DSW_project_name

Organization DS-Review

Created by Nikita Bardatskii (n.bardatskii@gmail.com)

Based on Common DSW Knowledge Model, 2.4.4 (dsw:root:2.4.4)

Project Phase Before Submitting the Proposal

Project Tags dsw_project_tag

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus aliquam enim a

posuere rhoncus. Pellentesque quis hendrerit lacus. Integer lectus diam, efficitur et ipsum

et, congue aliquet felis. Integer eget elit sit amet dui varius porttitor ut sit amet enim.

Aliquam erat volutpat. Nulla in velit viverra, aliquam tellus non, imperdiet orci. Duis ut maximus enim. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean congue pretium nisl, sed faucibus lorem egestas in.

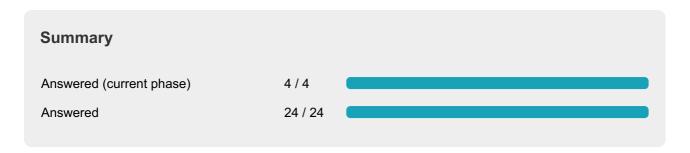
Created at 10 May 2023



Summary Report

Summary		
Answered (current phase)	29 / 29	
Answered	56 / 85	
Metric	Score	
Interoperability	0.60	
Reusability	1.00	
Good DMP Practice	0.33	
Openness	1.00	

I. Administrative information

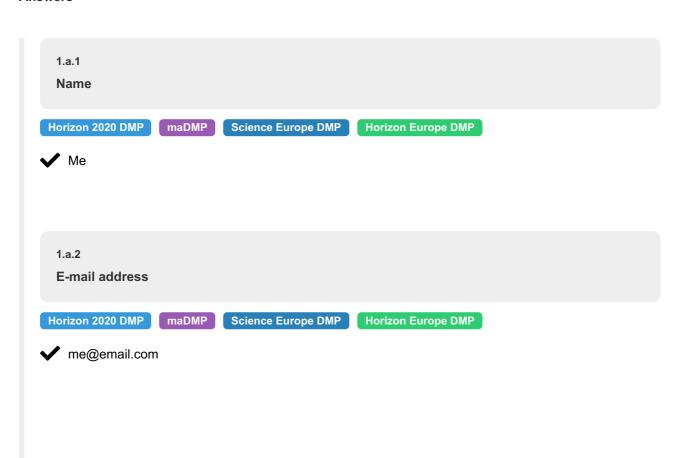


Questions



Each person contributing to creating or executing the data management plan should be added as a contributor. A project probably should have a Contact Person, and a Data Curator.

Answers



1.a.3
ORCID Identifier
Horizon 2020 DMP maDMP Science Europe DMP Horizon Europe DMP
✓ imagine here some orcid identifier
imagine here donne droid identation
1.a.4
Affiliation
7 time to 1
Horizon 2020 DMP Science Europe DMP Horizon Europe DMP
•
Just some affiliation here
1.a.5
Role
Note
Horizon 2020 DMP maDMP Science Europe DMP Horizon Europe DMP
Delegain a marie at about discourse of the course of affined by detection
Roles in a project should be given as they are defined by <u>datacite</u> .
You should specify at least one "Contact Person". If your project has a work package for data management,
identify the leader of that work package as "Data Curator".
✓ Contact Person
✓ Data Collector
Data Curator
✓ Data Manager
Data Protection Officer
✓ Data Steward
Distributor
Editor
Producer
Project Leader
Project Manager
Project Member
Researcher Rights Uniden
Rights Holder
Sponsor Supervisor

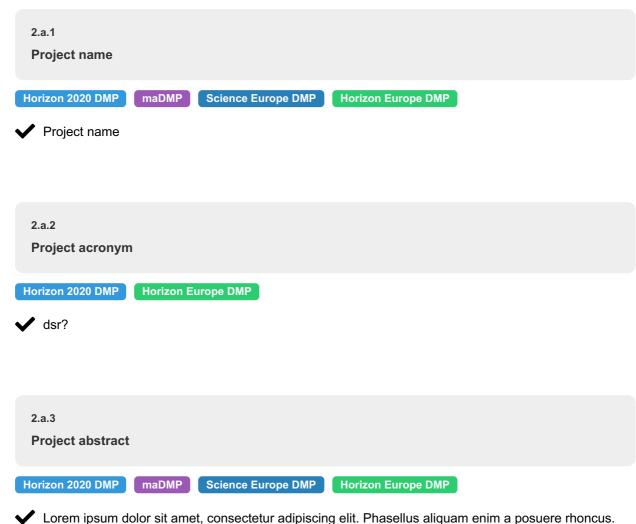
Work Package Leader
Other

2
Research Project(s)

Horizon 2020 DMP maDMP Science Europe DMP Horizon Europe DMP

Add each of the research project(s) that you are (or will be) working on and for which the data and work are described in this DMP. Give each project a small identifying name for yourself.

Answers



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus aliquam enim a posuere rhoncus. Pellentesque quis hendrerit lacus. Integer lectus diam, efficitur et ipsum et, congue aliquet felis. Integer eget elit sit amet dui varius porttitor ut sit amet enim. Aliquam erat volutpat. Nulla in velit viverra, aliquam tellus non, imperdiet orci. Duis ut maximus enim. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean congue pretium nisl, sed faucibus lorem egestas in.

2.a.4

Link to a project proposal or another description of the methods used in the project

X This question has not been answered yet!

2.a.5 Starting date of the project

Horizon 2020 DMP

Science Europe DMP maDMP

Horizon Europe DMP

2023-03-15

2.a.6

Ending date of the project

Horizon 2020 DMP

Science Europe DMP maDMP

Horizon Europe DMP

2023-09-08

2.a.7

Funding

Horizon 2020 DMP

maDMP Science Europe DMP Horizon Europe DMP

Add all the funding that are part of this project.

Answers

2.a.7.a.1

Funder

Horizon 2020 DMP

Science Europe DMP maDMP

Horizon Europe DMP

Specify the name of the funder that you ask for funding for your project. If the funder is not present in the suggested list, please specify a complete URL to the funder web site.



National Science Foundation Crossref http://dx.doi.org/10.13039/100000001 2.a.7.a.2 **Funding status** Horizon 2020 DMP Science Europe DMP maDMP Horizon Europe DMP c. Granted 2.a.7.a.3 **Grant number** Horizon 2020 DMP maDMP Science Europe DMP Horizon Europe DMP To execute the DMP, is additional specialist expertise required? 🗸 a. No Do you require hardware or software in addition to what is usually available in the institute? Horizon 2020 DMP Horizon Europe DMP ✓ b. Yes

4.b.1

What specific hard/software do you need, and why?

Horizon 2020 DMP

Science Europe DMP Horizon Europe DMP

✓ Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus aliquam enim a posuere rhoncus. Pellentesque quis hendrerit lacus. Integer lectus diam, efficitur et ipsum et, congue aliquet felis. Integer eget elit sit amet dui varius porttitor ut sit amet enim. Aliquam erat volutpat. Nulla in velit viverra, aliquam tellus non, imperdiet orci. Duis ut maximus enim. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean congue pretium nisl, sed faucibus lorem egestas in.

5

Describe national / funder / sectorial / departmental policies and procedures for data management that you will be using.

Horizon Europe DMP

Answers

5.a.1

Name of the policy or process

Horizon Europe DMP

no policy

5.a.2

Give a link/reference to the policy or process

Horizon Europe DMP

https://somepolicylinkg.com

5.a.3

Description of how and why you are using this policy or process

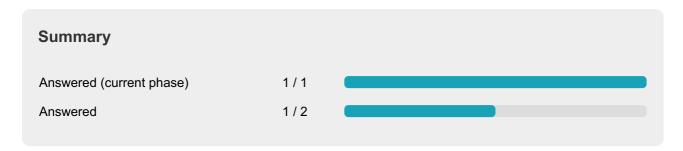
Horizon Europe DMP

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus aliquam enim a posuere rhoncus. Pellentesque quis hendrerit lacus. Integer lectus diam, efficitur et ipsum et, congue aliquet felis. Integer eget elit sit amet dui varius porttitor ut sit amet enim. Aliquam erat volutpat. Nulla in velit viverra, aliquam tellus non, imperdiet orci. Duis ut maximus enim. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean congue pretium nisl, sed faucibus lorem egestas in. Aenean vel purus enim. Proin tellus velit, rutrum eu neque nec, sodales sodales nulla. Aenean rutrum eros bibendum dapibus finibus. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc ut lobortis urna, ac mollis enim. Vivamus maximus efficitur turpis, sit amet venenatis dui malesuada et. Quisque sed tempor eros. Nam lacinia ante sed eleifend condimentum. Pellentesque pharetra hendrerit massa, eu suscipit ante gravida nec. Praesent hendrerit elit vitae justo placerat, sed auctor nisi maximus. Nulla dictum, elit ut pellentesque rhoncus, arcu tortor pretium justo, sed blandit quam massa eget erat. Vestibulum ipsum augue, sollicitudin in molestie quis, fermentum non ante.

II. Re-using data

Before you decide to embark on any new study, it is good practice to check all options to re-use existing available data, either collected or generated by yourself in an earlier project, or data from others (Barend Mons calls this "Other PEople's Data And Services" or OPEDAS). This can include reusable data that have been created for an earlier study, and also so-called "reference data" which is used by many projects.

It is not because we can generate massive amounts of data that we always need to do so. Creating data with public money is bringing with it the responsibility to treat those data well and (if potentially useful) make them available for re-use by others. And the circle is only complete if such data is actually re-used.



Questions

Do you need guidance to find existing data?

Research funding organisations more and more demand that you search for existing data sets that could have information that you need, before assuming you need to collect all data yourself. You are asked to list what you have been able to locate, and whether you have found it suitable to use in your research. Do you need pointers to find such existing data sets?

X This question has not been answered yet!



Are there any data sets available in the world that are relevant to your planned research?

Data Stewardship for Open Science: <u>atq</u>

External Links: <u>Google dataset search</u>, <u>Datacite Search</u>, <u>RDMkit on Reusing Data</u>, <u>RDMkit on Existing</u>



It is possible that you have overlooked existing data! This question is not only about data sets that are similar to what you want to determine yourself, but also reference data or data that should be mined from the existing literature. Further, it is very likely that you will refer to related data, e.g. other databases where you usually "quickly look something up", but that could maybe be properly integrated, especially if you need to do such lookups multiple times.

III. Creating and collecting data

In this chapter we describe all the sources of data: they can e.g. come from instruments or from questionnaires; data can be newly collected as part of the current project, but it can also be pre-existing data that may need proper contracts with the maintainer, some pre-processing, and quality checks. It can also be reference data that is part of curated resources and (public) databases.

For more information see Collecting in RDMKit



Questions

Are you running the project in a collaboration between different groups or institutes?

X This question has not been answered yet!

Will you be collecting physical samples?

Will you be collecting artefacts like specimens, minerals, biological samples?

Data Stewardship for Open Science: <u>kuz</u>

🖊 a. No

3

How will you do file naming and file organization?

Horizon 2020 DMP

Horizon Europe DMP

Putting some thoughts into file naming can save a lot of trouble later.

External Links: RDMkit on data organisation

X This question has not been answered yet!

4

Do you need guidance on what data formats/types to use?

In many research fields there are standard file formats and types that have been established by disciplinary standards organisations. Using such standards can help increase the interoperability of your data with other data and tools. Do you need help finding the right standards to use in your research?

X This question has not been answered yet!

What existing data formats/types will you be using?

Horizon 2020 DMP

Science Europe DMP Horizon Europe DMP

Have you identified types of data that you will use that are used by others too? Some types of data (for example "images" or "tables") are used by many different projects. For such data, often common standards exist (in our example "JPG" and "CSV" [comma separated values]) that help to make these data reusable. Are you using such common data formats?

Please make sure you list all the data types that are important for your project. You should make sure also to list the formats used in any data sets that you are re-using.

Data Stewardship for Open Science: njy

Answers

5 a 1

Data format/type

Horizon 2020 DMP

Science Europe DMP

Horizon Europe DMP



ISO/TS 19115-3:2016 Geographic information -- Metadata -- Part 3: XML schema implementation for fundamental concepts



https://fairsharing.org/10.25504/FAIRsharing.9de4c3

Is this a standard data format widely used by researchers in this field?

Horizon 2020 DMP

Science Europe DMP Horizon Europe DMP



🖊 a. No

5.a.2.a.1

Why are you using a non-standard format?

Horizon 2020 DMP

Science Europe DMP Horizon Europe DMP



a. There is no standardized format for this data type

5.a.3

Does this data format enable sharing and long term archiving?

Horizon 2020 DMP

Science Europe DMP

Horizon Europe DMP

Complicated (binary) file formats tend to change over time, and software may not stay compatible with older versions. Also, some formats (e.g. DOC, XLS) hamper long term usability by making use of patents or being hampered by restrictive licensing.

Ideally a format should be simple, text only, completely described, not restricted by copyrights, and implemented in different software packages.

🖊 b. Yes

5.a.4

What volume of data of this type will you be working with?

Horizon 2020 DMP

Science Europe DMP Horizon Europe DMP



a. So small that it is not a problem

5.a.5

Is this data format completely described?

Formats like XLS or SQL are very flexible; they can be adapted to many different uses, and this makes them good for interoperability. However, their flexibility also makes that it is not immediately obvious from the file structure how it can be used. The data needs a proper description in order for others (or yourself at a later time) to be able to unambiguously understand what it contains.

X This question has not been answered yet!

6

Do you need guidance on what encodings/terminologies/vocabularies/ontologies to use?

This question has not been answered yet!

7

What existing encodings/terminologies/vocabularies/ontologies will you be using?

Horizon 2020 DMP

Horizon Europe DMP

This question has not been answered yet!

Will you be using new types of data?

Horizon 2020 DMP

Horizon Europe DMP

Sometimes the type of data you collect can not be stored in a commonly used data format. In such cases you may need to make your own, keeping interoperability as high as possible.

Data Stewardship for Open Science: ikk

a. No, all of my data will fit in common formats

9

How will you be collecting and keeping your metadata?

Science Europe DMP

Horizon 2020 DMP

Horizon Europe DMP

For the re-usability of your data by yourself or others at a later stage, a lot of information about the data - for example how it was collected and how it can be used - should be stored with the data. Such data about the data is called **metadata**, and this set of questions are about this metadata.

Data Stewardship for Open Science: <u>rhm</u>

[7] External Links: RDMkit on documentation and metadata, Metadata Standards Catalogue (RDA)

X This question has not been answered yet!

10

Will you be acquiring data using measurement equipment?

Horizon 2020 DMP

Science Europe DMP

Horizon Europe DMP



🖊 a. No

11

Do you have any non-equipment data capture?

Horizon 2020 DMP | Science Europe DMP | Horizon Europe DMP

Does the data you collect contain non-equipment captured data such as questionnaires, case report forms, electronic patient records?

Data Stewardship for Open Science: <u>ybw</u>



🗸 a. No

12

Is there a data integration tool that can handle and combine all the data types you are dealing with in your project?



b. Yes

12.b.1

What software will you be using to collect all data?

X This question has not been answered yet!

13

Will you collect any data connected to a person, "personal data"?

Horizon Europe DMP

Very many kinds of data are connected to people. If there could be a way for someone, including yourself, to find out who that person is, that is considered personal data.

Simple examples are name, birth day or address; but there are many other data that can be personal: for example a voice recording, a combination of a location and a time (traffic flow), genetic information, or an X-ray of the skull.

- External Links: <u>RDMkit about Sensitive Data</u>, <u>RDMkit on Human Data</u>, <u>RDMkit on Data Protection</u>
- X This question has not been answered yet!

14

How is the ownership of the collected data arranged?

Horizon 2020 DMP

Science Europe DMP

Horizon Europe DMP

X This question has not been answered yet!

15

Will you monitor data integrity once it has been collected?

Working with large amounts of heterogenous data in a larger research group can have implications for the data integrity. How do you make sure every step of the workflow is done with the right version of the data? How do you handle the situation when a mistake is uncovered? Will you be able to redo the strict minimum data handling?

Data Stewardship for Open Science: <u>spg</u>

[7] External Links: RDMkit on Data Quality

IV. Processing data

In the processing phase, the data will be undergoing the mostly automated steps for processing, before the analysis and interpretation.

In this chapter, many questions are focusing on the compute environment that is used to process the data and make it available for interpretation by project partners. Some of those questions (e.g. on workflow systems and data provenance) are also relevant for the work in the interpretation phase.



Questions

Will you be using a shared working space to work with your data?

Horizon 2020 DMP | Science Europe DMP | Horizon Europe DMP

Will you be using a working space containing data and software specific to the project that is shared between all the people working on the data in the project? Sometimes such a system is called a Virtual Research Environment.

✓ a. No, participants in the project each have different collections of data and tools.

Are data that project members and partners store themselves adequately backed up and traceable?

Science Europe DMP

✓ b. Yes, protected against both equipment failure and human error

2

Data storage systems and file naming conventions

Science Europe DMP

It is a good idea to pre-define how data will be organised in the project work space, and to set conventions for how any data files and folders will be named.

External Links: <u>RDMkit on data organisation</u>, <u>RDMkit on data storage</u>

X This question has not been answered yet!

Workflow development

It is likely that you will be developing or modifying the workflow for data processing. There are a lot of aspects of this workflow that can play a role in your data management, such as the use of an existing work flow engine, the use of existing software vs development of new components, and whether every run needs human intervention or whether all data processing can be run in bulk once the work flow has been defined.

a. This has been arranged

How will you make sure to know what exactly has been run?

External Links: <u>RDMkit on Data Analysis</u>

X This question has not been answered yet!

5

How will you validate the integrity of the results?

Horizon 2020 DMP

Horizon Europe DMP

6

Do you need to do compute capacity planning?

If you require substantial amounts of compute power, amounts that are not trivially absorbed in what you usually have abailable, some planning is necessary. Do you think you need to do compute capacity planning?



Is the risk of information loss, leaks and vandalism acceptably low?

Horizon 2020 DMP Science Europe DMP Horizon Europe DMP

There are many factors that can contribute to the risk of information loss or information leaks. They are often part of the behavior of the people that are involved in the project, but can also be steered by properly planned infrastructure.

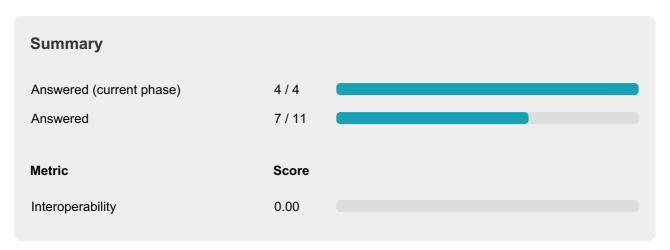
X This question has not been answered yet!

Do you have a contingency plan?

What will you do if the compute facility is down?

V. Interpreting data

The interpretation of the data consists of the last steps of processing (often with manual interventions), visualisation, and data integration. In this chapter many questions about data interoperability will come up.



Questions

1

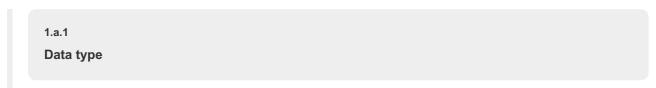
List the data formats you will be using for interpretation and describe their structure

Give each type of data a name that you recognise.

If you have data in many different structures, integrating the data may be more challenging.

External Links: <u>RDMkit on Machine Actionability</u>, <u>RDMkit on Data Processing</u>

Answers





Dictyostelium Phenotype Ontology



https://fairsharing.org/10.25504/FAIRsharing.9c1p18

1 a 2

How is this data structured?

✓ d. Images, audio or video tracks, with some structured metadata attached

2

Will you be doing integration or linking of different data types?

If you are getting different types of data from different sources and want to use them together it is likely that you will need to match items and glue everything together. This can be done with traditional table database technology, but it is also possible to use Linked Data and RDF.



3

Will you be using common ontologies?



4

Will there be potential issues with statistical normalization?

X This question has not been answered yet!

5

Will you be integrating different data sources to get more samples or more data points?

6

Will you be integrating different data sources in order to get more information for each sample or data point?

X This question has not been answered yet!

7

Do you have all tools to couple the necessary data types?



8

Will you be using a federated analysis approach?

In some cases it is not practical to bring all data together:

- It may be legally hard to collect the data in one place for analysis
- It may be technically hard to transport data to a single place for analysis

In such cases, a Federated analysis approach may be applicable. Examples of such techniques are DataShield and the Personal Health Train. Secure multi-party computation may be useful too to prevent information leaking between parties.



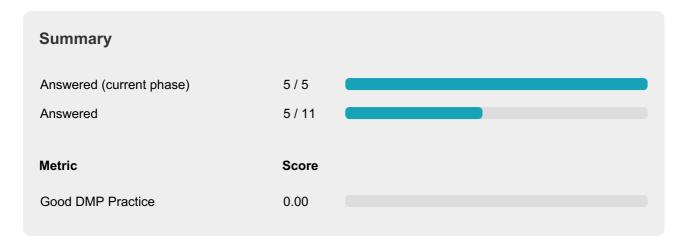
9

Will you be doing (automated) knowledge discovery?

Data Stewardship for Open Science: <u>bzu</u>

VI. Preserving data

In this chapter, issues regarding data publication and long term archiving are addressed.



Questions

1
Specify a list of data sets you will be producing

Horizon 2020 DMP maDMP Science Europe DMP Horizon Europe DMP

Add all the data sets you will be producing. Give each a short name, sufficient for yourself to know what data it is about. It is useful to think about a data set as some collection of data that will be ending up in the same place.

- External Links: <u>RDMkit on Collecting Data</u>, <u>RDMkit on Data Preserving</u>
- X This question has not been answered yet!

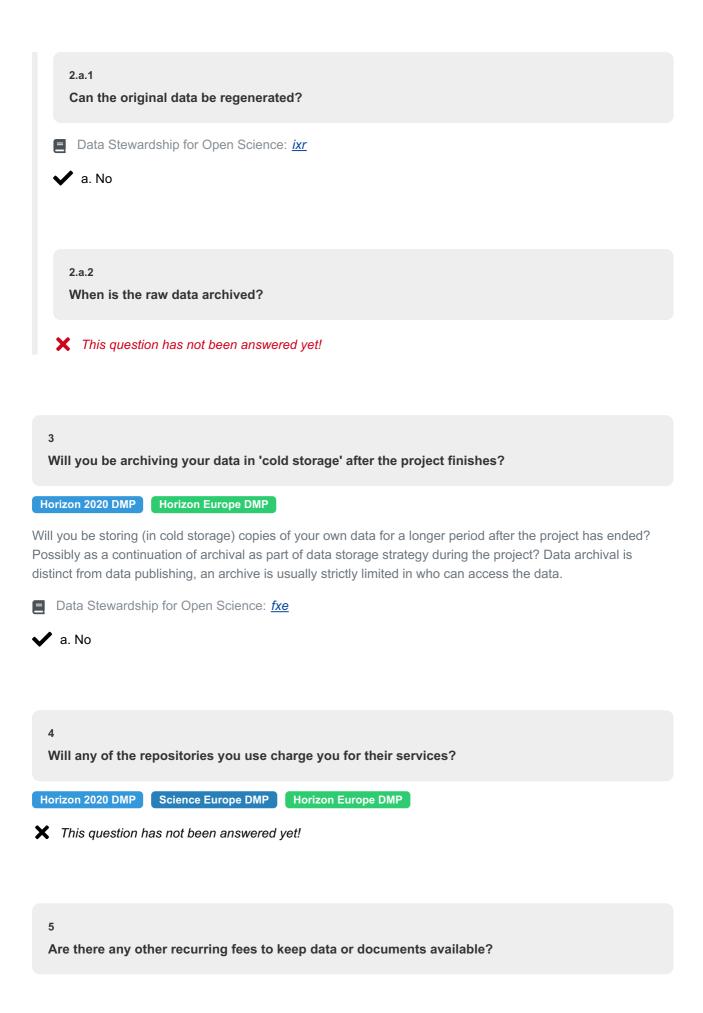
Will you be archiving data (using so-called 'cold storage') for long term preservation already during your project?

Horizon 2020 DMP Horizon Europe DMP

Much of the raw data you have will need to be archived for your own later use somewhere. This is often done off-line on tape, not on the disks of the compute facility. Please note that this does not refer to the data publication.

Data Stewardship for Open Science: kjp

🗸 a. No



Are you using any commercially licensed products to keep data, software or documents available, for which a regular fee must be paid?



6

Did you budget for the time and effort it will take to prepare the data for publication?

Horizon 2020 DMP Science Europe DMP

Horizon Europe DMP

This question has not been answered yet!

Will you be making sure that blocks of data deposited in different repositories can be recognized as belonging to the same study?

X This question has not been answered yet!

Specify a list of software packages you will be publishing

Specify a short name for each software package.

This question has not been answered yet!

Will reference data be created?

Will any of the data that you will be creating form a reference data set for future research (by others)?

Much of todays data is used in comparison with reference data. You may be comparing your own data with a "standard set" which is maintained as a collection by someone else. Or you could be determining differences to a standard (for example in bioinformatics, a genome is often compared with a reference genome to identify genomic variants). Will you be creating any data that will be reference data for other researchers?

Data Stewardship for Open Science: <u>rbz</u>

VII. Giving access to data

This chapter deals with the information needed by people who will re-use your data, and with the access conditions they will need to follow.



Questions

The FAIR principles do not contain any direction towards "Openness". This is done on purpose, because there can be compelling reasons not to make data "Open", such as privacy, other sensitive data, or intellectual property protection.

The true goal of funding agencies is to create the maximum value for society from their investments. They therefore often add "As open as possible, as closed as necessary" to the requirements for funding.

Data Stewardship for Open Science: <u>ivm</u>



2
Can all of your data become completely open over time?

Horizon 2020 DMP maDMP Science Europe DMP Horizon Europe DMP

Some data may be subject to a temporary embargo, or need to stay closed for specific reasons.



Will you use temporary restrictions on the reuse of the data (embargo)?

Horizon 2020 DMP | Science Europe DMP | Horizon Europe DMP



a. No, all data will be opened immediately

Will metadata be available openly?

Horizon Europe DMP

Horizon Europe demands that metadata is completely open, e.g. provided under a CC0 license, and that any deviation from this is clarified.



b. Yes

4.b.1

Will metadata contain instructions how to get access to the data?

Horizon Europe DMP

X This question has not been answered yet!

4.b.2

Will the metadata be available in a form that can be harvested and indexed?

Horizon Europe DMP

Repositories often allow search engines and catalogues to index the metadata in an automated way. Will this be the case for your data?