CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	How to be FAIR with your data - A teaching and training handbook for higher education institutions
Authors	Engelhardt, Claudia;Biernacka, Katarzyna;Coffey, Aoife M.;Cornet, Ronald;Danciu, Alina;Demchenko, Yuri;Downes, Stephen;Erdmann, Christopher;Garbuglia, Federica;Germer, Kerstin;Hellström, Margareta;Hettne, Kristina;Hibbert, Dawn;Jetten, Mijke;Karimova, Yulia;Kryger Hansen, Karsten;Kuusniemi, Mari Elisa;Letizia, Viviana;McCutcheon, Valerie;McGillivray, Barbara;Ostrop, Jenny;Petersen, Britta;Petrus, Ana;Reichmann, Stefan;Rettberg, Najla;Reverté, Carmen;Rochlin, Nick;Saenen, Bregt;Schmidt, Birgit;Scholten, Jolien;Shanahan, Hugh;Straube, Armin;Van den Eynden, Veerle;Vandendorpe, Justine;Venkataram, Shanmugasundaram;Wiljes, Cord;Wuttke, Ulrike;Yeomans, Joanne;Zhou, Biru
Publication date	2022-01
Original Citation	Engelhardt, C., Biernacka, K., Coffey, A., Cornet, R., Danciu, A., Demchenko, Y., Downes, S., Erdmann, C., Garbuglia, F., Germer, K., Hellström, M., Hettne, K., Hibbert, D., Jetten, M., Karimova, Y., Kryger Hansen, K., Kuusniemi, M. E., Letizia, V., McCutcheon, V., McGillivray, B., Ostrop, J., Petersen, B., Petrus, A., Reichmann, S., Rettberg, N., Reverté, C., Rochlin, N., Saenen, B., Schmidt, B., Scholten, J., Shanahan, H., Straube, A., Van den Eynden, V., Vandendorpe, J., Venkataram, S., Wiljes, C., Wuttke, U., Yeomans, J. and Zhou, B. (2022) How to be FAIR with your data - A teaching and training handbook for higher education institutions. Available at: https://doi.org/10.5281/zenodo.5665492 (Accessed: 27 January 2022)
Type of publication	Book
Link to publisher's version	https://doi.org/10.5281/zenodo.5785252 - 10.5281/zenodo.5665492
Rights	© 2021, the Authors. This work is licensed under a Creative Commons Attribution 4.0 International License. - https://creativecommons.org/licenses/by/4.0/

Download date	2026-10-24 21:24:26
Item downloaded from	https://hdl.handle.net/10468/12492





Project Title	Fostering FAIR Data Practices in Europe
Project Acronym	FAIRsFAIR
Grant Agreement No	831558
Instrument	H2020-INFRAEOSC-2018-4
Topic	INFRAEOSC-05-2018-2019 Support to the EOSC Governance
Start Date of Project	1st March 2019
Duration of Project	36 months
Project Website	www.fairsfair.eu

D7.4 How to be FAIR with your data

A teaching and training handbook for higher education institutions

Work Package	WP7
Deliverable	D7.4 FAIR Competences Adoption Handbook for Universities
Lead Author (Org)	Claudia Engelhardt (University of Göttingen)
Contributing Author(s) (Org)	Katarzyna Biernacka (Humboldt-Universität zu Berlin), Aoife Coffey (University College Cork), Ronald Cornet (Amsterdam University Medical Centre), Alina Danciu (Sciences Po Paris), Yuri Demchenko (University of Amsterdam), Stephen Downes (National Research Council of Canada), Christopher Erdmann (American Geophysical Union), Federica Garbuglia (European University Association), Kerstin Germer (Humboldt-Universität zu Berlin), Kerstin Helbig (Humboldt-Universität zu Berlin), Margareta Hellström (Lund University and ICOS Carbon Portal), Kristina Hettne (Leiden University Libraries), Dawn Hibbert (University of Northampton), Mijke Jetten (Dutch Techcentre for Life Sciences and Health-RI), Yulia Karimova (Institute for Systems and Computer Engineering, Technology and Science), Karsten Kryger Hansen (Aalborg University), Mari Elisa Kuusniemi (University of Helsinki), Viviana Letizia (Elsevier), Valerie McCutcheon (University of Glasgow),

	Barbara McGillivray (King's College London and The Alan Turing Institute), Jenny Ostrop (University of Bergen), Britta Petersen (Christian-Albrechts-Universität zu Kiel), Ana Petrus (University of Applied Sciences of the Grisons), Stefan Reichmann (TU Graz), Najla Rettberg (University of Göttingen), Carmen Reverté (Institute of Agrifood Research and Technology), Nick Rochlin (University of British Columbia), Bregt Saenen (European University Association), Birgit Schmidt (University of Göttingen), Jolien Scholten (Vrije Universiteit Amsterdam), Hugh Shanahan (Royal Holloway, University of London), Armin Straube (University of Limerick), Veerle Van den Eynden (KU Leuven), Justine Vandendorpe (ZB Med – Information Centre for Life Sciences), Shanmugasundaram Venkataram (DCC and OpenAIRE), Cord Wiljes (Universität Bielefeld), Ulrike Wuttke (University of Applied Sciences Potsdam), Joanne Yeomans (Leiden University), Biru Zhou (McGill University)
Editorial Team	Raisa Barthauer, Yuri Demchenko, Claudia Engelhardt, Federica Garbuglia, Kerstin Germer, Margareta Hellström, Valerie McCutcheon, Hugh Shanahan, Armin Straube, Shanmugasundaram Venkataram, André Vieira, Joanne Yeomans, Biru Zhou
Illustrations	Patrick Hochstenbach (Ghent University)
Due Date	31.12.2021
Date	11.01.2022
Version	1.1 REVISED DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION
DOI	10.5281/zenodo.5665492

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

This work is licensed under a Creative Commons Attribution 4.0 International License.

Abstract

The handbook was written and edited by a group of about 40 collaborators in a series of six book sprint events 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 and teaching by providing practical material, such as competence profiles, learning outcomes and lesson plans, and supporting information. It incorporates community feedback received during the public consultation which ran from 27 July to 12 September 2021.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Versioning and contribution history

Version	Date	Authors	Notes
0.1	10.06.2021	Claudia Engelhardt, all book sprint contributors (authors and editors)	Draft for internal review
0.2	23.07.2021	Claudia Engelhardt, Editorial Team	Content ready for community review
0.3	23.07.2021	Claudia Engelhardt	Version for community review
0.4	12.09.2021	Community reviewers (see Acknowledgements)	Comments and suggestions for consideration by the authors and editors
0.5	11.10.2021	Claudia Engelhardt, all book sprint contributors (authors and editors)	Revised version incorporating comments and suggestions received during the community review
0.6	24.11.2021	Claudia Engelhardt, Editorial Team	Revised version for project-internal review
0.7	09.12.2021	Maria Johnsson (Lund University), Sandro Fiore (University of Trento)	Project-internal review
1.0	21.12.2021	Claudia Engelhardt, Raisa Barthauer	Final version incorporating comments received from internal review
1.1	11.01.2022	Claudia Engelhardt	List of contributing authors amended

Disclaimer

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.

Acknowledgements

This handbook underwent a community review from 26 July to 12 September 2021. We are grateful to all contributors for their valuable, much appreciated feedback.

We would like to extend special thanks for their extensive and thorough review and contribution to:
Romain David, Hervé L'Hours, Karsten Peters, Esther Plomp, Muriel Swijghuisen Reigersberg, Francesco Varrato, Niklas Zimmer

Furthermore, we thank:
Esther Asef, Bill Ayres, Noemi BC, Fay Campbell, Leyla Jael Castro, Julien Colomb, Philipp Conzett, Antica Culina, Stefanie De Bodt, Vilém Děd, Julian Dederke, Mary Donaldson, Christina Elsenga, Jeanine Finn, Vinciane Gaillard, Marjan Grootveld, W H, Simon Kerridge, Ilja Kocken, Ellen Leenarts, Allyson Lister, Lachlan MacLeod, Izaskun Mallona, Paula Martinez Lavanchy, Janice Masud-Paul, Joke Meeus, Gene Melzack, Megan O'Donnell, Lisanna Paladin, Limor Peer, Robin Rice, Jürgen Rohrwild, Susanna-Assunta Sansone, Gabriele Schwiertz, Yasmeen Shorish, Shelley Stall, Alexander Steckel, Liz Stokes, Annette Strauch, Ádám Száldobágyi, Rick Thompson, Christophe Trefois, Enrique Wulff, as well as everyone who contributed anonymously.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Table of contents

1 – Motivation **7**

2 – About this book **11**

2.1 How this book came about 11

2.2 What is FAIR? 12

2.3 Why make data FAIR? 14

2.4 Who will find this book useful and why? 15

3 – FAIR skills and competences **16**

3.1 The FAIRsFAIR Competence Framework and Body of Knowledge for Higher Education 17

3.2 FAIR competence profiles for the bachelor, master and doctoral level 18

3.3 Learning outcomes 21

4 – Teaching and training designs for FAIR **32**

4.1 Introduction 32

4.2 Elemental phases in course design 33

5 – FAIR lesson plans **48**

6 – Implementing FAIR **50**

6.1 Introduction 50

6.2 Getting to FAIR institutional policies 50

6.3 Data management planning 54

6.4 Data processing and documentation 54

6.5 Support infrastructure 55

6.6 Data publication 58

6.7 Data reuse 61

7 – References **63**

8 – About the authors & facilitators **69**

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Appendix A – Resources **72**

Appendix B – Target audience personas **74**

Appendix C – Data Stewardship Competence Groups (CF-DSP) and enumeration (according to FAIRsFAIR Deliverable D7.3) **80**

Appendix D – Draft Body of Knowledge (supplement to FAIR Competence Framework) **87**

Appendix E – Knowledge units and corresponding learning outcomes for bachelor, master and PhD level **101**

Appendix F – Lesson plans **112**

Lesson plan 1: FAIR in a nutshell 114

Lesson plan 2: Data management plans (DMP) 117

Lesson plan 3: Documentation 123

Lesson plan 4: Data creation 125

Lesson plan 5: File formats 128

Lesson plan 6: Metadata 131

Lesson plan 7: Data standardisation and ontologies 134

Lesson plan 8: Persistent identifiers (PIDs) 138

Lesson plan 9: Licences, copyright and intellectual property rights (IPR) issues 141

Lesson plan 10: Finding and reusing data 146

Lesson plan 11: Repositories 150

Lesson plan 12: Dealing with confidential, personal, sensitive and private data and ethical aspects 155

Lesson plan 13: Data access 163

Lesson plan 13: Additional material – data availability statements 166

Lesson plan 14: FAIR software/citable code 167

Lesson plan 14: Additional material on software citation 170

Lesson plan 15: Research data management – overview and best practices 174

Lesson plan 16: Data management and governance in industry and research 177

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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) in their educational programmes. Managing, curating, and preserving research data following the FAIR principles has undoubtedly acquired strategic importance in the institutional agendas of universities. Higher Education institutions across Europe and the world recognise that practicing good RDM is key to stay on par with the digital transition in the production and dissemination of scientific knowledge and, at the same time, to drive the shift towards the mainstreaming of Open Research, commonly known as Open Science.

This handbook offers a practical tool that supports universities in this endeavour, providing guidelines and model lesson plans for universities to integrate RDM and FAIR data-related content in bachelor, master and doctoral education programmes. It will also be of interest to other stakeholders wishing 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 about FAIR.

Survey data gathered from universities across 36 European countries illustrated a gap between recognising the strategic importance of research data skills and securing their presence in university programmes (Morais et al. 2021). While between 55% and 70% of 272 universities surveyed from 26 October 2020 to 15 January 2021 acknowledged the strategic importance of RDM and FAIR practices, the data indicated a substantial gap with their implementation. High levels of implementation were in fact reported by only 15-25% of the surveyed institutions. This gap was not limited to the level of institutional policies or infrastructure but was also evident with regard to 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 universities 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 confronted with the need to tackle challenges they are facing in the context of (open) research data. These are mainly determined by a general lack of awareness among their research communities of what the FAIR principles are and a widespread shortage of skills and competences related to how they can be put into practice (Morais et al. 2021).

New policies and frameworks are arising at the international, European and national levels to promote the mainstreaming of Open Research. They present universities with opportunities to confront the aforementioned challenges and receive financial and capacity support to develop their

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

own initiatives in support of Open Research practices. However, efforts to effectively leverage these opportunities will fall short of their potential if research and support staff are not equipped with adequate skills and competences.

At the European level, the FAIR principles will be a cornerstone of the European Open Science Cloud (EOSC) implementation. The EOSC will federate existing research data infrastructures from EU Member States and Associated Countries into a new, shared virtual environment, aiming at providing the scientific community with seamless access to FAIR research data and services. In this way, 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 in support of the development of 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 at present and in the future, and to contribute to its mission and implementation. At the same time, universities cannot, and should not, do this just by themselves, 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 in 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, is also crucial for 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 requirement 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 the EU funded programmes 2021-2027 will also require data produced by 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 demand that grant holders deliver reusable and accessible data arising from their funded research projects. Whilst funders have previously mainly encouraged data sharing, it is increasingly mandated that data created by publicly-funded research projects are 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, both national and funder as well as institutional policies play an

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

important role, and are constantly in flux (Sveinsdottir et al. 2021). Enhancing teaching and training provisions for RDM and FAIR data will be instrumental to address these new expectations on how research data should be managed and hence to ensure the continued access of institutions to European, national 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 getting ready 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 license costs to more qualitative costs related 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, it is expected that the benefits of having FAIR data will 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, and improve collaboration across scientific communities (ibid.). While the economic argument is part of the discussion around FAIR data, universities should 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 the demand for data-skilled professionals is growing exponentially (OECD 2020). Researchers also need to be equipped with a

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

basic level of data management skills allowing them to work efficiently within their research teams, where the distribution of competencies with their support staff is becoming increasingly variable (ibid.). However, at the doctoral level, general training will not be enough and will have to be coupled with a discipline-specific approach.

In conclusion, a growing number of international, European and national initiatives are emerging to embed Open Research practices in the standard way of conducting research. Investing in new and better training for RDM and FAIR data skills will be the key to fully take 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 in universities, requiring higher education institutions to take the lead in bringing forward 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 in bringing forward the transition to 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. The Handbook also presents didactic approaches on how to teach FAIR, equipping readers with knowledge for getting started with designing their own courses and training activities to be implemented at their institutions.

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

2 – About this book

2.1 How this book came about

This handbook was written first in a book sprint that was organised by the EU-funded FAIRsFAIR project under the lead of the University of Göttingen and subsequently finalised by an editorial process. It brought together a variety of experts in the field of RDM and teaching. The aim of FAIRsFAIR, which is running 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 pool of expertise to draw on beyond the project partners involved in this task (University of Göttingen, European University Association, University of Amsterdam and University of Minho), the approach of a book sprint was chosen – which can be very successful, as shown by recent examples such as 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.), *The FAIR Cookbook,* for the Life Sciences (https://fairplus.github.io/the-fair-cookbook/content/home.html), or the *Top 10 FAIR Data & Software Things* (Martinez et al. 2019).

The book sprint consisted of six three-hour sessions which were held between 1st and 10th June 2021: a kick-off meeting, four dedicated sprint sessions, and a wrap-up meeting. Because of the ongoing COVID-19 pandemic, everything took place 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. Although from diverse disciplinary backgrounds, they all possess ample relevant expertise in terms of RDM and the FAIR principles and in most cases also experience in teaching and training and/or lesson, course, or curriculum design. Including the FAIRsFAIR colleagues, about 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 in order to further improve the first version. This was followed by a revision from the Editorial Team, the presentation of the revised draft in a workshop on 12 October 2021, and its subsequent finalisation. The Handbook was published in December 2021.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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.

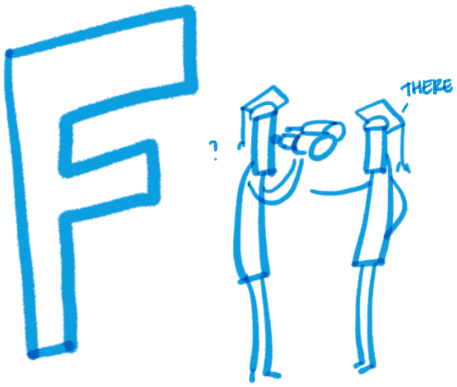
While FAIR and Open data are overlapping but distinctive concepts, both focus on data sharing, ensuring that data are made available in ways that promote access and reuse (Higman et al. 2019). While Open Research promotes a cultural change towards sharing research outputs, FAIR concentrates on how to prepare data in a way 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, due, for example, to 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 data that are FAIR compliant are not necessarily of high quality, and the issue of quality assurance of the data is a separate one, 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[1]).

The FAIR principles are typically translated into concrete complementary actions that should be taken by researchers, infrastructure providers, research funders and other actors (European Commission 2018; Science Europe 2021). They are increasingly becoming a requirement through European and national 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, etc.), which provide guidance on what they expect researchers to implement during the course of their projects, e.g. DMP templates, checklists to identify FAIR-compliant repositories, etc. (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.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558



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

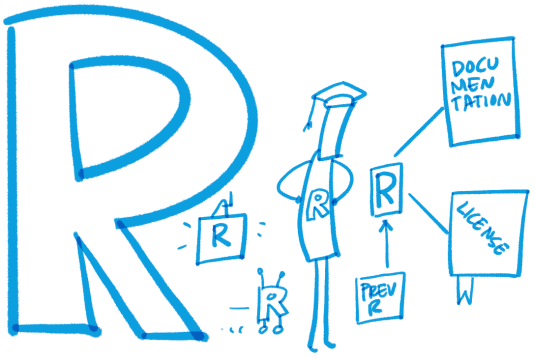
To be Reusable:

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



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

2.3 Why make data FAIR?

Upholding integrity and reproducibility are key to any good research, and best practice in RDM is an essential part of the effort to accomplish this. Open Research and especially the FAIR principles are a set of guidelines that could be viewed as a gold standard for RDM. There are many reasons beyond those of research integrity and reproducibility that should be taken into account for why adopting the FAIR principles should be encouraged and embraced. For researchers, both those that own or produce the data and those that reuse data provided by others, the ability to find, retrieve and reuse data will simply make their lives easier and will increase the visibility of the data so that its value is increased. In addition, FAIR data enable easier data integration within and across disciplines, supporting worldwide, multi- and interdisciplinary research endeavours addressing global challenges such as climate change, health emergencies or the realisation of the sustainable development goals. When considering the financial implications, especially for publicly funded research, reduction of double efforts and increasing reuse of existing data is a key motivator and there have been studies that show the implications of data management that are not FAIR-compliant (e.g. EC 2019): 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 fulfil the obligations of making their data FAIR compliant. Meanwhile, DMPs are also becoming increasingly important and mandatory and many templates explicitly provide guidance for the components of the FAIR principles, such as templates and guidelines provided by Horizon 2020 until recently, 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 the alignment of the two concepts where possible, with the notable exception of those data that cannot be made open for reasons such as their ethical sensitivity, copyright, cultural protocols, or commercial licensing. However, even for those data, metadata should be made available for discoverability, which can then be requested and shared in a safe manner through access control mechanisms where appropriate. Not only does this aid in data reuse but also increase 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, 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 covers only a part of best practice in RDM and Open Research (e.g. data curation practices, data services, data visualisation).

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 concerns a number of roles which contribute to this process at various levels.

To get 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 informed the development of the handbook structure and content. For more information about the procedure and the outcomes of the persona exercise, please refer to Appendix C.

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

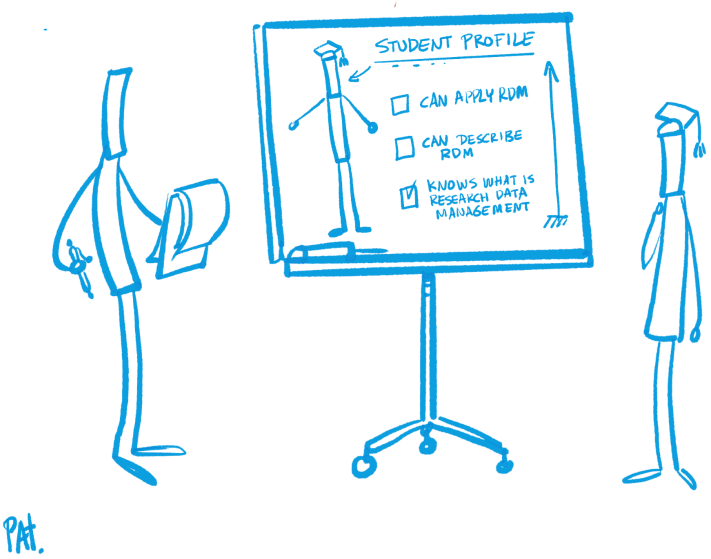
Area of activity	**Roles concerned (examples)**	**Most relevant chapters**
lesson and course planning, creation, and teaching	lecturers, professors, trainers, support staff	3 - FAIR Skills and Competences 4 - Teaching and training designs 5 - FAIR lesson plans
design, implementation, and adaptation of 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

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

3 – FAIR skills and competences

Before actually implementing topics related to FAIR in curricula and teaching, one first has to define which knowledge and competencies students at different educational levels should acquire. Here, we are suggesting a core set of Knowledge Units and associated learning outcomes for each of the bachelor, master and PhD levels.[2] These sets are discipline-agnostic and might 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).

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[3] (see Appendix E) created by the FAIRsFAIR project (which both in turn heavily build on the EDISON Data Science Framework, EDISONcommunity, 2020). They will be summarised below (section 3.1) before we describe the approach to (and result of) creating the competence profiles and learning outcomes (sections 3.2 and 3.3).



[2] Initially, we had considered six roles in total: In addition to bachelor, master and PhD 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.

[3] 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.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 that are relevant for Data Stewardship. It defines Competence Groups for the 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 D (taken from Demchenko et al. 2021, pp. 70 et sqq.).

The accompanying Body of Knowledge (BoK) breaks the Competence Groups of the FAIR Competence Framework down into a number of Knowledge Units (with each Knowledge Unit covering a specific aspect or topic), 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). There is 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 and so was the (draft) Body of Knowledge. However, at the time the book sprint took place, the FAIRsFAIR BoK was still a work in progress – with only one Knowledge Area Group having been updated yet compared to the original EDISON version: Data Management, which is the area that covers most of the Knowledge Units that are of importance for 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 E.

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

3.2 FAIR competence profiles for the bachelor, master and doctoral level

Method

The scope of the competencies 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 competencies 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 each the bachelor, master and PhD level by assigning one of five ranges (irrelevant, basic, intermediate, advanced or professional) to them.[4] These 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.

This step also involved the exclusion or merging of Knowledge Units that were considered irrelevant or redundant, e.g. a number of concepts relating strongly 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 the other areas of the BoK. In this case, the respective Knowledge Unit was added to the table. Some were not represented by any existing Knowledge Unit. In this case, a new item was created and added.[5]

In a second step, the group discussed and agreed upon which of the selected Knowledge Units could be considered entry-level content (or in other words compulsory topics). This was again done for the Bachelor, Master and PhD levels. The competence profiles that were defined this way are presented in the table below.

[4] 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, master and PhD students. These were reviewed by the other participants before the next session. In the following session, the group discussed each individual item, approved or amended the classification. Knowledge Units that were deemed irrelevant or redundant were removed. This way, the table was collaboratively consolidated.

[5] In addition to bachelor, master and PhD 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 focusing on the most relevant target audiences of HEI teaching.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Competence profiles

Table 2: Competence profiles for the bachelor, master and doctoral level

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
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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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
Data interoperability and metadata 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	basic	basic	intermediate	no

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Regulation) compliance				
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, for those Knowledge Units deemed entry-level content[6], 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 process of course design (more information about this will be provided in chapter 4).

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.

[6] Again, this limitation was due to capacity reasons. It would be desirable to create learning outcomes for the other Knowledge Units as well at a later point in time.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

		- [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.
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	basic	- [b] Can describe what a data management plan (DMP) is.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

compliance		- [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).
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, indigenious 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.

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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

Topic	**Required level**	**Learning outcomes** [b]=basic, [i]=intermediate, [a]=advanced]
General principles and concepts in data management – overview	inter-mediate	- [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	inter-mediate	- [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	inter-mediate	- [b] Can paraphrase the concept of Open Research. - [b] Can describe the benefits of Open Research. - [a] 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.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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	inter-mediate	- [b] Can describe what a backup is and give 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. - [i] Can explain institutional backup solutions and apply them to own files.
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	basic	- [b] Can explain aspects of interoperability

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

metadata management		(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	inter-mediate	- [b] Can summarise and explain ethical principles and responsible data use (e.g. CARE, indigenious 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	inter-mediate	- [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	inter-mediate	- [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	inter-mediate	- [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.

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

		- [i] Can use ontologies to describe resources (Apply).

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

Topic	**Required level**	**Learning outcomes** [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	inter-mediate	- [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, metadata 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.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

		- [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.
Open Science/Research, Open Access, Open Data	advanced	- [b] Can paraphrase the concept of Open Science. - [b] Can describe the benefits of Open Science. - [a] Can describe Open Access and Open Data as areas of Open Science. - [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. - [a] Can plan publication of Open Access publications and FAIR data.
Persistent Identifiers (PID), Open Researcher and Contributor ID (ORCID), Research Organization Registry (ROR)	inter-mediate	- [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. - [i] Can apply PIDs to their own research outputs. - [i] Can use PIDs to collaborate with others.
FAIR (Findable, Accessible, Interoperable, Reusable) principles in data management	inter-mediate	- [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. - [i] Can plan for FAIR research outputs. - [i] Can write and develop a research data management plan. - [i] Can apply the principles to their own work. - [i] Can evaluate the FAIRness of their own work or the work of others.
Master data management, data dictionaries	inter-mediate	- [b] Can develop a data management plan for their own work.

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

		- [b] Can identify different types of data documentation. - [b] Can explain the purpose of the documentation. - [b] Can use existing documentation. - [i] Can modify existing documentation. - [i] Can evaluate and prioritise data management activities.
Data security and protection	inter-mediate	- [b] Can define different levels of data security (user, folder, files). - [b] Can explain different ways of data protection (physical, encryption etc.). - [i] Can use different levels of security for their own work. - [i] Can apply data protection methods like password protection and encoding. - [i] Does share and collaborate in a secure way.
Data backup	advanced	- [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. - [i] Can explain institutional backup solutions and apply them to own files. - [a] Can analyse and evaluate backup. - [a] Can solve backup problems independently or with further assistance from support staff.
Personal data protection, GDPR compliance	inter-mediate (depen-ding on discipl.)	- [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. - [i] Can analyse compliance to legal regulations for sensitive data. - [i] Can apply mechanisms to protect data appropriately.
Data management planning, FAIR data management and compliance	inter-mediate	- [b] Can describe what a data management plan (DMP) is. - [b] Can explain why data management planning is a step towards FAIR.

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

		- [i] Can tell which areas should be covered in a DMP. - [i] Can sketch a DMP for their own research project.
Data interoperability and metadata management	inter-mediate	- [b] Can explain aspects of interoperability (Knowledge). - [b] Can relate metadata management to interoperability (Understand). - [i] Use domain-relevant standards, models and formats for interoperable data (Apply). - [i] Can relate metadata management to interoperability (Apply).
Data provenance, data lineage	inter-mediate	- [b] Can illustrate with an example what data provenance/data lineage means. - [i] Can transfer how data provenance/data lineage plays a role in their own research project. - [i] Can apply data provenance good practices to their own data and ensure that an unbroken data lineage is established for their work.
Responsible data use, data privacy, ethical principles, IPR and legal issues	advanced	- [b] Can summarise and explain ethical principles and responsible data use (e.g. CARE, indigenious 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. - [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.
Data quality management, best practices and frameworks, data quality metrics	advanced	- [b] Can summarise best practices ensuring data quality. - [i] Can describe how to recognise quality data. - [a] Can use best practices and frameworks on their own data to ensure their quality.
Trusted data repositories and certification	inter-mediate	- [b] Can explain what a trusted data repository is and how to find it (re3data.org and FAIRsharing).

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

		- [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	inter-mediate	- [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).

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

4 – Teaching and training designs for FAIR

4.1 Introduction

In higher education and research, the topic of FAIR has gained considerable interest. Teaching FAIR can be positioned in the broader discussion around 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.

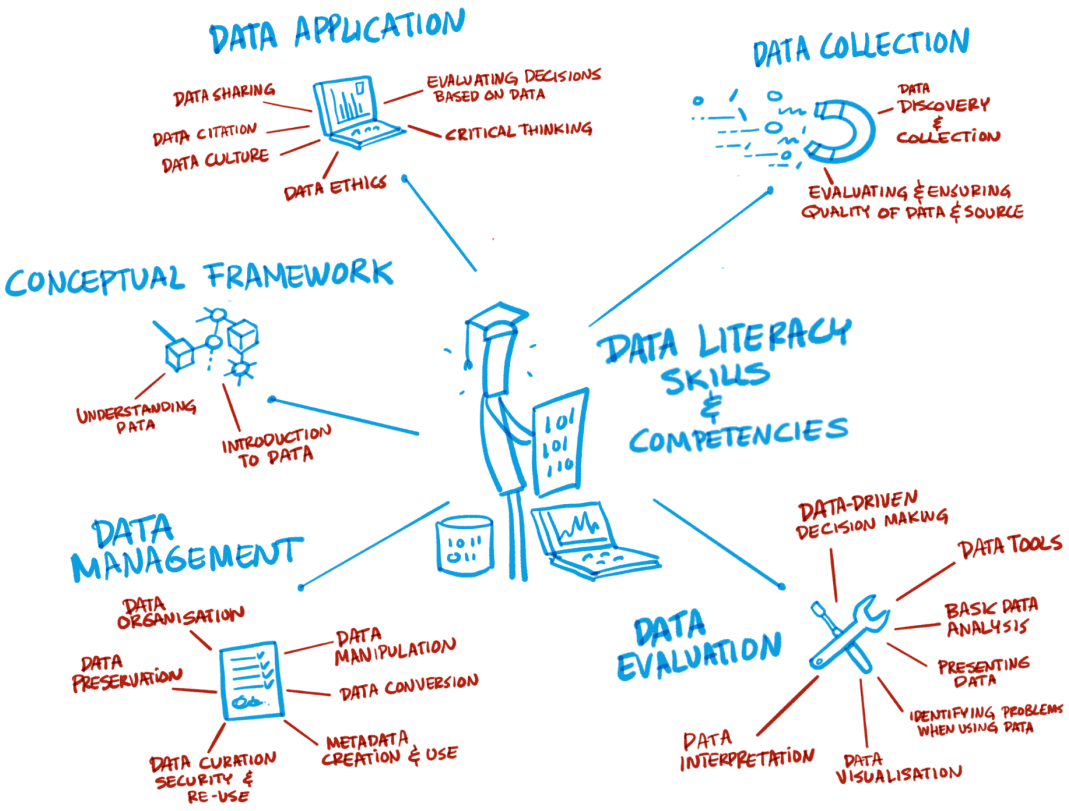


Figure 1: Schematic representation of data literacy skills and competencies by Patrick Hochstenbach, based on Guler (2019, p. 15), originally adapted from Ridsdale et al. (2015, p. 38).

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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 of this chapter help you to integrate current practices in FAIR data in your own teaching, without having to organise a separate course for these topics (but they also allow you to have a full course 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 of 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 and allows 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.
- Help your students (or other persons whom it may concern) to use open resources combined with disciplinary theories and models.

Before thinking about and working on the structure and the 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. In chapter 4.2, Step 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 the learners have been identified, the next steps towards the design of the course 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:

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- "Learning Experience (LE): any setting or interaction in or via which learning takes place: e.g., a lecture, game, exercise, role-play, etc." (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 the 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.

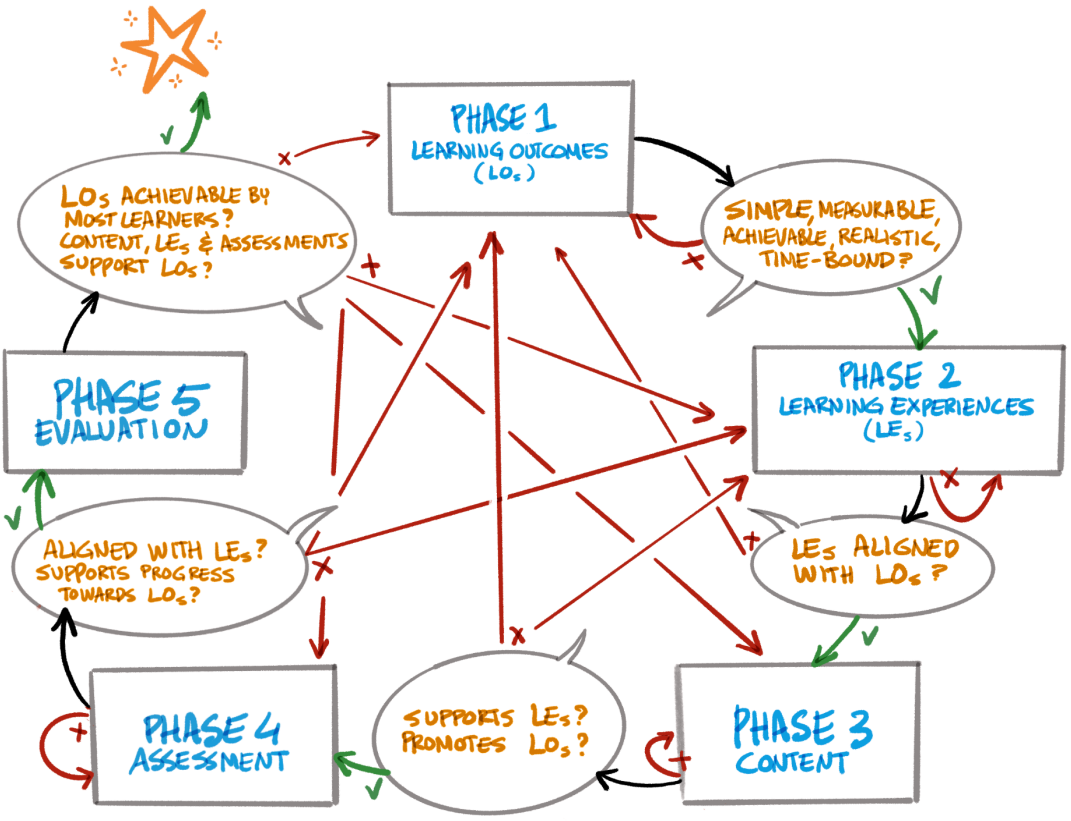


Figure 2: Nicholls' phases of curriculum design & their dependencies by Patrick Hochstenbach, adapted from Via et al. (2020, p. 4). The diamonds show the key considerations in each phase. While red arrows represent revisions in case the requirements resulting from the considerations have not been met yet, 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).

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 the 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 the driver of decision making when developing training and teaching (cf. 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 are essential to consider when developing measurable learning outcomes:

- be specific and well-defined
- be realistic
- rely on active verbs
- focus on learning products, not the learning process[7]
- be simple[8]
- be appropriate in number
- support assessments that generate actionable evidence

To summarise: Learning Outcomes should be based on competencies 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 that 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, you might also wish to include how the learning will be achieved (this part is more about the content), and why (this part is more about the incentives).

A helpful tool for formulating learning outcomes are taxonomies, like the taxonomy of educational objectives by Benjamin Bloom (known as Bloom's taxonomy or BT) that defines cognitive levels of learning outcomes (Bloom et al. 1956) and its revised version by Andersen and Krathwohl (Andersen and Krathwohl 2001) that 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 (such as: for a whole course, a specific session, for a part of a session, macro and micro-goals). As a general rule, one session might have around 3-5 individual learning outcomes (this can be discussed and adapted to the context in question, but one should be careful to not aim for more than can be achieved in the time available).

[7] 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).
[8] This means not to combine several pieces of knowledge, skills or abilities in one Learning Outcome.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

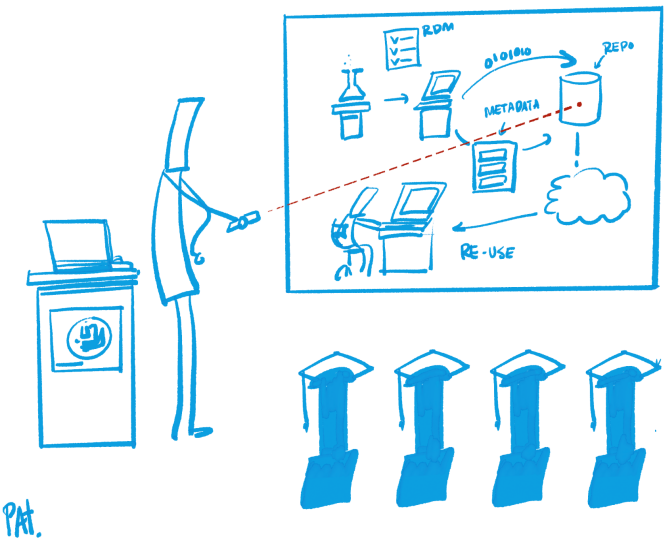
On a more generic level, the following learning outcomes could for instance be formulated, using the verbs of the Bloom's taxonomy to make learning outcomes actionable:

- Students can recognise and define the FAIR principles.
- Students can explain and interpret the FAIR principles.
- Students can apply the FAIR principles.
- Students can analyse and critically discuss the FAIR principles.
- Students can evaluate commonly used data repositories regarding their compliance with the FAIR principles with the aim to use them for their research area.

Furthermore, on a more granular level, learning outcomes may be formulated. For example:

- For each of the FAIR elements;
- On how these FAIR elements relate to the different stages of the data or research lifecycle;
- 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 fit the methods used to the time slot available, experiences and skills of the target group and their expectations. If the course is a part of a curriculum, students most likely do not 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 you want to develop and enhance the skills of staff members. The reasons for

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

participants to join informal training events might vary significantly. 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, basic training for new affiliate researchers, etc.). 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 the 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 for 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 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[9] material combined with an interactive workshop. Starting a lecture with researchers telling about their experiences, how they have implemented the elements of FAIR to 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, and therefore, it is good practice to include discussions, other activating methods and hands-on exercises following the lecture to consolidate the key learning points. 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 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 the content delivery in a dedicated time. Going beyond a dedicated lecture on FAIR, with a bit of planning, instructors may be able to fully incorporate FAIR teaching in any existing course (e.g., introduction to research methods).

[9] 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).

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 topic of FAIR or they can be more general. For example, in a "What should I know about FAIR" workshop, participants can discuss what FAIR means for them. In a "Where should I deposit my data to be FAIR" workshop, participants can choose a repository and deposit a data set. In a "How to write a DMP" workshop, participants can write their own DMP. The workshop can also focus on a research method, where you can embed tasks around FAIR, such as local institutional data storage options, documentation, file naming conventions, etc.

By arranging a workshop, you get an opportunity to discuss and find out 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 additional add-on workshops that would vary from year to year to meet the specific needs of the audience.

Pros: Workshops are ideal for delivering single topic content or for a targeted group of 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 topics of FAIR in one single workshop. Therefore, teachers or training providers may design and conduct a workshop series covering various topics of FAIR. Sometimes, learners might miss out on important topics covered in separate workshops due to self-selection biases (e.g. I am only attending the workshops that I deem interesting) or time constraints. Making connections from one workshop to another with brief recaps or highlighting key points of prior 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's 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 where you can 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 welcome 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 the willing to push the FAIR agenda forward.

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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 some 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 the flexible self-paced independent nature of the learning format. 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 (e.g. high enrolment rate but low completion rate). While traditional courses can 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 to this could be to make part of your online course synchronous by providing weekly or bi-weekly live office hours. Using interactive learning contents (e.g. https://h5p.org/) embedded in the LMS will also facilitate the retention of learner's interest..

Self-learning material

Self-learning material is an important part of any training format. This material is a reference that learners consult as a recommended information source after the course or event. Self-learning material can also be used separately to learn the basics of FAIR or 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 on organising the contents so that it is easy for users to browse and find the information they are looking for. The inventory of the self-learning materials will grow over time, thus a clear table of contents or glossary is an essential tool for users.

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 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) as references.

Cons: The learning experience of using self-learning materials is rather passive and it is not easy to track learning progress and outcomes. Many learners will fall into the scenario where: "I will look at it later" means "Never". Because it is relatively easy to produce and compile a large number of

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

self-learning materials, without proper attention to the organisation and maintenance of the most up-to-date information, self-learning materials could very quickly become a mess, which in turn could create difficulties for learners trying to find and access relevant information.

Training interventions[10]

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 realise that there is a clear knowledge gap, and these situations might be the right entry points to provide specific/customised information about FAIR and start a discussion about the topic to promote FAIRness in data management, not only for meeting grant application and policy requirements but also for improving 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 answer stakeholders' needs.

Cons: This service model is operating on a case-by-case basis, which might be time-consuming to reach all the stakeholders in your institute and not scalable, especially if you are operating with a very small service provision team.

[10] 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)

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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

Type of learning experience	Pros	Cons
Lectures	• Great delivery format for experienced and motivated learners; • Possibility to fully incorporate FAIR contents in any existing course.	• Could be time consuming; • Learner engagement is key.
Workshops	• Ideal for delivering single topic content or for a targeted group; • Flexible, short and easy to organise.	• To cover all topics of FAIR, a workshop series may be required; • Learners might miss out on important topics due to self-selection biases.
Events	• Most suitable for promotional purposes; • Great for networking; • Ideal to provide information for follow-up services and/or direct links to access existing materials.	• Limited time; • Messages need to be clear and concise.
Online courses	• Flexible, self-paced, independent. • Easy to manage and update at the back-end with minimal impact on learner experiences.	• Need to be aware of issues of student retention in asynchronous learning; • Lack of live interactions and low engagement; • Could include weekly or bi-weekly live office hours to include partial synchronous learning; • Use interactive course contents to facilitate learner engagement.
Self-learning materials	• Could be used and referenced in conjunction with other training formats;	• Passive learning; • Not easy to track learning progress and outcomes;

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

	• Could be used by learners, teachers, trainers, and research support staff as references.	• Need to pay attention to the organisation and maintenance of the most up-to-date information in self-learning materials.
Training interventions	• Great way to keep in touch with the research community by directly addressing their knowledge gaps.	• Case-by-case consultation; • Might be time-consuming.

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, utilise institutional, local, regional, national and/or international resources and make allies with the 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 different delivery formats mentioned above:

> With the joint efforts between 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 online self-paced learning program 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 faculty monthly meetings to promote various service offerings to researchers. Because 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.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Step 3: Select content that is relevant to the learning outcomes

The content of a course is the specific subject that it covers. Because FAIR encompasses a wide range of sub-topics, it needs to be broken down into individual content blocks, such as "copyright law", "metadata", "data repositories", etc. For a more comprehensive list, see chapter 5 which can be used as inspiration and a source to combine into your own course formats. Obviously, your selection of content and teaching format will depend on your audience and the available time.

With teaching the FAIR principles – as with most other topics – there is a very real danger of cramming too much content into too little time. Therefore, you should drop all content that is not aligned with the learning outcomes. If you identify essential content that needs to be covered, e.g. an existing institutional data policy, that does not support the learning outcomes, 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 next two steps.

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

1. FAIR activities by the letters

Topics may be presented in the order in which they are relevant for the four main topics of FAIR: Findable, Accessible, Interoperable, Reusable, in the order of the letters. This approach is especially relevant 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) are relevant for more than one principle, and it is usually helpful to build on already existing knowledge of students, you might also take one of the following three approaches into consideration. If you apply one of the following in your course, we recommend a special learning unit on FAIR in the overall curriculum, that connects the topics with the four key FAIR principles.

2. Along the research data lifecycle

The research data lifecycle[11] provides a generalised, structured look at the individual steps of how research projects handle research data. While it is obviously an idealised model, it has proven useful in teaching RDM, most notably the writing of data management plans (DMPs).

[11] Due to different disciplines and contexts, a great variety of such models exists (see, for example, Ball 2012).

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

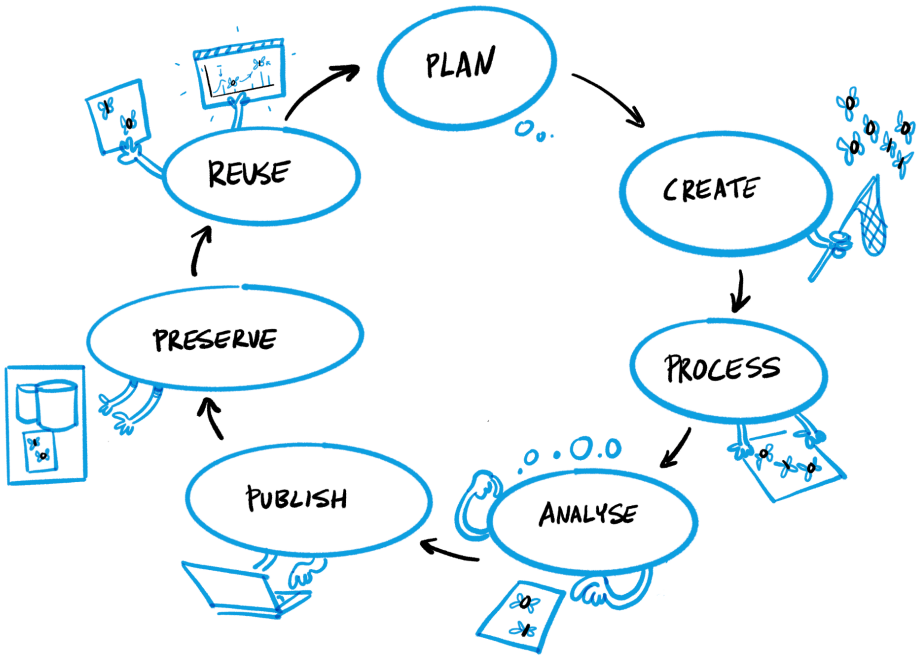


Figure 3: Research Data Lifecycle by Patrick Hochstenbach, adapted from UK Data Service, n.d.

The process starts with a research question and selecting possible approaches. Ideally, in this early stage, one explores what data already exists and might (partially) be reused (all letters of FAIR are relevant here). After drafting how data will be managed (ideally supported by a data management plan), data are collected, stored, described and analysed. Which data are preserved for the long-term depends on a range of conditions (ethical and legal restrictions, plans for further use, hardware costs, etc.), and finally the data may be prepared for publication and possible reuse by others or as input to a future project.

3. Linking FAIR practices to data management plans and planning

A data management plan (DMP) provides guidance through 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. Therefore, following the topics of a DMP template (e.g. Science Europe 2021) is a good 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 writing 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.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

4. Connecting 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, 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 help instructors to improve and update content (Via et al. 2020). As illustrated in Table 2 below, different assessments can be conducted at different levels and serve different purposes.

Table 7: Approaches to the assessment of progress towards learning outcomes

Assessment goals	Participants	Stakeholders	Example
• To identify the learning progress of the learners • To evaluate whether learners can apply their FAIR knowledge readily in a hands-on situation	Learners	Learners & Trainers	Ask learners to draft a DMP for their projects. This could be for an ongoing project or an example project. Work with learners on the DMP together to identify immediate knowledge gaps and provide relevant feedback and resources to learners via the drafted DMP. 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).
• To collect information from learners on what they like or dislike about the contents and format of delivery • The ultimate goal is to improve educational experiences for future learners	Learners	Trainers	Teaching and training evaluation surveys are often used in this context. Teaching and training programme developers might want to consult their institution's Teaching and Learning Services (or equivalent) for designing suitable evaluation survey questions. 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).

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

			Samples of course evaluation questions bank for a full course or a comprehensive learning programme (Consultation with your local teaching and learning services is recommended): 1. UC Berkeley: https://teaching.berkeley.edu/course-evaluations-question-bank 2. McGill University: ● English: https://www.mcgill.ca/mercury/files/mercury/course-evaluation-questionnaires-en-final.pdf ● French: https://www.mcgill.ca/mercury/files/mercury/course-evaluation-questionnaires-fr-final.pdf
● Primarily for administrative purposes ● To perform programme evaluation at the institutional level ● To budget and plan for resources	Trainers	Trainers and Institutions	Sample key performance indicators (KPI): ● Enrolment rates and completion rates of a specific programme ● Learner satisfaction survey ● Working hours used

Step 5: Evaluate course effectiveness

The final step is to evaluate if the course led learners to the learning outcomes that were defined in the beginning. The results will help to identify problems with the course design in order to make adjustments that can 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: e.g. for a full semester of seminars for students (Wiljes and Cimiano 2019), course evaluation 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 good basis to evaluate if students have acquired the knowledge, skills and abilities as defined by the learning outcomes. In addition, this also allows you to identify problems with specific topics and narrow this

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

down to the specific methods (i.e. learning experiences) that were used. To give an example: If the topic of “Metadata” is presented as a talk and the final evaluation of students’ DMPs reveals that they are not able to apply the contents properly, another teaching method for this topic should be tested. For example, you could 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.

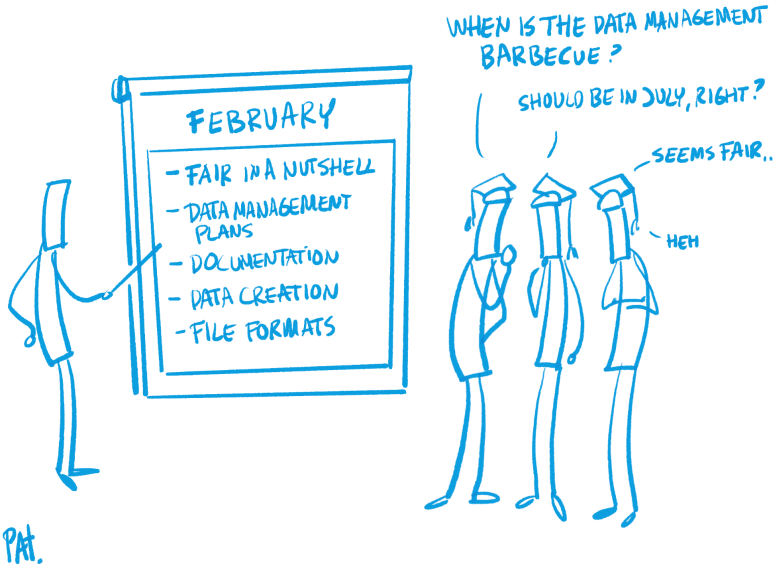
With shorter courses, e.g. a 4-hour workshop, evaluating course effectiveness is more challenging. We recommend leaving enough room for students’ questions and also writing them down. Usually, a lively discussion is a good sign that students are progressing.

To some extent, the metrics to assess learners’ progress as described in Step 4 might also help to evaluate overall course effectiveness. However, you should note that these metrics might also need to be improved iteratively.

Doing an anonymous survey on student satisfaction can complement the evaluation of course effectiveness. However, it should be interpreted with care because student satisfaction might 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).

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 (see list below). This list of lesson plans is not an exhaustive list and can be updated.

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

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

All lesson plans follow the same scheme[12] which includes the FAIR elements concerned, the learning outcomes, a summary of tasks/actions, material/equipment needed, references and take home tasks. More detailed information on the implementation of FAIR aspects (i.e. the practical application of the content taught through the lesson plans) is provided in chapter 6.

[12] The lesson plan template used here was developed based on this template: https://www.class-templates.com/support-files/lpt_word_001-printable_lesson_plan_template.pdf

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Table 8: Mapping of lesson plans to FAIR principles

Lesson	F1	F2	F3	F4	A1	A2	I1	I2	I3	R1
FAIR in a nutshell	X	X	X	X	X	X	X	X	X	X
Data management plans	X	X	X	X	X	X	X	X	X	X
Documentation										X
Data creation	X	X	X	X	X	X	X	X	X	X
File formats	X	X	X	X	X	X	X	X	X	X
Metadata	X	X	X	X	X	X	X	X	X	X
Data standardisation and ontologies	X				X	X	X			X
Persistent identifiers	X	X	X	X	X	X				
Licences, copyright and intellectual property rights										X
Finding and reusing data										X
Repositories	X	X	X	X	X	X	X	X	X	X
Dealing with confidential, personal, sensitive and private data and ethical aspects					X	X				X
Data access	X	X	X	X	X	X	X	X	X	
FAIR software/citable code	X	X	X	X	X	X	X	X	X	X
Research data management – overview and best practices	X	X	X	X	X	X	X	X	X	X
Data management and governance in industry and research	X	X	X	X	X	X	X	X	X	X

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

6 – Implementing FAIR



6.1 Introduction

Researchers can not do the heavy lifting in data management according to FAIR principles alone, they need to rely on support services provided by their institutions. This chapter therefore shifts the perspective from the individual researcher or research projects to the institution: How can they support their researchers with FAIR data management? Which support services are necessary, which infrastructure needs to be 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 as well as 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 is very much beyond the scope of this handbook.

6.2 Getting to FAIR institutional policies

Adopting an institutional research data policy that embraces the FAIR principles can give recognition, energy and resources towards the implementation of good practices. Implementing FAIR does require the reshaping and alignment of existing policies.

This section will look at key stakeholders and ways to cultivate an institution-wide FAIR research data environment.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 towards 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. An 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 either at the institutional or the departmental level (for example for clinical research practices). 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 as contradicting or impeding the implementation of FAIR for some research projects. Striving towards FAIR data management can make the task of protecting personal 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, ethics, etc.) to develop a cohesive FAIR research data framework at the institutional level that 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.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 the currentness of all policies. Every individual policy will then have a primary owner. The owner has to maintain the policy, supervise compliance and needs to organise 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 is currently occupying 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.
6. Policy owners forward 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 will therefore only provide a rough overview of potential stakeholders who might be involved in policy implementation or update:

Research offices look out for compliance with funder requirements and can be a key driver of institutional adoption of FAIR principles. Other involvements might 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, hosting of institutional repositories, and maintaining cyber and IT security.

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, reviewing of DMPs, RDM training and consultations.

Ethics boards need to give approval to a range of research proposals. Processes and procedures around research data are an important aspect in 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 by applicable privacy laws and regulations, such as the EU General Data Protection Regulation (GDPR) and

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 via the safeguarding of 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 to maintain 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 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 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 and private data and ethical aspects
Lesson Plan 16: Data management and governance in industry and research

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

6.3 Data management planning

Utilising Data Management Plans (DMPs) is a way to ensure the quality and consistency of data management throughout the data lifecycle and is 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 the requirements. Topics in DMPs usually include:

- **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. In terms of data creation, collection, and description, support and guidance can be provided.
- **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 depend on 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 is a key task 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, have sample plans, include the DMP templates from many funding bodies, and researchers can 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 knowing and understanding the context of data creation is 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 not having undergone any processing following their creation.

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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, data cleaning, data anonymisation, describing data using metadata, choosing appropriate data formats, and choosing appropriate repositories. Research fields differ (sometimes) markedly in all these parameters, e.g. in 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 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/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 re-usable 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 in order to reuse data and translate that information into documentation following appropriate metadata standards (if applicable);
- make sure to appropriately document all steps taken from raw, to processed data and then research-ready data;
- save & backup documentation alongside important iterations of both primary and processed data;
- think about the appropriateness of file formats in use and for sharing/publications.

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

To enable FAIR data practices within a Higher Education Institution resources and infrastructure are required. Some of the elements in FAIR can be met through open source solutions, however, in order to ensure that the potential for data to be reused is maximised, and to enable academics to optimise FAIR data, investment in both staffing and platforms is recommended.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Systems for storage, backup and collaboration

Researchers increasingly depend on 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 available for collaboration with partners while being appropriately protected. A common back up 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 by third parties using a variety of open source or proprietary technologies hosted. Whatever the selected technical infrastructure investment is required to enable academics to optimise FAIR data to maximise their potential for reuse.

In addition to back up 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 (control on permissions) mechanisms can facilitate collaboration and data sharing within teams. Access control is also critical to protect sensitive data such as personal information about data subjects which has both legal and ethical implications.

Questions about storage, backup and data security are included in many DMP templates 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 life cycle and in enabling FAIR. They provide a persistent identifier, make the descriptive metadata available and give access to the data (if applicable). Some repositories offer preservation, 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 around metadata, ontologies, software etc. These repositories can provide technical infrastructure for ongoing storage, resource discovery and access of (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 that 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 on ongoing migration to long term formats), supporting technologies, and associated metadata and ontologies as they change over time. The FAIRsFAIR project has developed a capability/maturity approach that aligns repository

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 on the level of support required by their data types and that offered by the repository. These can range from basic storage, to resource discovery to managing access and use of sensitive data, supporting the peer review of data associated with publications or services around digital preservation.

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

1. Best choice (if available) is to use a dedicated (external) data archive or repository already established that caters specifically to the research domain to preserve the data according to recognised discipline specific standards.
2. 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 might 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 at FAIRsharing.

Cost-free 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 in developing institutional capacities in this dynamic field, for example in regard to digital preservation. It is therefore recommended that where funding/resources are available that Higher Education Institutes invest in their own data repositories. 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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 utilised to ensure the 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 or 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* (https://public.ccsds.org/pubs/650x0m2.pdf) ensures that data are secure, findable and usable for as long as it is needed. Not only do many research funders require that datasets are either 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 research data generated after a period of time are still accessible, even if the software and technology in use at that time is now outdated.

The *DPC Rapid Assessment Model* (https://www.dpconline.org/digipres/implement-digipres/dpc-ram) has been designed to enable organisations for 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 (https://www.dpconline.org/handbook) with 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 the researchers' contributions is fundamental to ensuring widespread adoption of FAIR principles. Once the data have been created, processed, analysed, and their preservation is ensured, a clear pathway to crediting the authors in all data-related publications needs to be established. At a minimum, datasets need to be cited like other references to provide credit to the researchers.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 data sets 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 re-usable. 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, making the work more shareable, discoverable, comprehensible, reusable, and reproducible.

Authors need to provide contextual information on the relevant data set, method, software code or another 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. Because these statements are often in free-text form, the identification of the level of data access and availability expressed in them is not a straightforward operation. However, a study on 531,889 research articles from PLOS[13] and BMC[14] (Colavizza et al. 2020) has shown that only 12-21% 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 the publication of the data in a repository and the reference to it in the research papers, dedicated data papers can also contribute to the increased visibility of the data and recognition of the researchers' work.

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

[13] Public Library of Science (PLOS)
[14] BioMed Central (BMC)

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

relevant raw data in a repository, 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* (www.nature.com/sdata), *GigaScience* (www.academic.oup.com/gigascience), *F1000Research* (https://f1000research.com) for scientific disciplines and the Journal of Open Humanities Data (https://openhumanitiesdata.metajnl.com) and *Research Data Journal for the Humanities and Social Sciences* (https://brill.com/view/journals/rdj/rdj-overview.xml) 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 the point of view of researchers, 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 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 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* (www.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 for example the integration between GitHub and Figshare/Zenodo. The Software Sustainability Institute offers advice (https://www.software.ac.uk/which-journals-should-i-publish-my-software).

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

Other forms of articles that focus on a specific component of the research or the research process include such on hardware and lab resources as well as microarticles and visual case discussions (see Elsevier n.d.).

Learn more:
Lesson Plan 13: Data access
Lesson Plan 14: FAIR software/citable code

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 have 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 will look 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. Data sharing agreements define the purpose of data sharing, govern what happens to data at each stage of the research process, specify 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 be integrated in 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, such as "information sharing agreement", instead of data sharing agreement is in use.

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 for e.g. data protection principles.[15]

Enhancing discoverability

Researchers following the FAIR data principles will have documented their data with rich metadata. To make data sets findable, these metadata need to be available as widely 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 data set 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

[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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

conscious effort by the repository. The 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 providing information via OAI-PMH.

Another way of enhancing the discoverability of data sets 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. On a more advanced level is the interlinking of data sets 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 as an integral part of their research process. Activities supporting this aim include:

- Showcasing examples of successful reuse of data sets 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
- Creating experimental and collaborative spaces like Data Labs (Open a Glam Lab offers advice how to approach this task)

Learn more:
Lesson Plan 9: Licences, copyright and intellectual property rights (IPR) issues
Lesson plan 10: Finding and reusing data

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

7 – References

Zotero group: https://www.zotero.org/groups/4273178/fairsfair_booksprint

All references accessed 17 December 2021.

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 [Accessed 10 June 2021]

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. *The Open Science Training Handbook*. Zenodo, 2019. 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., Masia, B. B., 1956. T*axonomy of educational objectives. 1: Cognitive domain*. New York: McKay.

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, 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*. Accessed 7 June 2021. Available from: https://casrai.org/rdm-glossary/. [Accessed 7 June 2021].

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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., HrynaszkiewiczI, 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. Data FAIRport Conference. Jointly Designing a Data FAIRport* [online]. Available from: https://www.datafairport.org/component/content/article/8_news/9_item1/index.html [Accessed 21 July 2021].

Davidson, J., Engelhardt, C., Proudman, V., Stoy, L., and Whyte, A., 2019. *D3.1 FAIR Policy Landscape Analysis* [online] (v1.0). FAIRsFAIR. 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]. 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.

German Research Foundation, 2019. *Guidelines for Safeguarding Good Research Practice. Code of Conduct* [online]. https://doi.org/10/gg927v.

Donaldson, M., Schwamm, H. and Campbell, Fay, 2017. EPSRC DMP Assessment Rubric v2.0 [online]. Zenodo. https://doi.org/10.5281/zenodo.247087.

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.

Elsevier. *Research Elements*. Elsevier. Available from: https://www.elsevier.com/authors/tools-and-resources/research-elements.

European Open Science Cloud, 2021. *Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC)* [online]. Version 1.0 European Open Science Cloud. Available from: 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:

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

https://op.europa.eu/en/publication-detail/-/publication/9c7f3c1e-2aea-11eb-9d7e-01aa75ed71a1.

European Commission, 2019. *Cost-Benefit Analysis for FAIR Research Data* [online]. Available from: https://op.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1/language-en.

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 [Accessed 23 June 2021].

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: https://gdpr-info.eu/ [Accessed 10 June 2021].

FAIR Cookbook: an online resource for the Life Sciences to make and keep data FAIR, created by academics (including ELIXIR) and pharmas [online]. Available from: https://fairplus.github.io/the-fair-cookbook/content/home.html [Accessed 26 July 2021]

FAIRsharing, interlinking standards, repositories and policies [online]. Available from: https://fairsharing.org and https://doi.org/10.25504/FAIRsharing.2abjs5 [Accessed 26 July 2021].

Faniel, I. M. and Jacobsen, T. E., 2010. Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. *Computer 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/learning/use-open-data-in-teaching/#/id/5b06ca3ba8216e322da360b6f [Accessed 20 July 20219.

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.

Higman, R., Bangert, D. and Jones, S., 2019. Three Camps, One Destination: The Intersections of Research Data Management, FAIR and Open. *Insights the UKSG Journal* 32 (18), https://doi.org/10/gf4jhr.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

ICPSR - Preserving Respondent Confidentiality [online]. Available from: https://www.icpsr.umich.edu/web/pages/deposit/confidentiality.html [Accessed 17 June 2021].

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-Assessent. *American Journal of Pharmaceutical Education* [online] 82 (5) 6299. https://doi.org/10.5688/ajpe6299.

Koninkrijksrelaties, Ministerie van Binnenlandse Zaken en, 2021. *Wet medisch-wetenschappelijk onderzoek met mensen* [online]. Wet. Available from: https://wetten.overheid.nl/BWBR0009408/2021-05-26. [Accessed 10 June 2021]

———. *Wet op de dierproeven* [online]. Wet. Available from: https://wetten.overheid.nl/BWBR0003081/2019-01-01 [Accessed 10 June 2021].

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., Giaretta, 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://dx.doi.org/10.21428/16ac48ec.f54af6ae.

Martinez, P. A., Erdmann, C., Simons, N., Otsuji, R., Labou, S., Johnson, R., Castelao, G. et al., 2019. Top 10 *FAIR Data & Software Things* [online]. https://doi.org/10/gkbnxv.

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]. Brussels: European University Association. 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*. 0 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]. https://doi.org/10.6084/M9.FIGSHARE.1354039.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 [Accessed 17 June 2021].

RDMkit [online]. Available from: https://rdmkit.elixir-europe.org/ [Accessed 2 June 2021].

Registry of Research Data Repositories [online]. Available from: https://www.re3data.org/ [Accessed 10 June 2021].

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]. N.p.: FAIRsFAIR. 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]. The Turing Way. Available from: https://the-turing-way.netlify.app/community-handbook/bookdash.html [Accessed 12 July 2021].

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]. 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 [Accessed 23 July 2021].

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.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

8 – About the authors & facilitators

Name, Affiliation & ORCID	**CRediT Roles** (https://credit.niso.org/)
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
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

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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
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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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
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, 0000-0002-8217-4025	Writing – original draft
Biru Zhou, McGill University, 0000-0001-6914-1432	Writing – original draft Writing – reviewing and editing

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Appendix A – Resources

This is not an exhaustive list but a starting point.

Glossaries

- CASRAI RDM Terminology
- Terms4FAIRSkills

DMP (and other) tools and guides

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

Guides/Practices

- ELIXIR RDMkit
- CESSDA Data Management Expert Guide
- 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 Aware Tool
- FAIR Cookbook, practical recipes with applied examples
- The Turing Way

Support for license 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/

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Appendix B – Target audience personas

This appendix is the documentation of 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 into the position of a reader of the book that they were going to collaboratively write, and think about the needs and requirements regarding the handbook from the recipient's perspective. The outcomes informed the discussions about 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. And, most importantly, by stepping into the persona's shoes, they tried to answer the questions:

- 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?

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

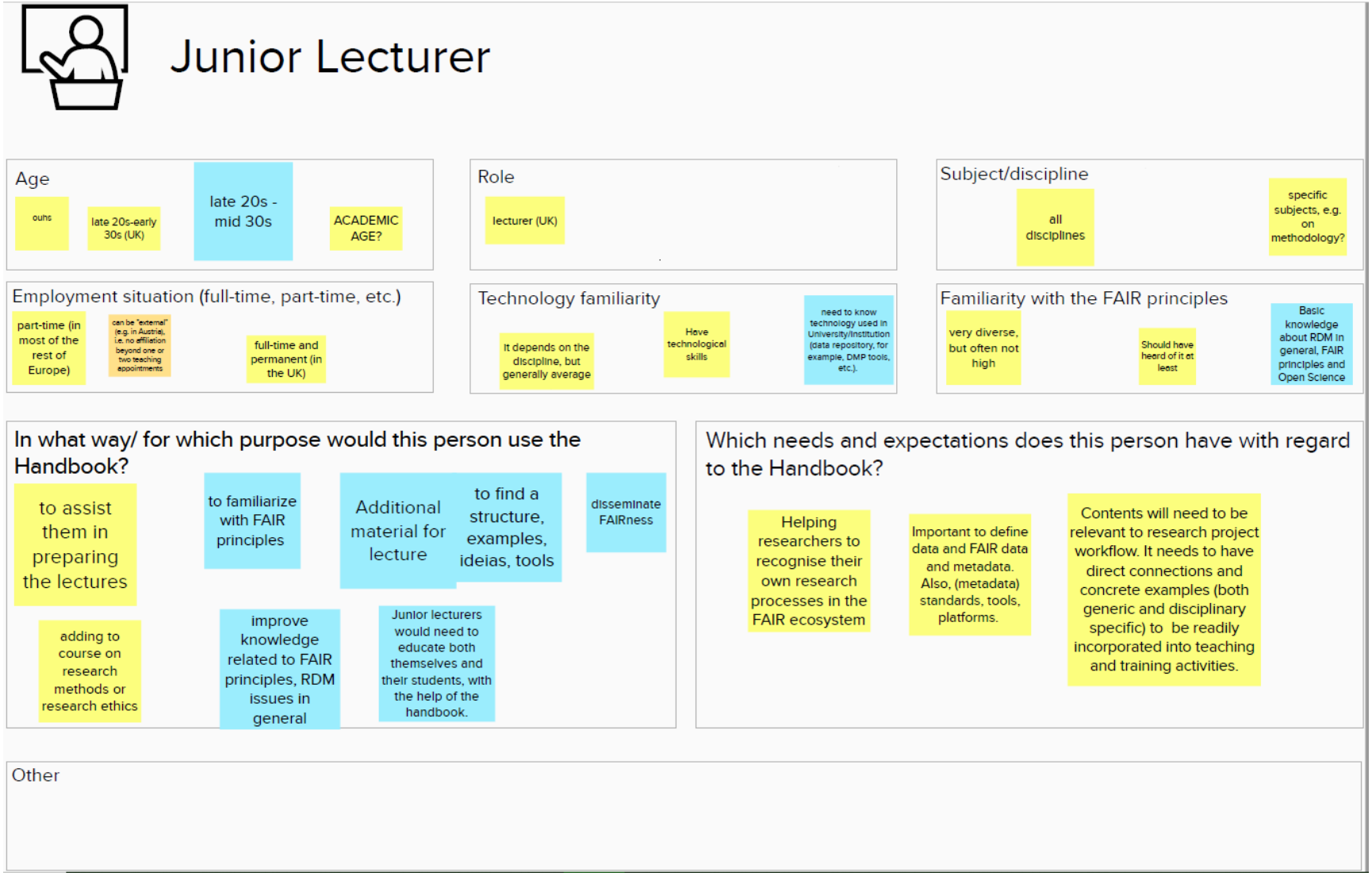
Here is a brief summary of the breakout session outcomes.

In terms of the purposes, there are strong similarities between the junior lecturer, the professor and the support staff member. All three are envisioned 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 inform and teach themselves about the FAIR principles (all three groups), to get advice for grant applications (professor), or use it as a reference for good FAIR practice and to check FAIR compliance (support staff). The doctoral programme manager is seen using the handbook for the 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 the 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 that they are using it as a tool for in the best way possible. It should therefore be easy to navigate and to understand, the content should be accurate and 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 for tailoring courses to a specific audience.

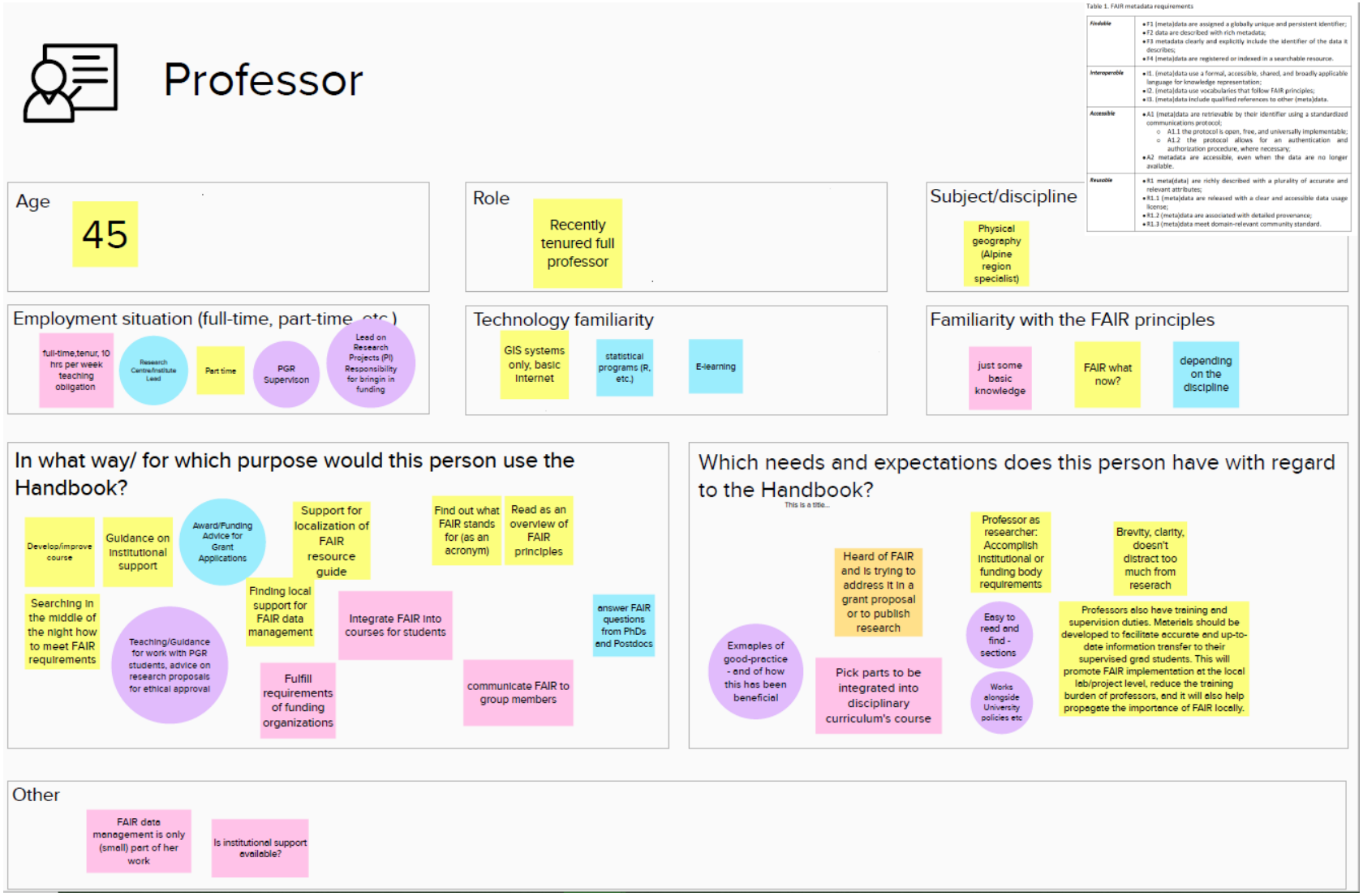
FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION



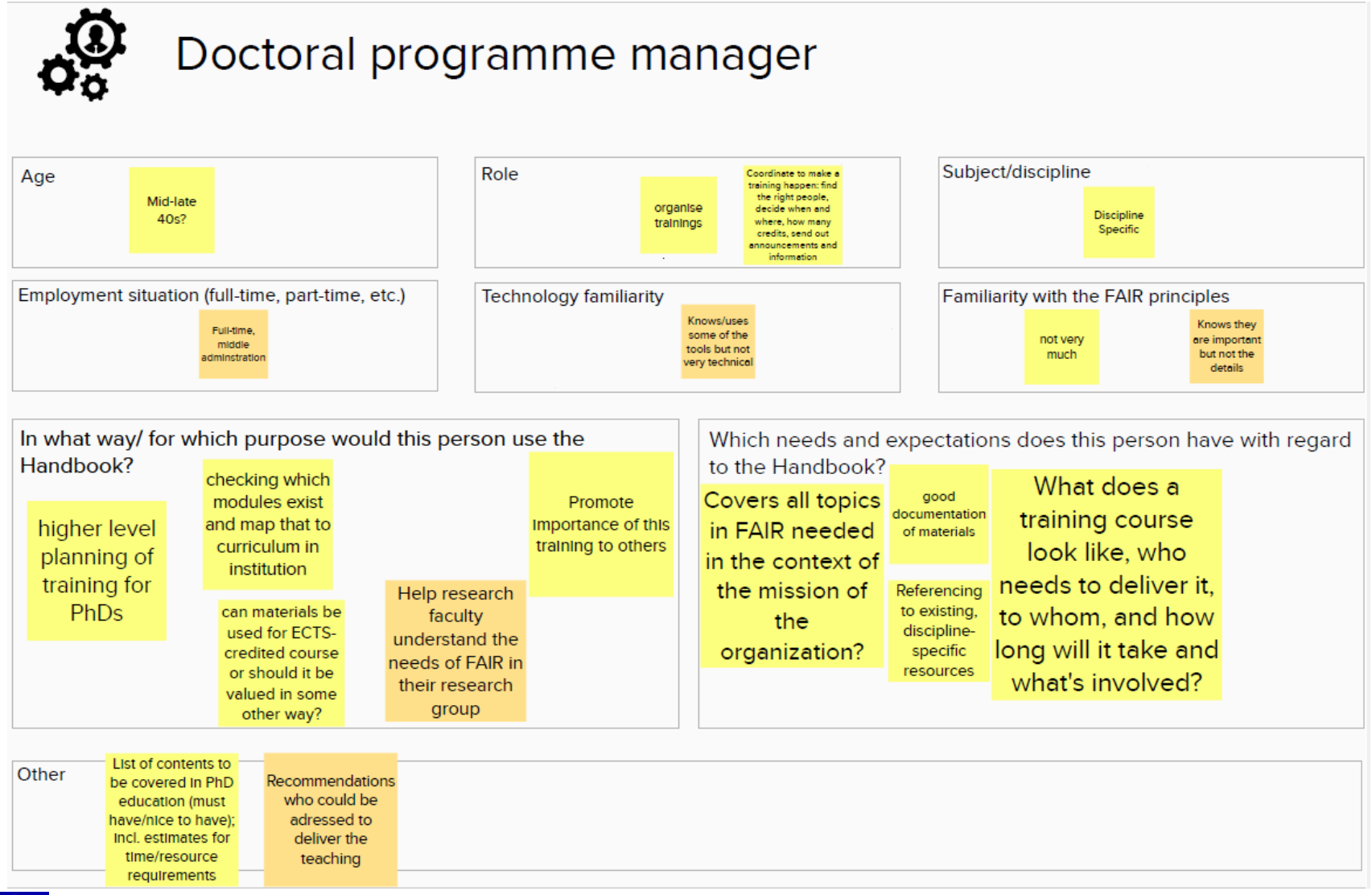
FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION



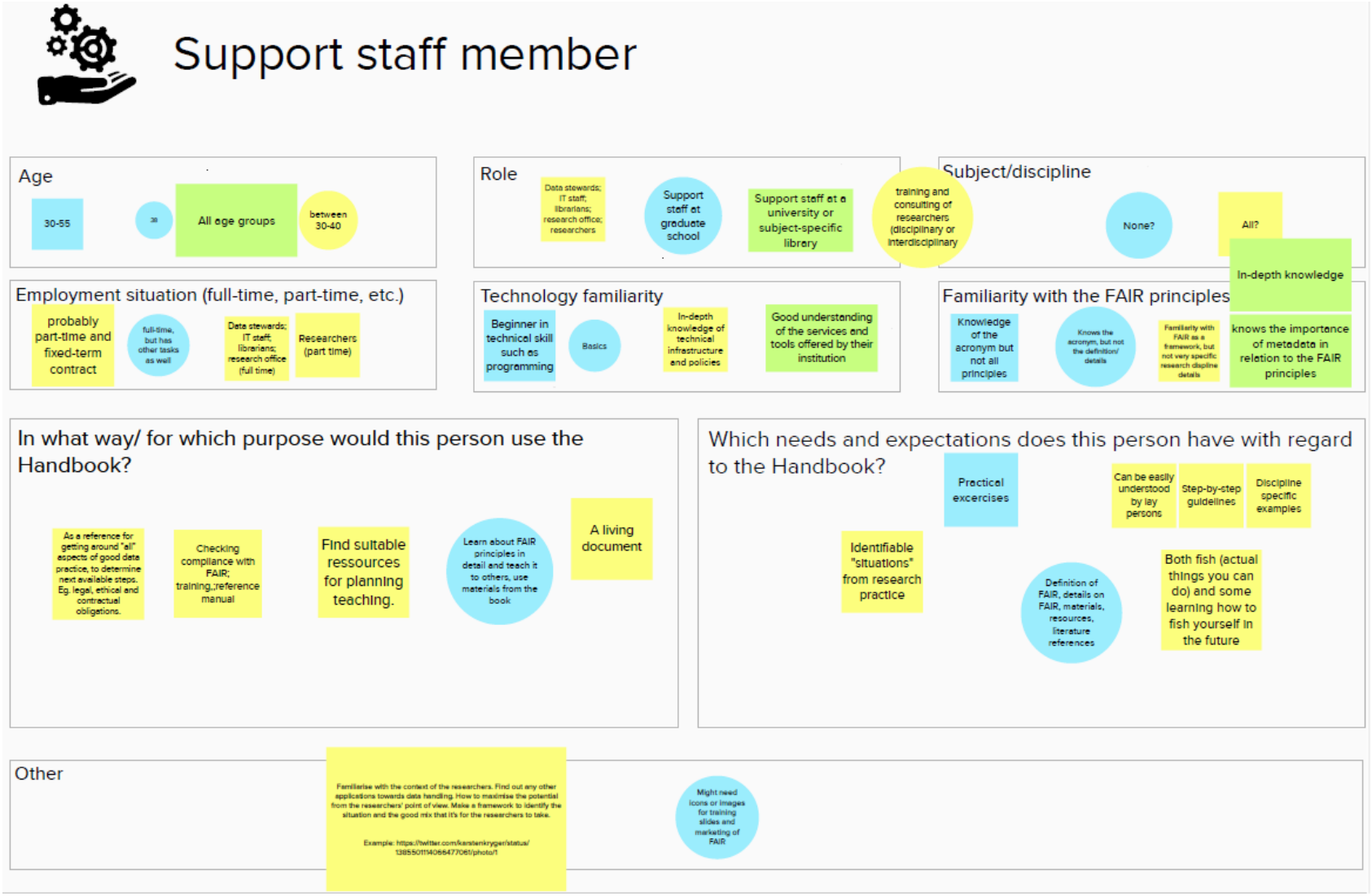
FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION



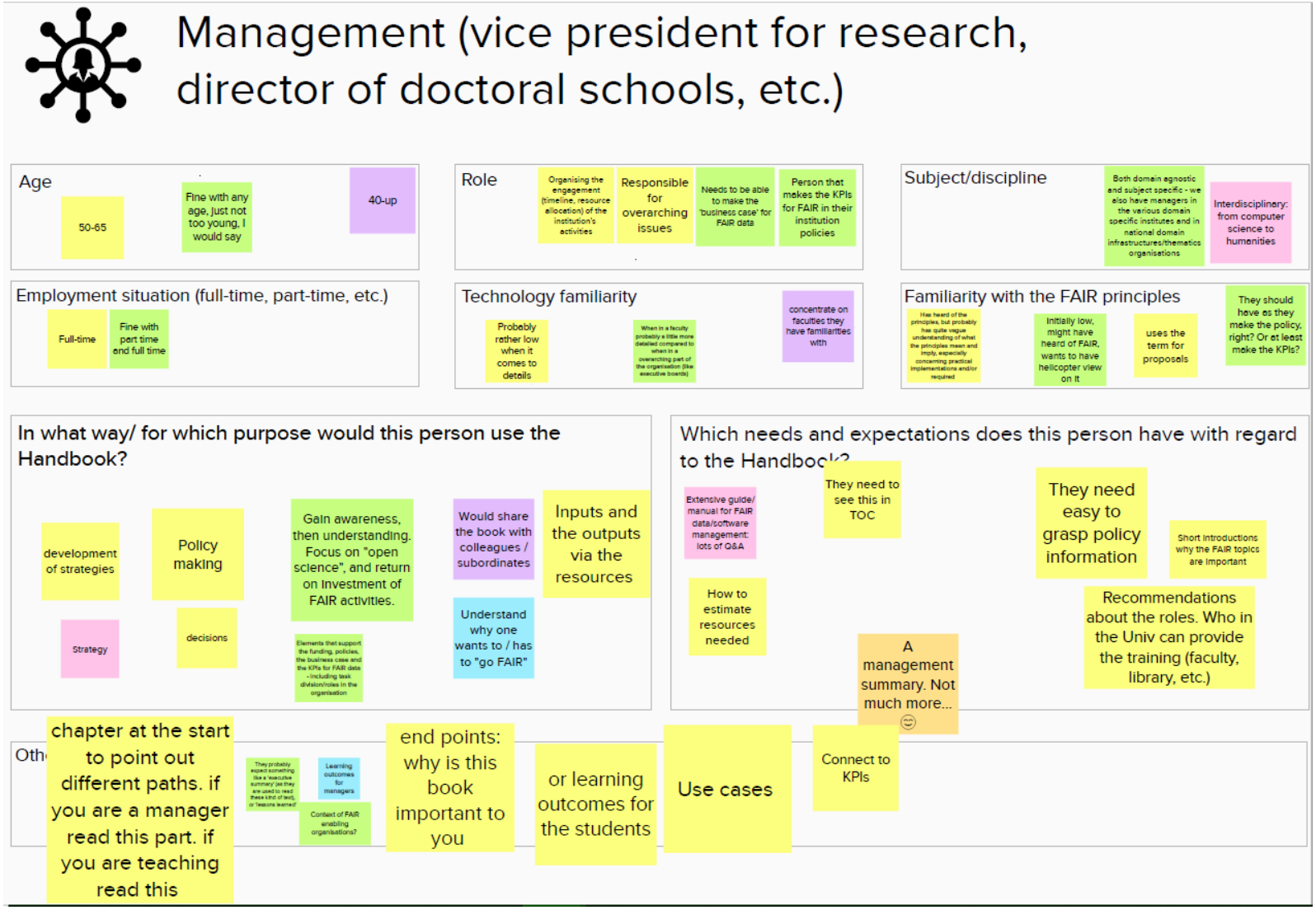
FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Appendix C – Data Stewardship Competence Groups (CF-DSP) and enumeration (according to FAIRsFAIR Deliverable D7.3)[16]

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 in chapter 3.

Data Management (DSDM)	**Data Science Engineering (DSENG)**	**Data Science Research Methods and Project Management (DSRMP)**	**Data Science Domain Knowledge (DSDK) as Business Process Management (DSBA)**
DSDM – extended, relevant Develop and implement data management strategy for data collection, storage, preservation, and availability for further processing, Ensure compliance with FAIR data principles.	DSENG – no changes, generally relevant 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.	DSRMP – revised, generally relevant 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	DSDK – generally relevant 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

[16] Demchenko et al., 2021, pp. 70 et sqq.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

		• Base research on collected scientific facts and collected data	
DSDM01 – extended, essential 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	DSENG01 – no changes, low relevance Use engineering principles (general and software) to research, design, develop and implement new instruments and applications for data collection, storage, analysis and visualisation	DSRMP01 – generally relevant Create new understandings, discover new relations by using the research methods (including hypothesis, artefact/experiment, evaluation) or similar engineering research and development methods	DSBA01 – relevant for organisation processes and data 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 • Data management and Quality Assurance of organisational data assets
DSDM02 – extended, essential 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.	DSENG02 – no changes, low relevance 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	DSRMP02 – generally relevant Direct systematic study toward the understanding of the observable facts, and discovers new approaches to achieve research or organisational goals	DSBA02 – relevant for organisation processes and data Operationalise fuzzy concepts to enable key performance indicators measurement to validate the business analysis, identify and assess potential challenges

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

• Ensure metadata compliance with FAIR requirements • Be familiar with the metadata management tools	technologies for large datasets and cloud based data analytics platforms		• Specify requirements/develop data models for organisational data
DSDM03 – extended, essential Integrate heterogeneous data from multiple sources and provide them for further analysis and use • Perform data preparation and cleaning • Match/transfer data models of individual datasets	DSENG03 – extended, relevant 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 • Develop new tools and applications, ensure support of the data FAIRness requirements by existing and new tools and applications	DSRMP03- extended, essential 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,	DSBA03 – generally relevant 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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

DSDM04 – extended, highly essential Maintain historical information on data handling, including reference to published data and corresponding data sources • Publish data, metadata and related metrics • Perform and maintain data archiving • Develop necessary archiving policy, comply with Open Science and Open Access policies if applicable • Perform data provenance and ensure continuity through the whole data lifecycle, ensure data provenance	DSENG04– extended, essential Develop, deploy and operate data infrastructure, including data storage and processing facilities, using different distributed and cloud based platforms. • Implement requirements for data storage facilities to comply with the data management policies and FAIR data principles in particular.	DSRMP04 – generally relevant 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	DSBA04 – relevant for organisation processes and data Analyse opportunity and suggest the use of historical data available at organisation for organisational processes optimisation • Coordinate implementation of FAIR data principles for collected data, ensure proper lineage and provenance of collected data
DSDM05 – extended, essential Develop policy and metrics for data quality management (e.g. Altmetrix),	DSENG05– extended, relevant Consistently apply data security mechanisms and controls at each stage of the data processing, including data anonymisation,	DSRMP05 – extended, essential Design experiments which include data collection (passive and active)	DSBA05 – relevant for organisation processes and data Analyse customer relations data to optimise/improve interaction with the

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

maintain data quality and compliance to standards, perform data curation Interact/Collaborate with data providers and data owners to ensure data quality	privacy and IPR protection, ensure standards and corresponding data protection regulation compliance, in particular GDPR. • Define and implement (coordinate) data access policies for different stakeholders and organisational roles	for hypothesis testing and problem solving • Work with Data Science, Data Stewardship and data infrastructure teams to develop project/research goals.	specific user groups or in the specific business sectors
DSDM06 – extended, essential Develop and manage/supervise policies on data protection, privacy, IPR and ethical issues in data management, address legal issues if necessary. • Ensure GDPR compliance in data management and access • Develop data access policies and coordinate their implementation and monitoring, including security breaches handling	DSENG06– extended, essential 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 • Define, implement and maintain data model, reference data, master data definitions, implement consistent metadata	DSRMP06 – extended, essential Develop and guide data driven projects, including project planning, experiment design, data collection and handling	DSBA06 – relevant for organisation processes and data Analyse multiple data sources for marketing purposes; identify effective marketing actions

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

DSDM07* - added new, essential 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.			DSBA07 – added, essential Coordinate intra organisational activities related to data analytics, data management and data provenance/lineage along all data flow stages, ensure data FAIRness
DSDM08* - added new, essential 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.			

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

DSDM09* - added new, essential 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.			

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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

as of 27 May 2021

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

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

				KU1.01.16	Mathematical software and tools
KAG1-DSDA: Data Science Analytics	DSDA.02/ML	KA01.02	Machine Learning	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
KAG1-DSDA: Data Science Analytics	DSDA.03/DM	KA01.03	Data Mining	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

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

				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
KAG1-DSDA: Data Science Analytics		DSDA.04/ TDM	KA01.04	Text Data Mining	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
KAG1-DSDA: Data Science Analytics		DSDA.05/PA	KA01.05	Predictive Analytics	KU1.05.00	Predictive analytics methods overview

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

				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
KAG1-DSDA: Data Science Analytics	DSDA.06/ MODSIM	KA01.06	Computational modelling, simulation and optimisation	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
KAG1-DSDA: Data Science Analytics	DSDA.07/ DAVIZ	KA01.07	Data Analytics Visualisation and Story Telling	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
KAG2-DSENG: Data Science Engineering	DSENG.01/ BDI	KA02.01	Big Data Infrastructure and Technologies	KU2.01.00	Big Data Infrastructure Technologies and tools overview

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

				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
KAG2-DSENG: Data Science Engineering	DSENG.02/ DSIAPP	KA02.02	Infrastructure and platforms for Data Science applications	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

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

				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, steaming, parallel)
				KU2.02.07	Enterprise information systems
				KU2.02.08	Data security and protection
KAG2-DSENG: Data Science Engineering	DSENG.03/ CCT	KA02.03	Cloud Computing technologies for Big Data and Data Analytics	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
KAG2-DSENG: Data Science Engineering	DSENG.04/ SEC	KA02.04	Data and Applications security	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

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

KAG2-DSENG: Data Science Engineering	DSENG.05/ BDSE	KA02.05	Big Data systems organisation and engineering	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
KAG2-DSENG: Data Science Engineering	DSENG.06/ DSAPPD	KA02.06	Data Science (Big Data) applications design	KU2.06.01	Data analytics, data handling software requirements and design
				KU2.06.02	Applications engineering management

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

				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
KAG2-DSENG: Data Science Engineering	DSENG.07/IS	KA02.07	Information systems (to support data-driven decision making)	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
KAG3-DSDM: Data Management	DSDM.01/ DMORG	KA03.01	General principles and concepts in Data Management and organisation	KU3.01.00	General principles and concepts in Data Management - Overview
				KU3.01.01	Overview Data type, data type registries, data formats

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

				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
KAG3-DSDM: Data Management	DSDM.02/ DMS	KA03.02	Data management systems	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

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

				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
KAG3-DSDM: Data Management	DSDM.03/ EDMI	KA03.03	Data Management and Enterprise data infrastructure	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
KAG3-DSDM: Data Management	DSDM.04/ DGOV	KA03.04	Data Governance	KU3.04.00	Overview Data governance principles and organisation

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

				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
KAG3-DSDM: Data Management	DSDM.05/ BDSTOR	KA03.05	Big Data storage (large scale)	KU3.05.00	Big Data storage Overview

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

				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
KAG3-DSDM: Data Management	DSDM.05/ DLIB	KA03.06	Data archives and data libraries	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
KAG4-DSRMP: Research Methods and Project Management	DSRMP.01/ RM	KA04.01	Research Methods and Research data	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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

				KU4.01.08	Use cases analysis: research infrastructures and projects
KAG4-DSRMPM: Research Methods and Project Management	DSRMP.02/ PM	KA04.02	Research Project Management	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
KAG5-DSBPM: Business Analytics	DSBA.01/BAF	KA05.01	Business Analytics Foundation	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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

				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
KAG6-DSBA: Business Analytics	DSBA.02/ BAEM	KA05.02	Business Analytics organisation and enterprise management	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 Intelligence), organisational models and requirements
				KU5.02.10	Data quality management, FAIR data principles for organisational data

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Appendix E – Knowledge units and corresponding learning outcomes for bachelor, master and PhD level

Content/topic [from FAIR Competences BOK], based on EDISON Data Science Framework	**basic learning outcomes**	**intermediate learning outcomes (include and build on basic learning outcomes)**	**advanced learning outcomes (include and build on intermediate learning outcomes)**	**Bachelor**	**Master**	**PhD**	**Entry-Level Content?**
General principles and concepts in data management – overview	- Can define Research Data Management (RDM) and can describe its relevance and benefits.	- Can describe RDM measures to be taken (including explaining why) at different stages of the research process.	- Can practically apply theoretical knowledge about proper RDM measures to be taken at different stages to their own research process/project.	basic	inter-mediate	advanced	Yes
Overview of data types, data type registries and data formats	- Can describe what types of data exist (Knowledge). - Can explain what data type registries are (Knowledge). - Can identify data formats (Knowledge). - Can search and find data formats in registries.	- Can determine proper data types for a resource (Analyse). - Can use a data type registry (Apply). - Can use proper data formats to express resources (Apply).	None.	basic	basic	inter-mediate	Yes

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Metadata, metadata formats, standards and registries	- 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. - Can search metadata registries to find resources.	- Can design rich metadata to describe a resource. - Can use proper metadata formats and models to express these metadata. - Can deposit metadata in a repository.	basic	inter-mediate	advanced	Yes
Open Research, Open Access, Open Data	- Can paraphrase the concept of Open Research. - Can describe the benefits of Open Research. - Can describe Open Access and Open Data as areas of Open Research.	- Can recognise if a publication is open access. - Can discover platforms for Open Access/Open Data. - Can articulate what is required to make research outputs open. - Can contrast FAIR and Open.	- Can plan publication of Open Access publications and FAIR data.	basic	inter-mediate	advanced	Yes
Metadata management, registries and publication	- Can explain aspects of metadata management and the publication process in metadata registries.	- Can perform basic steps related to metadata management. - Can execute steps in metadata publication.	- Can select appropriate metadata formats and a metadata registry appropriate for the subject domain of a research project.	basic	basic	inter-mediate	No

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Persistent Identifiers (PID), Open Researcher and Contributor ID (ORCID), Research Organization Registry (ROR)	- Can recognise PIDs and explain the different use cases for PIDs. - Can explain the importance of PIDs for FAIR data. - Can use PIDs to access data or other resources.	- Can apply PIDs to their own research outputs. - Can use PIDs to collaborate with others.	None	basic	basic	inter-mediate	Yes
FAIR (Findable, Accessible, Interoperable, Reusable) principles in data management	- 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.	None	basic	basic	inter-mediate	Yes
FAIR metadata management and tools for FAIR metadata management	- Can name aspects related to FAIR metadata management. - Can give an example of tools for FAIR metadata management.	- Can describe aspects of metadata management to comply with FAIR. - Can work with one of the FAIR metadata management tools.	- Is able to support FAIR metadata management for the selected subject domain. - Can assess and select tools for FAIR metadata management.	basic	inter-mediate	advanced	No
Databases and database management systems, data modelling	- Can explain what a database is, including common database terminology.	- Can identify basic database classifications and discuss their differences.	- Is able to design and implement own databases.	basic	basic	basic	No

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

	- Can explain and list some of the advantages and disadvantages of using databases. . Can distinguish between databases and spreadsheets. - Can recall basic concept of data modelling.	- Can recall the most common database models and discuss their usage. - Understands how a relational database is designed, created, used, and maintained. - Is able to build and assess data-based models.					
Data structures	- Understands and can restate the fundamentals of basic data structures. - Is able to implement and apply data structures.	- Is able to describe the usage of various data structures algorithms. - Is able to explain and summarise the advantages and disadvantages of various data structures implementations.	- Is able to analyse the performance characteristics of algorithms using mathematical and measurement techniques. - Is able to design and apply appropriate data structures for solving computing problems.	basic	basic	basic	No
Master data management, data dictionaries	- Can develop a data management plan for their own work. - Can identify different types of data documentation. - Can explain the purpose of the documentation. - Can use existing documentation.	- Can modify existing documentation. - Can evaluate and prioritise data management activities.	None	basic	basic	inter-mediate	Yes
FAIR data management	- Can name the main stakeholders or parties that potentially mandate	- Can identify the FAIR and RDM requirements that are	- Can plan proper measures for RDM and making data FAIR (without support).	irrelevant	basic	inter-mediate	No

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

requirements and compliance	FAIR compliance and data management measures. - Can list FAIR data management requirements.	relevant for the own research context. - Can explain where to get support with regard to RDM and the FAIR principles. - Can plan proper measures for RDM and making data FAIR (with support if necessary). - Can apply proper measures for RDM and making data FAIR (with support if necessary).	- Can apply proper measures for RDM and making data FAIR (without support).				
Data management, including reference and master data	- Can define reference and master data. - Understands the critical roles reference and master data play in data management.	- Can describe different Master Data Management (MDM) architectures and their suitability for different needs.	- Is able to design a Data Governance Framework and to manage master and reference data.	irrelevant	basic	basic	No
Data storage and operations	- Can identify different options for data storage and their operational aspects . - Can state different types and functions of storage systems.	- Can specify and explain requirements regarding data storage for specific data or organisational processes.	- Can compare different storage options. - Can select and justify a data storage solution for a project or organisation.	basic	inter-mediate	advanced	No
Data infrastructure, data registries and data factories	- Can list existing infrastructure elements and services required to support consistent data	- Can specify and explain requirements with regard to the data infrastructure and its components for specific data or organisational data.	- Can compare different infrastructure solutions. - Can select and justify a data storage solutions for a project or organisation.	basic	basic	inter-mediate	No

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

	management and handling.		- Understands the role and functions of the data factories.				
Data security and protection	- Can define different levels of data security (user, folder, files). - Can explain different ways of data protection (physical, encryption etc.).	- Can use different levels of security for their own work. - Can apply data protection methods like password protection and encoding. - Does share and collaborate in a secure way.	None.	basic	basic	inter-mediate	Yes
Data backup	- Can describe what a backup is and tell reasons for backup creation. - Can explain the 3-2-1 rule and apply it to their own files. - Can identify institutional backup solutions.	- Can explain institutional backup solutions and apply them to own files.	- Can analyse and evaluate backup. - Can solve backup problems independently or with further assistance from support personnel.	basic	inter-mediate	advanced	Yes
Personal data protection, GDPR compliance	- Can explain reasons for data protection. - Knows basic rules and legal regulations for sensitive data (e.g. GDPR). - Knows how to comply with these rules and laws.	- Can analyse compliance to legal regulations for sensitive data. - Can apply mechanisms to protect data appropriately.	None	basic	basic	inter-mediate (depending on discipline)	Yes

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Data anonymisation/ pseudonymisation	- 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.	None	irrelevant	basic (depending on discipline)	inter-mediate (depending on discipline)	No
Data management planning, FAIR data management and compliance	- Can describe what a data management plan (DMP) 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.	- Can develop a 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 progress ("living document"). - Can apply principles to protect personal sensitive data and develop Data Protection Impact Assessment, if required. (depending on discipline)	basic	basic	inter-mediate	Yes
Data integration and interoperability, data preparation and data cleaning	- Can explain aspects related to data interoperability and integration. - Can explain aspects of data preparation and cleaning.	- Can perform basic tasks in data interoperability and integration. - Can perform basic tasks in data preparation and cleaning.	- Can select best solutions/standards for data interoperability. - Can select appropriate tools and methods for data integration. - Can select appropriate methods and tools for data preparation and cleaning.	basic	inter-mediate	advanced	No

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Data interoperability and metadata management	- Can explain aspects of interoperability (Knowledge). - Can relate metadata management to interoperability (Understand).	- Use domain-relevant standards, models and formats for interoperable data (Apply) - Can relate metadata management to interoperability (Apply).	None	basic	basic	inter-mediate	Yes (very basic)
Organisational roles in data governance, data stewardship	- Can define data governance and name its components. - Can name different roles involved in data governance.	- Can name roles and structures in data governance and knows how they work together. - Can recall goals and added value of data governance.	- Can develop strategies to successfully embed data governance in an organisation.	basic	basic	inter-mediate	No
Data provenance, data lineage	- Can illustrate with an example what data provenance/data lineage means.	- Can transfer how data provenance/data lineage plays a role in their own research project. - Can apply data provenance good practices to their own data and ensure that an unbroken data lineage is established for their work.	- Can use tools for data provenance management.	basic	basic	inter-mediate	Yes

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Responsible data use, data privacy, ethical principles, IPR and legal issues	- Can summarise and explain ethical principles and responsible data use (e.g. CARE, indigenious data). - Can describe legal issues around data use and management (e.g. licences, patents, policies, contracts etc.).	- Can analyse if ethical principles or legal issues play a role in their own work.	- 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.	basic	inter-mediate	advanced	Yes
Data quality management, best practices and frameworks, data quality metrics	- Can summarise best practices ensuring data quality.	- Can describe how to recognise quality data.	- Can use best practices and frameworks on their own data to ensure their quality.	basic	inter-mediate	advanced	Yes (basic concept)
Data protection policies (including personal data), data access policies, GDPR compliance	- Can state general requirements on data protection and access control. - Can give examples of policies related to data protection and access control. - Can list the main aspects related to the GDPR.	- Can explain general requirements on data protection and access control. - Can explain content and use of policies related to data protection and access control. - Can explain what are the main aspects related to GDPR in organisational data management.	- Can write specific policies related to data protection and access. - Can analyse compliance to GDPR in organisational data management.	basic	basic	inter-mediate	No

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Trusted data repositories and certification	- Can explain what a trusted data repository is and how to find it (re3data.org and FAIRsharing). - Can compare different certifications for data repositories (e.g. CoreTrustSeal, CLARIN certification).	- Can discover trusted repositories and identify those that are certified.	- Can use a trusted repository to share research output.	basic	basic	inter-mediate	Yes (basic concept)
Data discovery (published data), data selection and use in research (added from DRSMPM Knowledge Area Group)	- Can explain the importance of data discovery and reuse.	- Can discover published datasets in their discipline. - Can cite data.	- Can develop a strategy to search for data. - Can articulate criteria for data selection. - Can extract datasets and build their own work on them.	basic	inter-mediate	advanced	Yes (basic concept)
Research data lifecycle (added)	- Can explain the steps of the research data lifecycle. - Can compare different lifecycle models.	- Can apply the research data lifecycle to their own work.	None.	basic	basic	inter-mediate	Yes

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Ontologies, controlled vocabularies (added)	- Can explain the role of ontologies and vocabularies (Knowledge). - Can recognise the use of ontologies and vocabularies (Knowledge) - Can identify a few domain-relevant ontologies (Knowledge). - Can search and find terminologies in registries.	- Can use ontologies to describe resources (Apply).	- Can use ontologies for search and analysis (Apply).	basic	inter-mediate	advanced	Yes

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Appendix F – Lesson plans

Mapping of the competence profile topics to the lesson plans

Topic	**Number of relevant lesson plan** (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)
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)

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Lesson plan 1: FAIR in a nutshell

FAIR element(s):

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 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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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): UG, Masters, PhD

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]?
        i. Findable
        ii. Accessible
        iii. Interoperable
        iv. 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]
        i. Intersections between the terms
        ii. 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
        - ☐ Upload to and share your data via a repository
        - ☐ Describe your data with as much detail as possible

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- ☐ Apply a persistent identifier

b. A is for making data accessible
   - ☐ Consider what can and will be shared under which conditions
   - ☐ Obtain participant consent and perform risk management

c. I is for making data interoperable
   - ☐ Use open, standardised and common formats
   - ☐ Consistent vocabulary
   - ☐ Apply common metadata standards
   - ☐ Linked data

d. R is for making data reusable
   - ☐ Consider permitted use
   - ☐ Apply appropriate license
   - ☐ Add sufficient provenance information
   - ☐ When using data of others, give credit by data citation

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, Rosie, et al. “Three Camps, One Destination: The Intersections of Research Data Management, FAIR and Open.” *Insights the UKSG Journal*, vol. 32, May 2019, p. 18, doi:10/gf4jhr.
Jones, Sarah & Grootveld, Marjan. (2017, November). How FAIR are your data?. Zenodo. http://doi.org/10.5281/zenodo.3405141
Knight, Gareth. Preparing Data for Sharing: The FAIR Principles. Presentation, 1 December 2015. Available at: https://www.slideshare.net/lshtm/preparing-data-for-sharing-the-fair-principles
Library Carpentry. The Top 10 FAIR Data and Software things. https://librarycarpentry.org/Top-10-FAIR/, also: https://doi.org/10/gkbnxv)
PARTHENOS. PARTHENOS Guidelines to FAIRify Data Management and make data reusable. 2019. https://doi.org/10.5281/zenodo.3368858
Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 9. Doi: https://doi.org/10.1038/sdata.2016.18

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Lesson plan 2: Data management plans (DMP)

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

Primary audience(s): UG, Masters, PhD

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 progress ("living document")
- Can apply principles to protect personal sensitive data and develop 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 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

The good Data Management Plan touches all ***FAIR principles (Findable, Accessible, Interoperable, Reusable)***[17].

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. Moreover, a DMP also contains information about persistent identifiers (e.g. DOI), 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,

[17] https://www.go-fair.org/fair-principles/

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 makes it possible to describe the data with more detail and accuracy, making it easier to understand by others. Moreover, during DMP creation, it is necessary to indicate the information that is needed to prepare the data for sharing and reuse with appropriate licenses and rules, namely, how the data can be reused, for whom the data can be valuable.

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



Source: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

c) When do I need to create DMP, at what stage of the project?

2) Content of the 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 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)

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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
    i) What does data quality mean in research data management?
    ii) Quality assurance guidelines (data description, metadata standards, documentation, data checking, etc.)
    iii) Ensuring quality control (curation processes, data entry programs, use of standardised data formats, etc.)
        (1) documenting the calibration of instruments
        (2) taking duplicate samples or measurements
        (3) standardised data capture, data entry or recording methods
        (4) data entry validation techniques
        (5) methods of transcription
        (6) peer review of data
        (7) etc.
    iv) Data quality for publishing in repositories (Completeness, Uniqueness, Timeliness, Validity, Accuracy, Consistency)
    v) 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)

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

   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, etc.) (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., et al.: The Cookbook, Engaging Researchers with Data Management (2019). https://doi.org/10.11647/OBP.0185
- Michener WK (2015) Ten Simple Rules for Creating a Good Data Management Plan. PLoS Comput Biol 11(10): e1004525. doi:10.1371/journal. pcbi.1004525
- Dominik Schmitz, Daniela Hausen, Ute Trautwein-Bruns: Content of a Data Management Plan. RWTH Aachen University. 2020. Available at DOI: 10.18154/RWTH-2019-10064, https://youtu.be/fcCj6sNvoOw
- Research Data Netherlands: The what, why and how of data management planning, 2014, https://youtu.be/gYDb-GP1CA4
- Juran, Joseph M., and A. Blanton Godfrey. Juran's quality handbook: Fifth Edition. McGraw-Hill Education, 1998. Available at: https://gmpua.com/QM/Book/quality%20handbook.pdf
- Chapman, Arthur D. Principles of data quality. GBIF, 2005. https://docs.niwa.co.nz/library/public/ChaArPrindq.pdf
- OpenAire. https://www.openaire.eu/when-do-i-have-to-create-a-data-management-plan
- Miksa T, Simms S, Mietchen D, Jones S (2019) Ten principles for machine-actionable data management plans. PLoS Comput Biol 15(3): e1006750. https://doi.org/10.1371/journal.pcbi.1006750
- Practical Guide to the International Alignment of Research Data Management, https://scienceeurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management/

Tools

- DMP Online
- Argos DMP

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- 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
Checklist for DMP
Framework for DMP creation
Massachusetts Institute of Technology (MIT) has a Total Data Quality Management Program (TDQM)
Data Quality Review
DLCM DMP tools

Use Cases/Examples of DMP

- CESSDA Qualitative Questions for DMP
- CESSDA Quantitative Questions for DMP
- Cancer research
- Clinical research
- Population research
- Biological research
- Karimova Y., Ribeiro C., David G. (2021) Institutional Support for Data Management Plans: Five Case Studies. In: Garoufallou E., Ovalle-Perandones MA. (eds) Metadata and Semantic Research. MTSR 2020. Communications in Computer and Information Science, vol 1355. Springer, Cham. https://doi.org/10.1007/978-3-030-71903-6_29
- Barbosa, Susana & Karimova, Yulia. (2020). SAIL Data Management Plan (Version 1.0.0). Zenodo. DOI: 10.5281/zenodo.4286210
- Biodiversity and Ecological Research Data: Diepenbroek, M., et al. (2014). Towards an integrated biodiversity and ecological research data management and archiving platform: the

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

German federation for the curation of biological data (GFBio). In: Plödereder, E., Grunske, L., Schneider, E. & Ull, D. (Hrsg.), Informatik 2014. Bonn: Gesellschaft für Informatik e.V.. (S. 1711-1721)

- Best Practices for Biomedical Research Data Management
- Harvard Longwood Medical Area Research Data Management Working Group

Use cases/Examples of Data Quality processes

- Biodiversity:
  - Data quality. (2017). doi: 10.1787/9789264279889-9-en
  - Chapman, A. D. 2005. Principles of Data Quality, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen. Url: https://docs.niwa.co.nz/library/public/ChaArPrindq.pdf
  - Chapman, A., Belbin, L., Zermoglio, P., Wieczorek, J., Morris, P., & Nicholls, M. et al. (2020). Developing Standards for Improved Data Quality and for Selecting Fit for Use Biodiversity Data. Biodiversity Information Science And Standards, 4. doi: 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 and Script
    - Sensor data correction
    - Geo-referencing Data. NSS Post-processing

Take Home Tasks

1) Analysis of Existing Metadata Standards https://rdamsc.bath.ac.uk/scheme-index and https://fairsharing.org/standards
2) 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
3) Analysis of the DMP examples for scientific domain relevant to learners
4) Analysis of the examples of the Data Quality procedures
5) Datasets validation from data quality perspective
6) Creation of a data quality policy for an specific use case
7) Creation of the DMP for a project relevant for learners
8) Preparation of a presentation for defence

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Lesson plan 3: Documentation

FAIR element(s):

Re-usable
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): UG, Masters, PhD

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?
        (1) Project-level
        (2) File-level
        (3) Item-level
        (4) 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?
        (1) Clear articulation of how this will be done and by whom
        (2) Standardised process for accurate, consistent, and complete documentation
3) Depending on discipline/domain of the group, introduce relevant documentation formats

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

a) Readme file
b) Data dictionary
c) Codebook
d) Commented code
e) Lab/field notebook (including Jupyter Notebooks, R markdown, electronic lab notebooks, etc.)
    i) If introducing multiple formats, outline similarities/differences and use cases
    ii) 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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Lesson plan 4: Data creation

FAIR element(s):

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 needs to be integrated with other data. In addition, the data needs 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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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): UG, Masters, PhD

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. Participants create the research data lifecycle: Participants receive cards with key terms of the lifecycle. In groups, they shall arrange the cards discussing what the terms might mean. In the end, they 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
        i. Participants 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
        i. Quantitative
        ii. Qualitative
    b. Research instruments
        i. Questionnaires/Surveys
        ii. Interviews
        iii. Field observations
        iv. Other
    c. Data planning (see also Data Management Plans)
        i. Participants write a short Data Management Plan according to a template. It doesn't have to be very detailed. Important for the task is that the participants think about the data and write down their initial thoughts in bullet points.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

d. Locate existing research data
    i. See task Reuse of existing data (2,b,i)
e. Collect new research data
f. Capture and create metadata
    i. Create a board (e.g. Padlet, Miro or a flipchart) and let the participants 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., Wuttke, U. (2020). Train-the-Trainer Concept on Research Data Management (Version 3.0): Zenodo. http://doi.org/10.5281/zenodo.4071471

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Lesson plan 5: File formats

FAIR element(s):

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 needs to be integrated with other data. In addition, the data needs 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 audienc(s)e: UG, Masters, PhD

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Learning outcomes:

- Knows which formats support FAIR data
- Understands what the differences between open and proprietary formats are
- 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 to do if you still (need to) use proprietary formats?
        i) How to convert file formats?
        ii) How to export files into a different format?
        iii) 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 (answer quiz, do exercises)
    a) Questionnaire: Open or not? Which of these file formats support FAIR data?
        i) Which of these text formats are suitable for long-term archiving? (Multiple Choice)
            (1) txt
            (2) docx
            (3) odt
            (4) html
        ii) Which of these tabular formats are suitable for long-term archiving? (Multiple Choice)
            (1) xlsx
            (2) csv
            (3) tsv
            (4) spss portable
        iii) Which of these image formats are suitable for long-term archiving? (Multiple Choice)
            (1) jpg
            (2) png
            (3) tiff
            (4) gif

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

b) The trainees choose a random folder from their directory. They check the stored file formats in terms of the FAIR principles and try to export or convert the file in another 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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 needs to be integrated with other data. In addition, the data needs 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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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): UG, masters, PhD

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).
        i. Metadata are:
            1. Standardised
            2. Structured
            3. Machine and human readable
            4. They are 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:
        i. Data Documentation Initiative (DDI) – social science

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

   ii. Ecological Metadata Language (EML) - ecology
   iii. Flexible Image Transport System (FITS) – astronomy
  b. Minimum information standards
3. Use metadata catalogues/registries and search for suitable standards

Metadata are at the heart of machine and human readable description of data, whether this is 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 there is the possibility to use metadata at different granularities and domain specificity with more generalist metadata not providing as much usefulness and value to the underlying data than those that are domain specific.

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 program: https://osf.io/bhzf8/
  - https://books.google.es/books?hl=en&lr=&id=mgS3CgAAQBAJ&oi=fnd&pg=PR5&dq=metadata+typology+&ots=kPQwq65xF3&sig=nW4SNn4sw8Y0MVxT1mA0dXYHGbU&redir_esc=y#v=onepage&q=metadata%20typology&f=false
  - https://www.worldscientific.com/doi/abs/10.1142/9789812836304_0002
- 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

- Create the metadata for a dataset
  - 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/
  - How to create a metadata profile or template
    - FAIRplus example
- Encode the data in a dataset using controlled vocabularies/ontologies
  - FAIRsharing terminology artifacts
  - 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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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

Data usually needs to be integrated with other data. In addition, the data needs 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, eg, 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 therefore reuse.

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

Primary audience(s): UG, masters, PhD - not with 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.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- Understands usage scenarios of ontologies during data collection, data analysis and when making data available through repositories, API's 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 reuse of data also across disciplinary boundaries. 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[18]
    c. Standard geospatial representations, eg 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, including now also DNA barcoding[19]
    f. Standards for dates and times (ISO 8601), for countries (ISO 3166), for geographical names (Getty Thesaurus)
2. You could also show a bad example of how not using standards makes things more difficult, or means more work to clean and translate data, for example:
    a. Survey data where standardised responses are still captured as 'text' rather than as numerical codes (dataset with 'male', 'female' rather than numeric codes)
    b. Datasets where units of variables are not defined, so one does not know whether 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?
4. What ontologies are and their function in the semantic web. Learn types of Ontologies.

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

- F2. Data are described with rich metadata - utilising ontologies is part of good metadata practice

[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)

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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

References

Use cases:

- BioSharing registry: https://biosharing.org/
- Specifications of Standards in Systems and Synthetic Biology: https://www.degruyter.com/document/doi/10.1515/jib-2018-0013/html
- 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. Developing data interoperability using standards: A wheat community use case [version 2; peer review: 2 approved]. F1000Research 2017, 6:1843. Doi: 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:

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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.pd
- https://bartoc.org/
- 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 what standards 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…)
  - Data FAIRification tools: https://fairplus.github.io/the-fair-cookbook/content/recipes/interoperability/rdf-conversion.html

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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, 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 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): UG, masters, PhD

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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 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, institutions
    b. Define together what persistent identifier are
    c. Explain the difference between persistent identifiers and authority files
2. Show the different types of PIDs and show how their syntax can look:
    a. DOI
    b. Crossref
    c. ORCID
    d. ROR
    e. RAID
    f. other
3. Explain how you can recieve an 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 to use PIDs in relation to different versions of a dataset or dynamic datasets?
    a. Versioning exercise
6. Introducing PID graphs and their importance
    a. Explain the importance of PID graphs with an 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/dois.html
- https://search.crossref.org/
- https://www.doi.org/
- https://orcid.org/
- https://ror.org/
- https://www.raid.org.au/

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- FAIRsharing list of community-used identifier schemas
- 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
- RDA recommendations: https://www.rd-alliance.org/system/files/RDA-DC-Recommendations_151020.pdf
- https://dl.acm.org/doi/10.1145/3311790.3396660
- https://librarycarpentry.org/lc-fair-research/02-findable/index.html
- Research Data Netherlands. (2014). Persistent identifiers and data citation explained [Video]. Retrieved from https://youtu.be/PgqtiY7oZ6k
- Martin Fenner, Joe Wass, Tom Demeranville, Sarala Wimalaratne, & Richard Hallett. (2019). D2.2 PID Metadata Provenance. Zenodo.
- Introducing the PID Graph. https://doi.org/10.5438/jwvf-8a66

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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): UG, masters, PhD

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 there are countries without copyright law);
- Know different types of rights (economic and moral);
- Know different types of licences and understand what actions you can perform with them (e.g. CC, ODbL, ACA, OGL, LGPL);
- Know the meaning of non-commercial and commercial licence (e.g. CC BY-NC);
- Understand the existence of different types of restrictions;
- Know tools and guides to choose the correct licence;
- Apply the acquired knowledge in practice (e.g. answer quiz, do 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 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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

issues help to prepare your data for professional reuse with or without restrictions, with appropriate licence, protect you as proprietary and avoid unpleasant situations with reusing the 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)
      i) Is (research) data protected by copyright law in the same way as other works?
         (1) Let the participants define research data they work with
         (2) 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 you contact the copyright holder?
      i) what you will be using (amount and content)
      ii) the context their work will be used within
      iii) where you will be using the work (e.g. publicly online)
      iv) for what purpose (e.g. educational, commercial, personal)
      v) 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;
      i) CC0 - No Rights Reserved
      ii) Attribution CC BY
      iii) Attribution ShareAlike CC BY-SA
      iv) Attribution-NoDerivs CC BY-ND
      v) Attribution-NonCommercial CC BY-NC
      vi) Attribution-NonCommercial-ShareAlike CC BY-NC-SA
      vii) Attribution-NonCommercial-NoDerivs CC BY-NC-ND
   b) Software licences
      i) Public domain
      ii) Permissive

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

iii) LGPL(GNU)
iv) Copyleft
v) Proprietary

c) Open Source Licences
i) Apache License 2.0
ii) BSD 3-Clause "New" or "Revised" license
iii) BSD 2-Clause "Simplified" or "FreeBSD" license
iv) GNU General Public License (GPL)
v) GNU Library or "Lesser" General Public License (LGPL)
vi) MIT license
vii) Mozilla Public License 2.0
viii) Common Development and Distribution License
ix) Eclipse Public License version 2.0

d) Other type of licences
i) ODbL
ii) ACA
iii) OGL

e) Orphan works and search guidance for applicants
f) Remember: Licence-free is not the same as a free licence

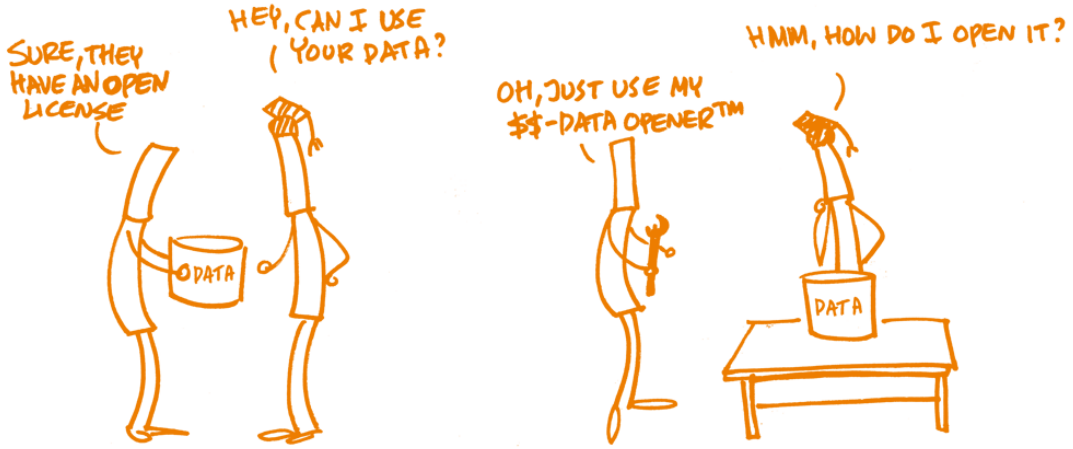


Image source:
https://open-science-training-handbook.gitbook.io/book/open-science-basics/open-licensing-and-file-formats

5) Tools that 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 the ownership of the data may be regulated by employment and service contracts

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 card game
9) Application of knowledge in practice (answer quiz, do exercises);
    a) E.g. Which licence may you grant if you want to combine data with the following licences:
        i) CC BY and CC BY-SA?
        ii) CC BY-SA and CC BY-NC?
        iii) CC BY and CC BY-ND?
10) Do 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. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/how-guides
- https://opendefinition.org/licenses/
- Nemlioglu, Ilayda. "A comparative analysis of intellectual property rights: a case of developed versus developing countries." Procedia Computer Science 158 (2019): 988-998. https://doi.org/10.1016/j.procs.2019.09.140
- Margoni, Thomas, & Tsiavos, Prodromos. (2018). Toolkit for Researchers on Legal Issues. Zenodo. DOI:10.5281/zenodo.2574619
- International Copyright Basics
- CESSDA 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, Lucie, and Andreas Wiebe. *Safe to Be Open*. 2013, DOI:10.17875/gup2013-160.
- Burrow, Sheona; Margoni, Thomas and McCutcheon, Valerie (2018), Information Guide: Introduction to Ownership of Rights in Research Data. CREATe, University of Glasgow. http://eprints.gla.ac.uk/171314/

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- Burrow, Sheona; Margoni, Thomas and McCutcheon, Valerie (2018), Information Guide: Making Research Data Available. CREATe, University of Glasgow. http://eprints.gla.ac.uk/171315/
- Burrow, Sheona; Margoni, Thomas and McCutcheon, Valerie (2018), Information Guide: Choosing a Licence for Research Data. CREATe, University of Glasgow. http://eprints.gla.ac.uk/171316/
- Burrow, Sheona; Margoni, Thomas and McCutcheon, Valerie (2018), Information Guide: Using Research Data. 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

1) 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). Zenodo. DOI: 10.5281/zenodo.4286210
2) Examples of Usage Rights
3) Copyright Examples
4) What can be Copyrighted (Examples)
5) Ownership of WUR research data

Take Home Tasks

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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Lesson plan 10: Finding and reusing data

Being able to reuse data and analyse secondary data can not only 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[20]', the ultimate goal of FAIR is to optimise the reuse of data. In order to be reusable, data should respond, 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): Masters, PhD, 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 data set 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 teh importance of data citation when reusing data

Summary of Tasks / Actions:

- Speaking about "good scientific practice" : why is it important to use secondary data instead of collecting primary data ?
- Identify a strategy to find data appropriate for a specific research project.

[20] *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).*

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- Identify “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 re-use.
- Look at data citation models: 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
Computers & Internet

Resources:

Why are research data managed and reused?
https://rwww.fsd.tuni.fi/en/services/data-management-guidelines/why-are-research-data-managed-and-reused/

An interesting point on good scientific practice is made on this blog post of the Finish Data Archive. Also, benefits of data reuse are briefly described :

> “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.”

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

The author describes the reuse of available data as a “complicated process” (Kim and Yoon, 2017; Zimmerman, 2007).

> “**Finding and downloading data are only the first of many steps in this process. Downloading is followed by the often necessary appraisal of a dataset on completeness, trustworthiness, and appropriateness, as these characteristics are key factors for re users to reuse the dataset in the end as a basis for their scientific work** (Faniel et al., 2015; Yoon, 2016; Kim and Yoon, 2017). If the dataset seems trustworthy, and there is the intention to use it (Kim and Yoon, 2017) the data may still need filtering, cleaning

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

and may need to be integrated with other information, before it is fit for reuse purposes. The fact that a time investment is needed to reuse a dataset, means that less time profit is made at reuse. At each of these activities the choice can be made to discard the dataset after all and not reuse it. Expectedly, actual reuse of datasets is a fraction of the downloads and views of datasets (e.g. see data from Fear, 2013)."

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 in 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, 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, whether it is 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. Or 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 CESSDA -- European consortium for promoting the results of social science research (https://datacatalogue.cessda.eu/) with guidelines on discovering data

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

(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 are pillars to data reuse. FAIRsharing maps the landscape of community-developed standards defines 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 relevant standards to your domain and discipline

References:

- 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, 18(1), p.10. DOI: 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. 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, 2(2).

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Lesson plan 11: Repositories

FAIR elements: All

Primary audience(s): UG, Masters, PhD

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. 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 objects (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[21]. 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 with your local institutional requirements on what you should address. Are you obliged to upload your research outputs locally?

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.

[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.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 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. Most often, 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, the assignment of the resulting data to a subject area may be difficult. Cross-disciplinary/generic repositories offer a solution here. 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 is 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 support, 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.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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.[22]
            - 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, tell 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, back-up 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
    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".

[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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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: https://www.re3data.org. Here you may 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 or 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 have provided 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 section, How to choose a repository, and also OpenAire's guidance, use FAIRsharing or re3data to find a trustworthy repository in political science. What did you find?
    b. From the section,FAIR principles and repositories, use the Zenodo Sandbox to upload test data (e.g. an example text file) and assign a licence. What did you find?

7. **Take home task:**
    a. 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
- Web browser / Internet
- Access for different repositories (e.g. credentials)

References

1. OpenAIRE (n.d.) How to select a data repository? Accessed 23 July 2021. https://www.openaire.eu/opendatapilot-repository-guide.
2. Biernacka, Katarzyna, Maik Bierwirth, Petra Buchholz, Dominika Dolzycka, Kerstin Helbig, Janna Neumann, Carolin Odebrecht, Cord Wiljes, and Ulrike Wuttke. *Train-the-Trainer Concept on Research Data Management* (version 3.0). Zenodo, 2020. https://doi.org/10.5281/zenodo.4071471.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

3. 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.
4. USGS (n.d.) Repositories. Accessed 23 July 2021. https://www.usgs.gov/products/data-and-tools/data-management/repositories
5. Library Carpentry (n.d.) Library Carpentry - FAIR Data and Software. Accessed 23 July 2021. https://librarycarpentry.org/lc-fair-research/07-assessment/index.html
6. AGU (n.d.) Data & Software for Authors. Accessed 23 July 2021. https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Data-and-Software-for-Authors
7. 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.
8. Stall, Shelley, Martone, Maryann E., Chandramouliswaran, Ishwar, Crosas, Mercè, Federer, Lisa, Gautier, Julian, Hahnel, Mark, Larkin, Jennie, Lowenberg, Daniella, Pfeiffer, Nicole, Sim, Ida, Smith, Tim, Van Gulick, Ana E., Walker, Erin, Wood, Julie, Zaringhalam, Maryam, & Zigoni, Alberto. (2020). *Generalist Repository Comparison Chart.* Zenodo. https://doi.org/10.5281/zenodo.3946719

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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): UG, Masters, PhD

This lesson plan contains ideas for teaching students and researchers 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, for example 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; or information that might have 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 routinely be 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.

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Sharing confidential data will often come down to restricting access to a dataset. In this lesson plan we discuss lesson objectives and activities that can be done in class to discuss aspects that are useful to consider 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 applicable legislation. The idea here is that readers who wish to use this lesson plan could tailor this lesson towards legislation that is applicable to the research context in which the students are working. The main message is that data that cannot be shared freely for a reason or another can still be made FAIR and strategies can be implemented to do so are fairly similar, no matter what the reason is.

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 ways 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 for archiving/publishing 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

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- 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 students identify what types of data they are working with and how they should deal with those regarding relevant legislation:
        i) Take the relevant 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 (e.g. 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.
            (1) 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're affiliated to and to the country/countries in which they're carrying out their research
                (a) Europe: GDPR
                (b) Canada: PIPEDA
                (c) Australia: Privacy Act 1988
                (d) UK: Common law duty of confidentiality
            (2) Medical legislation
                (a) Netherlands: WMO (legal framework for medical scientific research)
                (b) U.S.: HIPAA
            (3) Animal testing regulation
                (a) 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 those and what this means for their workflow when it comes to making data available to others.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

2) Explain the rationale behind the legislation, protocols and guidelines that you discuss, so that students understand why they are there and they don't consider them just boxes that need to be ticked. You can ask students to reflect on the relevant requirements by means of statements and discuss those in the group.
3) Introduce security measures that students can take to protect research participants and sensitive data related to them:
    a) Preventing unauthorised access by means of reliable verification methods (passwords, two-step authentication)
    b) Pseudonymisation of personal data
    c) Storing key files in a location separate from other research data
    d) Encryption (full disk and folders and files)
    e) Access rights to those who are 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 of 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 that 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 UG and MA students; it is up to the teacher/instructor to assess whether this part should be included or excluded):
    a) Explain which practical things 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 are things that are often not under 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.
        i) Technical:
            (1) The location where data are stored should have the option to restrict access, so that only authorised people can access the data.
            (2) There should be a contact point where data access requests can be sent.
        ii) Organisational:
            (1) There needs to be someone who can receive data access requests and reply to them.
            (2) There needs to be someone who has the authority to decide whether a data access request will be granted or denied.
            (3) A set of criteria needs to be available on the basis of which access will be granted or denied.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 Terms of use of the PsychData repository and the template for a data user agreement from Open Brain Consent):
   i) Data may only be used for the purpose of academic research and instruction.
   ii) Data may not be forwarded to third parties.
   iii) Any publication based on the data must cite the dataset.
   iv) No attempts may be made to re identify or contact participants.
   v) Data needs to be stored in a secure work environment. Anyone re-using the data must provide the technical specification of the secure environment.
   vi) 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 one containing the non-confidential materials that could be valuable to other researchers, for example protocols, syntaxes.
   b) Provide examples of such cases, so that students are presented with a tangible form of what a dataset with different access restrictions could look like:
      i) FEM growth and yield data monocultures - Grand fir in DANS Data Station Life, Health and Medical Sciences. Plot data book, tree maps atlas and README file are publicly accessible, while for the other files you need to request permission for access.
      ii) 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:

8) 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:
      i) The Ultimate consent form from Open Brain Consent, or the GDPR edition.
      ii) 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 that is suitable for their discipline. Ask them to study the template and to find out if there are any statements relating to making data available to others after the project.
9) 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:

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

i) '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 they would like to deposit their data in. Ask them to find out if this repository has any instructions on de-identifying data. Discuss the findings with the group.

10) 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:

i) 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.

ii) Pseudonymisation in small-scale quantitative research - This overview presents nine basic steps for pseudonymising data.

iii) 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

iv) Guide to Social Science Data Preparation and Archiving - On p. 42-43 you find concrete steps for de identifying data

v) Anonymisation section in the CESSDA Data Management Expert Guide - This section provides practical steps for making data about people less identifiable.

vi) Anonymisation postcard - This postcard illustrates that even with very little and general information, individuals can be identified, depending on the context.

vii) Privacy risks matrix - The matrix on p. 4-5 explains the risk levels for re-identification of data about people. P. 6 presents examples of various levels of de-identification.

viii) 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 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 assess if their data can be anonymised.

c) Show examples of strategies and tools that are available for pseudonymising and anonymising data, for example:

i) Anonymising qualitative data, UK Data Service - Advice for how to deidentify various types of qualitative data: text, transcripts and audio-visual data.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

ii) Anonymising quantitative data, UK Data Service - Advice for how to de identify quantitative data, for example by removing or aggregating variables or reducing the precision of a variable.
iii) 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.

11) 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 another location than the file(s) with the research data, so that linking the two files is not trivial.

Dealing with ethical aspects:

12) Explain that sharing or publishing data shouldn't harm individuals, which could for example be the case if the data have been collected among vulnerable groups, or when individuals show unique abnormalities. Refer students to the ethics committee or review board in their institution, so that this committee can 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 > Anonimysation, Consortium of European Social Science Data Archives
- Guide to Social Science Data Preparation and Archiving, ICPSR (Inter-university Consortium for Political and Social Research)

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- 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 ask them to discuss it with a privacy expert in their institution and adjust it 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 to assess 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 that they learn how these techniques affect their data.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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): UG, masters, PhD

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)

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 API or on the web, with a range of access options. Whilst open access to data may be ideal, there can be genuine reasons why that is not possible.
Data access categories[23,24] 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*".[25]
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 data are not accessible, except maybe to regulators.

2. Explain the criteria that can influence access decisions[26]:

- Presence of personal information in the dataset, that can be used to identify an individual person
- 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)
    - Commercial interests of a company
- Intellectual Property, where the early release of the data can adversely affect patents or valorisation routes
- Confidentiality agreement which means that access to and sharing of data is restricted to the contracting parties.

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 can be:

- Capture data in an anonymous way

[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/

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- Anonymise information in a dataset, so individuals (people, animals,...) cannot be identified from the information they have contributed in the research
- Gain permission from people to make data open, even if the data contains personal or sensitive information (informed consent)
- 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://zenodo.org/record/3820473#.YL859ofiuUk
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

Additional Information: Data Availability Statements:
https://docs.google.com/document/d/1Jee_GuilqQA6q1VdrAyqKDIVZqNiiBI_awFvbpKOdeA/edit#

Take Home Tasks
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)

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 http://doi.org/10.15000/a789457"
Embargoed data	"All data underpinning this publication will be available from the University of FAIR-Data Repository at http://doi.org/10.15002/a1234a56 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 http://doi.org/10.15000/a1234b56"
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 http://doi.org/10.15129/a1234b56"
Physical data	"Physical data supporting this publication are stored by the University of FAIR-Data. Details of the data and how it can be accessed are available from the University of FAIR-Data Repository at http://doi.org/10.15129/a1234b56"
Secondary data	"Pre-existing data underpinning this publication are openly available from UKDS at http://doi.org/10.12345/54321. Further information about data processing, and additional new supporting data are available from the University of FAIR-Data Repository at http://doi.org/10.15129/a1234b56"
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 http://doi.org/10.15129/a1234b56"
No data	"This work is entirely theoretical, there is no data underpinning this publication."

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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): Masters, PhD

Learning outcomes

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

Summary of Tasks / Actions

1. Define research software by:
   - 1.1. Giving a definition of research software
   - 1.2. Giving examples and counterexamples (e.g. word processing software) of research software; be sure to include a breadth of examples including scripts and workflows.
   - 1.3. Identify similarities and differences between research data and software with regard to the application of the FAIR principles.
   - 1.4. 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, in each case provide a concrete example of how to carry out the principle.
   - 2.1. Findable - F: The software, and its associated metadata, should be 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 must be assigned distinct identifiers representing different levels of granularity.
       - F1.2 Different versions of the same software must be assigned distinct identifiers.
     - 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.
   - 2.2. Accessible - The software, and its metadata, must be retrievable via standardised protocols
     - A1. Software is retrievable by its identifier using a standardised communications protocol.
       - A1.1 The protocol is open, free, and universally implementable.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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

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

2.3. Interoperable - I: The software interoperates with other software through exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs).

I1: Software reads, writes and exchanges data in a way that meets domain-relevant community standards.

I2: Software includes qualified references to other objects.

2.4. Reusable - R: The software is both usable (it can be executed) and reusable (it 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 must have 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
   - 3.1. Quality of the form *vs.* quality of the function of a research software
   - 3.2. Testing for code maintainability
   - 3.3. Validation of the functional correctness
   - 3.4. Security measures
   - 3.5. Computational efficiency

4. Recognise software citation as key to recognising research software as a first class research output
   - 4.1. Software citation principles
   - 4.2. 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
- FAIR4RS WG - Draft of FAIR4RS principles
- Lamprecht, Anna-Lena et al. 'Towards FAIR Principles for Research Software'. 1 Jan. 2020 : 37 – 59.
- S. Hettrick et al., UK Research Software Survey 2014, Zenodo, 2014. doi:10.5281/zenodo.14809.
- Library Carpentry: FAIR Data and Software - Software

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Best practices

- Library Carpentry: FAIR Data and Software - Software
- Lamprecht, Anna-Lena et al. 'Towards FAIR Principles for Research Software'. 1 Jan. 2020 : 37 – 59.
- Five recommendations for FAIR software

Limits

- Lamprecht, Anna-Lena et al. 'Towards FAIR Principles for Research Software'. 1 Jan. 2020 : 37 – 59.
- RDA - FAIR for Research Software (FAIR4RS) WG

Software Citation

- Katz, Daniel S. et al., 'Recognizing the value of software: a software citation guide'
- Chue Hong, Neil P. et al. 'Software Citation Checklist for Developers'
- Citation file format: CITATION.cff
- Additional material on software citation produced during the FAIRsFAIR book sprint

Further resources

- A fresh look at FAIR for Research Software
- 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
- LibraryCarpentry - Top 10 FAIR Data & Software Things for specific disciplines
- LibraryCarpentry - Research Software
- CodeRefinery - Reproducible Research: Sharing code and data

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Lesson plan 14: Additional material on software citation

It is appropriate to consider software in the context of FAIR given 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. 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., arXiv 1905.08674 [cs.CY]). Nonetheless 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. F1000Research 9:1257, 2021. https://doi.org/10.12688/f1000research.26932.2) . 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 optimizing the reusability of scholarly content by developing best-practice recommendations for tagging content in JATS XML aims to support the various ways people may cite software.
This recommendation is not necessarily the optimum way or editorial policy of individual journals. Authors are exploring different ways to make their content, source materials, and methodology accessible to the readers, and throughout this recommendation, we try to indicate where software citation initiatives are promoting change and development.

The following is the minimal requirement 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, preferentially, an archive or repository that provides persistent identifiers.
- Date: the date the software was published.
- Identifier: a resolvable pointer to the software, preferentially, 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 the PDF. DOIs are preferable, and other examples of PIDs include Handles, RRIDs, ASCL IDs, swMath IDs, Software Heritage IDs, ARKs, etc. If there is no PID for the software, a URL to where the software exists may be the best identifier available.

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 a bracketed description of the citation (e.g., Computer software) to be included.

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Recommendation

Minimal requirement 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 so would require 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).,

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

@designator: Used on such elements as edition number (<edition>) and version (<version>) to hold an unadorned numeric or alphabetic 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:BIOMD0000000156">
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>

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

</element-citation>
</ref>

Additional reading

Software Metadata Recommended Format Guide (SMRF)
Katz DS, Chue Hong NP, Clark T et al. Recognizing the value of software: a software citation guide [version 2; peer review: 2 approved]. F1000Research 2021, 9:1257 (https://doi.org/10.12688/f1000research.26932.2)

Guidance for:

- journals: https://doi.org/10.12688/f1000research.26932.1
- authors: https://doi.org/10.5281/zenodo.3479199
- software developers: https://doi.org/10.5281/zenodo.3482769
- software repositories and registries: https://arxiv.org/abs/2012.13117
- software citation use cases: https://peerj.com/articles/cs-86/#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 could be different from the authors of a research paper associated with the code. This recommendation gives 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.

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 students or professional audiences of vocational education and training.

Learning outcomes:

1) Understanding the RDM process and main use cases
2) Understanding Open Research and Open Data (Definition, Standards, Open Data use and reuse, open government data, European policies and initiatives)
3) Understanding FAIR principles in Research Data Management, maturity model and compliance
4) Working with sensitive, personal or private data (General Data Protection Regulation [GDPR] and its requirements, Ethics approval process and form)
5) Understand what a Data Management Plan is, its purpose and benefits for a project or organisation
6) Know tools, guides, templates to support RDM, metadata management, DMP creation
7) Apply the acquired knowledge in practice, namely be able to create a DMP, create and publish data and metadata
8) 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 form of tutorial, webinar or self-paced self-study course
Required time: 2 sessions of lecture (1.5 hrs) and 1 session practice (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):

A. Use cases for research data management and stewardship
   - Preserving the Scientific Record

B. Data Management elements (organisational and individual)
   - Goals and motivation for managing your data
   - Data formats, Metadata, related standards
   - Creating documentation and metadata, metadata for discovery
   - Using data portals and metadata registries

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- Tracking Data Usage, data provenance, linked data
- Handling sensitive data
- Backing up data, backup tools and services

C. Responsible Data Use (Citation, Copyright, Data Restrictions)

- Data privacy and GDPR compliance

D. FAIR principles in Research Data Management, supporting tools, maturity model and compliance

E. Data Management Plan (DMP)

F. Data Stewardship and organisational data management

- Responsibilities and competences
- DMP management and data quality assurance

G. Open Research and Open Data (Definition, Standards, Open Data use and reuse, open government data)

- Research data and open access
- Repository and self-archiving services
- Research Data Alliance (RDA) products and recommendations: Persistent Identifiers (PID), data types, data type registries, others
- ORCID identifier for data and authors
- Stakeholders and roles: engineer, librarian, researcher
- Open Data services: ORCID.org, Altmetric Doughnut, Zenodo

Practice:

Hands on practice including the following topics:

a) Data Management Plan design, templates and tools

b) Metadata and tools, metadata registries

c) Selection of licences for open data and contents (e.g. Creative Common, and Open Database)

Materials / Equipment

1) Collection of DMP templates
2) Example metadata for research data and publications
3) Collection of links to RDM tools, metadata registries,

References

- General Data Protection regulation - https://gdpr-info.eu/
- LIcense 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&p=2138498
- Towards FAIR principles for research software - https://content.iospress.com/articles/data-science/ds190026

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- FAIR Cookbook, developed by Life Sciences academics and pharmas, 2021 https://fairplus.github.io/the-fair-cookbook/content/home.html
- FAIRsharing for (meta)data standards and interlinked repositories

Take Home Tasks

Organisational Data Management Plan creation (using provided template and/or using online tools)

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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

FAIR elements: all

Primary audience(s):
This lesson is targeted 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 Computer or information science background.

Learning outcomes:

1) Understanding the Enterprise Data Management and Governance process and main use cases. DAMA (Data Management Association) Data Management Body of Knowledge (DMBOK)
2) Understanding European Data Spaces concept and initiatives, European policies and regulations, GDPR (General Data Protection Regulation)
3) Understanding elements of the enterprise data management infrastructure and services: Data Warehouses, cloud based storage, data lakes
4) Understanding data modelling process, data models, data structures. Master data management
5) Understanding FAIR principles in Research Data Management and their applicability to industrial use cases
6) Understanding data management maturity frameworks and best practices
7) Understand what a Data Management Plan is, its purpose and benefits for a project or organisation
8) Apply the acquired knowledge in practice, namely be able to create a DMP, assess organisational data security and compliance
9) 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 a form of lecture+practice, tutorial or self-paced self-study course.
Suggested time: 2 sessions of lecture (1.5 hrs) and 1 session practice (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 Advanced course

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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 are the main topics that can 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. 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 organisation's data security and compliance requirements
d) Advanced: Data Modelling, Relational data model creation

Materials / Equipment

1) Collection of DMP templates
2) Example metadata for research data and publications
3) Collection of links to enterprise Data Management and Governance practices and recommendations

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

References

1. DAMA Data Management Body of Knowledge (DMBOK), DAMA International, 2017
2. GO FAIR Initiative [online] - https://www.go-fair.org/go-fair-initiative/
3. General Data Protection regulation - https://gdpr-info.eu/
4. DMP Templates - https://guides.lib.umich.edu/c.php?g=283277&p=2138498
5. Towards FAIR principles for research software - https://content.iospress.com/articles/data-science/ds190026
6. A European strategy for data COM(2020) 66 final, 19.02.2020 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066
7. European Data Governance https://ec.europa.eu/digital-single-market/en/european-data-governance
8. EU/Parlament 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
9. GAIA-X – A Federated Data Infrastructure for Europe - https://www.gaia-x.eu/
10. FAIR Cookbook, developed by Life Sciences academics and pharmas, 2021 https://fairplus.github.io/the-fair-cookbook/content/home.html

Take Home Task

Organisational Data Management Plan creation (using provided template and/or using online tools)

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558