

AGAPE: An introductory course to open science
for early career researchers

An Opening Doors initiative

Contents

Introduction	5
Structure of the book	6
What did we leave out?	6
About Opening Doors project	6
Meet the authors	6
Acknowledgement	6
How to contribute to this project	6
1 Introduction to open science	7
2 Open research, open data, open access	9
3 Pros and cons	11
4 Research data lifecycle	15
5 FAIR principles	19
5.1 What is FAIR?	19
5.2 Test your understanding	25
5.3 Recommended activities	25
6 Data centers and data repositories	27
7 Data ethics	29
8 Coding and other skills	31

9	Communication and ethics of open science	33
10	Opening your research	35
11	Conclusion	37
12	How to contact us	39
13	Data centers and data repositories	41
14	Conclusion	43
15	Final Quiz	45



Introduction

Greetings, fellow early career researcher or open science-curious friend!

In this course, we would like to introduce you to the world of open science. Whether you are familiar with some of its concepts and resources or the open science movement doesn't ring any bells, we believe that what you learn here will be interesting for you and at the same time highly useful for your future career.

We ourselves are PhD students who first met during the course focusing on open and collaborative research. And because we felt that what we learned was very helpful and other students should have an opportunity to get familiar with these concepts too, we decided to create Agape. Agape means wide open, such as open science we want to promote. The word *agapē* originates from Greek and means love that is unconditional, such as our love for science. Under Agape we aim to disseminate open science between students, starting with this course and continuing with a series of workshops where we can learn, exchange our opinions and experiences and together contribute to the future.

With this course, Agape would like to open doors for you into the world of open science and introduce various concepts that we think are very important but we were not told about. Whilst we all heard about scientific integrity and open access publishing at some point in our studies, the domain of open science encompasses a much larger area. Given its extent, this course does by far not cover the whole scope of open science. However, during the course, we provide you with useful links to other resources should you wish to learn more and start practising open science.

The course is structured into chapters that are written to expand on various topics. We think that the order they follow is logical and the latter chapters are building on knowledge acquired in previous chapters. But you can decide to go through them in whatever order you like by clicking on different chapters in the menu on the left or to return to some of them should you find something is not clear or you forgot it in the meantime. At the end of each chapter are MCQs where you can test your freshly acquired knowledge.

And now, without any further delay, let's quench that thirst for knowledge!

Structure of the book

How to read this course

Try these toolbar features located near the top of your browser:

- Menu
- Search
- Font to adjust text size and display
- View source code on GitHub (if available)
- Download book files (if available)
- Shortcuts (arrow keys to navigate; **s** to toggle sidebar; **f** to toggle search)
- Social Media
- Share

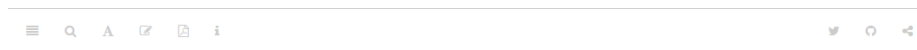


Figure 1: Toolbar features in open-access web edition

What did we leave out?

About Opening Doors project

Meet the authors

Acknowledgement

How to contribute to this project

Disclaimer

The information in this book is provided without warranty. The authors and openingdoor team have neither liability nor responsibility to any person or entity related to any loss or damages arising from the information contained in this book.

```
bookdown::serve_book()
```

Chapter 1

Introduction to open science

Bullet points: Defining science: “systematic enterprise that builds and organise knowledge in the form of testable explanations and predictions about the universe” [...] science performs knowledge validation “through (the) sharing of findings and data and through peer review”. That is, sharing these knowledge bits with other scientists *is* the method for evaluating their truthiness and validity. This process is called *peer-reviewing*. Isn’t science open already? What is the current “openness” status and how it changed over time? When and where does the term open science come from? Why we do NOT need open science: The misuse of openly available dual-edge knowledge The misinterpretation and misunderstanding of openly available knowledge from the amateurs and the general public. The degradation of the peer-review process and the advent of predatory “open” journals. The commodification of scientific knowledge, where knowledge is only apparently /free/ but the user becomes the product The monopoly of the western world in designing open science. The costs of higher accessibility to data and publications. The absence of metrics in evaluating and recognizing one’s efforts in generating and maintaining openly available knowledge. The risks of extreme rigour hindering exploratory research. Why we DO need open science: The open science currents: opening what and how? Developing and building infrastructure that help scientists practice open science Implement new evaluation processes for scientists to progress in their career based also upon their degree of “openness” Dedicating time to improve knowledge divulgation, comprehensibility, and accessibility Removing legal barriers that limit a complete access to the generated knowledge Pragmatic approach (“as open as possible as close as necessary”) The key challenges Going past the traditional scientific community Knowledge validation without economical barriers or conflict of interest Multilingual knowledge Educating a new generation of scientists Costs and infrastructures Monitoring the status of open science

Chapter 2

Open research, open data, open access

Loading...

Chapter 3

Pros and cons

3.0.1 Coverted using word2md

Lorem ipsum

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum ac elit mattis, rutrum diam sit amet, pretium turpis. Aliquam id commodo mauris. Nam sagittis purus quis molestie auctor. Morbi hendrerit maximus nibh vitae pharetra. Integer id nisl a orci faucibus dapibus. Donec blandit bibendum nisl, at cursus ex tincidunt id. Suspendisse et nulla et dui pharetra ultrices. Fusce pulvinar quis justo vel posuere.

Nullam finibus at urna ultricies porttitor. Proin a sapien vitae magna scelerisque efficitur et eu justo. Nunc eu tempus lorem. Sed metus nisl, condimentum vel ante nec, semper fringilla odio. Fusce blandit libero nec justo dignissim mattis. Integer id turpis ultricies, lacinia neque a, feugiat lorem. Sed at lorem eget eros vestibulum pellentesque. Integer vitae tortor ut urna auctor tincidunt. Vestibulum gravida condimentum enim non congue. Integer mollis augue vitae felis pretium luctus. Sed condimentum sagittis odio. Nunc a volutpat mi, eget eleifend neque. Sed sed lorem cursus, mollis libero nec, dapibus risus. Nullam volutpat elit non lectus dignissim, at tristique leo vulputate.

Lorem ipsum

Sed vestibulum hendrerit nisl at tempor. Sed mollis suscipit purus id vulputate. Nulla vulputate sollicitudin tortor sit amet sagittis. Phasellus interdum lectus a orci cursus, non ullamcorper velit dictum. Donec non pharetra dui. Pellentesque nec euismod nisl. Donec porta iaculis lectus sit amet scelerisque. Quisque at ipsum lacinia ante tincidunt vestibulum vitae in justo. Maecenas vel elit et leo finibus gravida suscipit vitae mi. Proin libero leo, dignissim id ante ut, sodales laoreet felis. Maecenas massa augue, rutrum nec luctus et, fermentum eu ligula.

Orci varius natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec pellentesque sagittis laoreet. Maecenas at quam neque. Etiam sed lorem a dolor tempus lacinia. Aenean eget nibh a erat pulvinar egestas et a odio. Aenean non auctor eros. Nulla a pharetra velit. Sed porta, orci quis aliquam fermentum, augue dui volutpat lacus, in faucibus elit massa nec dui. Nam posuere suscipit vehicula. Fusce et mauris ante. Cras lorem nisi, pretium at laoreet id, mattis non sem. Vivamus dapibus ultricies magna. Sed sed leo vitae est tincidunt pulvinar.

Lorem ipsum

Nullam dapibus, metus ut tempus molestie, erat eros dignissim est, quis dapibus dui diam eu dui. Aliquam vitae viverra felis, ut sodales nulla. Vestibulum ac lorem in justo mollis venenatis et ut leo. Nullam facilisis aliquam augue nec convallis. Aliquam ac ante interdum augue mollis aliquam. Integer sed convallis felis. Mauris placerat, purus at ullamcorper porta, enim mauris tristique mi, in feugiat justo nulla quis felis. Praesent hendrerit commodo ex, vitae eleifend ligula condimentum eget. Integer gravida ornare tortor malesuada tincidunt. Morbi non mauris at elit imperdiet dignissim in interdum ex. Mauris sed ligula pulvinar massa aliquam pellentesque vitae sit amet arcu. Morbi ultrices pulvinar laoreet. Aliquam eget nisl efficitur, molestie lorem non, vulputate elit. Sed faucibus enim et lectus tempor, et ullamcorper metus imperdiet. Sed id gravida ante, quis aliquet diam. Pellentesque id ipsum velit.

Vestibulum faucibus justo a enim vehicula, id egestas mi mollis. Vivamus facilisis dui quis libero maximus, nec tempus ipsum interdum. Etiam nec justo pretium, molestie diam ac, iaculis enim. Sed aliquet neque id ex euismod, eget volutpat tortor pellentesque. Donec consequat luctus ante, vitae vestibulum ante. Integer suscipit iaculis mauris at ornare. Suspendisse vehicula eros nulla, ac ullamcorper dolor tristique quis. Nam consectetur faucibus finibus. Curabitur vel massa vel eros congue facilisis. Nullam elit ante, congue ultrices purus sed, bibendum sodales sapien. Donec luctus dolor ut porta varius. Aenean dapibus laoreet turpis, sed suscipit eros lobortis quis. Ut tristique, dui vitae aliquet semper, lectus velit volutpat magna, at aliquet mauris eros eu nunc. In non augue et tortor ultricies maximus. Nullam quis ex ut velit tempus mattis non id ante.

Activities:

- Cras ut hendrerit risus. Suspendisse potenti.
- Praesent at egestas lorem.
- Nunc gravida vestibulum orci, nec porta sapien mollis non.
- Aliquam dapibus, erat sed fermentum consectetur, urna lacus ultricies libero, ut bibendum mauris quam eget metus.
- Phasellus neque sapien, scelerisque quis velit id, venenatis pellentesque risus.

- Nam euismod faucibus euismod.
- Mauris vitae vehicula magna, ut posuere tellus. Aliquam bibendum feugiat egestas. Aliquam pretium vitae tellus sed viverra. In hac habitasse platea dictumst. Sed quis pulvinar libero, et cursus diam. Mauris sollicitudin ultrices convallis.

Activities:

Donec nec lectus at nisi rhoncus condimentum.

Nulla egestas scelerisque diam, id condimentum mi rutrum iaculis.

Duis semper arcu ac sagittis placerat.

Proin ut arcu id tortor luctus rhoncus id ut diam.

Activities:

Donec nec lectus at nisi rhoncus condimentum.

Nulla egestas scelerisque diam, id condimentum mi rutrum iaculis.

Duis semper arcu ac sagittis placerat.

Proin ut arcu id tortor luctus rhoncus id ut diam. Sed quis ultricies nulla. Nulla sit amet purus tempor, sagittis metus at, efficitur arcu. Pellentesque sodales mi metus, vel maximus tellus sodales quis. Quisque et malesuada justo.

In a sem quis dolor pulvinar commodo.

- Donec sem felis, sodales vel sem id, aliquam elementum nibh.
- Curabitur vel dignissim elit. Sed at arcu eu odio facilisis tristique.
- Curabitur purus est, maximus in felis luctus, consequat facilisis ex.
- Sed vitae libero nec nunc euismod scelerisque ac vitae metus.
- Pellentesque at enim pharetra, blandit dui sed, sodales nisi.
- Praesent facilisis nunc orci, vitae feugiat velit consequat eget.
- Phasellus pretium orci sed tempor interdum.
- Nulla ac lectus nulla. In vitae ligula nulla.
- Integer congue sapien id nibh semper fringilla.
- Cras vel odio quis mauris cursus vehicula.
- Curabitur in tortor et lacus scelerisque finibus.

- Sed consequat sem id mi consectetur, accumsan viverra ex fringilla.

Sed in bibendum libero. Maecenas porttitor arcu nec elit congue, et ornare mauris dapibus. Nunc volutpat lectus a metus eleifend, ac consectetur felis condimentum. Quisque eleifend mauris vestibulum nunc pretium, at auctor enim feugiat. Integer efficitur sed risus eu mollis. Curabitur dapibus sodales magna, eget hendrerit quam mattis id. Donec nunc ipsum, condimentum ac tristique a, interdum eu quam. Vestibulum vestibulum nulla ac justo convallis, laoreet semper nulla malesuada. Aliquam nisl mi, luctus id porta at, varius vel quam. Cras suscipit pulvinar magna. Donec sit amet porta velit. Mauris rutrum feugiat vestibulum. Vivamus vitae dolor lobortis, gravida mi eu, imperdiet orci.

Praesent gravida magna eget est placerat, quis pharetra velit mollis. Praesent in convallis mi. Cras dignissim posuere sem, sed facilisis ipsum venenatis id. Morbi eu lectus magna. Nullam ultricies sodales arcu at faucibus. Maecenas eu ultricies libero. Nunc quis libero dolor.

Nunc imperdiet faucibus laoreet. Vivamus malesuada ligula vel metus pharetra, eget ornare ligula cursus. Proin id maximus justo. Fusce non consequat velit. Pellentesque luctus lacus vitae massa ultricies, ac euismod lectus condimentum. Quisque fermentum rhoncus erat imperdiet mattis. Sed eget ante ut mi tempus ultrices. Sed porttitor ipsum a dapibus vehicula. Curabitur orci leo, sagittis et orci sit amet, scelerisque iaculis libero.

Mauris vitae nisi rutrum, interdum nulla sed, euismod leo. Etiam ut tortor semper ligula consectetur pulvinar. Mauris nibh tortor, vestibulum non gravida et, scelerisque in tellus. Nam sed sollicitudin massa, id vestibulum orci. Nunc faucibus, sapien et gravida sollicitudin, purus turpis sollicitudin ante, vitae gravida nibh erat at turpis. Aenean condimentum vulputate maximus. Integer ex eros, venenatis sed sodales non, ultrices vel purus.

Chapter 4

Research data lifecycle

4.0.0.1 Tested using pandoc

Lorem ipsum

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum ac elit mattis, rutrum diam sit amet, pretium turpis. Aliquam id commodo mauris. Nam sagittis purus quis molestie auctor. Morbi hendrerit maximus nibh vitae pharetra. Integer id nisl a orci faucibus dapibus. Donec blandit bibendum nisl, at cursus ex tincidunt id. Suspendisse et nulla et dui pharetra ultrices. Fusce pulvinar quis justo vel posuere.

Nullam finibus at urna ultricies porttitor. Proin a sapien vitae magna scelerisque efficitur et eu justo. Nunc eu tempus lorem. Sed metus nisl, condimentum vel ante nec, semper fringilla odio. Fusce blandit libero nec justo dignissim mattis. Integer id turpis ultricies, lacinia neque a, feugiat lorem. Sed at lorem eget eros vestibulum pellentesque. Integer vitae tortor ut urna auctor tincidunt. Vestibulum gravida condimentum enim non congue. Integer mollis augue vitae felis pretium luctus. Sed condimentum sagittis odio. Nunc a volutpat mi, eget eleifend neque. Sed sed lorem cursus, mollis libero nec, dapibus risus. Nullam volutpat elit non lectus dignissim, at tristique leo vulputate.

Lorem ipsum

Sed vestibulum hendrerit nisl at tempor. Sed mollis suscipit purus id vulputate. Nulla vulputate sollicitudin tortor sit amet sagittis. Phasellus interdum lectus a orci cursus, non ullamcorper velit dictum. Donec non pharetra dui. Pellentesque nec euismod nisl. Donec porta iaculis lectus sit amet scelerisque. Quisque at ipsum lacinia ante tincidunt vestibulum vitae in justo. Maecenas vel elit et leo finibus gravida suscipit vitae mi. Proin libero leo, dignissim id ante ut, sodales laoreet felis. Maecenas massa augue, rutrum nec luctus et, fermentum eu ligula.

Orci varius natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec pellentesque sagittis laoreet. Maecenas at quam neque. Etiam sed lorem a dolor tempus lacinia. Aenean eget nibh a erat pulvinar egestas et a odio. Aenean non auctor eros. Nulla a pharetra velit. Sed porta, orci quis aliquam fermentum, augue dui volutpat lacus, in faucibus elit massa nec dui. Nam posuere suscipit vehicula. Fusce et mauris ante. Cras lorem nisi, pretium at laoreet id, mattis non sem. Vivamus dapibus ultricies magna. Sed sed leo vitae est tincidunt pulvinar.

Lorem ipsum

Nullam dapibus, metus ut tempus molestie, erat eros dignissim est, quis dapibus dui diam eu dui. Aliquam vitae viverra felis, ut sodales nulla. Vestibulum ac lorem in justo mollis venenatis et ut leo. Nullam facilisis aliquam augue nec convallis. Aliquam ac ante interdum augue mollis aliquam. Integer sed convallis felis. Mauris placerat, purus at ullamcorper porta, enim mauris tristique mi, in feugiat justo nulla quis felis. Praesent hendrerit commodo ex, vitae eleifend ligula condimentum eget. Integer gravida ornare tortor malesuada tincidunt. Morbi non mauris at elit imperdiet dignissim in interdum ex. Mauris sed ligula pulvinar massa aliquam pellentesque vitae sit amet arcu. Morbi ultrices pulvinar laoreet. Aliquam eget nisl efficitur, molestie lorem non, vulputate elit. Sed faucibus enim et lectus tempor, et ullamcorper metus imperdiet. Sed id gravida ante, quis aliquet diam. Pellentesque id ipsum velit.

Vestibulum faucibus justo a enim vehicula, id egestas mi mollis. Vivamus facilisis dui quis libero maximus, nec tempus ipsum interdum. Etiam nec justo pretium, molestie diam ac, iaculis enim. Sed aliquet neque id ex euismod, eget volutpat tortor pellentesque. Donec consequat luctus ante, vitae vestibulum ante. Integer suscipit iaculis mauris at ornare. Suspendisse vehicula eros nulla, ac ullamcorper dolor tristique quis. Nam consectetur faucibus finibus. Curabitur vel massa vel eros congue facilisis. Nullam elit ante, congue ultrices purus sed, bibendum sodales sapien. Donec luctus dolor ut porta varius. Aenean dapibus laoreet turpis, sed suscipit eros lobortis quis. Ut tristique, dui vitae aliquet semper, lectus velit volutpat magna, at aliquet mauris eros eu nunc. In non augue et tortor ultricies maximus. Nullam quis ex ut velit tempus mattis non id ante.

Activities:

- Cras ut hendrerit risus. Suspendisse potenti.
 - Praesent at egestas lorem.
 - Nunc gravida vestibulum orci, nec porta sapien mollis non.
 - Aliquam dapibus, erat sed fermentum consectetur, urna lacus > ultricies libero, ut bibendum mauris quam eget metus.
 - Phasellus neque sapien, scelerisque quis velit id, venenatis > pellentesque risus.
 - Nam euismod faucibus euismod.
 - Mauris vitae vehicula magna, ut posuere tellus. Aliquam bibendum > feugiat egestas. Aliquam pretium vitae tellus sed viverra. In > hac habitasse platea dictumst. Sed quis pulvinar libero, et > cursus diam. Mauris sollicitudin ultrices convallis.
-

+=====+

+-----+

In a sem quis dolor pulvinar commodo.

- Donec sem felis, sodales vel sem id, aliquam elementum nibh.
- Curabitur vel dignissim elit. Sed at arcu eu odio facilisis > tristique.
- Curabitur purus est, maximus in felis luctus, consequat facilisis > ex.
- Sed vitae libero nec nunc euismod scelerisque ac vitae metus.
- Pellentesque at enim pharetra, blandit dui sed, sodales nisi.
- Praesent facilisis nunc orci, vitae feugiat velit consequat eget.
- Phasellus pretium orci sed tempor interdum.
- Nulla ac lectus nulla. In vitae ligula nulla.
- Integer congue sapien id nibh semper fringilla.
- Cras vel odio quis mauris cursus vehicula.
- Curabitur in tortor et lacus scelerisque finibus.
- Sed consequat sem id mi consectetur, accumsan viverra ex fringilla.

Sed in bibendum libero. Maecenas porttitor arcu nec elit congue, et ornare mauris dapibus. Nunc volutpat lectus a metus eleifend, ac consectetur felis condimentum. Quisque eleifend mauris vestibulum nunc pretium, at auctor enim feugiat. Integer efficitur sed risus eu mollis. Curabitur dapibus sodales magna, eget hendrerit quam mattis id. Donec nunc ipsum, condimentum ac tristique a, interdum eu quam. Vestibulum vestibulum nulla ac justo convallis, laoreet semper nulla malesuada. Aliquam nisl mi, luctus id porta at, varius vel quam. Cras suscipit pulvinar magna. Donec sit amet porta velit. Mauris rutrum feugiat vestibulum. Vivamus vitae dolor lobortis, gravida mi eu, imperdiet orci.

Praesent gravida magna eget est placerat, quis pharetra velit mollis. Praesent in convallis mi. Cras dignissim posuere sem, sed facilisis ipsum venenatis id. Morbi eu lectus magna. Nullam ultricies sodales arcu at faucibus. Maecenas eu ultricies libero. Nunc quis libero dolor.

Nunc imperdiet faucibus laoreet. Vivamus malesuada ligula vel metus pharetra, eget ornare ligula cursus. Proin id maximus justo. Fusce non consequat velit. Pellentesque luctus lacus vitae massa ultricies, ac euismod lectus condimentum. Quisque fermentum rhoncus erat imperdiet mattis. Sed eget ante ut mi tempus ultrices. Sed porttitor ipsum a dapibus vehicula. Curabitur orci leo, sagittis et orci sit amet, scelerisque iaculis libero.

Mauris vitae nisi rutrum, interdum nulla sed, euismod leo. Etiam ut tortor semper ligula consectetur pulvinar. Mauris nibh tortor, vestibulum non gravida et, scelerisque in tellus. Nam sed sollicitudin massa, id vestibulum orci. Nunc faucibus, sapien et gravida sollicitudin, purus turpis sollicitudin ante, vitae gravida nibh erat at turpis. Aenean condimentum vulputate maximus. Integer ex eros, venenatis sed sodales non, ultrices vel purus.

Chapter 5

FAIR principles

5.1 What is FAIR?

In a nutshell, FAIR is a set of guiding principles to make data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable.

The FAIR principles were first launched in 2014 at a Lorentz workshop and officially published in 2016 with the focus on the EU's goal of increasing sharing and reusing of research data. The implementation of the FAIR principles for research data is a requirement imposed by the EU, alongside the EU's request on Open Science & Open Data. It is noteworthy that the FAIR principles are not a standard.

What is in it for you if you make your data FAIR?

The FAIR principles have multiple advantages for researchers. In general, by working in line with the FAIR principles, you can make your research more transparent, collaborative and sustainable and meanwhile facilitate your data management and protect your data's value for future use.

More specifically, you can expect the following by working with the FAIR principles:

- Greater impact and visibility of your research
- Opportunities for new research collaborations
- More credit for yourself as a researcher
- A more efficient data management plan
- Possibilities for future research

What is in for science if you make your data FAIR?

The FAIR principles also bring great benefits to the research community and thereby a fulfilling sense of community commitment to you as a researcher. FAIR principles:

- Enhance scientific enquiry and debate
- Enable innovation and new data use
- Increase the efficiency of research due to reusability and replication studies
- Provide a valuable resource for education and training
- Encourage the improvement and validation of research methods
- Enable scrutiny of research results
- Facilitate transparency and accountability

Key concepts to start with if you want to FAIRify your data

- **PID (persistent identifier)**

A PID is a long-lasting reference to a document, a file, a web page or another object. It is usually used for digital objects that are accessible over the Internet, but can also be used for physical objects. For example, the PID for a book can be its ISBN (International Standard Book Number). The use of PID can effectively slow or prevent the damage of "link rot" in citations, which means that the cited URLs "go dead" because the contents are removed for different reasons.

You can encounter all kinds of PIDs in your research work. Here are two of the most frequently used types:

1. DOI (digital object identifier)

The use of DOI is to identify academic and professional information, such as research articles, reports, datasets, publications – and in some cases also government documents and commercial videos.

Archiving your data with data DOI as the PID will allow you to be compliant with the FAIR principles and enhance the impact of your research through increased visibility, leading to more citations.

You can read more about DOI on the official website of the International DOI Foundation (IDF).

1. ORCID (open researcher and contributor ID)

How to find the work of one specific researcher among all the baffling names? ORCID might be your answer. ORCID provides a persistent identity for humans, so that a particular author's contributions to the literature or publications in the humanities can be easily and clearly recognized.

- **Metadata**

In short, you can simply define Metadata as "data about data".

There are multiple categories of metadata by different definitions, while the following three are the most relevant to the FAIR principles:

- **Descriptive metadata** are data that allow people to discover and identify them through the context or content, including title, author, abstract, keywords, etc.
- **Structural metadata** are data about the project's internal structure and relationships to other objects, including the unit of analysis, data collection method, sampling procedure, etc.
- **Administrative metadata** are data that are relevant for managing the project, including provenance, licence, creation date, file type, etc.

Metadata are not set from the beginning. Instead, they are subject to changes and updates. Remember to add or modify your metadata continuously throughout the project.

Metadata can help you to play better with the FAIR principles, because metadata are machine-readable and, especially when they have a PID, search engines can easily find them.

Let's FAIR up!

- **The principles**

The FAIR principles are quite straightforward. Below are the guidelines and you can read about the details for each at FAIR principles website.

1. **Findable**

- F1. Metadata and 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. Metadata and data are registered or indexed in a searchable resource.

1. **Accessible**

A1. Metadata and 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 authorization procedure, where necessary.

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

1. Interoperable

I1. Metadata and data use a formal, accessible, shared and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles:

- Ontologies
- Vocabularies
- Taxonomies

I3. Metadata and data include qualified references to other (meta)data.

1. Reusable

R1. Metadata and data are richly described with a plurality of accurate and relevant attributes:

R1.1. Metadata and data are released with a clear and accessible data usage licence;

R1.2. Metadata and data are associated with detailed provenance;

R1.3. Metadata and data meet domain-relevant community standards.

• Step by step

There are six FAIRification practices you can do to make your data FAIR.

- Documentation
- File formats
- Metadata
- Access to data
- PID (persistent identifiers)
- Data licences

Documentation

Documentation of data usually happens on two levels:

1. Data-level documentation. At this level you should include information such as data type, data processing procedures, structure of the data, e.g. questions, variables, concepts, etc.
2. Project-level (or study-level) documentation. At this level you should include information such as when, how and why the data were generated and by whom, how the data were processed, what quality assurance measures have been used, etc.

It is noteworthy that the lists are not exhaustive - other information or data files are often included at both levels.

When it comes to publishing and reserving data, FAIR documentation enables you as a researcher to show how the data was generated and for what purpose by including information such as the following:

- Methodology descriptions
- Codebooks
- Questionnaires
- Scripts like editor- and do-files (STATA)
- Laboratory notebooks and experimental protocols
- Software syntax and output files
- Database schemes
- Provenance information about secondary data
- The finalised data management plan

Publishing the documentation together with your data in a repository will boost the reusability of your data and the likelihood of your data being cited - thus more FAIR data.

File formats

Different file formats have different characteristics and properties and thus there can be some limitations to some formats. It is a good idea to decide the purpose of a file first – for example, data collection/processing/analysis, reuse, or preservation – as it helps to determine which format to use. Sometimes it can be handy to keep some data files in multiple formats.

When it comes to publishing and reserving data, you have to consider whether the file formats used for data collection, processing and analysis are also appropriate formats for long-term preservation. Furthermore, in the spirit of the FAIR principles, choose the right file format for publishing and preserving so that you and others can access and use the data later.

Here are some examples of preferred FAIR file formats for preservation:

- Containers: TAR, GZIP, ZIP
- Databases: XML, CSV, JSON
- Geospatial: SHP, DBF, GeoTIFF, NetCDF
- Video: MPEG, AVI, MXF, MKV
- Sounds: WAVE, AIFF, MP3, MXF, FLAC
- Statistics: DTA, POR, SAS, SAV
- Images: TIFF, JPEG 2000, PDF, PNG, GIF, BMP, SVG
- Tabular data: CSV, TXT
- Text: XML, PDF/A, HTML, JSON, TXT, RTF
- Web archive: WARC

Metadata

Earlier in this chapter, we learned about the concepts and categories of metadata, which play an important role in making your data FAIR. Remember to add metadata continuously to your research data, not just at the beginning or at the end of your project. You can read more about metadata standards and ontologies at Dublin Core and RDA Metadata Standards Directory Working Group. Don't forget that your metadata must have a findable PID (persistent identifier) that is typically assigned when a digital resource is placed in a data repository.

Access to data

Always consider the following before you make your data accessible:

- Who are the data available for and under which conditions?
- How are the data backed up?
- How is the above documented?
- How may the Intellectual Property Rights (IPR) agreements restrict access to the data sets both during the collection and after finalising the project?
- Do you and your collaborators agree on the 4 points above and the standard procedures and documents?

It might sound a little surprising that sensitive data, which include, for example, personal or confidential information, can also be FAIR without being open. Common practice is to anonymise (change to impersonal ID's) or de-identify (remove ID's) the data. However, this often comes with some limitations. For example, old, de-identified data cannot be added to new data after a certain period of time, which limits the reusability of the data.

PID (persistent identifiers)

Previously in this chapter, we have gained an understanding about PID (persistent identifiers), but how can you get a PID for your data and metadata? You can start with finding a repository that will provide a PID.

- You may find something interesting and suitable for you on the list of repositories recommended by the European Research Council.
- You can visit Re3data, which is a global registry of research data repositories from various academic disciplines.
- FAIRsharing allows you to discover databases grouped by domain, species or organisation.
- You can also check whether your institution has a local repository that can provide a PID for research data stored at their own local repository.
- And there is a lot more if you still want to browse for other repositories, such as OpenAIRE, Figshare, ROAR, etc.

Data licences

A data licence is a legal arrangement between the creator of the data and the end-user/the place for data depositing, which specifies what users can do with the data. The most commonly used data licences are the suite of Creative Commons (CC) copyright licences, which concern reusability of the data and are irrevocable. Another widely known licence is Copyright.

5.2 Test your understanding

Loading...

5.3 Recommended activities

In a recommend activities section like this one, we will recommend the activities to increase your understanding of the concepts and improve your practical knowledge.

Find an online dataset and investigate how FAIR the dataset is:

- Findable – Do PID's exist? Are metadata searchable?
- Accessible – Are metadata and data retrievable?
- Interoperable – Using open file format? API?
- Reusable – Can you find the provenance, licence and description of data?

Try to answer the following questions:

- Will you consider making your data FAIR? Why/why not?
- What do you think the advantages/disadvantages could be?

Chapter 6

Data centers and data repositories

Chapter 7

Data ethics

Chapter 8

Coding and other skills

Chapter 9

Communication and ethics of open science

Chapter 10

Opening your research

Chapter 11

Conclusion

Chapter 12

How to contact us

Chapter 13

Data centers and data repositories

Chapter 14

Conclusion

Chapter 15

Final Quiz

The demo quiz for the final assessment Loading...