Joint Programming Initiative

A HEALTHY DIET FOR A HEALTHY LIFE

FAIR data principles Procedure

January 2018

Guidelines FAIR data principles

1. Background

The JPI HDHL coordinates research with the impact of diet and lifestyles on health, significantly contributing to the construction of a fully operational European Research Area for the prevention of diet-related diseases and strengthening the leadership and competitiveness of research activities in this field.

The researchers involved in the JPI HDHL calls must ensure open access to project results and information about the project. Beyond an obligation to inform the public, effectively and carefully planned dissemination activities can serve as a powerful advertisement for the future products, processes or services resulting from the projects and can pave the way for future market entry.

The JPI HDHL has already presented in September 2013 a **Quick Guide** which provides an overview of the Intellectual Property and Open Access issues to be taken into account within the JPI-HDHL Joint Actions (http://www.healthydietforhealthylife.eu/index.php/hdhl-documents/key-documents).

To further improve the impact of research investments the Management Board of the JPI HDHL adopted the FAIR data Principles and agreed that all new funded research by the JPI HDHL should apply these principles. With the adoption of the FAIR data principles the JPI HDHL contributes to an open and sustainable European Data Infrastructure and stimulates the circulation and reuse of scientific data. In addition the JPI HDHL contributes to the vision of Commissioner Moedas for EU research and innovation policy: Open Innovation, Open Science and Open to the World.

2. FAIR data management principles[1]

The FAIR Data Principles consist of a concise and measureable set of principles for data sharing and is already endorsed by a diverse set of stakeholders (academia, industry, funding agencies, and scholarly publishers) with the intent to act as a guideline for those wishing to enhance the reusability of their data holdings. The FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. FAIR refers to the findability, accessibility, interoperability and reusability of data.

2.1 Findability

- (meta) data are assigned with a global unique and persistent identifier
- Data are described with rich metadata
- Metadata clearly and explicitly include the identifier of the data it describes
- (meta) data are registered or indexed in a searchable resource

2.2 Accessibility

- (meta) data are retrievable by their identifier using a standardized communications protocol
- Protocol is open, free and universally implementable
- Protocol allows for an authentication and authorized procedure, where necessary
- Metadata are accessible, even when the data are no longer available

2.3 Interoperability

[1] Wilkinson M.D. et al. **The FAIR Guiding Principles for scientific data management and stewardship.** SCI Data 2016. 15,33:160018. Doi 10.1038/s data.2016.18.

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

2.4 Reusability

- (meta)data are richly described with a plurality and relevant attributes
- (meta)data are released with clear and accessible data usage license
- (meta)data are associated with detailed provenance
- (meta)data meet domain-relevant community standards

3. Data management plan

Proper data management is essential to ensure that data will be FAIR at the end of a projects. A data management plan (DMP) is a tool for performing data management. The DMP describes the data you expect to acquire or generate during the course of a research project, how you will manage, describe, analyse, and store those data, and what mechanisms you will use at the end of your project to share and preserve your data.

The JPI HDHL developed a DMP template (see Annex 1 for the guideline of this template) to support researchers by their data management. By filling out this template you will create a DMP, and the executing of this plan will ensure that the research results and underlying data can be verified and reused.

Please note:
Although the JPI HDHL in general supports open access of data, JPI HDHL recognises that under some circumstances (parts of) data gathered in project should be kept (temporarily) closed. In case researchers believe that their data cannot be open access, they should explain the reason in the DMP / data management strategy.

4. More information & support

FAIR DATA

More information about FAIR can be found here:

- FORCE11
- Fair data principles (article in nature)

SUPPORT AND TOOLS IN DEVELOPING YOUR DMP

- To draft a DMP/ data management strategy it might be helpful to ask a data management expert for assistance. Many universities and research institutions will have a data manager, ICT expert or research support staff that could advise you.

- ENPADASI
  JPI HDHL funded ENPADASI which is an open access research infrastructure for both observational/epidemiological and intervention/mechanistic nutritional studies. ENPADASI makes it possible to store and reuse your data according to the FAIR data principles and compare your data which similar studies or with studies with a similar outcome.

  Please note that using the ENPADASI system will not mean that your data will become automatically open access, it means that you have to indicate who has access under which conditions. Your metadata and data can be stored in the ENPADASI system (DASH-IN) either on your own server or one that is hosted by ENPADASI partners. The individual data may be

shared as raw data or as summarized data depending on the legal and ethical agreements underlying the data. If you indicate in the data management plan under 1.6 that you will use the ENPADASI system as facility to store your data, you can use the ENPADASI standard data management plan (see www.ENPADASI.eu), in which most of the fields already have been filled out. The data management costs will be needed to support the maintenance of the ENPADASI system, if you decide to make use of it. ENPADASI also has information on legal and ethical regulations for data sharing in the partner countries.

- Metadata standards directory of the Research Data Alliance: can be searched for discipline-specific standards and associated tools

- DataCite: provides persistent identifiers (DOIs) for research data.

- Digitial Curation Center (DCC): s an internationally-recognised centre of expertise in digital curation with a focus on building capability and skills for research data management. The DCC provides expert advice and practical help to research organisations wanting to store, manage, protect and share digital research data.

Annex 1 Guideline JPI HDHL DMP template

Structure of the template

The DMP template is a set of questions that you should answer. The template is arranged based on the principles of the FAIR data. You can work on the plan during the various phases of the project which will follow the research data life cycle (collecting and analysing data, storing data during and after the project and publishing data).

The questions in the DMP are intended to inform you about data management and to record specific important information regarding your project. Not every section of the DMP will be answerable at the start of the project. Rather, the DMP is a living document in which information can be added or changed over the life course of the project.

Below the main sections of the DMP template are outlined.

1. General features of the project and data collection (7 questions)

This section describes the purpose of the data collection/generation and provides a number of basic features of your data collection. This forms the starting-off point for planning your approach to data management, this includes:

- The origin of the data
- The size of the data collection
- The type of data

2. Making data findable (3 questions)

The information covered in this section of the DMP is related to the FAIR principle of **Findable.** It will help you to ensure that new data collection will be **findable** at the conclusion of your project (according to FAIR principles)

Data collections can be found in online metadata catalogues or web portals, along with added descriptive information about the complete data collection (metadata). Usually such databases or archives have an advanced search engine, which you can use to find reusable data collections.

Please note that this section does not deal with storage of the actual data in an archive or respiratory (this is covered in section 5, Making data reusable).

This section produces the following *key characteristics*:

- The **location** where you have placed your data collection (a metadata catalogue, archive or web portal)
- A **persistent identifier** (for example DOI code) making the data collection traceable and citable

3. Making data openly accessible (3 questions)

The information covered in this section of the DMP is related to the FAIR principle of **Accessible** data.

If you enable **open access** to your collection, the data, once registered in an archive or repository, will be freely accessible to any interested parties.

Restricted access is also possible. This means that you, as the data producer, attach certain **conditions** to access to the collection. Interested parties can submit a request for a data set.

4. Making data interoperable (2 questions)

The information covered in this section of the DMP is related to the FAIR principle of **Interoperable** data. Data which are Interoperable can be combined, exchanged and integrated with data from other

data collections. This enriches the collection, and provides interesting opportunities for new avenues of research.

If you wish to combine (or exchange, or integrate) different data sets, you must ask yourself whether this is *allowed* and whether this is *possible*.

Is it allowed? In research involving human subjects, the question is if combining data is permissible in an *ethical-legal* sense.

Is it possible? This involves the question whether it is *technically* possible to combine, exchange or integrate the data. To make sure that it is, you must choose a metadata standard enabling the exchange of data at record levels.

Finally, the **quality** of the data collection determines whether or not the data are interoperable and reusable. Quality (that is, the full and correct conservation of data) is a point of attention throughout the duration of the project.

5. Making data reusable (4 questions)

The information covered in this section is related to the FAIR principle of **reusable data**. The goal of data management is to ensure that data collections are verifiable and reusable (according to FAIR principles). All sections of this DMP format contribute to this goal.

To make sure that your data remain **accessible and reusable in the long term**, you must take proper care of the quality of the collection, document the data acquisition process, and select the right data for long-term archival and sustained storage. In the interest of long-term usability, also keep in mind the necessary maintenance to the collection and updates to the required software.

6. Allocation of resources (3 questions)

Costs are related to data management and long-term storage of data. This section summarizes the estimated costs and the allocation of resources to cover these costs.

Within the JPI HDHL national researchers are often funded by their national funding agencies. Rules and regulations regarding the eligibility of such cost in the research project may vary. Therefore, you are strongly advised to discuss this issue with your national funding agency and check if (part of these cost) may be included in the research budget.

7. Legalisation and regulations (3 questions)

This section is aimed at informing you about **laws and regulations** applying to scientific research, and the conditions imposed by these laws on the collection, use, and publication of data.

You have to describe any potential ethical aspects of the work to be carried out, and how the project will fulfil applicable requirements in institutional, national and European Union directives and law.

Data management plan template JPI HDHL

1 General information of the project and the data

1.1 Describe the purpose of the data collection/generation and its relation to the objectives of the project.

1.2 This project will:

☐ Use existing data
☐ Generate new data
☐ Combine and/or integrate datasets
☐ Add new data to existing datasets
☐ Other, namely: ……..

1.3 Describe for each project partner the data that he/she will bring into this project

1.4 In case you will re-use existing data and/or will combine/integrate existing data sets please indicate if you have permission to use the data.

☐ Yes, I have permission to use the data for this project.
☐ No, I have not yet the permission to use the data for this project.
☐ No permission is required, since the data are openly accessible.

1.5 Indicate the expected size of the data in terms of subject numbers and giga-/terabytes.

1.6 Describe the facilities that will be used for storage and backup of the data during the project.

1.7 Describe the expected end products of this project and indicate to whom this might be useful. Also indicate if the expected end products will be available for future research and reproduction research. Examples of possible end products are: raw data, processed data, biobank, metadata, syntaxes, software.

2. Making data findable

2.1 Indicate the repository or archive that will be used for post-project archiving/publishing purposes.

2.2 Indicate if the metadata of the project will be registered or indexed in a searchable resource:

☐ Yes the metadata will be registered in the respiratory of storage.
☐ Yes the metadata will be registered in an online web portal, namely:……..
☐ I have not decided yet where the metadata will be registered
☐ Other, namely:……..

2.3 Indicate if a persistent and unique identifier will be used to ensure that the data will be findable and locatable.

☐ Yes, the Digital Object Identifier (DOI-code) will be used
☐ Yes, the follow Persistent Identifier will be used:……
☐ No persistent and unique identifier will be used (please explain):……

3. Making data openly accessible

3.1 Indicate if (parts of) the data produced and/or used in the project will be made openly accessible as the default

☐ Yes, all the data will be openly accessible
☐ Yes partly, the following datasets will be openly accessible:……..
☐ No, (parts of) the data will be restricted available (indicate which datasets will have restricted access and explain the reason for restricted access):……

3.2 Indicate which part of your data will directly available or after an embargo period (in case of an embargo period, please explain why and how long this will apply).

3.3 In case of restricted access:
- Indicate if the terms of use are developed in collaboration with a legal advisor

- Indicate which of the following conditions will apply to the use of your data.:
☐ Agreements on methodology
☐ Allowance to link the dataset with other datasets
☐ Sharing of data for commercial purposes (taking into account provision of state aid law)
☐ Agreements on publication and authorship
☐ The manner in which the dataset can be accessed
☐ The permitted period during which the data set can be used
☐ Data security
☐ Reimbursement of cost (for example for obtaining the data)
☐ Other conditions, namely:…….

4. Making data interoperable

4.1 Describe the standard data (or file) format that will be used for recording the data to make your data interoperable (explain if the data format will be machine actionable).

4.2 To provide descriptive information (metadata) on the data that have been used and/or collected during my research I will use:

☐ A metadata schema specific of my research discipline, namely:……..
☐ A standard metadata schema, namely:……
☐ I have not decided yet which metadata scheme will be used in the project

5. Making data re-usable

5.1 Describe how you will ensure that the research data will be of sufficient quality to allow other researchers to interpret and reuse them. Please include:
- Documentation of the research process (for replication research)
- Quality of research data in terms of completeness, correctness and consistency of the data
- Documentation of used software

5.2 Indicate if you have selection criteria to determine which part of the data should be preserved once the project has finished.

☐ Yes
☐ Not yet
☐ (Part of) the data need to be destroyed after the project

5.3 Indicate the choice for long-term storage at the end of the project.

☐ Archive with a data seal of approval, namely:……..
☐ Archive with another certificate, namely: ……
☐ Archive with no certificate (please explain how your data will remain accessible and reusable in long term): ……
☐ No yet decided: ……..

5.4 Indicate the intended data preservation period. Please indicate if this in accordance with international recommendations.

6. Allocations of resources

6.1 Indicate the estimated data management cost and how these costs are covered (researchers should discuss this issue with national funding agencies to check if part of the cost can be included in the research budget).

6.2 Indicate the cost for long-term storage of your data and indicate how these costs will be covered.

6.3 Indicate who will be responsible for data management in the project.

7. Legalisation and regulations

7.1 Indicate if your research includes human subjects.

☐ Yes human subjects
☐ No

7.2 In case the project will include human subjects please specify the relevant ethical or legal regulation concerning privacy-sensitive data. Please specify the legislations for the different countries that are involved in the project.

7.3 In case the research includes human subjects:

- indicate if you have secured permission from the participants to use the data (specify the form of permission)
- indicate if you will anonymize the data or use pseudonymous data.
- indicate if informed consent for data sharing and long term preservation is included in the data collection.

Guidance Data Management Plan JPI HDHL

This document provides more guidance and background information to the questions of the Data Management Plan. The data management plan can be created online (DMP-online or JPI HDHL website) and the guidance for the questions will be added in the online environment.

1. General information of the project and the data

Question	Guidance	Additional information
1.1 Describe the purpose of the data collection/generation and its relation to the objectives of the project.	Describe which kind of data will be collected/generated to perform the study (studies). Give a brief description of the data that will be created, noting its content and coverage.	
1.2 the project will: - using existing data - generate new data - combine and/or integrate datasets - add new data to existing datasets - other		
1.3 Describe for each project partner the data that he/she will bring into this project	Describe for each project partner the data that he/she will bring into this project. Specify the types of data (e.g. quantitative, qualitative; generated from surveys, clinical measurements, interviews, medical records, electronic health records, administrative records, genotypic data, images, tissue samples) for each data set that will be used.	
1.4 In case you will re-use existing data and/or will combine/integrate existing data sets please indicate if you have permission to use the data.	Please check the conditions for reuse of data, e.g. costs, achieving informed consent for connecting databases) and indicate if you will have the permission to use the data.	
1.5 Indicate the expected size of the data in terms of subject numbers and giga-/terabytes.	Estimate the volume of data in GB/TB and how this will grow to make sure any additional storage and technical support required can be provided.	
1.6 Describe the facilities that will be used for storage and backup of the data during the project.	Describe how the data will be stored and backed-up to ensure the data and metadata are securely stored during the lifetime of the project. Storing data	

	on laptops, computer hard drives or external storage devices alone is very risky. The use of robust, managed storage with automatic backup, for example that provided by university IT teams, is preferable. Questions to consider: • Where will the data be stored? • How will the data be backed up? i.e. how often, to where, how many copies, is this automated… • Who will be responsible for storage and backup? • Do you have access to enough storage or will you need to include charges for additional services?	
1.7 Describe the expected end products of this project and indicate to whom this might be useful. Also indicate if the expected end products will be available for future research and reproduction research.	Outline the types of data that are expected to be produced from the project and to whom this might be useful.	
2. Making data findable		
2.1 Indicate the repository or archive that will be used for post-project archiving/publishing purposes.	Researchers should consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. To aid data preservation, sharing and reuse the use of established data repositories, community databases and related initiatives is recommended. An international list of data repositories is available via Databib or Re3data.	
2.3 Indicate if the metadata of the project will be registered or indexed in a searchable resource.	By providing some basic metadata online (i.e. the title, author, subjects, keywords and publisher of the dataset) potential new users can find out about your data and identify whether it could be suitable for their research purposes. Please indicate if (basic) metadata of your dataset will be registered online and indicate where the metadata will be registered.	
2.4 Indicate the unique identifier that will be used to ensure that the data will be findable and locatable	Persistent identifiers supports simple and effective methods of data citation, discovery, and access. Indicate if and which persistent Identifier will be used. In case no persistent identifier will be used please explain how you will ensure that your data and metadata will be findable and locatable.	DataCite

3. Making data openly accessible		
3.1 Indicate if (parts of) the data produced and/or used in the project will be made openly accessible as the default.		
3.2 Indicate which part of your data will directly available or after an embargo period (in case of an embargo period, please explain why and how long this will apply).	JPI HDHL requirement is for timely data sharing, with the understanding that a limited, defined period of exclusive use of data for primary research is reasonable according to the nature and value of the data, and that this restriction on sharing should be based on simple, clear principles.	
3.3 In case of restricted access: - Indicate if the terms of use are developed in collaboration with a legal advisor - conditions will apply to the use of your data.	Restrictions to data sharing may be due to participant confidentiality, consent agreements or IPR. Strategies to limit restrictions may include: anonymising or aggregating data; gaining participant consent for data sharing; gaining copyright permissions; and agreeing a limited embargo period.	
4. Making data interoperable		
4.1 Describe the standard data (or file) format that will be used for recording the data to make your data interoperable (explain if the data format will be machine actionable).	Outline and justify your choice of format e.g. SPSS, Open Document Format, tab-delimited format, MS Excel. Decisions may be based on staff expertise, a preference for open formats, the standards accepted by data centres or widespread usage within a given community. Using standardised and interchangeable or open lossless data formats ensures the long-term usability of data.	
4.2 Indicate the metadata standard that will be used to provide descriptive information of the data.	Metadata should be created to describe the data and aid discovery. Describe the types of documentation that will accompany the data to provide secondary users with any necessary details to prevent misuse, misinterpretation or confusion. This may include information on the methodology used to collect the data, analytical and procedural information,	

	definitions of variables, units of measurement, any assumptions made, the format and file type of the data. Researchers are strongly encouraged to use community standards to describe and structure data, where these are in place. The DCC offers a catalogue of disciplinary metadata standards.	
5. Making data re-usable		
5.1 Describe how you will ensure that the research data will be of sufficient quality to allow other researchers to interpret and reuse them. Please include: • Documentation of the research process (for replication research) • Quality of research data in terms of completeness, correctness and consistency of the data • Documentation of used software	Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies. Outline how the data will be collected/generated and which community data standards (if any) will be used at this stage. Indicate how the data will be organised during the project, mentioning for example naming conventions, version control and folder structures. Consistent, well-ordered research data will be easier for the research team to find, understand and reuse.	
5.2 Indicate if you have selection criteria to determine which part of the data should be preserved once the project has finished.	Research data are preserved to reuse the date or to verify or validate research results. However, in some cases the preservation of data is not necessary (for example in some cases it might be cheaper to regenerate the data or it might be easier due to privacy issue of data). Therefore researchers should consider which data should be preserved beyond the period of funding. This should be based on what has long-term value and is economically viable to keep. Information on the selection criteria for the preservation data can be found in the DCC guide: How to appraise and select research data for curation.	
5.3 Indicate the choice for long-term storage at the end of the project.	In section 2 you already indicated the repository or archive that will be used for post-project archiving/publishing purposes. In this section we come back to this issue with a specific focus on the sustainability of the chosen respiratory. To ensure a sustainable long-term storage of your data some international guideline are developed. Such international guideline is the DATA Seal	***key characteristics***

	Approval. The data seal approval is developed to ensure that archived data can still be found, understood and used in the future. In case you store your data at a respiratory with data seal approval you have the assurance that your data and associated materials will be stored in a reliable manner and can be reused. It is recommended to check if the respiratory of your choice has a data seal approval or to which extent the respiratory comply with criteria of the data seal approval.	
5.4 Indicate the intended data preservation period. Please indicate if this in accordance with international recommendations.	Researchers should consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. The long-term preservation period may depend on the type of data. As a guideline researchers should strive to retain data for a minimum of 10 years from the end of the project. For data that by their nature cannot be re-measured, efforts should be made to retain them indefinitely.	
6. Allocation of resources		
6.1 Indicate the estimated data management cost and how these costs are covered (researchers should discuss this issue with national funding agencies to check if part of the cost can be included in the research budget).	Carefully consider any resources needed to deliver the data management plan. Please take into account: - Any relevant technical expertise, support and training that is likely to be required - any hardware or software which will be purchased or additional storage and backup costs that may be charged by IT services. - Possible charges applied by data repositories	
6.2 Indicate the cost for long-term storage of your data and indicate how these costs will be covered.		
6.3 Indicate who will be responsible for data management in the project.	Outline the roles and responsibilities for all activities e.g. data capture, metadata production, data quality, storage and backup, data archiving & data sharing. Individuals should be named where possible. For collaborative	

	projects you should explain the co-ordination of data management responsibilities across partners.	
7. Legalisation and regulations		
7.1 Indicate if your research includes human subjects.		
7.2 In case the project will include human subjects please specify the relevant ethical or legal regulation concerning privacy-sensitive data. Please specify the legislations for the different countries that are involved in the project.		
7.3 In case the research includes human subjects: - indicate if you have secured permission from the participants to use the data (specify the form of permission) - indicate if you will anonymize the data or use pseudonymous data. - indicate if informed consent for data sharing and long term preservation is included in the data collection.	**Informed consent** In order to make sure that research data can be made available for future reuse, it is important that consent for future reuse of the data by other researchers is sought from participants. Participants should be informed how research data will be stored, preserved and used in the long-term, and how confidentiality can be protected when needed. Consent procedures must be tailored for the specific research context, methods and sample, the nature of the data (personal, sensitive, level of detail), the format of the data (surveys, written, recordings) and the planned data uses and handling. This will influence the type of consent and consent process used. More detailed information and examples of informed consent form can be found at uk data service.	