



# Science Europe Guidance Document

Presenting a Framework for Discipline-specific  
Research Data Management

JANUARY 2018



**SCIENCE  
EUROPE**  
Shaping the future of research

January 2018

'Science Europe Guidance Document Presenting a Framework for  
Discipline-specific Research Data Management': D/2018/13.324/1

Author: Science Europe

Co-ordination: Science Europe Working Group on Research Data

Editor: Peter Doorn (Netherlands Organisation for Scientific Research),  
Chair of the Science Europe Working Group on Research Data

For further information, please contact [office@scienceeurope.org](mailto:office@scienceeurope.org)

© Copyright Science Europe 2018. This work is licensed under a Creative Commons Attribution 4.0 International Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited, with the exception of logos and any other content marked with a separate copyright notice. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.





# Contents

|   |           |
|---|-----------|
| <b>Introduction</b>   | <b>5</b>  |
| <b>A Framework for Research Data Management</b>                     | <b>6</b>  |
| Domain Data Protocols   | 6         |
| The Framework   | 7         |
| Proof of Concept  | 7         |
| <b>A Framework for Domain Data Protocols</b>                        | <b>8</b>  |
| Components of the Framework   | 9         |
| 1. Formal Minimum Conditions  | 10        |
| 2. Applicable Laws and Regulations                                  | 11        |
| 3. FAIR Principles  | 13        |
| 4. Applicable Standards   | 14        |
| 5. Templates and Examples   | 15        |
| 6. Support Resources  | 15        |
| <b>Proofs of Concept from different Communities</b>                 | <b>18</b> |
| Humanities (general): DARIAH  | 20        |
| Humanities – Archaeology: PARTHENOS/ARIADNE                         | 22        |
| Linguistics – Language Data: CLARIN                                 | 24        |
| Social Sciences – Survey data: CESSDA                               | 26        |
| Social/Behavioural Sciences – Psychology                            | 28        |
| Social Sciences – Family of Studies on Longitudinal Ageing          | 30        |
| Life Sciences – Bio-informatics: ELIXIR and Force11/RDA FAIRSharing | 32        |
| Plant Sciences: ERA-CAPS  | 35        |
| Climate Research: ICOS  | 38        |
| <b>Notes and References</b>   | <b>40</b> |
| <b>Annex A – DCC Default DMP Template</b>                           | <b>42</b> |
| <b>Glossary</b>   | <b>46</b> |





## Introduction

To improve research data management (RDM), Research Funding Organisations and Research Performing Organisations increasingly require researchers to develop a Data Management Plan (DMP) for their project proposals or their evaluation.

While researchers recognise the benefits of better RDM, they often see these new requirements as an additional burden imposed on them by funders or employers. Funders and research organisations, for their part, are unsure about the practical possibilities and the best ways to implement the relevant policies. They would appreciate a system that makes it easier for them to assess, compare, and evaluate DMPs.

In the policy context of the Open Science agenda, DMPs are also gaining importance. Science Europe believes in the development of Open Science in a way that recognises the driving role of the scientific communities in shaping and adopting Open Science practices, such as data sharing and re-use. However, discussions around Open Science usually focus on topics such as infrastructures and governance, while data management remains an unsolved issue. Various DMP templates exist across funders, research institutions, countries, and disciplines, which vary significantly.

*The Research Data Working Group would like to acknowledge the contribution of Patrick Aerts from DANS<sup>2</sup> for his valuable contributions to the report.*



# A Framework for Research Data Management

The Science Europe Working Group (WG) on Research Data has made it its mission to develop and advocate a framework for ‘disciplinary research data management protocols’, also called Domain Data Protocols (DDPs) as a pragmatic solution to ensure proper implementation of individual DMPs.

Based on this general Framework, scientific communities are encouraged and enabled to set up protocols according to their specific needs. Individual researchers can then use the protocols as template for their DMPs in any given research field. This will make RDM planning and evaluation easier for all parties involved (funding agencies, research performing organisations, and the research communities), support good research practice, and ensure research integrity.<sup>1</sup>

The approach of the WG has been to give ownership of such a framework to research communities, while taking into account the needs of funders and employers. The Framework presented in this document merges a bottom-up approach that places researchers in the driving seat with a top-down vision on how funders can ensure the efficacy and accountability of publicly funded research. It will simplify the process of DMP creation and evaluation and prevent DMPs from becoming a bureaucratic imposition on researchers, creating a win-win situation for researchers, funders, and the wider society.

## Domain Data Protocols

The core idea of the approach is that research communities will use this Framework to formulate ‘protocols’ for the collection and management of data within their disciplinary domain or community. Instead of having to evaluate and monitor many individual DMPs, funders and research organisations would simply require project proposers to comply with the relevant protocol. This would result in much shorter DMPs on average, reducing the time needed to review and evaluate them, as well as the time needed for researchers to create them.

An RDM protocol would contain the usual elements of a DMP. It would pay special attention to standards and guidelines for data management that are relevant for a specific field or research community that shares similar data collection and processing methods. It would be a public document that could be properly referenced and should be considered a template that is already mostly filled in, possibly offering alternatives from which a researcher can choose, depending on the particularities of their research project.

An RDM protocol would not replace the responsibility of individual researchers to have and work according to a DMP. Instead, it is a building block to start from, or a DMP in its own right, with only a few parameters left to be filled in. Researchers should also not be required to blindly obey such a protocol (deviations from the protocol should be possible as long as an explanation is provided), but it will allow them to state that their DMP complies with the protocol of their field.

That protocol will be required to incorporate minimum conditions set in the Framework terms of reference, formulated by research funding and performing organisations.

Efforts to developing protocols, guidelines, best practices, and templates for DMPs are already taking place in several areas. However, such recommendations and guidelines do not usually have a formal status. The WG intends for RDM protocols developed by communities to be peer-reviewed, openly published, kept up to date, and formally recognised by research funding and research performing organisations. The protocols will have to comply with certain criteria and quality standards and their development will require the expertise of scientific communities.

## The Framework

The Framework, presented in the first part of this document, is designed to support research performing organisations and funders as well as researchers in their efforts to improve RDM. It provides guidance on how to establish domain-specific DDPs. It sets a number of minimum requirements for disciplinary or community data protocols that closely resemble the requirements of current DMPs. Communities are encouraged to develop their Protocols based on the Framework. Funders can then rely on them, knowing that individual DMPs that are based on the Framework and the respective domain-specific protocol comply with general acknowledged requirements. When funders encourage researchers to use their specific DDPs when submitting a DMP, the use of this structured approach will become standard. The Framework will fit perfectly in data policies that have already been or are still being formulated. It includes three conceptual key messages:

1. Treat Software Sustainability on equal footing with RDM at the policy level (at the practical level several aspects need to be dealt with differently).
2. Consider data and software explicitly as value objects.
3. Distinguish different stakeholder groups when addressing RDM and Software Sustainability.

This document will not explicitly re-address these issues, except for the remark that what applies to RDM, the Framework, and the protocols should be applied to Software Sustainability as well. For the sake of the readability of this document, software protocols will not be mentioned separately. Thus, when 'Domain Data Protocol' (DDP) is mentioned, the reader should consider that the same applies for the software equivalent, a 'Domain Software Protocol'.

## Proof of Concept

In developing the Framework, the WG has contacted various external scientific communities who were asked for proof of concept of the approach taken by the WG. Reactions from the different communities regarding this approach were positive overall and several communities expressed their interest to collaborate for the development of first RDM protocols. The detailed responses and feedback received from these different communities are given in the 'Proofs of concept' part of this document. The intention of this second part is to give an idea of the state of play in various communities and their attitudes towards a more structured RDM approach. Organisations working on their own protocols can find good practices and food for thought for their own approaches.



# A Framework for Domain Data Protocols

The Framework's set of minimum requirements (or terms of reference) encompasses matters such as implementation of applicable laws and regulations, references of standard data formats and software principles. It also deals with references to FAIR data<sup>3</sup> and elements that allow for funding agencies and governments to be properly accountable for the funds spent on research. This Framework should be considered as the basis for the development of DDPs by the various scientific communities.

In this document, DDPs are defined as generally agreed-upon guidelines, or predefined written procedural methods. One might also conceive a DDP as a 'model DMP' for a given domain or community that shares common methods.

These protocols will make life easier for researchers. They can refer to the data protocol that is to be followed, instead of 'reinventing the DMP wheel'. The protocols will also raise quality standards of DMPs, making them a stronger tool and more useful for research communities.

The protocols will also benefit funders. Instead of checking thousands of individual DMPs, funders can endorse the much smaller amount of disciplinary/domain/community protocols. This makes DMPs a tool that can easily be checked for compliance, rather than a bureaucratic burden.

The particularity of this Framework and of future protocols based on it, is that they are and should be developed from both a bottom-up and top-down approach:

- ▶ **Bottom-up:** it will be largely left to the research communities to design their own ways of handling their research data during and after a project (including the software that is necessary to read the data). This should be based on their needs and requirements, their experiences, their workflows, and their way of communicating about their work, data, and software. The issues of RDM and Software Sustainability would become domain-specific. The description of the relevant matters concerning data and software is documented in what will be referred to as a DDP.
- ▶ **Top-down:** By agreeing on a set of minimum requirements for DDPs, funders and research performing organisations contribute to achieve a basic level of uniformity across Europe, to achieve added-value goals regarding data openness, re-usability, and cross domain exchangeability, and to ensure that the process will start and finish within a designated period of time.

The advantages of an approach that is both bottom-up and top-down, are:

- ▶ It prevents situations where scientific domains or scholarly communities find that top-down requirements or templates for DMPs are not applicable or useful for their field or research. DMPs will be better accepted by researchers and will lead to better researcher engagement in RDM.
- ▶ The costs for processing DMPs will be reduced, as will be the burdens for funders. This will allow a stronger focus on and better assessment of deviating RDM solutions.

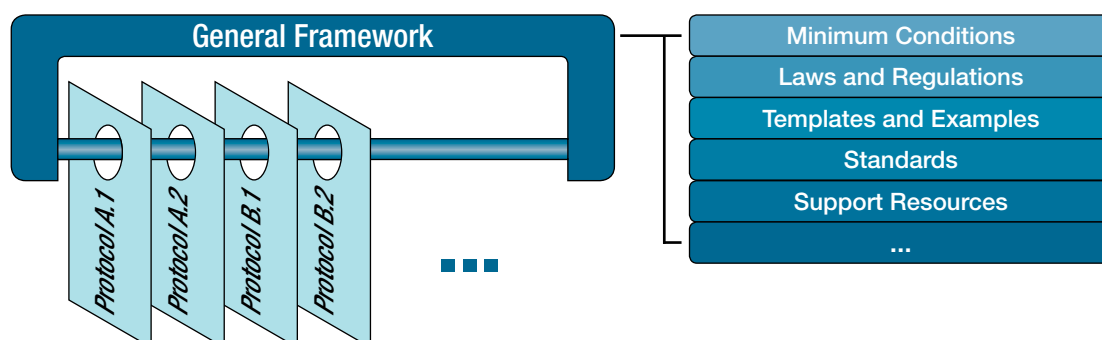


- ▶ Details regarding the actual requirements can vary according to the needs of different disciplines or research areas. Even a very generic protocol or 'model DMP' will be helpful in this regard. Protocols cannot be mandatory, but as they provide support for the parties involved, they should be as broadly applied as possible. If researchers feel that the protocol does not fit their work in their particular case, they can still decide to deviate from it, but will need to explain why they chose to do so.
- ▶ The different communities will need to decide on the level of detail that they find useful (within the margins of the Framework). There may be even alternative DDPs for different purposes (depending on size of project, type/volume of data, and so on) within one domain.

## Components of the Framework

The Framework sets the minimum conditions for each and every DDP. It contains the following elements that should be included in DDPs when appropriate:

1. Formal minimum conditions
2. Applicable laws (national and European)
3. Applicable regulations (local, national, and European)
4. FAIR Principles
5. Applicable standards
6. Templates and examples
7. Support resources



## 1. Formal Minimum Conditions

The formal minimum conditions are applicable to all DDPs and set a minimum standard for their quality and uniformity. The DDPs that conform to these minimum conditions should be generally accepted by the organisations who impose these minimum conditions. Research communities should build on these general requirements when creating DDPs with specific requirements that the research community imposes on itself.

The following list includes common minimum requirements that funding agencies and academic organisations, who use this Framework, should require as elements of RDM:

### General

1. All projects should appoint<sup>4</sup> a person who shall be referenced as the 'Data Officer'.<sup>5</sup> This person would:
  - a) act as the project's contact point to account for all data- and software-related matters that might arise during the project; when the project is finished, it should be clear whether this Data Officer remains responsible for preservation of and access to the data, or whether this responsibility is transferred to another person or institution;
  - b) participate in the selection of an existing DDP, the compilation of a basic DMP, as part of the grant application and their implementation and/or adaptation if so required during the project's lifetime;
  - c) take measures to guarantee data and software integrity, and report on any unwanted or unforeseen situations that might harm the integrity and/or the continued existence of the data or software; and
  - d) take measures to guarantee the safety of any personal or privacy-related data, and report on any unwanted or unforeseen situations or data breaches.
2. Each process leading to scientific data that are potentially relevant to the outcome of the research, should be documented at the earliest possible moment, preferably before the process actually starts.
3. Each grant application for projects involving data or software creation or adaptation should describe the nature of the data (type of data) and their estimated volume.
4. Data generation, acquisition, and retrieval should make use of standard data formats accepted in the field of research and of which publicly available descriptions exist. If, in individual cases, this is not a viable option, other formats may be used provided the reasons for this exception are well explained and the formats used are described in publicly accessible documents.<sup>6</sup>
5. If a larger volume of raw data is reduced to a smaller dataset for further research, the reduction or transformation process must be described. This includes cases where data are deleted from samples. Where data reduction or transformation is dependent on specific software, measures need to be taken to guarantee that the procedures or routines used are clearly understandable and preferably reproducible.

6. All data involved in the research behind a scientific publication need to be stored for a period to be defined by the respective communities.<sup>7</sup>
7. Along with each dataset, the background, provenance data, and metadata need to be stored as well.
8. As part of proper academic conduct, researchers should properly reference, cite, and credit re-used data and software so as to allow a future impact analysis of data and software. Principal Investigators are considered responsible for fostering this proper conduct in the course of the projects.

### **Concerning Domain Data Protocols**

1. Each DDP should require that data will be handled according to the FAIR Principles. In cases where parts of these principles cannot be applied due to legal, privacy, or other serious constraints, this must be clearly explained. Necessary metadata to ensure findability should be provided in all circumstances and in generally accepted metadata standard formats, even where the FAIR Principles cannot be otherwise applied.
2. Each DDP should require that appropriate measures are taken with a trusted party<sup>8</sup> to host the project data during the life of the project and for long-term preservation at the end of the project. This will ensure that the data are and remain secure and FAIR. Arrangements to host the data should preferably be made in writing and as early as possible.
3. DDPs should not violate any national or European laws, or formally agreed regulations. In case of the treatment of personal data, this must be mentioned explicitly in the DDP.

### **Exceptions**

1. There may be reasons for exceptions to the rules for minimum conditions, as science and technology constantly evolve. In these cases, the funder must confirm in writing that the documented exception will be accepted.

## **2. Applicable Laws and Regulations**

The following list of applicable laws (national and European), regulations (local, national, and European), and other sets of rules is not exhaustive, but gives a broad overview of regulatory aspects that need to be taken into consideration when developing a DDP.

This section contains existing laws and regulations (and laws in preparation) that provide insights in external regulatory elements that may be applicable to data collection, archiving, including privacy and security measures.

DDPs are required to respect all relevant applicable laws and regulations and help researchers take these into account.

An overview of applicable laws and regulations for data and software should be provided to researchers in the form of a webpage. This should be kept up-to-date with evolving legislation. Data Service providers or infrastructure providers could give further information to support researchers.

### Privacy

- ▶ **Personal Data Protection Acts** are present in all European countries and concern general laws regulating the protection of personal data. They are based on European Directive 95/46/EC.<sup>9</sup> This Directive will be replaced in the near future by the General Data Protection Regulation (GDPR),<sup>10</sup> which all EU Member States will have to implement in their national legislation by May 2018.
- ▶ **Obligations to Report Data Leakage Acts** are additions to the Personal Data Protection Acts. They deal with the publication of personal data and contain sanctions in the form of penalties.
- ▶ **Medical Treatment Agreement Acts** regulate the use and preservation of personal (patient) data in and for medical research.
- ▶ **Scientific Medical Research with Humans Acts** regulate scientific research in the medical field, in particular how to handle personal health-related data. These make ethical reviews compulsory for all medical research projects.

### Intellectual Property Rights

- ▶ **Copyright Acts** regulate the rights of the creator of a work. One distinguishes between exploitation rights and personal intellectual rights ('moral rights').
- ▶ The **Database Rights Act** recognises the investments made in creating and/or compiling a database. It is based on European Directive 96/9/EC.<sup>11</sup>
- ▶ **Related Rights Acts** or **Neighbouring Rights Acts** mostly refer to the rights of performers, phonogram producers, and broadcasting organisations.
- ▶ **Patent Acts** are for the protection of patents. Publication of research results (including data) is restricted during the application stage of a patent.

### Public data

- ▶ **Public Records Acts** (Public Archives Acts) oblige all public administration offices and services to preserve their documents and transfer these, after appraisal and selection, to public archives.
- ▶ **Public Sector Information Acts** (concerning re-usability of public data) are based on European Directive 2013/37/EU<sup>12</sup> that focuses on the economic aspects of the re-use of public information. It encourages Member States to make as much of this information as possible available for re-use. This also covers content held by museums, libraries, and archives, but does not apply to the educational, scientific, and broadcasting sectors.

- ▶ **Freedom of information Acts** regulate and enable citizen access to documents held by public authorities or companies carrying out work for a public authority. They do not specifically deal with access to research data.
- ▶ **Heritage Acts** are relevant for archaeological research data in so far as that they regulate ownership of documentation (data) from archaeological excavations.
- ▶ **Statistical Information Acts** regulate the competencies of the statistics authorities in data gathering as well in access to data.
- ▶ **Land Registry Acts** (cadastral information) regulate the competencies of the national land registries and access to their data, with special provisions concerning personal data contained in their various databases.

#### **Codes of Conduct/Ethical Issues**

- ▶ **Codes of Conduct**, where these exist on a national level or in an institution, should be taken into account in DMPs. They contain the general principles of good academic teaching and research.
- ▶ **Codes of Practice** for the use of personal data in scientific and scholarly research are based on the Personal Data Protection Acts<sup>13</sup> and prescribe how to handle personal data in research practice.
- ▶ **Codes of Conduct** for Medical Research regulate how researchers should handle medical personal data. They may be based on Medical Treatment Agreement Acts.

### **3. FAIR Principles**

The FAIR Principles form accepted basic rules within the research sector and are a ‘mind-set’ framework for conducting science properly and responsibly. These principles should be applied to all research involving data and/or software creation and so be included in all DDPs.

#### **The FAIR Principle essentials**

The FAIR Principles provide a guideline for those wishing to enhance the re-usability of their data holdings: these principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its re-use by individuals.

#### **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 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

### **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
- R2.** (meta)data meet domain-relevant community standards

### **FAIR implementation**

The FAIR Principles do not present readily implementable procedures, and the practicalities of their implementation and application are subject to debate in many disciplines. However, by making their inclusion a main requirement of the Framework, better re-usability across disciplines is encouraged. DDPs should take into consideration that (a) data re-use will not be restricted to users from within a community or discipline; and (b) different interpretations and implementations of the FAIR Principles exist.

## **4. Applicable Standards**

Implicit to the FAIR Principles is the use of standard formats for data and archives to enable interoperability and optimise re-usability. Likewise, software should be written in such a way that it can be maintained and re-used at minimum cost and effort.

Standardised open formats will not always be available and in some cases, the most commonly used formats may be proprietary ones. In addition, technological development may make previous standards obsolete and introduce new ones. This should not hinder scientific development, and so good reasons may exist for using new or non-standardised formats. The researcher should provide an explanation for this in the DMP.

Several organisations provide information on preferred file formats for research data (or digital objects in general), such as the UK Data Archive,<sup>14</sup> DANS (Netherlands),<sup>15</sup> and the Library of Congress (US).<sup>16</sup> They also provide supporting information that will be of use to the communities authoring DDPs. The DDPs are to refer to the applicable standard or preferred file formats for that community.

**Formatting Policy:** There are two elements to the data or file formatting policy: the introduction of preferred formats and of a list of acceptable formats:

- ▶ **Preferred formats** are file formats of which one can be confident that they will offer the best long-term guarantee in terms of usability, accessibility and sustainability. Depositing research data in preferred formats should always be accepted by data service hosts.
- ▶ **Acceptable formats** are file formats that are widely used in addition to the preferred formats, and which will remain moderately to reasonably usable, accessible and robust in the long term. Of course, the use of preferred formats is favoured, but acceptable formats are in most cases also to be accepted without discussion.

## 5. Templates and Examples

Templates and examples of DDPs will be a good starting point for developing new DDPs, and to prevent re-inventing the wheel. This should help research communities to compile DDPs more quickly. An initial, generic template that has very much in common with the approach taken by the WG, is provided in Annex A.

## 6. Support Resources

DDPs should provide a limited set of resources for support to researchers, giving background information and links to existing resources on topics that are of particular relevance to that domain or community. This will make it easier for scientists and research groups to find information on this topic.







# Proofs of Concept

---

From Different Communities



## Proofs of Concept from different Communities

Different scientific communities are already dealing with the topic of research data management (RDM). As a starting point, the Science Europe Working Group (WG) on Research Data has approached several of these communities based on existing contacts and compared their practices with its own approach for a proof of concept of its method. These proofs are based on a combination of desk research, prior knowledge about various research fields by the members of the WG, and direct feedback from selected domains and research communities. The prospects for developing DDPs are briefly described for the following areas:

| Domain   | Community                                     | Respondents  | WG Contact  |
|--|---|--|---|
| Humanities                                     | DARIAH  | Jennifer Edmond<br>Laurent Romary                            | Peter Doorn<br>DANS/NWO, The Netherlands  |
| Humanities:<br>Archaeology                     | PARTHENOS-<br>ARIADNE                         | Franco Niccolucci<br>Julian Richards<br>Andres Sparre Conrad | Peter Doorn<br>DANS/NWO, The Netherlands  |
| Linguistics:<br>Language Data                  | CLARIN  | Franciska de Jong<br>Dieter Van Uytvanck                     | Peter Doorn<br>DANS/NWO, The Netherlands  |
| Social Sciences:<br>Survey Research            | CESSDA  | Ron Dekker<br>Ivana Ilijasic Versic                          | Peter Doorn<br>DANS/NWO, The Netherlands  |
| Social and Behavioural<br>Sciences: Psychology | Psychology<br>departments and<br>associations | Sander Nieuwenhuis   | Peter Doorn<br>DANS/NWO, The Netherlands  |
| Social Sciences:<br>Ageing Studies             | SHARE, TILDA                                  | Margaret Foley<br>TILDA                                      | Patricia Clarke<br>HRB, Ireland   |
| Life Sciences:<br>Bio-informatics              | ELIXIR, FORCE11/<br>RDA FAIRSharing           | Michael Ball<br>BBSRC<br>Susanna A. Sansone<br>FAIRSharing   | Karl Gertow<br>VR, Sweden<br>Geraldine Clement-Stoneham<br>RCUK, United Kingdom |
| Plant Science                                  | ERA-CAPS (former<br>WG on RDM)                | Paul Wiley<br>Michael Ball<br>BBSRC                          | Vasco Vaz<br>FCT, Portugal  |
| Climate Research                               | ICOS  | Ari Asmi<br>Eija Juurola                                     | Jyrki Hakapää<br>AKA, Finland   |

**Disclaimer:** The following information reflects the feedback given by members of different scientific communities. It represents their personal assessment, not consolidated feedback from their respective community.

The WG approached these communities at the end of 2016 and in January 2017 with a description of the work done by the WG and asked the following questions to receive feedback on its approach to develop a Framework for DDPs:

- 1.** Do you consider the approach [of the WG] useful and feasible for your domain?
- 2.** Please take a look at the Digital Curation Centre (DCC)<sup>17</sup> general DMP template.<sup>18</sup> It is clear that this template is addressing individual researchers and is not intended for a broader community: not all elements can be described generically. There will always be individual elements in a DMP, starting with the subject/content and volume of the data (Q1.1). However, other topics, like those on specific data formats or preferred software can perhaps be answered at the level of the research community.
  - 2.1** Could you indicate which questions in the DMP template you think are not answerable for the community you represent?
  - 2.2** Do you miss particular questions, or would you want some of them to be phrased differently?
  - 2.3** Which important building blocks do you find superfluous or do you miss in the template?
- 3.** Are you willing, on behalf of your community, to compile a first draft of a generic protocol for your community?

The following structure is used to describe the responses given with respect to developing a DDP by the community:

- 1.** Short characteristic of the domain or community
- 2.** Existing situation with respect to research data management in the community
- 3.** Interest of the community in participating in the effort to develop domain protocols
- 4.** Suggestions and comments of the community on protocol elements to take into consideration
- 5.** How to proceed?

The following chapters present the responses received by the WG from the domains and research communities that were contacted for proof of concept.

# Humanities (general): DARIAH

## **1. Short characteristic of the domain or community**

DARIAH<sup>19</sup> is a distributed, pan-European infrastructure for Arts and Humanities scholars working with computational methods. It supports digital research as well as the teaching of digital research methods. DARIAH ERIC is a Landmark on the 2016 ESFRI Roadmap.<sup>20</sup> The network connects several hundreds of scholars and dozens of research facilities in currently 17 European member countries. In addition, DARIAH has several co-operating partner institutions in non-member countries, and strong ties to many research projects across Europe. People in DARIAH provide digital tools and share data as well as know-how. They organise learning opportunities for digital research methods, like workshops and summer schools, and offer training materials for the field of Digital Humanities.

## **2. Existing situation with respect to research data management in the community**

The DARIAH community works together in working groups, one of which focuses on the development of guidelines and standards (GiST). The working groups cluster around 'Virtual Competency Centres' (VCCs). The VCC 'Scholarly Content Management' deals with the various stages of the scholarly content life cycle, from creation, curation, and dissemination, through to the pooling of scholarly digital resources and results for re-use. It offers services and resources for the representation and management of data, as well as for the management of associated legal and organisational issues. It thereby aims to enhance data quality, preservation, and deep interoperability, as well as furthering a culture of data sharing in the Arts and Humanities.

Among the key infrastructure concepts contributed by this VCC are relevant standards, reference licenses, and best practice guidelines. Its products and support services address a diverse target community including Arts and Humanities data centres and research networks, as well as individual researchers.

## **3. Interest of the community in participating in the effort to develop domain protocols**

DARIAH Directors Laurent Romary and Jennifer Edmond expressed their interest in the approach of domain protocols. They "support any initiative that makes data management easier for Humanities researchers, and would be happy to look over or contribute to what Science Europe is planning to produce."

## **4. Suggestions and comments of the community on protocol elements to take into consideration**

According to Jennifer Edmond, the DCC guidelines are "useful in that they are broad and inclusive – if the proposed RDM protocol can maintain this spirit, it could be a very useful tool."

DARIAH's central efforts are currently focused on consolidating a conversation about data re-use (in the form of a DARIAH Data Re-Use Charter<sup>21</sup>) and the relationship between cultural heritage institutions and researchers. Although this approach differs from what Science Europe proposes,



<http://www.dariah.eu>

the goal is largely the same: to make access to and preservation of data important to its community easier. Edmond considered the Science Europe approach “a slightly different take on the problem, looking to facilitate widespread agreement around core values such as interoperability and reciprocal sharing, and provide more convenient ways of communicating about these issues.”

### **5. How to proceed?**

Edmond said: “It may well be that somewhere in the DARIAH ecosystem, either in a working group or in a national node, something more exactly aligned with your work is going on, but I have not been able to put a name to it in my digging (the current GiST work looks to be pointing far more at the use of specific standards and the development of a ‘standards survival kit’ than the creation of wider protocols).” She mentioned Sara di Giorgio at CLARIN,<sup>22</sup> who leads a task in the shared PARTHENOS<sup>23</sup> Cluster on research data management. Possible contacts could be DARIAH-FR, DARIAH-NL, and the GiST working group.

# Humanities – Archaeology: PARTHENOS/ARIADNE

## 1. Short characteristic of the domain/community

Over the years, archaeology has become a highly protocolled domain, especially since the acceptance of the Valletta Treaty (formally the European Convention on the Protection of the Archaeological Heritage, also known as the Malta Convention) by the Council of Europe. This 1992 treaty aims to protect European archaeological heritage “as a source of European collective memory and as an instrument for historical and scientific study.”<sup>24</sup>

Supported by (or part of) DARIAH, a succession of European Research Infrastructure projects making archaeological research data available in a sustainable way have been carried out, starting with the ARENA portal (now part of ARIADNE), CARARE,<sup>25</sup> and ARIADNE.<sup>26</sup> The PARTHENOS project incorporates and continues elements of the earlier projects in a wider Humanities and heritage context.<sup>27</sup>

## 2. Existing situation with respect to research data management in the community

In most European countries, almost all archaeological information, either during fieldwork, afterwards, or both, is recorded digitally. In the Netherlands, digital work in archaeology is promoted by the specification of a Quality Standard for Dutch Archaeology (KNA),<sup>28</sup> which explicitly requires that basic information on each project should be transferred in a uniform way to appointed physical and digital depots. These and other specifications are maintained by SIKB,<sup>29</sup> a network organisation in which the private and the public sector strive to continuously and structurally enhance the standards of activities relating to soil management in the Netherlands.<sup>30</sup> To support the digital deposit of archaeological find material, a validation tool is available, allowing the excavator to monitor the correctness of files for deposit automatically. The tool monitors the use of the required fields and the codes of the various domain tables, but it has no control over the archaeological accuracy of the file contents. A reporting tool helps the excavator or the custodian to consult the contents of the digital delivery note.

The E-Depot for Dutch Archaeology (EDNA)<sup>31</sup> was established by DANS<sup>32</sup> and the Cultural Heritage Agency (RCE)<sup>33</sup> to archive digital research data of Dutch archaeologists in a sustainable manner and to make them available for re-use. Since 2007, the KNA obliges archaeologists to deposit and store their digital data for re-use in the archiving system of DANS. EDNA contains data of archaeological research (GIS data, field drawings, data tables, photographs) and the final reports on this research. This concerns research in the broadest sense: from field survey to excavation, from specialist research to dissertation. The archived reports and datasets can be found per archaeological organisation or per specific project, and are accessible for other scientists. More than 80% of the archaeological data in EDNA is publicly accessible.



<http://www.parthenos-project.eu>



<http://www.ariadne-infrastructure.eu>

In England and Scotland, almost all regions require archaeological work to be reported via OASIS,<sup>34</sup> an online form whereby fieldwork reports are uploaded and made available through Open Access by the Archaeology Data Service (ADS).<sup>35</sup>

### **3. Interest of the community in participating in the effort to develop domain protocols**

According to the ARIADNE and PARTHENOS co-ordinator Franco Niccolucci, the invitation to develop protocols for data management in the Humanities in general and for archaeology in particular “is a very important initiative. The approach is correct and useful, possibly the only way of achieving good results. [...] This is something of great interest for all [Humanities] researchers.” Julian Richards, Director of the Archaeological Data Service (ADS) in the UK and deputy coordinator of ARIADNE, agrees that “this is a valuable development; it is important that such domain protocols are harmonised at an international level.”

### **4. Suggestions and comments of the community on protocol elements to take into consideration**

In addition to what is already mentioned above about Dutch protocols, in the UK the ADS has been using the DCC template for archaeological data management planning. It has been a requirement of funding councils for several years, and many institutions have adapted it for their own purposes. The DMP online tool is also very useful and has been adapted for archaeology requirements. Specific training materials for data management in archaeology were developed by ADS in collaboration with Cambridge University library.<sup>36</sup>

### **5. How to proceed?**

In the context of the PARTHENOS project, work on a template for a Humanities DMP has already started. This is taking place in direct consultation with DARIAH and the template will have special clauses or requirements for archaeologists. The draft template will line up with the Science Europe invitation to develop a protocol for Humanities RDM in the context of PARTHENOS/DARIAH and an ARIADNE-inspired one for archaeology.

# Linguistics – Language Data: CLARIN

## 1. Short characteristic of the domain or community

CLARIN (Common LAnguage Resources and technology INfrastructure) makes digital language resources available to scholars, researchers, students and citizen scientists from all disciplines, especially in the Humanities and Social Sciences. CLARIN offers long-term solutions and technology services for deploying, connecting, analysing and sustaining digital language data and tools. CLARIN, which became an ESFRI ERIC in 2012, supports scholars who want to engage in cutting-edge data-driven research, contributing to a truly multilingual European Research Area.

The CLARIN mission is enabled through a networked federation of centres<sup>37</sup> that are fully operational in many European countries: language data repositories, service centres, and knowledge centres with single sign-on access for all members of the academic community in all participating countries. Tools and data from different CLARIN centres are interoperable, so that data collections can be combined and tools from different sources can be chained to perform complex operations to support researchers in their work.

## 2. Existing situation with respect to research data management in the community

In order to provide its services well, CLARIN has defined strict requirements for its centres.<sup>38</sup> Specifically, the Service Providing Centres, or CLARIN B-Centres, have to comply with stable technical and institutional criteria. Most of these comply with the Data Seal of Approval,<sup>39</sup> demanding the sustainability of data storage and long-term accessibility. Given its experience with formulating requirements for the centres, CLARIN seems well-positioned to formulate a protocol for data management for the researcher community, primarily in the area of language data.

In the US, the Linguistic Data Consortium (LDC) is an open consortium of universities, libraries, corporations and government research laboratories, hosted by the University of Pennsylvania.<sup>40</sup> It was formed in 1992 to address the critical data shortage then facing language technology research and development. Initially, LDC's primary role was as a repository and distribution point for language resources. Since that time, and with the help of its members, LDC has grown into an organisation that creates and distributes a wide array of language resources. LDC also supports sponsored research programmes and language-based technology evaluations by providing resources and contributing organisational expertise, and offers expertise in data management.





### **3. Interest of the community in participating in the effort to develop domain protocols**

CLARIN Directors Franciska de Jong (Executive Director) and Dieter Van Uytvanck (Technical Director) expressed their willingness to co-operate with Science Europe in the formulation of a protocol in the domain of language data.

### **4. Suggestions or comments of the community on protocol elements to take into consideration**

The subjects in the generic DCC template for DMP were deemed relevant for the CLARIN community.

### **5. How to proceed?**

The Executive and Technical Directors of CLARIN will co-ordinate the formulation of a protocol for the area of language data. A possible link with the LDC in the US mentioned above will be explored.

# Social Sciences – Survey data: CESSDA

## 1. Short characteristic of the domain/community

A large part of the social sciences uses quantitative data and data from surveys, registries, internet panels, which may include context data (geographic data, historic data) or Life Science data (genome sequences). Qualitative data also play an important role and have different setups and structures. For both types of data, the subjects are often individual respondents and questions include sensitive data and/or might reveal private information, despite anonymisation. This puts high standards on the secure (re-)use of the data. The data can be very large (registries) and very complex (multi-level, longitudinal panels) and this requires good data descriptions for new users.

Several pan-European research infrastructures support this domain, among which the Consortium of European Social Science Data Archives (CESSDA)<sup>41</sup> is the most generic. CESSDA provides large-scale, integrated and sustainable data services to the Social Sciences. It brings together Social Science data service providers across Europe, with the aim of promoting the results of Social Science research and supporting national and international research and co-operation. CESSDA is one of ESFRI's Landmark Infrastructures on its 2016 Roadmap, and has become an ESFRI ERIC in 2017. Norway is host to CESSDA and its main office is located in Bergen.

## 2. Existing situation with respect to research data management in the community

There are well-founded data collection methods and there is a long-standing tradition among social science researchers to use standardised ways of documenting survey data files. The data archives supporting the domain have always played an important role in setting or supporting standards, such as the Data Documentation Initiative (DDI)<sup>42</sup> and widely accepted practices of making codebooks, preferred formats for data storage and exchange, and so on.

The CESSDA consortium also provides training on RDM and DMPs. In addition, CESSDA service providers offer online services, documents, webinars, and tutorials to support digital preservation, data archiving, and data sharing.

## 3. Interest of the community in participating in the effort to develop domain protocols

The Social Sciences are rather heterogeneous, consisting of multiple communities. Data producers are organised along the various disciplines or research domains. Furthermore, data users might also be organised by societal challenges.

In any case, DMPs should offer tools and help data producers to construct their metadata (information at general level, data structure, and so on, up to variables and values levels). For users, the DMPs should be informative, offering search filters, and be helpful to quickly scan the relevance of the data. Ideally, DMPs are machine-readable, in order to be usable in text and data mining.

For CESSDA, the Science Europe approach with its principles and protocols can be very helpful in quality control. For example: if data producers follow the protocol(s), then the data repositories can process their data efficiently and confirm to funders that the researcher complies with their grant regulations.

#### **4. Suggestions and comments of the community on protocol elements to take into consideration**

The following suggestions and comments were provided by CESSDA:

- Evidently, not one size fits all – even within Social Sciences there can be different templates.
- Other, already existing templates by different research funders might be useful for a protocol for social survey data. The way forward is to check whether these comply with the Science Europe principles and protocols, and support data producers ('no red tape').
- Add the affiliation (institute) of the Data Officer (to have an entry if the person has left).
- Request information on data provenance, and on the way (or extent) that the research data are reproducible from the raw/original data.
- Include information on mode of access: because of the sensitivity of the data, different modes for accessing the data may apply. This may include compliance with (disciplinary) Codes of Conduct.
- Work on international scale, for example by including the ICPSR (Interuniversity Consortium for Political and Social Research)<sup>43</sup> DMP approach.
- Missing/superfluous/alternative elements:
  - Ask for information about type of study.
  - Include questions about relevant policies that the survey should comply with on institutional/national/EU level.

#### **5. How to proceed?**

As a follow-up, CESSDA can elaborate on DMPs for Social Science data (especially for quantitative data), based on the principles and protocols from this Science Europe Framework document.

CESSDA could set up pilots with data provider communities and data users to test usability of these DMPs, as well as set up training programmes on how to set up and use DMPs during the research process.

# Social/Behavioural Sciences – Psychology

## 1. Short characteristic of the domain or community

Within the social and behavioural sciences, psychology is a large and varied field that over time has become a very data-intensive science. A wide range of research methods is used in psychology. These methods vary by the sources of information that are drawn on, how that information is sampled, and the types of instruments that are used in data collection. Methods also vary by whether they collect qualitative data, quantitative data, or both.

As the study of human behaviour and the mind, seeking to understand the role of mental functions in individual and social behaviour, while also exploring the physiological and biological processes that underlie cognitive functions and behaviours, the data collected and analysed in psychology almost by definition relate to human subjects. Because of the privacy sensitivity of a lot of psychological research, ethics codes and legal requirements traditionally play an important regulatory role in the field.

In recent years, the discipline was unsettled by a couple of high-profile cases of scientific misconduct and by allegations of a replication crisis, arguing that many findings in the field cannot be reproduced. Focus on these issues has led to more attention to data management practices and renewed efforts in the discipline to re-test important findings.

## 2. Existing situation with respect to research data management in the community

Record Keeping Guidelines were published by the American Psychological Association (APA) in 2007.<sup>44</sup> These guidelines were “designed to educate psychologists and provide a Framework for making decisions regarding professional record keeping.” Many recommendations relate both to paper and electronic records. Guideline 9 is dedicated to digital data: “Electronic records, like paper records, should be created and maintained in a way that is designed to protect their security, integrity, confidentiality, and appropriate access, as well as their compliance with applicable legal and ethical requirements.”

More recently, faculties of psychology at many universities have formulated more detailed protocols, guidelines, recommendations, templates, and regulations for data management. This is certainly the case in the Netherlands after the ‘Stapel Affair’,<sup>45</sup> where ‘data storage protocols’ are now in place in most faculties.

## 3. Interest of the community to participate in the effort to develop domain protocols

The above shows a relevance, and indeed a need, in the field of psychology to develop protocols for data management that are more detailed than the ethical codes and cover more ground than data storage. So far, only the psychologists at Leiden University have been contacted by the WG, who have responded with interest.

#### **4. Suggestions and comments of the community on protocol elements to take into consideration**

According to Professor Sander Nieuwenhuis, responsible for the 'Data Storage Protocol Psychology' of Leiden University, the list of topics in the default DMP template of the DCC seems "fairly complete." The question "What quality assurance processes will you adopt?" from the DCC template needs further clarification. He also referred to the DMP template developed at Leiden University, that was further elaborated to satisfy the needs of the Department of Psychology, and that all psychology students need to fill out.

Other protocols and guidelines have been found and seem useful for the formulation of a DMP protocol for psychology:

- Guidelines on research data management, Faculty of Psychology and Educational Sciences, Ghent University (November 2016)<sup>46</sup>
- Record Keeping Guidelines, American Psychological Association (2007)<sup>47</sup>
- Data management plan of the Social and Organizational Psychology (SOP) department, Utrecht University (draft, June 2013)<sup>48</sup>
- Oxford Libguide on Psychology – Managing your research data<sup>49</sup>
- Guidance on Human Subjects Research Data Storage & Retention, University of Delaware<sup>50</sup>
- Data storage protocol of the Heymans Institute for Psychological Research, University of Groningen<sup>51</sup>
- Data Storage Protocol Psychology, University of Amsterdam (December 2014)<sup>52</sup>

#### **5. How to proceed?**

Possible ways forward are:

- to build on the Dutch (draft) protocols and to ask an ad hoc working group of psychologists involved in the preparation of faculty/national protocols, templates and guidelines to draft an international protocol for the whole discipline.
- to request the support of the APA mentioned above, and of the European Federation of Psychologists' Associations (EFPA).<sup>53</sup> EFPA provides a forum for European co-operation in a wide range of fields of academic training, psychology practice and research. There are 36 member associations of EFPA representing about 300,000 psychologists. EFPA has produced many documents on ethics, among which a model code, which is however extremely brief on research data.
- to reach out to Associations for sub-disciplines of psychology, such as the European Association of Social Psychology (EASP) and/or the European Association of Developmental Psychology (EADP), which both have the aim to promote excellence in European research in their respective fields.

# Social Sciences – Family of Studies on Longitudinal Ageing

## 1. Short characteristic of the community

There is a growing international family of studies on Longitudinal Ageing that manage cross-nationally comparable population data. These data tackle issues such as health, disability, retirement, active ageing, family and social support.

SHARE-ERIC (Survey of Health, Ageing and Retirement in Europe) is a multidisciplinary and cross-national Research Infrastructure responsible for a European Longitudinal Ageing Survey. This study examines the different ways in which people aged 50 and older live in 28 European countries, making it the largest pan-European Social Science panel study. More than 293,000 interviews with approximately 123,000 individuals have been collected since 2004. SHARE is centrally coordinated by the Munich Center for the Economics of Aging (MEA), part of the Max Planck Institute for Social Law and Social Policy. Researchers can access the data collected and generated in the SHARE projects free of charge through the SHARE Research Data Center.<sup>54</sup>

One of its 'sister studies' is The Irish Longitudinal study on Ageing (TILDA) that analyses the process of population ageing in-depth within Ireland. TILDA involves interviews on a two-yearly basis with a sample cohort of 8,000+ Irish residents aged 50 and over, collecting detailed information on all aspects of their lives, including the economic (pensions, employment, living standards), health (physical, mental, service needs and usage), and social aspects (contact with friends and kin, formal and informal care, social participation). An additional cohort study, the Intellectual Disability Supplement to The Irish Longitudinal Study on Ageing (IDS-TILDA) directly compares the ageing of people with intellectual disability (aged 40 and over; sample size 750+) with the general ageing population.

A description of TILDA data generated using the tools and methods used to generate TILDA data is available online.<sup>55</sup> The TILDA dataset incorporating all core results is made available via the Irish Social Sciences Data Archive held at University College Dublin<sup>56</sup> and the Interuniversity Consortium for Political and Social Research (ICPSR) within the Institute for Social Research at the University of Michigan.<sup>57</sup>

SHARE and TILDA are both embedded in a global network of harmonised ageing studies. The Gateway to Global Aging Data RAND US,<sup>58</sup> sponsored by the National Institute on Aging, facilitates cross-national comparative studies on aging using the entire family of health and retirement surveys around the world.



<https://tilda.tcd.ie>

Harmonised Data Files for cross-country analysis are available from:

- RAND HRS (*Health and Retirement Study*)
- ELSA (*English Longitudinal Study of Ageing*)
- SHARE (*Survey of Health, Ageing and Retirement in Europe*)
- KLoSA (*Korean Longitudinal Study of Aging*)
- JSTAR (*Japanese Study on Aging and Retirement*)
- CHARLS (*China Health and Retirement Longitudinal Study*)
- LASI (*Longitudinal Aging Study in India*)
- MHAS (*Mexican Health and Aging Study*)
- TILDA (*The Irish Longitudinal Study of Aging*)
- CRELES (*Costa Rican Longevity and Health Aging Study*)

## **2. Interest of the community in participating in the effort to develop domain protocols**

A request has been issued to the International SHARE community to contribute to the development of domain protocols; TILDA is a national study, and not an equivalent community of scale to others identified for participation in domain protocols. Contact with TILDA has been on examining its working processes for research data management and its international connections.

SHARE has responded that they consider the Science Europe WG “useful” and deem its initiative “a very interesting project.” Nevertheless, SHARE declined the invitation to collaborate, prioritising their core activities: data collection and research, as well as their efforts to extend in order to include 28 countries in wave 7 of their extension, and in combination with some structural workforce re-organisations at the Max Planck Institute. “The improvement and efficiency of Data Management Plans is, of course, important to us”, SHARE said, “but nevertheless a secondary activity for which we have no capacity at the moment.” SHARE regretted that they could not participate. The TILDA national study was happy to engage.

## **3. How to proceed?**

As SHARE is more an international survey than an infrastructure, a possible way forward is to collaborate with CESSDA. If a protocol for survey research data is drafted, the feedback of SHARE can be asked.

# Life Sciences – Bio-informatics: ELIXIR and Force11/RDA FAIRSharing

## 1. Short characteristic of the community

The Life Science community is becoming increasingly more data-intensive and the need for all aspects of data management and bioinformatics are increasing in parallel. Large-scale datasets, for example those generated through next-generation sequencing, proteomics, expression analyses, and so on, are being produced and used in a range of areas such as medical sciences, veterinary sciences, marine sciences, plant sciences and agriculture, and environmental sciences.

ELIXIR is an intergovernmental organisation that since its establishment in 2013 brings together data-related Life Science across Europe. Resources include databases, software tools, training materials, cloud storage, and supercomputers. ELIXIR was considered as an ESFRI Landmark in its 2016 Roadmap.

The goal of ELIXIR is to co-ordinate these resources so that they form a single infrastructure. This infrastructure makes it easier for scientists to find and share data, exchange expertise, and agree on best practices. By co-ordinating resources, ELIXIR ensures that users – individual scientists, large consortia, or other research infrastructures – can easily access data resources that are sustainable, built on strong community standards, and safeguarded in the long term.

ELIXIR follows a Hub-and-Nodes model, with a single Hub located alongside EMBL-EBI at the Wellcome Genome Campus in Hinxton (Cambridge, UK) and a growing number of Nodes located at centres of excellence throughout Europe, which co-ordinate nationally the bioinformatics services within that country.<sup>59</sup>

FAIRSharing<sup>60</sup> is a curated, informative and educational resource on the inter-related data standards, databases, and policies in the Life, Environmental and Biomedical Sciences. Operating since 2011 and run by an operational team at the University of Oxford, FAIRSharing is driven by an international advisory board, collaborates with US National Institutes of Health (NIH)'s 'Big Data to Knowledge' (BD2K) Initiative, and has recently become an ELIXIR-UK Node resource part of the ELIXIR interoperability platform. FAIRSharing is endorsed by a community of 68 organisations, including publishers (embedded in the data policies of 600 Springer Nature's journals, also PLoS, EMBO press, BMJ, F1000Research, BioMedCentral, Oxford University Press, and Wellcome Trust Open Research<sup>61</sup>), standardisation groups, and research data management support initiatives and libraries (such as those at JISC, Stanford, Cambridge and the Oxford Universities). FAIRSharing also operates as an open working group under Force11 and the Research Data Alliance<sup>62</sup> and has recently released its recommendations, signed by a number of adopters.<sup>63</sup>

## 2. Existing situation with respect to research data management in the community

ELIXIR's Nodes, sited throughout the current 21 ELIXIR Member States, run the resources and services that are part of ELIXIR. These include: data deposition resources for depositing data safely and securely; added-value databases providing researchers with access to well-curated data; bio-compute centres for cloud computing and analysis; services for the integration of data, software, tools and resources; training; and standards, ontology and data management expertise. For example, BioTools,<sup>64</sup> the ELIXIR Tools and Service registry, is a discovery portal for researchers





<https://www.elixir-europe.org>



<https://fairsharing.org>

to access over 2,100 Life Science databases and analysis tools. Likewise, TeSS,<sup>65</sup> ELIXIR's training portal, provides users with access to hundreds of training courses, events, and online training materials, including for DMP training.

ELIXIR has compiled a list of databases that it recommends for the deposition of experimental data.<sup>66</sup> The purpose of this Deposition Databases list is to provide guidance to those who formulate policy and working practices on the appropriate repositories for publishing open data in the Life Sciences, and to help those generating data to store these in the appropriate archive. For example, for the deposition of protein structure data, PDBe acts as the recognised Deposition Database, whilst for raw sequence data, the European Nucleotide Archive (ENA) is the recognised database.

Many ELIXIR Nodes provide data management support to scientists and Life Science research projects within that country. For example, the Swiss Institute of Bioinformatics (SIB) runs an 'embedded bioinformatician' scheme, where research projects can apply for data management expertise from one of these experts and charge the costs to their grant. ERA-NET projects in the Life Sciences are increasingly advising the research consortia they fund to use ELIXIR services. For example, the E-RARE project<sup>67</sup> has dedicated text on its website that encourages applicants to use ELIXIR services for their data management needs, whilst the ICPeMed<sup>68</sup> and TRANSCAN-2<sup>69</sup> ERA-NETS will include reference to these in their forthcoming Guide for Applicants.

In ELIXIR, FAIRSharing is the metadata and standards registry; this is a much-needed resource for RDM, because in the broad Life, Environmental and Biomedical Sciences, over a thousand of reporting guidelines (or checklists), models/formats and terminologies exist. Collectively known as content standards for data and metadata, these ensure that the information is reported consistently, efficiently and meaningfully. Content standards open datasets to transparent interpretation, verification, exchange, integrative analysis, and comparison, supporting the FAIR principles. The general context on interoperability standards, the types and variety and their role in RDM is summarised by a review commissioned by the Wellcome Trust.<sup>70</sup>

FAIRSharing is both an informative and an educational resource part of the RDM toolkit each researcher needs. As an informative resource, FAIRSharing ensures that these content standards are findable and accessible. As an educational resource, FAIRSharing works to provide the indicators necessary to monitor the development, evolution and integration of standards. By interlinking standards, databases, and data policies (from funders, journals, and other organisations), FAIRSharing guides users to discover those standards that are implemented by databases, and to find the policies that refer to them, providing evidence of use and other important indicators that users take into consideration when selecting a resource. Working with and for researchers, developers, curators, funders, journal editors, librarians, and data managers, FAIRSharing helps producers of standards (databases and policies) to ensure their resources are findable by prospective users, and enable consumers to make an informed decision as to which standard (database or policy) to (re-)use or endorse.

### **3. Interest of the community to participate in the effort to develop domain protocols**

The ELIXIR representative was positive regarding possibilities to produce a reflection on the DCC guidelines from an ELIXIR perspective. However, the representative appeared hesitant/unconvinced regarding a 'generic protocol' for the community, except for at the very highest level.

The FAIRSharing representative was very interested in participating and FAIRSharing is already undertaking activities towards this end. First, it is in discussion with the DMPOnline/DMPTool teams to ensure that – as these tools are enhanced – a 'FAIRSharing look-up service' is created for researchers, data managers, curators, and so on, operating in the Life, Environmental, and Biomedical Sciences. This look-up functionality and the FAIRSharing inter-linked content (of standards, databases and policies) would help researchers to for example find, cite in their DMP, and ultimately use the most appropriate content standards to annotate their datasets and/or the most relevant databases for their data type, knowing which content standards these databases require. Second, the content standards (reporting guidelines, models/formats, and terminologies) in FAIRSharing can be considered as an actual DDP, focused on deeper and specific metadata to cover the what, who, when, where, how, and why of a dataset. This is because the content standards can be seen as templates to help report and describe all elements of a dataset, for example the fundamental biological entities (such as samples, genes, cells), the experimental components (such as conditions, cell lines), but also complex concepts (such as bioprocesses, tissues and diseases), the analytical process, and the mathematical models.

### **4. Suggestions and comments of the community to take into consideration (elements in a protocol to pay attention to for this domain)**

ELIXIR covers a broad set of scientific disciplines in the Life Sciences, so identifying individual elements of protocols to consider over others may be difficult. That said, ELIXIR supports many data repositories, so as a starting point it is worth considering that researchers may be interested in submitting their data to repositories at the end of their research.

FAIRSharing work and experience highlight the following consideration: (i) with over a thousand content standards and thousands of databases, curating and interlinking their descriptions and understanding their maturity is a lengthy process, especially as both the coverage and status of these resources must be verified with their respective communities; (ii) the use of content standards in DDPs should be made 'invisible' to the researchers, and this is not a trivial task. To address the latter point, FAIRSharing is interested to deliver methods, tools and practices to create content standards-based templates for describing datasets smarter and faster; some research activities in this directions are already being undertaken.

### **5. How to proceed?**

ELIXIR has a Data Management Plan Working Group that aims to:

- Identify existing resources for Data Management training
- Make new materials for Data Management training
- Advertise ELIXIR data management expertise to the research community
- Help researchers get the most out of their data with the least risk

Contacting the Working Group and FAIRSharing is a possible way forward.

# Plant Sciences: ERA-CAPS

## 1. Short characteristic of the domain or community

ERA-CAPS (the ERA-Net for Co-ordinating Action in Plant Sciences) is a self-sustained network that, through the aggregation of the scientific and economic capabilities of network members, focuses on the co-ordination of sustainable transnational plant science research programmes.

It comprises nine partners from eight European countries and the USA, and 11 observers (10 of them funding organisations) and is co-ordinated by the Biotechnology and Biological Sciences Research Council (BBSRC, United Kingdom).

The main objective is to develop a common agenda and shared vision for plant science research across the European Research Area (ERA) and create a joint research programme. ERA-CAPS also aims to facilitate data management, access and sharing solutions.

## 2. Existing situation with respect to research data management in the community

Given the extremely wide research scope of ERA-CAPS and its resulting variety of data outputs and practices (including the fact that a good amount of plant science data is of non-digital nature), there were practical difficulties to establish a set of standards, which it was intended to be developed during its phase as an FP7 project. Instead, ERA-CAPS promoted the creation of an Expert Working Group of plant scientists to identify the major data issues that face the different plant science communities and to develop a roadmap giving possible solutions to the problems that could be taken forward by funders at the national, European and international level. The finished roadmap document acknowledged that good data stewardship, along suitable data standards, may multiply the data's scientific and societal impact.

In March 2014, a data-sharing policy common to all ERA-CAPS funding partners was adopted. From that point on, the policy was incorporated into the conditions of grants awarded through the ERA-CAPS joint programme and applied to the data generated by the funded research projects. The sharing of raw data and associated metadata facilitates the re-use, reintegration, and repurposing of these data.

The aim of ERA-CAPS was not to replicate existing policies, but rather to consolidate and identify best practice from these policies while still looking for underlying common principles that could be used to frame ERA-CAPS's own policy.

A few common principles have surfaced, such as the view that publicly funded research data are a public good and should be openly available, the importance of interoperability, the need to respect subject specific standards, and the need to abide by ethical and legal standards.

# Plant Sciences: ERA-CAPS

### **3. Interest of the community to participate in the effort to develop domain protocols**

The ERA-CAPS project co-ordinator, Dr Paul Wiley, was approached to gauge the potential interest of ERA-CAPS to develop a domain protocol for plant sciences, building on the work previously done within ERA-CAPS regarding data management requirements and the establishment of commonly accepted data standards. He supported the initiative as highly interesting, welcoming “anything that brings the importance of having data management plans at the beginning of the planning stage for research, and making this easier for researchers.” He also considered the described approach useful and an evolution of what was done with the ERA-CAPS calls data management requirements. However, the Expert Working Group on data standards was only active while the network was funded by the European Commission and completed their work in 2015. Thus, there is no longer any data expert group or person acting within the ERA-CAPS backdrop that could fill the role as potential point of contact to represent the community for data-related issues. There is also no longer any specific work package or task related to data management topics. He also commented it would be “difficult to generate something that serves the whole plant community, as this is very broad (for example the data generated by a plant ecologist would be very different from those involved in high-throughput phenotyping, and different again from those involved in genomics – each would have different standards, repositories, and so on, and at different levels of maturity).”

### **4. Suggestions/comments of the community on protocol elements to be taken into consideration**

Dr Michael Ball, data expert at BBSRC, was asked to comment on the applicability and usefulness of the DCC general DMP template for the management of plant sciences-related data outputs.

As a general comment, Dr Ball noted that the overall guidance should establish the clear principles in such a way that they can be applied flexibly to different projects, which may have particular nuances about how they operate.

He referred to the DCC DMP template as having most of the correct questions on it. However, he states that it could be clarified and made a little more flexible. To that effect, he additionally suggested:

1. A (very brief) introductory text outlining the principles – these are encapsulated in the FAIR Principles
2. Dividing the principles into
  - a) those that are required to share and re-use the data (for example formats, volumes, metadata, restrictions); and
  - b) those that are required to manage the data (for example versioning, responsibilities for management, long term archiving/storage)



3. Controlling the length by having decision points that will ask deeper questions where relevant, but allow these to be skipped if not – for example, if there are no restrictions on data sharing, then a number of questions related to this can be skipped. This would be easier in an online/ electronic form.

He also suggested to have a series of generic questions that prompt data planning decisions (for example consent, usage, responsibilities) and then focus on data types. Many of the decisions regarding data will flow from the type of data being used and generated (for example format, standards, suitable places to deposit and store, community norms, metadata, and so on).

#### **5. How to proceed?**

Given that there was no longer any potential point of contact to represent the ERA-CAPS research community concerning data management topics, Dr Wiley suggested that further developments towards the establishment of plant sciences data management protocols would best be taken by active researchers in this field. Accordingly, he introduced the Science Europe WG and its objectives to some key members of the former Expert Working Group who could be potential partners for next steps.

# Climate Research: ICOS

## **1. Short characteristic of the domain or community**

There is an ample amount of research and data on climate change. However, scientists found that the data was often scattered, difficult to reach and to access. The quality and consistency of measurements was not guaranteed. Nor did the data turn into information that could easily be used by citizens and decision makers.

Integrated Carbon Observation System Research Infrastructure (ICOS RI) is a pan-European Research Infrastructure that provides data enabling analyses of emissions and sinks of greenhouse gases, ecosystem function, and related research in Europe and adjacent key regions of Africa and Eurasia. The backbones of ICOS RI are the national measurement stations and networks of ICOS atmospheric, ecosystem and ocean stations that have the specific tasks in collecting and processing the data.

## **2. Existing situation with respect to RDM in the community**

Since 2008, ICOS RI has brought together the European high-quality national research and measurement stations and, through co-ordination and support, constitutes a European-wide research infrastructure that serves both the scientists and the society.

ICOS Carbon Portal (CP) is part of ICOS ERIC and offers access to research data, as well as easily accessible and understandable science and education products. It is a virtual ICOS data centre from where ICOS data and ancillary data sets will be published and be accessible for the users. The CP is responsible for handling and providing ICOS data products. All measurement data available in the CP is quality-controlled through the ICOS thematic centres, divided into Ecosystem, Atmospheric and Ocean Thematic Centres and a Central Analytical laboratory.

The CP is a data platform: a 'one-stop shop' for all ICOS data products. As such, it is envisioned as a virtual data centre: a place where all relevant ICOS data and ancillary data sets from external sources will be published and be accessible through the facilities of the CP. All types of ICOS data need to be easily discoverable, accessed, visualised, and available for further analysis by all interested parties. In addition, provisions shall be made by ICOS to provide standardised and comprehensive synthesis products that summarise the ICOS data, for example on annual and seasonal basis.

## **3. Interest of the community to participate in the effort to develop domain protocols, and suggestions and comments of the community on protocol elements to be taken into consideration**

The two representatives displayed an interest in contributing to the endeavour to formulate domain data protocols. The Science Europe WG received references to the ICOS Data Policy in which the general ICOS data policy principles are described,<sup>71</sup> and the 2014–2015 Progress Report that includes data management issues.<sup>72</sup> These documents display a keen interest by the community



<https://www.icos-ri.eu>

to regulate its data management practices. The ICOS data policy already contains many elements described in the terms of reference outlined by the Framework for domain data protocols, including sections on legal aspects, data processing, archiving, IPR, data attribution and citation, and licensing.

#### **4. How to proceed?**

Possible ways forward include exploring what is necessary for the ICOS Data Policy to be developed into a domain protocol.





# Notes and References

1. Science Europe Briefing Paper 'on Research Integrity: What it Means, Why it Is Important and How we Might Protect it' (December 2015): <http://scieur.org/integrity-paper>
2. <https://dans.knaw.nl/en>
3. <https://www.force11.org/group/fairgroup/fairprinciples>
4. Appoint here does not mean such a person needs to be explicitly employed for this goal. It may just as well be any person, somehow involved in the project, up to the very Principal Investigator, as long as it is clear that (s)he carries the responsibility for that function.
5. Or alternatively 'Data Steward' or 'Data Custodian'.
6. Several organisations maintain lists of preferred formats for data preservation and sharing. See, for example: Data Archive – <http://www.data-archive.ac.uk/create-manage/format/formats-table>; DANS – <https://dans.knaw.nl/en/deposit/information-about-depositing-data/DANSpreferredformatsUK.pdf>; and Library of Congress – <https://www.loc.gov/preservation/resources/rfs/index.html>
7. In many cases, this period is defined as at least ten years. However, it has to be decided what period is scientifically and/or socially appropriate. There are cases where it has to be indefinite.
8. This may be an institutional or other designated repository complying with minimum quality standards and guarantees for data security.
9. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML>
10. [http://ec.europa.eu/justice/data-protection/reform/files/regulation\\_oj\\_en.pdf](http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf)
11. Directive 96/9/EC of the European Parliament and of the Council on the legal protection of databases: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML>
12. Directive 2013/37/EU of the European Parliament and of the Council amending Directive 2003/98/EC on the re-use of public sector information: <http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32013L0037&from=EN>
13. General Data Protection Regulation: [http://ec.europa.eu/justice/data-protection/reform/files/regulation\\_oj\\_en.pdf](http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf)
14. <http://www.data-archive.ac.uk/>
15. <https://dans.knaw.nl>
16. <https://www.loc.gov/>
17. Digital Curation Centre: an internationally recognised centre of expertise in digital curation with a focus on building capability and skills for research data management. The DCC has published a template for RDM.
18. This template was attached to the communication and is reproduced in Annex A. The online version can be found at <https://dmponline.dcc.ac.uk/>
19. <http://www.dariah.eu/>
20. [http://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=esfri](http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri)
21. <https://digitalintellectuals.hypotheses.org/3031>
22. <https://www.clarin.eu/>
23. <http://www.parthenos-project.eu/>
24. <https://rm.coe.int/168007bd25>
25. <https://pro.europeana.eu/project/carare>, now part of Europeana.
26. <http://www.ariadne-infrastructure.eu>
27. <http://www.parthenos-project.eu/>
28. <https://downloads.arqueo-ecuatoriana.ec/ayhpwxgv/estandares/ArqueoHolandia.pdf>
29. <https://www.sikb.nl/>
30. <http://www.sikb.nl/archeologie/pakbon-en-sikb0102>
31. <https://dans.knaw.nl/en/about/services/archiving-and-reusing-data/easy/edna>



32. <https://dans.knaw.nl/en>
33. <https://cultureelerfgoed.nl/>
34. <http://oasis.ac.uk/pages/wiki/Main>
35. <http://archaeologydataservice.ac.uk>
36. <http://archaeologydataservice.ac.uk/learning/DataTrain/>
37. <https://www.clarin.eu/content/clarin-centres>
38. <http://hdl.handle.net/1839/00-DOCS.CLARIN.EU-77>
39. <https://www.dataeofapproval.org/en/assessment/>
40. <https://www ldc.upenn.edu/>
41. <https://cessda.net>
42. <http://www.ddalliance.org/>
43. <http://www.icpsr.umich.edu/icpsrweb/index.jsp>
44. <http://www.apa.org/practice/guidelines/record-keeping.aspx>
45. <http://www.sciencemag.org/news/2012/11/final-report-stapel-affair-points-bigger-problems-social-psychology>
46. <http://www.ugent.be/pp/en/research/rdm>
47. <http://www.apa.org/practice/guidelines/record-keeping.aspx>
48. [https://www.surf.nl/binaries/content/assets/surf/nl/kennisbank/2013/UU\\_Data+Management+Plan+SOP\\_DRAFT+June+2013\\_publ\\_def.pdf](https://www.surf.nl/binaries/content/assets/surf/nl/kennisbank/2013/UU_Data+Management+Plan+SOP_DRAFT+June+2013_publ_def.pdf)
49. <http://libguides.bodleian.ox.ac.uk/c.php?g=422945&p=2888381>
50. <http://www1.udel.edu/research/preparing/datastorage.html>
51. <https://www.rug.nl/research/heyman-institute/?lang=en>
52. <http://psyres.uva.nl/scientific-integrity>
53. <http://www.efpa.eu/>
54. <http://www.share-project.org/>
55. [http://tilda.tcd.ie/publications/reports/pdf/Report\\_DesignReport.pdf](http://tilda.tcd.ie/publications/reports/pdf/Report_DesignReport.pdf)
56. <http://www.ucd.ie/issda/data/tilda/>
57. <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/34315>
58. <http://www.g2aging.org/>
59. For further information, please see <http://www.elixir-europe.org>
60. <https://fairsharing.org>
61. <https://fairsharing.org/recommendations>
62. <https://rd-alliance.org/group/biosharing-registry-connecting-data-policies-standards-databases-life-sciences.html>
63. <http://dx.doi.org/10.15497/RDA00017>
64. <https://bio.tools>
65. <https://tess.elixir-europe.org>
66. The list is available in full at <https://www.elixir-europe.org/platforms/data/elixir-deposition-databases>
67. <http://www.erare.eu/Infrastructures/elixir>
68. [http://www.icpermed.eu/media/content/ERA\\_PerMed\\_2018\\_pre-announcement\\_FV.pdf](http://www.icpermed.eu/media/content/ERA_PerMed_2018_pre-announcement_FV.pdf)
69. <http://www.transcanfp7.eu>
70. Sansone, S.A. and Rocca-Serra, P., Interoperability Standards – Digital Objects in Their Own Right. Wellcome Trust. 2016. <https://doi.org/10.6084/m9.figshare.4055496.v1>
71. <https://www.icos-ri.eu/sites/default/files/cmisiCOS%20RI%20Data%20Policy.pdf>
72. <https://www.icos-ri.eu/sites/default/files/cmisiCOS%20Carbon%20Portal%20Progress%20Report%202014-2015>



# Annex A – DCC Default DMP Template

This Digital Curation Centre (DCC) Template was sent to the scientific communities that were contacted for proof of concept, in order to give them an idea of the approach taken by the WG, which is very close to the DCC approach.

| Questions  | Issues to consider   | Guidance   |
|--|--|--|
| <b>Section A: Data Collection</b>                            |  |  |
| 1. What data will you collect or create?                     | What type, format and volume of data?<br>Do your chosen formats and software enable sharing and long-term access to the data?<br>Are there any existing data that you can re-use?                          | Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.  |
| 2. How will the data be collected or created?                | What standards or methodologies will you use?<br>How will you structure and name your folders and files?<br>How will you handle versioning?<br>What quality assurance processes will you adopt?            | Outline how the data will be collected/created and which community data standards (if any) will be used. Consider how the data will be organised during the project, mentioning for example naming conventions, version control and folder structures. 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.  |
| <b>Section B: Documentation and Meta-data</b>                |  |  |
| 3. What documentation and meta-data will accompany the data? | What information is needed for the data to be to be read and interpreted in the future?<br>How will you capture/create this documentation and meta-data?<br>What meta-data standards will you use and why? | Describe the types of documentation that will accompany the data to help secondary users to understand and re-use it. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed.<br><br>Documentation may also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards. |

| Questions   | Issues to consider   | Guidance  |
|---|--|---|
| <b>Section C: Ethics and Legal Compliance</b>                                   |  |   |
| 4. How will you manage any ethical issues?                                      | <p>Have you gained consent for data preservation and sharing?</p> <p>How will you protect the identity of participants if required, such as via anonymisation?</p> <p>How will sensitive data be handled to ensure it is stored and transferred securely?</p>        | <p>Ethical issues affect how you store data, who can see/use it and how long it is kept. Managing ethical concerns may include: anonymisation of data; referral to departmental or institutional ethics committees; and formal consent agreements. You should show that you are aware of any issues and have planned accordingly. If you are carrying out research involving human participants, you must also ensure that consent is requested to allow data to be shared and re-used.</p>   |
| 5. How will you manage copyright and Intellectual Property Rights (IPR) issues? | <p>Who owns the data?</p> <p>How will the data be licensed for re-use?</p> <p>Are there any restrictions on the re-use of third-party data?</p> <p>Will data sharing be postponed/restricted, such as to publish or seek patents?</p>                                | <p>State who will own the copyright and IPR of any data that you will collect or create, along with the licence(s) for its use and re-use. For multi-partner projects, IPR ownership may be worth covering in a consortium agreement. Consider any relevant funder, institutional, departmental or group policies on copyright or IPR. Also consider permissions to re-use third-party data and any restrictions needed on data sharing.</p>  |
| <b>Section D: Storage and Backup</b>  |  |   |
| 6. How will the data be stored and backed up during the research?               | <p>Do you have sufficient storage or will you need to include charges for additional services?</p> <p>How will the data be backed up?</p> <p>Who will be responsible for backup and recovery?</p> <p>How will the data be recovered in the event of an incident?</p> | <p>State how often the data will be backed up and to which locations. How many copies are being made? Storing data on laptops, computer hard drives or external storage devices alone is very risky. The use of robust, managed storage provided by university IT teams is preferable. Similarly, it is normally better to use automatic backup services provided by IT Services than rely on manual processes.</p> <p>If you choose to use a third-party service, you should ensure that this does not conflict with any funder, institutional, departmental or group policies, for example in terms of the legal jurisdiction in which data are held or the protection of sensitive data.</p> |

| Questions  | Issues to consider  | Guidance  |
|--|---|---|
| 7. How will you manage access and security?  | <p>What are the risks to data security and how will these be managed?</p> <p>How will you control access to keep the data secure?</p> <p>How will you ensure that collaborators can access your data securely?</p> <p>If creating or collecting data in the field how will you ensure its safe transfer into your main secured systems?</p> | <p>If your data is confidential (e.g. personal data not already in the public domain, confidential information or trade secrets), you should outline any appropriate security measures and note any formal standards that you will comply with e.g. ISO 27001.</p>  |
| <b>Section E: Selection and Preservation</b>   |   |   |
| 8. Which data are of long-term value and should be retained, shared, and/or preserved? | <p>What data must be retained/ destroyed for contractual, legal, or regulatory purposes?</p> <p>How will you decide what other data to keep?</p> <p>What are the foreseeable research uses for the data?</p> <p>How long will the data be retained and preserved?</p>   | <p>Consider how the data may be re-used e.g. to validate your research findings, conduct new studies, or for teaching. Decide which data to keep and for how long. This could be based on any obligations to retain certain data, the potential re-use value, what is economically viable to keep, and any additional effort required to prepare the data for data sharing and preservation. Remember to consider any additional effort required to prepare the data for sharing and preservation, such as changing file formats.</p> |
| 9. What is the long-term preservation plan for the dataset?                            | <p>Where e.g. in which repository or archive will the data be held?</p> <p>What costs if any will your selected data repository or archive charge?</p> <p>Have you costed in time and effort to prepare the data for sharing / preservation?</p>  | <p>Consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving. If you do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the grant.</p>  |

| Questions   | Issues to consider   | Guidance  |
|---|--|---|
| <b>Section F: Data Sharing</b>                            |  |   |
| 10. How will you share the data?                          | <p>How will potential users find out about your data?</p> <p>With whom will you share the data, and under what conditions?</p> <p>Will you share data via a repository, handle requests directly or use another mechanism?</p> <p>When will you make the data available?</p> <p>Will you pursue getting a persistent identifier for your data?</p>   | <p>Consider where, how, and to whom data with acknowledged long-term value should be made available. The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. If possible, mention earlier examples to show a track record of effective data sharing. Consider how people might acknowledge the re-use of your data.</p> |
| 11. Are any restrictions on data sharing required?        | <p>What action will you take to overcome or minimise restrictions?</p> <p>For how long do you need exclusive use of the data and why?</p> <p>Will a data sharing agreement (or equivalent) be required?</p>  | <p>Outline any expected difficulties in sharing data with acknowledged long-term value, along with causes and possible measures to overcome these. Restrictions may be due to confidentiality, lack of consent agreements or IPR, for example. Consider whether a non-disclosure agreement would give sufficient protection for confidential data.</p>  |
| <b>Section G: Responsibilities and Resources</b>          |  |   |
| 12. Who will be responsible for data management?          | <p>Who is responsible for implementing the DMP, and ensuring it is reviewed and revised?</p> <p>Who will be responsible for each data management activity?</p> <p>How will responsibilities be split across partner sites in collaborative research projects?</p> <p>Will data ownership and responsibilities for RDM be part of any consortium agreement or contract agreed between partners?</p> | <p>Outline the roles and responsibilities for all activities e.g. data capture, meta-data production, data quality, storage and backup, data archiving &amp; data sharing. Consider who will be responsible for ensuring relevant policies will be respected. Individuals should be named where possible.</p>   |
| 13. What resources will you require to deliver your plan? | <p>Is additional specialist expertise (or training for existing staff) required?</p> <p>Do you require hardware or software which is additional or exceptional to existing institutional provision?</p> <p>Will charges be applied by data repositories?</p>   | <p>Carefully consider any resources needed to deliver the plan, e.g. software, hardware, technical expertise, etc. Where dedicated resources are needed, these should be outlined and justified.</p>  |



# Glossary

|                 |   |
|-----------------|---|
| <b>ADS</b>      | Archaeology Data Service  |
| <b>BBSRC</b>    | Biotechnology and Biological Sciences Research Council                  |
| <b>CAPS</b>     | Collective Awareness Platforms for Sustainability and Social Innovation |
| <b>CESSDA</b>   | Consortium of European Social Science data Archives                     |
| <b>CHARLS</b>   | China Health and Retirement Longitudinal Study                          |
| <b>CLARIN</b>   | Common Language Resources and Technology Infrastructures                |
| <b>CP</b>       | Carbon Portal   |
| <b>CRELES</b>   | Costa Rican Longevity and Health Aging Study                            |
| <b>DANS</b>     | Data Archiving and Networked Services                                   |
| <b>DARIAH</b>   | Digital Research Infrastructure for the Arts and the Humanities         |
| <b>DCC</b>      | Data Curation Centre  |
| <b>DDI</b>      | Data Documentation Initiative   |
| <b>DDP</b>      | Domain Data Protocol  |
| <b>DMP</b>      | Data Management Plan  |
| <b>EADP</b>     | European Association of Developmental Psychology                        |
| <b>EASP</b>     | European Association of Social Psychology                               |
| <b>EDNA</b>     | E-Depot Dutch Archeology  |
| <b>EFPA</b>     | European Federation of Psychologists' Associations                      |
| <b>ELSA</b>     | English Longitudinal Study of Ageing                                    |
| <b>EMBL-EBI</b> | European Bioinformatics Institute                                       |
| <b>ERA</b>      | European Research Area  |
| <b>ERIC</b>     | European Research Infrastructure Consortium                             |
| <b>ESFRI</b>    | European Strategy Forum on Research Infrastructures                     |
| <b>FAIR</b>     | Findable, Accessible, Interoperable, Re-usable                          |
| <b>GDPR</b>     | General Data Protection Regulation                                      |
| <b>ICOS</b>     | Integrated Carbon Observation System                                    |
| <b>ICPSR</b>    | Interuniversity Consortium for Political and Social Research            |
| <b>IPR</b>      | Intellectual Property Rights  |
| <b>ISO</b>      | International Organisation for Standardisation                          |
| <b>JSTAR</b>    | Japanese Study on Aging and Retirement                                  |
| <b>KLoSA</b>    | Korean Longitudinal Study of Aging                                      |

|                  |   |
|------------------|---|
| <b>LASI</b>      | Longitudinal Aging Study in India   |
| <b>LDC</b>       | Linguistic Data Consortium  |
| <b>MEA</b>       | Munich Centre for the Economics of Ageing   |
| <b>MHAS</b>      | Mexican Health and Aging Study  |
| <b>NIH</b>       | National Institute of Health  |
| <b>PARTHENOS</b> | Pooling Activities, Resources and Tools for Heritage eResearch Networking, Optimisation and Synergies |
| <b>RAND HRS</b>  | Health and Retirement Study   |
| <b>RDM</b>       | Research Data Management  |
| <b>RI</b>        | Research Infrastructure   |
| <b>SHARE</b>     | Survey of Health, Ageing and Retirement in Europe   |
| <b>SIKB</b>      | Stichting Infrastructuur Kwaliteitsborging Bodembeheer  |
| <b>TILDA</b>     | The Irish Longitudinal Study of Aging   |
| <b>VCC</b>       | Virtual Competencies Centres  |
| <b>WG</b>        | Working Group   |



Science Europe is a non-profit organisation based in Brussels representing major Research Funding and Research Performing Organisations across Europe.

More information on its mission and activities is provided at [www.scienceeurope.org](http://www.scienceeurope.org).

To contact Science Europe, e-mail [office@scienceeurope.org](mailto:office@scienceeurope.org).

**Science Europe**  
Rue de la Science 14  
1040 Brussels  
Belgium

Tel +32 (0)2 226 03 00  
Fax +32 (0)2 226 03 01  
[office@scienceeurope.org](mailto:office@scienceeurope.org)  
[www.scienceeurope.org](http://www.scienceeurope.org)