  - a. Resource provenance
  - b. Metadata provenance
  - c. How can PIDs contribute to provenance?
5. How are PIDs used in relation to different versions of a dataset or dynamic datasets?
  - a. Versioning exercise
6. Introduce PID graphs and their importance
  - a. Explain the importance of PID graphs with a use case (real use cases can be found here: <https://github.com/datacite/freya/issues?utf8=%E2%9C%93&q=is%3Aissue+is%3Aopen+label%3A%22PID+Graph%22++label%3A%22user+story%22+>)

**Materials/Equipment:**

- Computer/laptop
- Internet/browser

**References:**

- <https://datacite.org/does.html>
- <https://search.crossref.org/>
- <https://www.doi.org/>
- <https://orcid.org/>
- <https://ror.org/>
- <https://www.raid.org.au/>
- FAIRsharing list of community-used identifier schemas
- Ball, A. and Duke, M., 2015. How to Cite Datasets and Link to Publications DCC How-to Guides [online]. Edinburgh: Digital Curation Centre. Available from: <http://www.dcc.ac.uk/resources/how-guides>
- Research Data Alliance (RDA). *RDA recommendations* [online]. Available from: [https://www.rd-alliance.org/system/files/RDA-DC-Recommendations\\_151020.pdf](https://www.rd-alliance.org/system/files/RDA-DC-Recommendations_151020.pdf)
- Ananthakrishnan, R., Chard, K., D’Arcy, M., Foster, I., Kesselman, C., McCollam, B., Pruyne, J., Rocca-Serra, P., Schuler, R. and Wagner, R., 2020. An Open Ecosystem for Pervasive Use of Persistent Identifiers. *Practice and Experience in Advanced Research Computing* [online], 99–105. <https://dl.acm.org/doi/10.1145/3311790.3396660>
- Library Carpentry: FAIR Data and Software - Findable. Available at: <https://librarycarpentry.org/lc-fair-research/02-findable/index.html>
- Research Data Netherlands, 2014. *Persistent identifiers and data citation explained/* [online]. [Video]. Available from: <https://youtu.be/PgqtiY7oZ6k>
- Fenner, M., Wass, J., Demeranville, T., Sarala Wimalaratne, S. and Hallett, R., 2019. *D2.2 PID Metadata Provenance*. Zenodo. <https://doi.org/10.5281/zenodo.3248652>
- Fenner, M., & Aryani, A., 2019. Introducing the PID Graph. Datacite Blog. <https://doi.org/10.5438/jwvf-8a66>

## Lesson plan 9: Licences, copyright and intellectual property rights (IPR) issues

### FAIR elements:

#### Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
  - R1.1. (Meta)data are released with a clear and accessible data usage license
  - R1.2. (Meta)data are associated with detailed provenance
  - R1.3. (Meta)data meet domain-relevant community standards

**Primary audience(s):** Bachelor's, master's, PhD degree students

### Learning outcomes:

- Understand what licences are, their purpose and relation with the FAIR principle
- Know how the data can be reused and shared with others
- Be able to identify the owner of the data for a project which may or may not have many partners
- Know what copyright and intellectual property rights are
- Be aware that different copyright rules exist in different countries (and that there are countries without copyright law)
- Know the different types of rights (economic and moral)
- Know the different types of licences and understand what actions can be performed with them, e.g. CC, ODbL, ACA, OGL, LGPL
- Know the meaning of non-commercial and commercial licences, e.g. CC BY-NC
- Know the different types of restrictions
- Know tools and guides to choose the correct licence
- Apply the acquired knowledge in practice, e.g. quiz, exercises



**Summary of tasks/actions:**

1. Introduction to licences and (re)use issues;
  - a. FAIR focus on reusability, namely on R1.1. (Meta)data are released with a clear and accessible data usage license point of the FAIR principles. Licences, copyright and IPR issues help to clarify the FAIR reusable principle. They help identify legal, ethical and usage rights, understand who owns the copyright and IPR. Moreover, these issues help to prepare your data for professional reuse with or without restrictions, with an appropriate licence, while protecting you as licence holder and avoiding unpleasant situations surrounding reuse of data.
  - b. What licences are, their purpose and importance
  - c. What type of digital object should and can be licensed (data, software, code, etc.)
  - d. Understand the differences between licences used for data and software
2. Copyright and intellectual property rights
  - a. Definition
  - b. Type of intellectual property rights, e.g. copyright, patents, trademarks, industrial design rights, plant varieties, trade dress, trade secrets, database rights
  - c. Purpose of copyright
  - d. Copyright protected works; examples (e.g. All rights reserved (fully copyrighted)
    - Is (research) data protected by copyright law in the same way as other works?
      - Let participants define research data they work with
      - Explain the difference between copyright protected works and works that are not copyright protected (like pure information or facts), and show examples
  - e. Copyright exceptions; examples (e.g. Copyright exceptions);
  - f. What information do you need to provide when contacting the copyright holder?
    - What you will be using (amount and content)
    - Context in which the work will be used
    - Where you intend to use the work, e.g. publicly online
    - For what purpose, e.g. educational, commercial, personal
    - How they will be attributed

3. Usage rights: what does it mean? (Brief description and examples);
  - a. Definition
  - b. Type of rights, e.g. economic and moral; non-exclusive rights of use and exclusive rights of use;
  - c. What permissions do you have with a licence? (e.g. distribute, remix, adapt, build upon a material)
4. Different types of licences
  - a. Creative Commons;
    - CC0 - No Rights Reserved
    - Attribution CC BY
    - Attribution ShareAlike CC BY-SA
    - Attribution-NoDerivs CC BY-ND
    - Attribution-NonCommercial CC BY-NC
    - Attribution-NonCommercial-ShareAlike CC BY-NC-SA
    - Attribution-NonCommercial-NoDerivs CC BY-NC-ND
  - b. Software licences
    - Public domain
    - Permissive
    - LGPL(GNU)
    - Copyleft
    - Proprietary
  - c. Open Source Licences
    - Apache License 2.0
    - BSD 3-Clause “New” or “Revised” license
    - BSD 2-Clause “Simplified” or “FreeBSD” license
    - GNU General Public License (GPL)
    - GNU Library or “Lesser” General Public License (LGPL)
    - MIT license
    - Mozilla Public License 2.0
    - Common Development and Distribution License
    - Eclipse Public License version 2.0
  - d. Other types of licences
    - ODbL
    - ACA
    - OGL
  - e. Orphan works and search guidance for applicants
  - f. Remember: Licence-free is not the same as a free licence

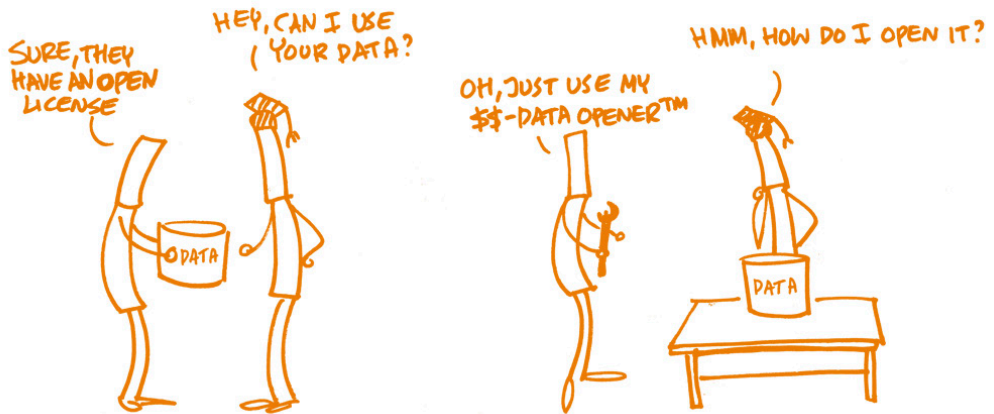


Image source: <https://open-science-training-handbook.gitbook.io/book/open-science-basics/open-licensing-and-file-formats>

5. Tools to help choose the right licence
  - a. EUDAT licensing tool/wizard
  - b. CC License chooser
  - c. Choose an open source license
  - d. CLARIN License Calculator
6. Ownership of data
  - a. Who owns the data?
  - b. Show the different ownership possibilities and explain that in many cases, ownership of data may be regulated by employment and service contracts
7. How to resolve FAIR compliance with IPR restricted data
  - a. Show examples of IPR, sensitive data, and other data that cannot be fully open. Explain how the metadata of this type of data can be open
8. Play Copyright the Card Game
9. Application of knowledge in practice (quiz, exercises)
  - a. Example: Which licence may you grant if you want to combine data with the following licences:
    - CC BY and CC BY-SA?
    - CC BY-SA and CC BY-NC?
    - CC BY and CC BY-ND?
10. Do an exercise related to searchability and licence issues, e.g. search for images on Google filtering by different licence types

**Materials/Equipment:**

- Computer/laptop
- Internet/browser
- Different tools for choosing licences, e.g. EUDAT License Selector

**References:****Definitions**

- Creative Commons Licences
- Open Data Commons Licences
- Ball, A., 2014. *How to License Research Data DCC How-to Guides* [online]. Edinburgh:Digital Curation Centre. Available from: <http://www.dcc.ac.uk/resources/how-guides>
- <https://opendefinition.org/licenses/>
- Nemlioglu, I. A comparative analysis of intellectual property rights: a case of developed versus developing countries. *Procedia Computer Science* [online] 158(2019): 988-998. <https://doi.org/10.1016/j.procs.2019.09.140>
- Margoni, T. and Tsiavos, P., 2018. *Toolkit for Researchers on Legal Issues* [online]. Zenodo. <https://doi.org/10.5281/zenodo.2574619>
- International Copyright Basics
- CEESDA Licensing your data
- Ownership of WUR research data
- How copyright protects your work
- Creative Commons Public Domain
- Using somebody else's intellectual property
- Open Science Training Handbook. Open Licensing and File Formats
- Guibault, L. and Wiebe, A. (Eds.), 2013. *Safe to be open: Study on the protection of research data and recommendations for access and usage*. Göttingen University Press <https://doi.org/10.17875/gup2013-160>.
- Burrow, S., Margoni, T. and McCutcheon, V., 2018. *Information Guide: Introduction to Ownership of Rights in Research Data* [online]. CREATE, University of Glasgow. . <http://eprints.gla.ac.uk/171314/>
- Burrow, S., Margoni, T. and McCutcheon, V., 2018. *Information Guide: Making Research Data Available* [online]. CREATE, University of Glasgow. <http://eprints.gla.ac.uk/171315/>
- Burrow, S., Margoni, T. and McCutcheon, V., 2018. *Information Guide: Choosing a Licence for Research Data* [online]. CREATE, University of Glasgow. <http://eprints.gla.ac.uk/171316/>

- Burrow, S., Margoni, T. and McCutcheon, V., 2018. *Information Guide: Using Research Data* [online]. CREATE, University of Glasgow. <http://eprints.gla.ac.uk/171317/>

### **Tools**

- Creative Commons. Choose a License
- Tool for choosing an open source license
- EUDAT License Selector
- CLARIN License Category Calculator

### **Examples**

- An example of the existence of different data owners in the same project. Barbosa, Susana, & Karimova, Yulia., 2020. *SAIL Data Management Plan (Version 1.0.0)* [online]. Zenodo. <https://doi.org/10.5281/zenodo.4286210>
- Examples of Usage Rights
- Copyright Examples
- What can be Copyrighted (Examples)
- Ownership of WUR research data

### **Take-home tasks:**

- Looking at your own research project (master's thesis, PhD thesis, etc.), work through the information provided and identify what permissions you will need, and also what licences or copyright you would like to publish your work under.
- Analyse different content with different licences, e.g. Flickr, YouTube, Wikimedia Commons, Vimeo, Wikipedia and the Internet Archive, Google.
- See and analyse examples of the CC0 licence (<https://creativecommons.org/2017/02/07/met-announcement/>). Identify the specificity of this licence.
- Find some examples of the real cases related to the licence, copyright and IPR issues, e.g. the case between Coca-Cola and Yotvata: [https://www.youtube.com/watch?v=2nyhjM2BDQU&ab\\_channel=EliLevineGoldberg](https://www.youtube.com/watch?v=2nyhjM2BDQU&ab_channel=EliLevineGoldberg).

## Lesson plan 10: Finding and reusing data

Being able to reuse data and analyse secondary data not only helps to save time and energy for researchers, it can also fast-track scientific discoveries with shared resources and perspectives, while adhering to the FAIR principles.

The FAIR elements that this lesson plan deals with focus on F (Findable), A (Accessible) and R (Reusable). As stated in the ‘FAIR Guiding Principles for scientific data management and stewardship’<sup>20</sup>, the ultimate goal of FAIR is to optimise the reuse of data. In order to be reusable, data should correspond, on a general level, to all of the FAIR principles, and in particular to the R ones:

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

**Primary audience(s):** Master’s and PhD degree students, researchers

### Learning outcomes:

- Can explain the importance of data discovery and reuse
- Can recognise the concept of ‘secondary data’ vs. collecting primary data
- Can discover published datasets in their discipline
- Can cite data
- Can develop a strategy to search for data
- Can articulate the criteria for data selection
- Can recognise the provenance of data they intend to use
- Can recognise the importance of the terms and conditions of data reuse
- Can recognise the importance of data citation when reusing data

### Summary of tasks/actions:

- Speaking of ‘good scientific practice’: Why is it important to use secondary data rather than collect primary data?
- Identify a strategy to find data appropriate for a specific research project.

---

<sup>20</sup> Wilkinson, M. D. et al. *The FAIR Guiding Principles for scientific data management and stewardship*. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).

- Identify a ‘trustworthy’ data repository: find relevant data in certified repositories, check measures taken by repositories to ensure that data are reusable. What are the criteria that ‘trustworthy’ data should meet?
- Look at some examples of datasets and how they express terms for reuse.
- Look at data citation models: Conduct a case study when wanting to cite multiple datasets from various repositories, providing different data citation models.
- Sensitise learners to further share *new knowledge* and *new data* created during this data reuse process.

**Materials/Equipment:**

- Computer
- Internet

**Resources:*****Why are research data managed and reused?***

An interesting point on good scientific practice is made on this blog post from the Finnish Social Science Data Archive, which also briefly describes the benefits of data reuse:

*“Reusing data is economic and saves resources. If suitable data are readily available, there is less need to spend time and money to collect new material. Data from large surveys often include material that has not been analysed in the original research. Data reuse helps to avoid duplication of data collection. It can also minimise collection on the hard-to-reach or the vulnerable. Valuable research data are of no use to the scientific community and future research if original data creators are the only persons to have any information on the data. If they relocate to other organisations or to other tasks, or retire, all information will disappear.”*

(<https://www.fsd.tuni.fi/en/services/data-management-guidelines/why-are-research-data-managed-and-reused/>)

**Time Efficacy Gain:**

Pronk, T.E., 2019. The Time Efficiency Gain in Sharing and Reuse of Research Data. *Data Science Journal* [online], 18(1), p.10. <http://doi.org/10.5334/dsj-2019-010>

The author uses a “mathematical model [...] to calculate the break-even point for time spent sharing in a scientific community, versus time gain by reuse” for a number of scenarios.

*“The results indicate that sharing research data can indeed cause an efficiency revenue for the scientific community. However, this is not a given in all modeled scenarios. The scientific community with the lowest reuse needed to reach a break-even point is one that has few sharing researchers and low time investments for sharing and reuse. This suggests it would be beneficial to have a critical selection of datasets that are worth the effort to prepare for reuse in other scientific studies. In addition, stimulating reuse of datasets in itself would be beneficial to increase efficiency in scientific communities.”*  
(Pronk 2019)

### **Review shared research data:**

CESSDA (Consortium of European Social Science Data)’s discovery section in the data management expert guide: <https://www.CESSDA.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/7.-Discover>

Including steps to take during the discovery process and a curated list of different types of social science data sources

### **Finding and citing data:**

Ball, A., & Duke, M. (2015). ‘How to Cite Datasets and Link to Publications’. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides>

Gregory, K., Groth, P., Scharnhorst, A., & Wyatt, S., 2020. Lost or Found? Discovering Data Needed for Research. *Harvard Data Science Review* [online], 2(2). <https://doi.org/10.1162/99608f92.e38165eb>

This study presents evidence from the largest known survey investigating how researchers discover and use data that they do not create themselves.

Surrey Repro Society – Finding and using secondary data (workshop slides) <https://osf.io/4yhtg/>

### **List of resources and data repositories for finding secondary data:**

An up-to-date list of available registered data repositories can be found at <https://www.re3data.org/> and at FAIRsharing.

Still, finding a trustworthy data repository that suits your research needs can be a challenge. A possible solution is to look for certified repositories, be it a core certification or a more formal one. For example, a core certification involves a minimally intensive process whereby data repositories supply evidence that they are sustainable and trustworthy. Alternatively, look for repositories that have been recommended by



your community and or research infrastructure in your discipline, such as ELIXIR for the Life Sciences.

The Core Trust Seal certified repositories: <https://www.coretrustseal.org/why-certification/certified-repositories/>

You could also look for the data catalogue of institutions, such as the data catalogue (<https://datacatalogue.CESSDA.eu/>) of the Consortium of European Social Science Data Archives (CESSDA), with guidelines on discovering data (<https://www.CESSDA.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/7.-Discover/Data-repositories-as-data-resources/>).

In general, repositories that have reusability and metadata assessment tools, such as Kaggle (<https://www.kaggle.com/datasets>) and KNB (<https://knb.ecoinformatics.org/>), are a valuable resource for data reuse.

### **List of data and metadata standards:**

Across the research disciplines there are thousands of standards that act as pillars for data reuse. FAIRsharing maps the landscape of community-developed standards, while defining the indicators necessary to monitor their development, evolution and integration, implementation and use in databases, and adoption in data policies by funders, journals and other organisations.

### **Take-home tasks:**

- Exercise on finding “trustworthy” data on a given topic during the class.
- Use the data found in the above as an example to practice data citation.
- Find standards relevant to your domain and discipline.

### **References:**

Finnish Social Science Data Archive. *Why are research data managed and reused*. <https://www.fsd.tuni.fi/en/services/data-management-guidelines/why-are-research-data-managed-and-reused/>

Pronk, T.E., 2019. The Time Efficiency Gain in Sharing and Reuse of Research Data. *Data Science Journal* [online] 18(1), p.10. <http://doi.org/10.5334/dsj-2019-010>

CESSDA (Consortium of European Social Science Data)’s discovery section in the Data management expert guide

Ball, A., & Duke, M., 2015. *How to Cite Datasets and Link to Publications*. DCC *How-to Guides* [online]. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides>

Gregory, K., Groth, P., Scharnhorst, A. and Wyatt, S., 2020. Lost or Found? Discovering Data Needed for Research. *Harvard Data Science Review* [online] 2(2). <https://doi.org/10.1162/99608f92.e38165eb>

## Lesson plan 11: Repositories

**FAIR elements:** All

**Primary audience(s):** Bachelor's, master's, PhD degree students

**Learning outcomes:**

- Can explain what repositories are, what they are useful for, and how they help with FAIR
- Can identify and understand the different types of repositories
- Can explain what a trusted data repository is and how to find it, e.g. via re3data.org or FAIRsharing
- Can compare different certifications for data repositories, e.g. CoreTrustSeal, CLARIN certification
- Can articulate different criteria that can be used to choose a repository
- Can discover trusted repositories and identify those that are certified
- Can use a trusted repository to share research output

**Summary of tasks/actions:**

1. Introduce the concept of repositories.
  - a. Explain the following:

Repositories are used to store, document and publish all kinds of digital objects. They are storage locations for digital (and physical) objects which enable the separate publication and archiving of digital objects.
  - b. Discuss: Why use a repository?

Data repositories can help make a researcher's data more discoverable and accessible, and lead to potential reuse. Using a repository can lead to increased citations of your work<sup>21</sup>. Data repositories can also serve as backups during rare events where data are lost to the researcher and must be retrieved. Depending on the discipline requirements – publisher, funders, institutional policies, national policies – researchers may be required to store their data in certain repositories.

Practical exercise: Check what you need to address with your local institutional requirements. Are you obliged to upload your research outputs locally?

---

21 Piwowar, Heather A., Vision, Todd J. 'Data reuse and the open data citation advantage' *PeerJ* 1:e175 (2013). <https://doi.org/10.7717/peerj.175>.

2. FAIR principles and repositories
  - a. **Findability:** Repositories can provide a persistent and unique identifier for data; help to add rich, clear and machine-readable metadata to data; make the data findable using web-based search engines.
  - b. **Accessibility:** Repositories can have open, free and standardised communication protocols with authentication and authorisation procedures; provide the existence of metadata independent of the availability of the data.
  - c. **Interoperability:** Repositories can use common semantic language, making data interoperable with applications or other workflows for analysis, storage and processing; help to provide metadata with vocabularies according to FAIR principles.
  - d. **Reusability:** Repositories can promote data reuse; help to provide rich, accurate relevant metadata with a data usage licence, detailed provenance, and using common standards.
  
3. Different types of repositories:
  - a. Identify discipline-specific vs. cross-discipline repositories: Repositories can be classified according to various aspects. In most cases, they are distinguished by whether they are discipline-specific, cross-discipline/generic, computing centre-based, or institutional. Discipline-specific or disciplinary repositories offer the benefits of visibility in the research community, research data management expertise, specialised tools, and are already established services in some disciplines. However, not all academic subject areas have established discipline-specific repositories.

Examples of free-to-use discipline-specific repositories:

    - ICPSR for the social sciences
    - PANGAEA for Earth and space science data
    - Crystallography Open Database (COD) for chemistry & crystallography

For interdisciplinary research, assignment of the resulting data to a subject area may be difficult. Cross-disciplinary/generic repositories offer a solution here as they accept data regardless of data type, format, content, or disciplinary focus. In some cases, however, they do not curate the data or offer other forms of quality control. This responsibility lies with the author/depositor.

Examples of cross-disciplinary, generic repositories that are free to use:

    - ZENODO (free to use, open source)
    - Figshare (free to use)
  - b. **Institutional repositories** are often free of charge and can be used for all of the institution's own subject areas. Many universities support research data management on campus through a central service. Research data services staff can be an excellent source of research data management sup-

port, including repository selection, and can help you comply with funder, publisher, and university requirements. Additionally, High Performance Computers (HPC) have infrastructure to support research using models and simulations, which may be involved in generating and/or analysing high-volume data. The IT operations team at the organisation may have recommendations for data management, storage and preservation.

4. Discuss how to choose a data repository
  - a. When selecting a repository, consider these factors:
    - Choose a repository early on when you start your data project. This can help you with efficiently structuring and preparing your data when it comes time to share it.
    - Consider how FAIR a repository is in terms of the services it offers you.<sup>22</sup>
      - The repository provides persistent identifiers, e.g. Digital Object Identifiers or DOIs. This is essential as it supports citation and linking to other research outcomes, e.g. papers, and grants.
    - Landing pages are provided for the digital objects with metadata that helps others find them, determine what they are, relate them to publications, and cite them. This allows your research to be more discoverable, reusable, and trackable via download statistics.
    - Responds to community needs, is preferably certified as a ‘trustworthy data repository’ (e.g. Core Trust Seal), and addresses long term sustainability.
    - Is ideally internationally recognised, commonly used and endorsed by the respective community.
    - Matches your particular data needs, e.g. formats accepted; access, backup and recovery, and sustainability of the service. Most of this information should be contained within the data repository’s policy pages.
    - Offers clear terms and conditions that meet legal requirements, e.g. for data protection, and allow reuse without unnecessary licensing conditions, e.g. restricted vs. open.
    - Provides guidance on how to cite the data that has been deposited.
    - Whether the repository charges for its services.
  - b. **There are a number of resources to help choose a repository.** This chart is designed to assist researchers in finding a cross-disciplinary/generic repository should no discipline-specific repository be available to preserve their research data: <https://doi.org/10.5281/zenodo.3946719>

---

22 COPDESS (2021) *Enabling FAIR Data - FAQs - Selecting a (FAIR) repository*. Accessed 24 June 2021. [http://www.copdess.org/enabling-fair-data-project/enabling-fair-data-faqs/#1\\_Selecting\\_a\\_Repository](http://www.copdess.org/enabling-fair-data-project/enabling-fair-data-faqs/#1_Selecting_a_Repository)

- c. Discuss using a **catalogue for data repositories**: In order to find an appropriate repository, the cross-disciplinary directory re3data (<https://www.re3data.org>) can be used. This is the outcome of a DFG-funded (Deutsche Forschungsgemeinschaft – German Research Foundation) project that lists German and international repositories for research data, with more than 2,580 entries at present. Another option is the RDA-endorsed FAIRsharing (<https://fairsharing.org/databases/>) which interlinks repositories, (meta)data standards and policies.
  - d. For managing sensitive data, see Lesson plan 12 on ‘Dealing with confidential, personal, sensitive and private data and ethical aspects’.
5. There is a wealth of data repositories out there. How do I find and choose an appropriate repository?
- a. You can find a suitable repository by consulting FAIRsharing and re3data.org. Here you can select the discipline, type of data, and/or country. It is also possible to filter by very detailed criteria, for example, for repositories that charge a fee for data upload or where data use is restricted. Filtering by software is also an option and can be helpful if you are using an application programming interface (API) with a programming language/library, e.g. Zenodo API and Python, R and Dataverse.
  - b. Discuss with the class how to select a FAIR-aligned repository. Some infrastructure providers offer overviews of how their services enable FAIR. Zenodo offers an overview of how the service responds to the FAIR principles: <https://about.zenodo.org/principles/>. Figshare also published a statement paper on how it supports the FAIR principles: <https://knowledge.figshare.com/publisher/fair-figshare>.
6. Apply your knowledge:
- a. Based on the ‘How to choose a repository’ section and OpenAire’s guidance, use FAIRsharing or re3data to find a trustworthy repository in political science. What did you find?
  - b. From the ‘FAIR principles and repositories’ section, use the Zenodo Sandbox to upload test data, e.g. an example text file, and assign a licence. What did you find?

### Take-home task:

- Understand how you can connect your research for better discovery. Read more about your digital presence: <https://data.agu.org/resources/digital-presence>

**Materials/Equipment:**

- Computer/laptop
- Internet/browser
- Access for different repositories, e.g. credentials

**References:**

- OpenAIRE (n.d.) How to select a data repository? Accessed 23 July 2021. Available from: <https://www.openaire.eu/opendatapilot-repository-guide>.
- Biernacka, K., Bierwirth, M., Buchholz, P., Dolzycka, D., Helbig, K. Neumann, J., Odebrecht, C., Wiljes, C. and Ulrike Wuttke, 2020. *Train-the-Trainer Concept on Research Data Management (version 3.0)* [online]. Zenodo. <https://doi.org/10.5281/zenodo.4071471>.
- Piwovar, H. A. and Vision, T. J., 2013. Data reuse and the open data citation advantage. *PeerJ* [online] 1:e175. <https://doi.org/10.7717/peerj.175>.
- U.S. Geological Survey (USGS) (n.d.) Repositories. Available from: <https://www.usgs.gov/products/data-and-tools/data-management/repositories>
- Library Carpentry (n.d.). *Library Carpentry - FAIR Data and Software*. Available from: <https://librarycarpentry.org/lc-fair-research/07-assessment/index.html>
- AGU (n.d.). *Data & Software for Authors*. Available from: <https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Data-and-Software-for-Authors>
- COPDESS (2021). *Enabling FAIR Data - FAQs - Selecting a (FAIR) repository*. Available from: [http://www.copdess.org/enabling-fair-data-project/enabling-fair-data-faqs/#1\\_Selecting\\_a\\_Repository](http://www.copdess.org/enabling-fair-data-project/enabling-fair-data-faqs/#1_Selecting_a_Repository).
- Stall, S. Martone, M. E., Chandramouliswaran, I., Crosas, M., Federer, L., Gautier, J., Hahnel, M., Larkin, J., Lowenberg, D., Pfeiffer, N., Sim, I., Smith, T., Van Gulick, A. E., Walker, E., Wood, J., Zaringhalam, Maryam, & Zigoni, A., 2020. *Generalist Repository Comparison Chart* [online]. Zenodo. <https://doi.org/10.5281/zenodo.3946719>

## Lesson plan 12: Dealing with confidential, personal, sensitive and private data and ethical aspects

### FAIR elements:

#### Accessible

Once the user finds the required data, they need to know how they can be accessed, possibly including authentication and authorisation.

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
  - A1.1 The protocol is open, free, and universally implementable
  - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

#### Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
  - R1.1. (Meta)data are released with a clear and accessible data usage license
  - R1.2. (Meta)data are associated with detailed provenance
  - R1.3. (Meta)data meet domain-relevant community standards

**Primary audience(s):** Bachelor's, master's, PhD degree students

This lesson plan contains ideas for teaching students and researchers on how to deal with the FAIR principles in relation to data that cannot be shared publicly. There are data types that cannot be freely shared, such as confidential information regarding trade secrets, information about human participants, sensitive information about endangered species, data under contractual agreements that prevent data users from further sharing, and information with potential ethical implications. For the purposes of this lesson plan, we will refer to all such data as 'confidential data'. Even though sharing confidential data is less straightforward than data that can be routinely shared, such data can nevertheless benefit from applying the FAIR principles so



that researchers working with confidential data can benefit from work that has been done before in their domain.

Sharing confidential data will often come down to restricting access to a dataset. In this lesson plan, we provide lesson objectives and activities that can be done in class to discuss aspects worth considering when making confidential data findable and accessible for others.

Since countries have their own legislation and guidelines for working with confidential data of the types described above, we will not provide any formal definitions of these types of data here. The lesson plan is general enough to be adjusted to local legislation. The idea here is that readers who wish to use this lesson plan can adapt it in line with legislation applicable to the research context in which the students are working. The main message is that data which cannot be shared freely for one reason or another can still be made FAIR by adjusting the strategies implemented just enough to suit the circumstances and ensure compliance with local legislation and guidelines.

### **Learning outcomes:**

#### ***General (confidential, personal, sensitive and private data)***

- Can explain reasons for data protection (confidential, personal, sensitive or private data)
- Knows basic rules and legal regulations for sensitive data, e.g. GDPR
- Can list the requirements students need to meet when working with these types of data, adhering to applicable laws and regulations related to the research context
- Can analyse compliance to protect data appropriately
- Can apply mechanisms to protect data appropriately (concrete steps that researchers can take during the research lifecycle to protect the confidentiality of their research data where necessary)
- Can define different levels of data security (user, folder, files)
- Can explain and apply different methods of data protection (physical, password protection, encryption, etc.)
- Can use different levels of security for their own work
- Can identify which repositories may be used to archive/publish confidential data
- Can recognise that metadata of confidential data can be made public
- Knows that if a researcher wants to control access to an archived dataset, an organisational body and technical infrastructure need to be in place to deal with data access requests. Recognises that a set of criteria needs to be available on the basis of which access will be granted or denied

- Can recognise that it is possible to split up a dataset from a research project and store/archive/publish the separate parts with different access restrictions, e.g. one with confidential data (restricted access) and one with non-confidential data (publicly accessible, for example protocols, syntaxes)

### ***Dealing with personal data***

- Knows that informed consent needs to be set up in a certain way to be able to publish/share personal data
- Knows that data repositories may require data to be de-identified to a certain extent before they may be uploaded there
- Can describe directly identifying attributes and detect them in data
- Can explain the difference between anonymisation and pseudonymisation
- Can anonymise/pseudonymise data by stripping identifying attributes
- Knows that for reasons of information security, a pseudonymised dataset and the corresponding key file should be archived separately

### ***Dealing with ethical aspects***

- Recognises ethical aspects a researcher needs to take into account when planning to publish/share their data

### **Summary of tasks/actions:**

#### ***General (confidential, personal, sensitive and private data)***

1. Outline what research confidentiality means:
  - a. Define research confidentiality and give examples of confidentiality requirements for sample projects involving human participants, industries, endangered species or protected natural resources.
  - b. Let learners identify what types of data they are working with and how they should deal with them in terms of applicable legislation:
    - Take the applicable legislation, protocols or guidelines for your country/region or discipline. Familiarise your students with the main principles, preferably communicated in such a way that it speaks to your audience, i.e. try to avoid explanations that are formulated in formal legal language, and relate these main principles to practical actions for your audience. The overview below lists some examples and is far from exhaustive, so make sure to discuss legislation relevant for your audience.
    - Privacy legislation (see this database on data protection and privacy laws of the world to find the relevant legislation for your lesson). Note that students need to take into account the legislation in the country

- of the institution they are affiliated to and to the country/countries in which they are carrying out their research
- Europe: GDPR
  - Canada: PIPEDA
  - Australia: Privacy Act 1988
  - UK: Common law duty of confidentiality
  - Medical legislation
    - Netherlands: WMO (legal framework for medical scientific research)
    - U.S.: HIPAA
  - Animal testing regulation
    - Netherlands: Wet op de dierproeven (legal framework for research with animal testing)
- c. Ask students to discuss their own research projects in groups with a focus on the types of data they are collecting, how the relevant laws and regulations apply to them, and what this means for their workflow when it comes to making data available to others.
2. Explain the rationale behind the legislation, protocols and guidelines you discuss so that students understand why they are there and do not simply consider them boxes that need to be ticked. You can ask students to reflect upon the relevant requirements by producing statements and discussing them in the group.
  3. Introduce security measures that students can take to protect research participants and sensitive data related to them:
    - a. Prevent unauthorised access by means of reliable verification methods (passwords, two-factor authentication)
    - b. Pseudonymisation of personal data
    - c. Store key files in a location separate from other research data
    - d. Encryption (full disk, folders, files)
    - e. Grant access rights to those authorised to access the data
  4. Ask students to search for repositories (at their institution or outside, use, e.g. <https://www.re3data.org/>) that are suitable for archiving personal data. Instruct them to read the repository policies and to decide if they could use them to store/archive/publish their data. Let them share their results with the rest of the group so they can inform each other about potentially useful repositories.
  5. Introduce the concepts of depositing data and providing a description of a dataset. Explain that even if a dataset cannot be shared because it contains confidential data, the metadata describing such a dataset can be made public. Show examples of such cases, for example:

- a. Sagres ship meteorological data (in INESC TEC research data repository)
  - b. The influence of screen time on sleep quality (in DANS Data Stations)
6. Practical issues around sharing confidential data (this task could be too advanced for bachelor's and master's degree students; it is up to the teacher/instructor to assess whether this part should be included or excluded):
- a. Explain practical aspects which need to be arranged if a researcher wants to have control over who has access to a dataset, both in a technical and organisational sense. Even though these things are often beyond a researcher's control, they do influence the choice of a repository and researchers need to be aware of these issues when they work with confidential data and are aiming to share their data in some way.
    - Technical:
      - The location where data are stored should have the option to restrict access so only authorised people can access the data.
      - There should be a contact point where data access requests can be sent.
    - Organisational:
      - There needs to be someone who can receive data access requests and reply to them.
      - There needs to be someone who has the authority to decide whether a data access request will be granted or denied.
      - A set of criteria needs to be available as a decision-making basis for granting or denying access.
  - b. Conduct an exercise in which researchers think about conditions under which they would like to share their data. First, give them some examples of conditions for reuse and let them formulate conditions they would like to work with afterwards. Examples of conditions (based on the Terms of Use of the PsychData repository and the template for a data user agreement from Open Brain Consent):
    - Data may only be used for the purpose of academic research and instruction
    - Data may not be forwarded to third parties
    - Any publication based on the data must cite the dataset
    - No attempts may be made to re-identify or contact participants
    - Data needs to be stored in a secure work environment. Anyone reusing the data must provide the technical specification of the secure environment

- Data will only be provided when the applicant has approval for their research project from the Institutional Review Board of the applicant's institution
7. Explain the strategy of storing two separate datasets:
    - a. Illustrate that in a research project, two data packages may emerge once the data are ready for storing and publishing: one containing the confidential data, and another containing the non-confidential materials that could be valuable to other researchers, for example protocols, syntaxes.
    - b. Provide examples of such cases so students are presented with a tangible form of what a dataset with different access restrictions could look like:
      - FEM growth and yield data monocultures - Grand fir in DANS Data Station Life, Health and Medical Sciences. The plot data book, tree maps atlas and README file are publicly accessible, while access rights are required for the other files
      - European Quality of Life Survey in UK Data Service. The integrated data file requires login, whereas the other files can be explored online without a login

### *Dealing with personal data*

1. On setting up informed consent to be able to share data:
  - a. Give examples of aspects that need to be included in an informed consent form to be able to share data at the end of a research project. You can use the examples provided here, or find examples relevant to your situation:
    - The Ultimate consent form from Open Brain Consent, or the GDPR edition
    - Tool Research Data Management Language for informed consent, Portage Network
  - b. Ask students to take an informed consent template that is used in their department or suited to their discipline. Ask them to study the template and to find out if there are any statements about making data available to others after the project.
2. Familiarise students with instructions that repositories might have for de-identifying data before they may be uploaded there:
  - a. Give examples of repositories' instructions to de-identify data to some extent. You can use the example provided here, or find examples relevant to your situation:
    - 'The practice of protecting confidentiality' in the Guide to Social Science Data Preparation and Archiving - On p. 42-43 you can read how direct

and indirect identifiers need to be treated when preparing a dataset for reuse.

- b. Ask students whether they have a repository in mind in which they would like to deposit their data. Ask them to find out if this repository has any instructions on de-identifying data. Discuss the findings with the group.
3. On de-identifying data:
- a. Where relevant, demonstrate the difference between pseudonymous and anonymous data.
  - b. Introduce background materials on pseudonymising and anonymising data, for example:
    - Anonymisation step-by-step, UK Data Service – Practical steps researchers can follow to find potentially identifiable information in their data, to assess the uniqueness of values in their data and the risks related to that, and to make the data less identifiable
    - Pseudonymisation in small-scale quantitative research – This overview presents nine basic steps for pseudonymising data
    - Report Dealing with pseudonymization and key files in small-scale research – This report describes the nine steps from the overview above in a more detailed way
    - Guide to Social Science Data Preparation and Archiving – On p. 42-43 you find concrete steps for de-identifying data
    - Anonymisation section in the CESSDA Data Management Expert Guide – This section provides practical steps for making data about people less identifiable
    - Anonymisation postcard – This postcard illustrates that even with very little and general information, individuals can be identified, depending on the context
    - Privacy risks matrix – The matrix on p. 4-5 explains the risk levels for re-identification of data about people. P. 6 provides examples of various levels of de-identification
    - Brain MRI data sharing guide (and see the interactive version as well) – This guide provides MRI researchers with practical information about the implications of the GDPR for MRI research. On slide 8 you can find practical advice on how to de-identify MRI data; some of the methods discussed there can be applied to other types of data as well.

Based on these sources (or other relevant sources), explain what it means for data to be pseudonymous and anonymous (depending on the applicable legislation) and, based on that, help students to find out which steps can be taken to pseudonymise data and to determine if their data can be anonymised.

- c. Show examples of strategies and tools that are available for pseudonymising and anonymising data, for example:
  - Anonymising qualitative data, UK Data Service – Advice on how to de-identify various types of qualitative data: text, transcripts and audio-visual data
  - Anonymising quantitative data, UK Data Service – Advice on how to de-identify quantitative data, for example by removing or aggregating variables or reducing the precision of a variable
  - Amnesia Anonymization tool – A data anonymisation tool that removes identifying information from data, both by removing direct identifiers and transforming indirect identifiers to avoid unique values in a dataset. Ask students to discuss in groups which de-identification techniques are useful for their own research data.
4. Illustrate that in the case of pseudonymised data, a key file may exist which enables people to link direct identifiers to the research data again. Explain that, for security reasons, this key file should be stored in a different location to the file(s) with the research data, making it difficult to link the two files.

### *Dealing with ethical aspects*

1. Explain that sharing or publishing data should not harm individuals, which could for example be the case if the data have been collected among vulnerable groups, or when individuals have a unique set of circumstances. Refer students to the ethics committee or review board in their institution to help them assess if data sharing or data publishing could potentially be problematic for the participants involved.

### **Materials/Equipment:**

- Computer/laptop
- Internet/browser

### **References:**

#### *Useful links*

- Database on data protection and privacy laws of the world

***Background information on personal data in research***

- Research data risk matrix
- Anonymization postcard, LCRDM (Dutch National Coordination point for RDM)
- Privacy risks matrix, LCRDM
- Basic steps for pseudonymization, LCRDM
- Pseudonymization report, LCRDM
- Research Data Management Language for informed consent, Portage Network

***Guides***

- Brain MRI data sharing guide
- CESSDA Data Management Expert Guide > Anonimisation, Consortium of European Social Science Data Archives
- Guide to Social Science Data Preparation and Archiving, ICPSR (Inter-university Consortium for Political and Social Research)
- Learning hub on Research Data Management with advice on anonymisation, UK Data Service
- Anonymising qualitative data, UK Data Service
- Anonymising quantitative data, UK Data Service
- Anonymisation step-by-step, UK Data Service
- Guide Publishing and sharing sensitive data, Australian National Data Service
- RDM Guidance for COVID-19 including data sharing, Portage Network
- Code of Conduct Toolkit for GODAN (Global Open Data for Agriculture & Nutrition)
- Human Participant Research Data Risk Matrix, Portage Network

***Tools***

- Amnesia Anonymization tool
- Registry of Research Data Repositories <https://www.re3data.org/>

***Use cases***

- Sagres ship meteorological data (in INESC TEC research data repository)
- The influence of screen time on sleep quality (in DANS Data Stations)
- FEM growth and yield data monocultures – Grand fir in DANS Data Station Life, Health and Medical Sciences.
- European Quality of Life Survey in UK Data Service



### ***Templates***

- Data User Agreement, Open Brain Consent
- Ultimate consent form, Open Brain Consent
- Ultimate consent form - GDPR edition, Open Brain Consent

### **Take-home tasks:**

1. Ask students to take the informed consent template they intend to use and then discuss it with a privacy expert in their institution, adjusting as and where necessary to be able to publish/share data in the way they envision at the end of their project.
2. Ask students to have a close look at their own data and determine if they can be anonymised and, if so, if the anonymised result is still worth publishing/sharing.
3. Ask students to practice pseudonymisation and anonymisation techniques with a sample of their dataset so they learn how these techniques affect their data.

## Lesson plan 13: Data access

### FAIR elements:

#### Findable:

The data access category should not influence the findability of data; all data should be findable irrespective of their access; the main thing is that the metadata should be openly accessible for data to be discoverable/findable.

F2. Data are described with rich metadata (defined by R1 below)

#### Accessible:

Irrespective of the data access category selected, there should be clear information on how data can be accessed (described in the metadata), and the protocol should be open, free and universally implementable. If data access is restricted then an authentication protocol can be used.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

#### Interoperable:

Open data are easier to use as linked data in an interoperable way, especially if available through an API. But interoperability may also require key identifiers to link separate datasets. If these identifiers can identify individual people, e.g. point coordinates of a house, social security number of a person, then access restrictions will be needed to allow such data to be linked.

I3. (Meta)data include qualified references to other (meta)data

**Primary audience(s):** Bachelor's, master's, PhD degree students

### Learning outcomes:

- Can state general requirements on data protection and access control
- Understands the different access options that exist for data/digital resources
- Understands the criteria that influence/define access conditions

- Can apply strategies to decide which access level is suitable for their data
- Can implement (alternative) research practices to achieve more open data
- Recognises how access is important to make data FAIR (all 4 letters)

### Summary of tasks/actions:

1. Introduce your audience to the different access options that exist.  
 Research data can be made available in data centres, data repositories, via an AP, or on the web, with a range of access options. While open access to data may be ideal, there can be genuine reasons why that is not possible.  
 Data access categories<sup>23,24</sup> can be:
  - Open access
  - Restricted access
  - Embargo
  - Closed access
 Open data can be defined as ‘*data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike*’.<sup>25</sup>  
 Access restrictions can require a contractual use agreement or data sharing agreement to be signed.  
 Embargo means that access is closed temporarily.  
 Closed access means that data are not accessible, except maybe to regulators.
2. Explain the criteria that can influence access decisions<sup>26</sup>:
  - a. Presence of personal information in the dataset which can be used to identify an individual
  - b. Sensitivity of information, where the release of the data can adversely affect
    - a person, e.g. information on political views, criminal activities;
    - biodiversity, e.g. the location of rare and endangered species;
    - a community, e.g. terrorism; and/or
    - commercial interests of a company.
  - c. Intellectual property, where early release of the data can adversely affect patents or valorisation routes
  - d. Confidentiality agreement, where access to and sharing of data is restricted to the contracting parties.

23 <https://data.blogs.bristol.ac.uk/bootcampsd/repositories/>

24 <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/6.-Archive-Publish/Publishing-with-CESSDA-archives/Access-categories>

25 <https://opendatahandbook.org/guide/en/what-is-open-data/>

26 <https://data.blogs.bristol.ac.uk/bootcampSD/what-counts/>

3. Show how a suitable access level can be decided, for example, using a decision tree. Example: Data Sharing guidelines – WUR
4. Explain that alternative research practices, or adaptations to research practices, could be used to enable more open data. Examples include the following:
  - a. Capture data in an anonymous way
  - b. Anonymise information in a dataset so individuals (people, animals, etc.) cannot be identified from the information they have contributed during the research
  - c. Gain permission from people to make data open, even if the data contain personal or sensitive information (informed consent)
  - d. Use citizen science and participatory research methods to co-create data that are then co-owned and can be released as open data

**Materials/Equipment:**

- Computer/laptop
- Internet/browser

**References:**

Research Data Bootcamp (Bristol) - Repositories for sensitive data: <https://data.blogs.bristol.ac.uk/bootcampsd/repositories/>

CESSDA Data Management Expert Guide: <https://doi.org/10.5281/zenodo.3820472>

Open Data Handbook: <https://opendatahandbook.org>

FOSTER Open Science: The Open Science Training Handbook | Zenodo (p18 onwards)

FAIR Cookbook: Declaring data's permitted uses

Data Sharing guidelines – WUR

**Take-home tasks:**

1. Do one of these exercises on data access:
  - Exercise: Data access and licensing (UK Data Service) (with answer)
  - Exercise: Licensing and Access Controls (UK Data Service) (with answer)
  - Data access exercise (FAIRsFAIR)

## Lesson plan 13: Additional material – data availability statements

The list below provides some example data availability statements. Please note that data access statements should be tailored to suit each publication, checking that they meet all funder and publisher requirements.

Statement type	Example statement
Openly available data	“All data underpinning this publication are openly available from the University of FAIR-Data Repository at <a href="http://doi.org/10.15000/a789457">http://doi.org/10.15000/a789457</a> ”
Embargoed data	“All data underpinning this publication will be available from the University of FAIR-Data Repository at <a href="http://doi.org/10.15002/a1234a56">http://doi.org/10.15002/a1234a56</a> from 01/02/2019 onwards, following the cessation of an embargo period.”
Restricted data	“Due to ethical/commercial issues, data underpinning this publication cannot be made openly available. Further information about the data and conditions for access are available from the University of FAIR-Data Repository at <a href="http://doi.org/10.15000/a1234b56">http://doi.org/10.15000/a1234b56</a> ”
Partially restricted data	“Due to the sensitive nature of this research, only a subset of the participants consented to their anonymised data being retained and shared. Anonymised interview transcripts and survey results from participants who provided consent, other supporting data, and further details relating to the restricted data, are available from the University of FAIR-Data Repository at <a href="http://doi.org/10.15129/a1234b56">http://doi.org/10.15129/a1234b56</a> ”
Physical data	“Physical data supporting this publication are stored by the University of FAIR-Data. Details of the data and how they can be accessed are available from the University of FAIR-Data Repository at <a href="http://doi.org/10.15129/a1234b56">http://doi.org/10.15129/a1234b56</a> ”
Secondary data	“Pre-existing data underpinning this publication are openly available from UKDS at <a href="http://doi.org/10.12345/54321">http://doi.org/10.12345/54321</a> . Further information about data processing, and additional new supporting data are available from the University of FAIR-Data Repository at <a href="http://doi.org/10.15129/a1234b56">http://doi.org/10.15129/a1234b56</a> ”
No new data created	“No new data were created during this study. Pre-existing data underpinning this publication were obtained from NPL and are subject to licence restrictions. Full details on how these data were obtained are available in the documentation available from the University of FAIR-Data Repository at <a href="http://doi.org/10.15129/a1234b56">http://doi.org/10.15129/a1234b56</a> ”
No data	“This work is entirely theoretical, there is no data underpinning this publication.”

## Lesson plan 14: FAIR software/citable code

**FAIR elements:** All (for details on how the FAIR principles can be applied to research software, see table 1 of Lamprecht, Anna-Lena et al. 2020).

**Primary audience(s):** Master's and PhD degree students

### Learning outcomes:

- Is able to explain how research software differs from other types of software
- Can understand the modified FAIR principles for software (FAIR4RS)
- Understands accepted best practices on the basis of FAIR4RS
- Can apply the principles of software citation

### Summary of tasks/actions:

1. Define research software:
  - a. Give a definition of research software
  - b. Give examples and counterexamples, e.g. word processing software, of research software; be sure to include a breadth of examples including scripts and workflows
  - c. Identify similarities and differences between research data and software with regard to application of the FAIR principles
  - d. Identify similarities and differences between FAIR software and Free and/or Open Source Software (FOSS)
2. Explore how the FAIR principles can be applied to software (Chue Hong et al. 2021)<sup>27</sup>, in each case providing a concrete example of how to carry out the principle
  - a. **Findable – F:** Software, and its associated metadata, is easy to find for both humans and machines.
    - F1. Software is assigned a globally unique and persistent identifier.
      - F1.1 Different components of the software representing different levels of granularity are assigned distinct identifiers.
      - F1.2 Different versions of the software are assigned distinct identifiers.

---

<sup>27</sup> Draft published in June 2021 by the FAIR4RS RDA working group (Chue Hong et al. 2021): <http://doi.org/10.15000/a789457>, reserved DOI for revised version currently in press: <https://doi.org/10.15497/RDA00068>

- F2. Software is described with rich metadata.
  - F3. Metadata clearly and explicitly include the identifier of the software they describe.
  - F4. Metadata are FAIR, and are searchable and indexable.
- b. **Accessible – A:** Software, and its metadata, is retrievable via standardized protocols.
- A1. Software is retrievable by its identifier using a standardized communications protocol.
    - A1.1 The protocol is open, free, and universally implementable.
    - A1.2 The protocol allows for an authentication and authorization procedure, where necessary.
  - A2. Metadata are accessible, even when the software is no longer available.
- c. **Interoperable – I:** Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.
- I1: Software reads, writes and exchanges data in a way that meets domain-relevant community standards.
  - I2: Software includes qualified references to other objects.
- d. **Reusable – R:** Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).
- R1. Software is described with a plurality of accurate and relevant attributes.
    - R1.1 Software is given a clear and accessible licence.
    - R1.2 Software is associated with detailed provenance.
  - R2. Software includes qualified references to other software.
  - R3. Software meets domain-relevant community standards.
3. (Advanced) Explore how software quality goes beyond the FAIR data principles
- a. Quality of the form vs. quality of the function of a research software
  - b. Test for code maintainability
  - c. Validation of the functional correctness
  - d. Security measures
  - e. Computational efficiency
4. Recognise software citation as key to recognising research software as a first-class research output
- a. Software citation principles
  - b. Ways to improve citability of own software, e.g. citation file format: CITATION.cff

## References:

### *Definition of research software*

- FAIR4RS subgroup 3 - Research software definition
- Chue Hong, N. P., Katz, D. S., Barker, M., Lamprecht, A.-L., Martinez, C., Psomopoulos, F. E., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A. and Honeyman, T., 2021. *FAIR Principles for Research Software (FAIR4RS Principles)*. <https://doi.org/10.15497/RDA00065> Revised version in press, reserved DOI: <https://doi.org/10.15497/RDA00068>
- Lamprecht, A.-L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., Dominguez Del Angel, V., van de Sandt, S., Ison, J., Martinez, P. A., McQuilton, P., Valencia, A., Harrow, J., Psomopoulos, F., Gelpi, J. Ll., Chue Hong, N., Goble, C. and Capella-Gutierrez, S., 2020. Towards FAIR principles for research software. *Data Science* [online] 3(1), 37–59. <https://doi.org/10.3233/DS-19002>.
- Hettrick, S., Antonioletti, M., Carr, L., Chue Hong, N., Crouch, S., De Roure, D., Emsley, I., Goble, C., Hay, A., Inupakutika, D., Jackson, M., Nenadic, A., Parkinson, T., Parsons, M. I., Pawlik, A., Peru, G., Proeme, A., Robinson, J. and Sufi, S., 2014. *UK Research Software Survey 2014* [Data set] [online]. Zenodo. <https://doi.org/10.5281/zenodo.14809>
- Library Carpentry: FAIR Data and Software - Software

### *Best practices*

- Library Carpentry: FAIR Data and Software - Software
- Lamprecht, Anna-Lena et al., 2020. Towards FAIR Principles for Research Software'. 2020. *Data Science* [online] 3(1), 37–59. <https://doi.org/10.3233/DS-190026>
- Five recommendations for FAIR software

### *FAIR for Research Software working group*

- RDA - FAIR for Research Software (FAIR4RS) WG
- Lamprecht, Anna-Lena et al., 2020. Towards FAIR Principles for Research Software. 2020. *Data Science* [online] 3(1), 37–59. <https://doi.org/10.3233/DS-190026>
- Chue Hong, N. P., Katz, D. S., Barker, M., Lamprecht, A.-L., Martinez, C., Psomopoulos, F. E., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A. and Honeyman, T., 2021. *FAIR Principles for Research Software (FAIR4RS Principles)* [online]. <https://doi.org/10.15497/RDA00065> Revised version in press, reserved DOI: <https://doi.org/10.15497/RDA00068>



**Software citation**

- Katz, D. S., Hong, N. P. C., Clark, T., Muench, A., Stall, S., Bouquin, D., Cannon, M., Edmunds, S., Faez, T., Feeney, P., Fenner, M., Friedman, M., Grenier, G., Harrison, M., Heber, J., Leary, A., MacCallum, C., Murray, H., Pastrana, E., ... and Yeston, J., 2021. Recognizing the value of software: A software citation guide (9:1257). *F1000Research* [online]. <https://doi.org/10.12688/f1000research.26932.2>
- Chue Hong, Neil P. et al. ‘Software Citation Checklist for Developers’
- Citation file format: CITATION.cff

**Further resources**

- Katz, D. S., Gruenpeter, M. and Honeyman, T., 2021. Taking a fresh look at FAIR for research software. *Patterns* [online] 2(3), 100222. <https://doi.org/10.1016/j.patter.2021.100222>
- Chapter 9 RDA COVID-19 group recommendations for research software
- EOSC Executive Board Working Group - Scholarly infrastructures for research software
- Software Sustainability Institute - FAIR software
- CarpentryCon2020 FAIR Software course
- Library Carpentry - Top 10 FAIR Data & Software Things for specific disciplines
- Library Carpentry - Research Software
- CodeRefinery - Reproducible Research: Sharing code and data

## Lesson plan 14: Additional material on software citation

It is appropriate to consider software in the context of FAIR due to the close relationship between data and software. Citing software is key to recognising it as a first-class research object in the same way data are. The FAIR4RS Working Group is at present adapting the FAIR principles to research software<sup>28</sup>. Providing mechanisms to cite software effectively is still very much in progress and has proved to be a complex problem (D.S. Katz et al., 2019, arXiv 1905.08674 [cs.CY]). Nevertheless, significant progress has been made over the last five years. The FORCE-11 Software Citation Implementation Working Group have developed checklists for (paper) authors and (software) developers, best practices for software repositories and registries (arXiv 2012.13117 [cs.DL]), and guidance for journals (D.S. Katz et al.). The CodeMeta project is developing a minimal metadata schema for science software and code in JSON and XML.

JATS4R (JATS for Reuse), a working group devoted to optimising reusability of scholarly content by developing best-practice recommendations for tagging content in JATS XML, aims to support the various ways in which people can cite software.

Authors are exploring different ways to make their content, source materials, and methodology accessible to readers, and throughout this recommendation, we try to indicate where software citation initiatives are promoting change and development.

The following are the minimum requirements for a software citation (followed by desirable):

Required:

- Creator(s): the authors or project that developed the software
- Title: the name of the software
- Publication venue: the publication venue of the software, ideally an archive or repository that provides persistent identifiers
- Date: the date on which the software was published
- Identifier: a resolvable pointer to the software, ideally a PID that resolves to a landing page containing descriptive metadata about the software, similar to how a Digital Object Identifier (DOI) for a paper points to a page about the paper rather than directly to a representation of the paper, such as a PDF file. DOIs are preferable, and other examples of PIDs include Handles, RRIDs, ASCL IDs, swMath IDs, Software Heritage IDs, and ARKs. If there is no PID for the software, a URL to where the software exists may be the best identifier available

---

28 Revised version in press, reserved DOI: <https://doi.org/10.15497/RDA00068>

Desirable:

- Version: the identifier for the version of the software being referenced. If the version is unidentified or unknown, the date of access should be used
- Type: some citation styles, e.g. APA, require the inclusion of a bracketed description of the citation, e.g. computer software.

### *Recommendation*

Minimum requirements for a software reference

1. **<mixed-citation> @publication-type="software"**. Software citations MUST use a value of "software" for the @publication-type attribute.  
[[Warning when @publication-type is "Software", "SOFTWARE", "softwares" or "software" with anything else in the value]]

Note: This maps to Datacite resourceTypeGeneral attribute "Software". JATS4R policy is to use lowercase for attribute values, in turn requiring crosswalk mapping of "software" to "Software"

2. **<pub-id>**. If there is a well-defined identifier for software, this element should be used, for example doi, accession number, or SWHID. As per existing JATS4R recommendations on data citations, this element should be used to hold both the repository ID for the software in the element content, and, if applicable, the full URL to the data in the @xlink:href attribute.

Note: GitHub/Bitbucket/GitLab is not considered a reliable authority for providing IDs, so a GitHub git commit ID is not considered a <pub-id>.

3. **@pub-id-type** on **<pub-id>**. In contrast to what is stated in the Tag Library ("Type of publication identifier or the organisation or system that defined the identifier"), this attribute should only be used to state the type of identifier, and not to specify the organisation or system that defined the identifier (for example, doi, SWHID, accession).
4. **@assigning-authority** on **<pub-id>**. When the given type of identifier can be assigned by more than one organisation, e.g. accession numbers biomodels.db, docker hub, and the organisation registering the identifier is known, you should include the @assigning-authority attribute on the <pub-id> element.

Note: DOIs do not require an assigning-authority because although there are different DOI registrants, the DOI organisation is a central resolver service.

**Context:**

Elements: <element-citation>, <mixed-citation> <person-group>, <name> / <string-name> / <collab>, <article-title>, <version>, <pub-id>, <ext-link>, <date-in-citation>, <publisher-name>, <source>

**Attributes:**

@publication-type: Type of Referenced Publication (for example, “book”, “letter”, “review”, “journal”, “patent”, “report”, “standard”, “data”, “working-paper”),

@person-group-type: Role of the persons being named in <person-group> element (for example, author, editor, curator),

@designator: Used on such elements as edition number (<edition>) and version (<version>) to hold an unadorned numerical or alphabetical value of the edition or version number for machine search, when the number is a phrase or textual value,

@pub-id-type: Type of publication identifier, such as a DOI or a publisher’s identifier,

@assigning-authority: Names the authority that assigned or administers an identifier used in this document, for example, Crossref, GenBank, or PDB.

**Examples:**

1. *Example of accession with assigning authority pair, so renderer can create link. Preferred option, but appreciate many renders will not create the link:*

```
<ref id="bib2">
<element-citation publication-type="software">
<source>BioModels</source>
<pub-id @assigning-authority="EBI"
@pub-id-type="accession"      xlink:href="https://identifiers.org/biomodels.db:BI-
OMD0000000156">
BIOMD0000000156</pub-id>
</element-citation>
</ref>
```

2. *Example of accession with assigning authority pair, with URL too (if concern renderer(s) will not generate the link):*

```
<ref id="bib2">
<element-citation publication-type="software">
<pub-id @assigning-authority="biomodels.db" xlink:href="https://www.ebi.ac.uk/
biomodels/BIOMD0000000156">BIOMD0000000156</pub-id>
</element-citation>
</ref>
```

### 3. Example of identifier as URL link only (least preferred)

Github example

```
<ref id="bib2">
<element-citation publication-type="software">
<person-group person-group-type="author">
<ext-link ext-link-type="uri" xlink:href="https://github.com/JATS4R/jats-validator-docker">https://github.com/JATS4R/jats-validator-docker</ext-link>
</element-citation>
</ref>
```

#### **Additional reading:**

- Software Metadata Recommended Format Guide (SMRF)
- Katz, D. S., Hong, N. P. C., Clark, T., Muench, A., Stall, S., Bouquin, D., Cannon, M., Edmunds, S., Faez, T., Feeney, P., Fenner, M., Friedman, M., Grenier, G., Harrison, M., Heber, J., Leary, A., MacCallum, C., Murray, H., Pastrana, E., ... and Yeston, J., 2021. Recognizing the value of software: A software citation guide (9:1257). *F1000Research* [online]. <https://doi.org/10.12688/f1000research.26932.2>

Guidance for:

- journals: Katz, D. S. et al., 2021. Recognizing the value of software: a software citation guide [version 2]. <https://doi.org/10.12688/f1000research.26932.2>
- authors: Chue Hong, N. et al., 2019. Software Citation Checklist for Authors. <https://doi.org/10.5281/zenodo.3479199>
- software developers: Chue Hong, N. et al., 2019. Software Citation Checklist for Developers. <https://doi.org/10.5281/zenodo.3482769>
- software repositories and registries: Task Force on Best Practices for Software Registries, 2020. Nine Best Practices for Research Software Registries and Repositories: A Concise Guide. <https://arxiv.org/abs/2012.13117>
- software citation use cases: Smith, A. M. et al., 2016. Software citation principles. (Table 2). <https://doi.org/10.7717/peerj.2394/table-2>

#### Note on authorship

We recognise the author names are often missing from Github readmes, and only user names and handles are available. Likewise, contributors to code repositories vary over time, and the authors of software may differ from the authors of a research paper associated with the code. This recommendation offers no guidance on how to manage policy decisions associated with these issues. However, it deals with the lack of actual names by allowing for user names and handles to be used in author tags.

## **Lesson plan 15: Research data management – overview and best practices**

**FAIR elements:** All

**Primary audience(s):**

This lesson is intended to deliver a concise overview of the research data management (RDM) principles and practices for master's degree students or professional audiences of vocational education and training.

**Learning outcomes:**

- Understanding the RDM process and main use cases
- Understanding Open Research and Open Data (definition, standards, Open Data use and reuse, open government data, European policies and initiatives)
- Understanding FAIR principles in research data management, maturity model and compliance
- Working with sensitive, personal or private data (General Data Protection Regulation [GDPR] and its requirements, ethics approval process and form)
- Understand what a data management plan is, its purpose and benefits for a project or organisation
- Know tools, guides, templates to support RDM, metadata management, DMP creation
- Apply the acquired knowledge in practice, namely be able to create a DMP, create and publish data and metadata
- Understand the key roles in RDM: Data Steward, Chief Data Officer, Data Protection Officer and other employees of the institution who can support the creation of DMP

**Delivery format:**

This lesson can be delivered in the form of a tutorial, webinar or self-paced self-study course.

Required time: 2 lecture sessions (1.5 hrs each) and 1 practice session (approx 1.5 hrs)

**Prerequisites:**

- Basic knowledge of computer software and applications
- Understanding of organisational and/or research process and data used or produced

**Lesson topics (Summary of tasks/actions):**

1. Use cases for research data management and stewardship
  - a. Preserving the scientific record
2. Data management elements (organisational and individual)
  - a. Goals and motivation for managing your data
  - b. Data formats, metadata, related standards
  - c. Creating documentation and metadata, metadata for discovery
  - d. Using data portals and metadata registries
  - e. Tracking data usage, data provenance, linked data
  - f. Handling sensitive data
  - g. Backing up data, backup tools and services
3. Responsible data use (citations, copyright, data restrictions)
  - a. Data privacy and GDPR compliance
4. FAIR principles in research data management, supporting tools, maturity model and compliance
5. Data management plan (DMP)
6. Data stewardship and organisational data management
  - a. Responsibilities and competences
  - b. DMP management and data quality assurance
7. Open Research and Open Data (definition, standards, Open Data use and re-use, open government data)
  - a. Research data and open access
  - b. Repository and self-archiving services
  - c. Research Data Alliance (RDA) products and recommendations: persistent identifiers (PIDs), data types, data type registries, etc.
  - d. ORCID identifier for data and authors
  - e. Stakeholders and roles: engineer, librarian, researcher
  - f. Open Data services: ORCID.org, Altmetric Doughnut, Zenodo

**Practice:**

Hands-on practice including the following topics:

- Data management plan design, templates and tools
- Metadata and tools, metadata registries
- Selection of licences for Open Data and contents, e.g. Creative Commons and Open Database

**Materials/Equipment:**

- Collection of DMP templates
- Example of metadata for research data and publications
- Collection of links to RDM tools, metadata registries

**References:**

General Data Protection Regulation – <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

Licence selector – <https://ufal.github.io/public-license-selector/>

DMP Online – <https://dmponline.dcc.ac.uk/>

DMP Templates – <https://guides.lib.umich.edu/c.php?g=283277&cp=2138498>

Lamprecht, A.-L. et al., 2020. Towards FAIR principles for research software. *Data Science* [online] 3(1), 37–59. <https://doi.org/10.3233/DS-190026>

FAIR Cookbook, 2021, developed by Life Sciences professionals in the academia and the industry sectors, including members of the ELIXIR community [online]. Available from: <https://w3id.org/faircookbook>

FAIRsharing for (meta)data standards and interlinked repositories

**Take-home tasks:**

- Organisational data management plan creation (using the provided template and/or online tools)



## **Lesson plan 16: Data management and governance in industry and research**

**FAIR elements:** All

**Primary audience(s):**

This lesson serves to deliver a concise overview of the data management and governance (DMG) practices in research and industry for master students or professional audiences of vocational education and training, primarily with a computer or information science background.

**Learning outcomes:**

- Understand the enterprise data management and governance process and main use cases according to the DAMA (Data Management Association) Data Management Body of Knowledge (DMBOK)
- Understand the European data spaces concept and initiatives, European policies and regulations, GDPR (General Data Protection Regulation)
- Understand elements of the enterprise data management infrastructure and services: Data warehouses, cloud-based storage, data lakes
- Understand data modelling processes, data models, and data structures. Master data management
- Understand FAIR principles in research data management and their applicability to industrial use cases
- Understand data management maturity frameworks and best practices
- Understand what a data management plan is, its purpose and benefits for a project or organisation
- Apply the acquired knowledge in practice, namely be able to create a DMP and assess organisational data security and compliance
- Understand the key organisational roles in DMG: Chief Data Officer, Data Steward, Data Protection Officer and other roles

**Delivery format:**

This lesson can be delivered in the form of lectures and practice, a tutorial or self-paced, self-study course. Suggested time: 2 lecture sessions (1.5 hrs each) and 1 practice session (approx 1.5 hrs)

**Prerequisites:**

- Basic knowledge of computer software and applications.
- Understanding of organisational processes (HR/staff, customers, products, shipments, orders, etc.) and data used or produced.
- Basic understanding of SQL for the advanced course

**Lesson topics (Summary of tasks/actions):**

The DMG course uses DAMA DMBOK as a general framework covering the majority of topics, extending them with data science and big data analytics platforms and enriching them with FAIR and industry best practices. The following main topics should be included in the course:

- Introduction. Big data infrastructure and data management and governance. European data spaces: definitions use cases. European policy on data governance, data protection, GDPR
- Data management concepts. Data management frameworks: DAMA data management framework, the Amsterdam Information Model (AIM). Extensions for big data and data science
- Enterprise data architecture. Data lifecycle management and service delivery model. Data management and data governance activities and roles
- Data science professional profiles and organisational roles, skills management and capacity building
- Data architecture, data modelling and design. Data types and data models. Metadata. SQL and NoSQL databases overview. Distributed systems: CAP theorem, ACID and BASE properties
- Enterprise big data infrastructure and integration with enterprise IT infrastructure. Data warehouses. Distributed file systems and data storage
- Big data storage and platforms. Cloud-based data storage services: data object storage, data blob storage, data lakes (services by AWS, Azure, GCP)
- Trusted storage, blockchain-enabled data provenance
- FAIR data principles and data stewardship, FAIR digital object and persistent identifier (PID)
- Data repositories, Open Data services, public services
- Data quality assessment. Data management maturity frameworks: DNV-GL data quality framework, DCC RISE, CIMM, etc.
- Big data security and compliance. Data security and data protection. Security of outsourced data storage. Cloud security and compliance standards and cloud provider services assessment

**Practice:**

Hands-on practice including the following topics:

- a. Data management plan design, templates and tools
- b. Metadata and tools, metadata registries
- c. Assessing an organisation's data security and compliance requirements
- d. Advanced: Data modelling, relational data model creation

**Materials/Equipment:**

- Collection of DMP templates
- Example metadata for research data and publications
- Collection of links to enterprise data management and governance practices and recommendations

**References:**

- Earley, S., Henderson, D., & Data Management Association (Eds.), 2017. *DAMA-DM-BOK: Data management body of knowledge* (2nd edition). Technics Publications.
- GO FAIR Initiative [online] – <https://www.go-fair.org/go-fair-initiative/>
- General Data Protection Regulation – <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
- DMP Templates – <https://guides.lib.umich.edu/c.php?g=283277&p=2138498>
- Lamprecht, A.-L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., Dominguez Del Angel, V., van de Sandt, S., Ison, J., Martinez, P. A., McQuilton, P., Valencia, A., Harrow, J., Psomopoulos, F., Gelpi, J. Ll., Chue Hong, N., Goble, C. and Capella-Gutierrez, S., 2020. Towards FAIR principles for research software. *Data Science* [online] 3(1), 37–59. <https://doi.org/10.3233/DS-190026>
- A European strategy for data COM(2020) 66 final, 19.02.2020 – <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066>
- European Data Governance Act <https://ec.europa.eu/digital-single-market/en/european-data-governance>
- EU/Parliament Regulation on European data governance (Data Governance Act) SEC(2020) 405 final, Nov 2020 – <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767>
- GAIA-X. *A Federated Data Infrastructure for Europe*. Available from: <https://www.gaia-x.eu/>

FAIR Cookbook, 2021, developed by Life Sciences professionals in the academia and the industry sectors, including members of the ELIXIR community. <https://w3id.org/faircookbook>

**Take-home task:**

- Organisational data management plan creation (using the provided template and/or online tools)



This handbook was written and edited by a group of about 40 collaborators in a series of six book sprints that took place between 1 and 10 June 2021. It aims to support higher education institutions with the practical implementation of content relating to the FAIR principles in their curricula, while also aiding teaching by providing practical material, such as competence profiles, learning outcomes, lesson plans, and supporting information. It incorporates community feedback received during the public consultation which ran from 27 July to 12 September 2021.



ISBN: 978-3-86395-539-7

Göttingen University Press