

E.1 Exercises – Qualitative data

1. John is starting up a new research project. He will investigate the perception of gender differences within the car manufacturing industry. Based on the literature, he has a good idea of the topics he wants to explore, e.g. how these perceptions influence decision-making processes. John knows that a lot of the data he is about to collect will contain personal data, e.g. interviews and some of his field notes. Other data will be confidential, e.g. internal market analyses and competitor profiles. To save time, he will outsource the transcription of interviews to an agency. He plans to write several articles and aims at publishing at least one in a FT50 journal like Organisation Science. John remembers that he should write a data management plan.

What information should he definitely include in his data management plan to plan for making his data more FAIR?

- A) *The research question, the complete list of references and the title of the journal he wants to publish in.*
- B) *The names of the companies he has collected market analyses of, the name of the transcription agency, and the names of the interviewees.*
- C) *A description of the data collection methods, an overview of where the data are stored and backed up, and the data processing agreement with the transcription agency.*
- D) *A categorisation of the data into personal, confidential and other, an overview of the analyses he plans to carry out, and his considerations of how and where the data will be archived for the long term.*

EXPLANATIONS TO APPEAR WHEN QUIZ IS DONE:

Option (a): NOT FAIR. This is very important information for the research project as a whole. However, it is only indirectly related to the generated data and therefore less pertinent in the context of making the data FAIR. It might just be relevant for John to check out his favourite journals' data requirements and include them in his considerations regarding data publication and preservation in the data management plan.

Option (b): NOT FAIR. This is information about the source of the data, not about the generation of the data. While it is important to document this information in the project context, information about how the data was generated and processed take precedence in order to make the data accessible and reproducible.

Option (c): FAIR. To make your data FAIR, it is important to plan and document their entire life cycle – from how they were generated, where they are stored and backed up to who can access them under which conditions. This information is therefore highly relevant.

Option (d): FAIR. Analysing and archiving data are both considered data processing and should therefore definitely be planned out. Categorising the data into different types provides a good basis for planning access management in the active research as well as after the project has ended. Both from a research ethical and from a legal perspective it is paramount to document what data is personal and what data is confidential and plan proper IT security measures accordingly.

2.John is done with the first round of data collection in the field and is now well underway with processing the collected data.

After two weeks he realises that he needs to document all data processing and assign metadata to his data in order to make them FAIR. How should John go about it?

A) *He starts logging all modifications of the data like data editing or cleaning processes in a data documentation file. He documents the coding procedures in a codebook.*

B) *He asks colleagues if they know of any relevant metadata standards that he could use.*

C) *He assigns structural metadata to the interviews he is coding.*

D) *He does not bother about metadata, because his interview data are qualitative. In the articles that he plans to publish he will add a note that researchers interested in the data are welcome to contact him for more details.*

EXPLANATIONS TO APPEAR WHEN QUIZ IS DONE:

Option (a): FAIR. This is a very important step of data FAIRification. Detailed data documentation is key to improve the data quality and make them reusable in the future. Data documentation includes information about data provenance (e.g. research question, data collection methods) as well as data processing (e.g. cleaning, coding, anonymisation, pseudonymisation). It is paramount to ensure the data documentation files, logbooks, and codebooks are clearly linked to the data files they relate to.

Option (b): FAIR. Assigning metadata is key to making your data FAIR. Many disciplines or research communities have established metadata standards. More experienced colleagues can point John to the relevant standards. Otherwise, he could check fairsharing.org or contact the RDM support at his home institution.

Option (c): FAIR. To assign structural metadata, John could use the DDI-metadata standard (see: [DDI Alliance](#)). He could create a DDI codebook that includes details on the data collection method, the generated dataset, and the concepts used in the analysis. He could record the administrative and descriptive metadata in a simple Read-me text file. This file would later be converted by the data repository into machine-readable metadata.

Option (d): NOT FAIR. This is not a FAIR approach, since only researchers with access to the published articles will know that the data exist and who to contact. To make his data FAIR, John has to generate data documentation and assign metadata - regardless of whether the data are qualitative or quantitative. Detailed documentation of the data provenance and all data processing is necessary to understand and reuse data in the future (the R in FAIR). Metadata enable all four FAIR-dimensions:

- *Findability – through persistent identifiers*
- *Accessibility – of the metadata and description of access procedures of the data (if relevant)*
- *Interoperability – through standardised metadata and data formats*
- *Reusability – through licences and domain-specific metadata.*

E.2 Exercises – Quantitative non-sensitive data

1. Georgios is starting up a new project to investigate structure-property relationships of a new composite material developed by his department. The project will combine finite element analyses with straightforward mechanical testing, infra-red imaging, and long-term humidity, temperature and fatigue analyses.

How can he ensure that his data management plan will address all his needs?

- A) Write a new DMP for his project detailing how he will ensure proper handling of all data types.
- B) Copy the DMP from a previous project into a new DMP.
- C) Hire a company to write his DMP for him.
- D) Hope that one of his collaborators will take care of the DMP.

EXPLANATIONS TO APPEAR WHEN QUIZ IS DONE:

Option (a): FAIR. This is the best option, which will enable Georgios to address all the different data related issues in his project.

Option (b): NOT FAIR. Not an optimal solution, but it's not a bad idea to check your previous DMP's for knowledge that you can transfer to your new project

Option (c): MAYBE FAIR. Could also be a correct option, but will cost money and time, since a company may not be aware of any institutional infrastructures available or policies regarding data management.

Option (d): NOT FAIR. While the PI can choose to hand over parts of the responsibility to collaborators, it is the overall responsibility of the PI to make sure that the DMP is adhered to and updated. Make sure to consult your institutional support staff for advice and counselling.

2. At a conference, Georgios meets a team of researchers working on similar fatigue studies in Austria and the United States. They decide to collaborate and share their measurement data to prepare one exhaustive study on how the entire class of composite materials respond to fatigue.

How should Georgios update his DMP now that he is working with collaborators at other institutions?

A) Every time Georgios and his collaborators have performed a new experiment, they plan to send a copy of the data to each other in an Excel file.

B) Georgios keeps track of all changes to his data, records why certain decisions were made, the provenance, the dates on which the data were recorded, and all documentation related to the experimental setup and calibration of lab equipment and updates his lab notebooks, metadata and DMP regularly.

C) At the start of his project Georgios hired a data manager to write his DMP and does not consider the DMP to be his responsibility. Georgios does update all his documentation and metadata regularly.

D) In addition to everything mentioned in (B), Georgios diligently maps his data according to a standard taxonomy, but his collaborators run out of time, and use their institution's internal structure. All collaborators share their documentation, DMP and metadata in a shared repository throughout the project and update them weekly.

EXPLANATIONS TO APPEAR WHEN QUIZ IS DONE:

Option (a): Not FAIR. Sending each other constant individual file updates will create a mess in terms of file versioning. Furthermore, to ensure you have FAIR data, you have to supplement the data with provenance, documentation, and metadata.

Option (b): FAIR. Perfectly in line with FAIR best practices for documentation and metadata

Option (c): Not FAIR. You must update your DMP when you make changes to your documentation plans, even when a third party helps you to set up your DMP.

Option (d): FAIR. So long as the system that Georgios' partners use is interoperable the data can still be FAIR. Georgios must of course update his DMP.

3. Georgios's research project is coming to an end, and there are just a few final steps he must take to make sure he publishes and preserves his data in a FAIR way.

What should Georgios do to make sure that his data will be FAIR?

A. *Georgios uploads his data to a trusted repository in non-searchable pdf format with no column headers or contextual information.*

B. *Georgios makes sure his data are accessible in a standardised format in a shared folder open to all members of his research group.*

C. *Georgios uploads his metadata to a trusted repository under a 6-month embargo, after which the data will also become available.*

D. *Georgios decides to store his data and metadata at his institutional repository where they will receive a DOI, even though he will be asked to convert his proprietary files to open file formats.*

EXPLANATIONS TO APPEAR WHEN QUIZ IS DONE:

Option (a): NOT FAIR. These data are OPEN, but not interoperable because all structure to the data is lost AND are not re-usable because even though you could manually copy over the data, the contextual information is missing.

Option (b): NOT FAIR. When your research project ends you must either publish or preserve your data. If the data are only shared with your research group, it cannot be called FAIR.

Option (c): FAIR. If it is clear in his documentation when and how the data will become available, the data can be called FAIR even though they are not accessible at first.

Option (d): FAIR. These data are OPEN, FAIR and, hopefully, useable long-term since they are being stored in a trusted repository supporting open file formats with the necessary metadata.

E.3 Exercises – Quantitative sensitive data

1. Rita is starting up a new project about psychological problems in patients diagnosed with HIV, and remembers that to make her project FAIR she needs a DMP. Her data will contain sensitive personal information. She will partner with researchers in the U.S.A. and Sweden. She is unsure how to cover all the necessary legal and FAIR aspects in her DMP. How should she proceed?

- A. *In her DMP Rita reports she will publish her data through a trusted repository at the end of the project.*
- B. *Rita plans to save her data in a proprietary format and does not reserve time in her project timeline to convert the data to an open format for dissemination.*
- C. *Rita plans to record all descriptive and administrative metadata in a single file in an open format, like a .txt file.*
- D. *Rita plans to postpone sharing her data until her work has been published in a top journal.*

EXPLANATIONS TO APPEAR WHEN QUIZ IS DONE:

Option (a): FAIR. Identifying a trusted repository at the start of your project will allow you to capture all relevant metadata required by the repository.

Option (b): NOT SO FAIR. If your data is stored in proprietary formats it is likely that they will not be re-usable if you lose access to the software or simply because the file format becomes outdated in 5 to 10 years.

Option (c): NOT SO FAIR. To have FAIR data your metadata must include structural metadata. While open file formats are preferred, for interoperability the relevant metadata must be structured in a meaningful way and, ideally, according to the standards in your field.

Option (d): FAIR. Data can be FAIR when it is private, when it is accessible by a defined group of people, or when it is accessible by everyone (open data). So long as you do plan to upload the metadata in a trusted repository it can be FAIR.

2. Rita is very aware of GDPR and chooses to be on the safe side to avoid questions of non-compliance. This implies that she makes the following decisions. How do you find the decisions?

A. *Rita intentionally discards relevant metadata associated to her sensitive data while she is still actively working with the data.*

B. *(FAIR:) During the project Rita shares her data in a standardised format with her research group through a shared folder.*

C. *(NOT FAIR:) Rita's project investigating a rare disease has suddenly run out of funds and is to be terminated. While finalising the project, she decides to delete her rare dataset.*

EXPLANATIONS TO APPEAR WHEN QUIZ IS DONE:

Option (a): NOT FAIR. Having transparency in your methods and data is an important aspect of research in general. Discarding metadata could render the data useless, hinder reproducibility, and/or can misguide others who hope to re-use the data. Storing and keeping legally collected data and metadata is in line with GDPR.

Option (b): FAIR. To be FAIR your data should be able to be found by appropriate people at appropriate times. This can include shared folders, drives, private databases, public databases or more, depending on the phase of research you are in. Remember to include this information in the consent form, if this is the legal basis for your data collection.

Option (c): NOT FAIR. Even though Rita may have worked according to the FAIR best practices in her project, she should not just delete her rare dataset. She should consider the cost of recreating the rare dataset and predict foreseeable future uses for her research. She should transfer the collected data to an organisation which are allowed to keep the data without consent from the participants, e.g. a national archive.

3. Rita wants to share the findings of her research and finds a repository where she can upload her data files. How FAIR are the following choices?

- A. *Before publishing or preserving her research data, Rita intentionally discards relevant metadata associated to her sensitive data.*
- B. *After the project, Rita uploads her quantitative data to a trusted repository in non-searchable pdf format with no column headers or contextual information.*
- C. *Rita shares her data in a standardised format with her research group through a shared folder.*
- D. *Rita sends her sensitive data to the National Archives available only under a restrictive end user licence.*
- E. *After the project, Rita uploads anonymised data files to her project web page.*
- F. *Rita uploads her metadata to a trusted repository under a 6-month embargo, after which the data will become available.*
- G. *Rita uploads her metadata to a trusted repository and provides her institutional contact information. She does not upload the sensitive data files. One month later she moves to a new institute.*

EXPLANATIONS TO APPEAR WHEN QUIZ IS DONE:

Option (a): FAIR: If this information must be discarded for legal reasons her data can still be FAIR if they can be re-used and if she documents her actions.

Option (b): NOT FAIR. These data are OPEN, but not interoperable because all structure to the data is lost AND are not re-usable because even though you could manually copy over the data, the contextual information is missing.

Option (c): NOT FAIR. At the end of your research project you must either publish or preserve your data. If the data are only accessible to your research group, they cannot be called FAIR.

Option (d): FAIR. Though sensitive data may never be publicly accessible and reusable they can still be FAIR if the terms for access are clear, the data is findable, and structured in a re-usable and interoperable way.

Option (e): NOT FAIR. Without a DOI her data may not be findable (or citable) even though they are published and accessible online.

Option (f): FAIR. If it is clear in her documentation when and how the data will become available, the data can be called FAIR even though they are not accessible at first.

Option (g): NOT FAIR. While publishing your metadata makes them findable, if your contact information immediately becomes outdated you cannot be reached, and therefore your data are not accessible. The contact information should be updated to point to your new institution or to someone else who can take over the responsibility.

4. Rita has just published the results of a 20-year study into the relationship between obesity and exercise in a top journal. Her work is ground-breaking, and many researchers are eager to dig into her data. However, her data contains the personal information of hundreds of individuals.

How could she ensure that her research data can still be FAIR?

- A) Rita stores all her data in a filing cabinet in her office, and allows fellow researchers to read it under supervision
- B) Rita deposits her un-anonymised data at the National Archives and publishes the metadata and documentation in an online repository which gives them a DOI.
- C) Rita stores her data in her institution's local secure storage solution and invites interested researchers to email her if they would like to collaborate.
- D) Rita anonymizes her data and publishes it along with the metadata and the documentation in an online repository linked to a DOI.

EXPLANATIONS TO APPEAR WHEN QUIZ IS DONE:

Option (a): NOT FAIR: Besides the fact that her data is not findable and not re-usable, this storage method is not recommended as it is not legal.

Option (b): FAIR: This way the data are searchable through a search engine and the data are citable because of the DOI. Data stored at the National Archives do not need to be anonymised, and informed consent is not necessary. For others to access the data, permission will need to be obtained from the Danish Data Protection Agency.

Option (c): NOT FAIR: Even though this is a common procedure for data handling after a project has been finalised, Rita's data will not be findable if they do not have a DOI.

Option (d): FAIR: Note that the data might lose its value when anonymised to the point that they could no longer be usable. Also be aware that some data cannot be fully anonymised, e.g. a data set with many variables and few patients combined with rare diseases or side effects.