



Organization	ELIXIR Norway
Created by	Ulli Rothweiler (ulli.rothweiler@uit.no)
Based on	RI gap analysis, 0.0.1 (elixir.no:ri-elixir-norway:0.0.1)
Project Phase	Before Finishing the Project
Created at	11 Mar 2021

I. Administrative details

Report

Indications

Answered (current phase)	7 / 11
Answered	10 / 14

Metrics

No metrics for this chapter.

Questions

1 Contributors

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.

Tags: *maDMP, Science Europe DMP*

✗ This question has not been answered yet!

2 RI

Add each of the project(s) that are you 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.

Tags: *maDMP, Science Europe DMP*

Answers

2.b.1 RI name

Tags: *maDMP, Science Europe DMP*

✓ *NorCryst*

2.b.2 Project short discription

Tags: *maDMP, Science Europe DMP*

✓ *The NORCRYST consortium has gathered all relevant academic institutions with expertise in crystallography in Norway, and can accommodate all national users from industry and academia with instruments, expertise and services within macromolecular crystallography and supporting techniques. Various specialized expertise are offered at each of the 4 nodes, with coordination by the Norwegian Structural Biology Centre (NorStruct) at UiT.*

2.b.3 Date the RI will started

Tags: *Science Europe DMP, maDMP*

✗ This question has not been answered yet!

2.b.4 Date the RI funding will end

Tags: *Science Europe DMP, maDMP*

✗ This question has not been answered yet!

2.b.5 Funding

Add all the funding that are part of this project.

Tags: *maDMP, Science Europe DMP*

Answers

2.b.5.b.1 Funder

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.

Tags: *Science Europe DMP, maDMP*

✓ *NFR*

2.b.5.b.2 Funding status

Tags: *Science Europe DMP, maDMP*

✓ *c. Granted*

2.b.5.b.3 Grant number

Tags: *maDMP, Science Europe DMP*

✓ *245828*

3 To execute the DMP, is additional specialist expertise required?

Tags: *Science Europe DMP*

✓ *b. Yes, we will be training existing staff*

3.b.1 What kind of training?

Tags: *Science Europe DMP*

✗ This question has not been answered yet!

4 Do you require hardware or software in addition to what is currently available in the participating institutions?

✓ *b. Yes*

4.b.1 What specific hard/software do you need, and why?

Tags: *Science Europe DMP*

✓ *Storage space for image files, mobile disks for transport.*

II. Re-using data

Before you decide to embark on any new study, it is nowadays 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.

Report

Indications

Answered (current phase)	7 / 11
Answered	8 / 13

Metrics

No metrics for this chapter.

Questions

1 Describe the utility of data produced at the RI; to whom might it be useful?

✓ *Drug design, structural biology, modeling, molecular dynamic simulations etc...*

2 Is there pre-existing data?

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

Tags: *maDMP, Science Europe DMP*

Data Stewardship for Open Science: [atq](#)

✓ *b. Yes*

2.b.1 Will you be using any pre-existing data (including other people's data)?

Will you be referring to any earlier measured data, reference data, or data that should be mined from existing literature? Your own data as well as data from others?

Tags: *maDMP, Science Europe DMP*

Data Stewardship for Open Science: [ezi](#)

✓ *b. Yes*

2.b.1.b.1 What reference data will you use?

Much of today's 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 (in bioinformatics, a genome is often compared with a reference genome to identify genomic variants). If you use reference data, there are several other issues that you should consider. What are the reference data sets that you will use?

Tags: *Science Europe DMP*

Data Stewardship for Open Science: [quc](#)

Answers

2.b.1.b.1.b.1 Source of reference data set

Give the name of the data set. You will be shown suggestions of data bases from FAIRSharing, but you can also type the name of a data set that is not in FAIRsharing

Tags: *Science Europe DMP*

✗ This question has not been answered yet!

2.b.1.b.1.b.2 What are the conditions of use for this data sets?

Tags: *Science Europe DMP*

✗ This question has not been answered yet!

2.b.1.b.1.b.3 Do you know in what format the reference data is available?

Do you know the data format of the reference data? Is this suitable for your work? Does it need to be converted?

Data Stewardship for Open Science: [*jxb*](#)

✗ This question has not been answered yet!

2.b.1.b.1.b.4 Is the reference data resource versioned?

Many reference data sets evolve over time. If the reference data set changes, this may affect your results. If different versions of a reference data set exist, you need to establish your "version policy".

Tags: *Science Europe DMP*

Data Stewardship for Open Science: [*rgy*](#)

✗ This question has not been answered yet!

2.b.1.b.1.b.5 How will you make sure the same reference data will be available to reproduce results of your users?

Will the reference data in the version you use be available to others?

✗ This question has not been answered yet!

2.b.1.b.2 Will you use non-reference data sets?

Tags: *Science Europe DMP*

✓ *b. No*

2.b.1.b.3 Will you couple existing (biobank) data sets?

✓ *a. No*

2.b.2 Do you need to harmonize different sources of existing data?

If you are combining data from different sources, harmonization may be required. You may need to re-analyse some original data.

Data Stewardship for Open Science: [*wht*](#)

✓ *a. No*

2.b.3 Will you be using data that needs to be (re-)made computer readable first?

Some old data may need to be recovered, e.g. from tables in scientific papers or may be punch cards.

📖 Data Stewardship for Open Science: [pth](#)

✓ a. No

III. Creating and collecting data

We will make sure that we know what data will be generated at the RI and when it will be generated. We also need to make sure that there will be adequate storage space to deal with it, and that all the responsibilities have been taken care of.

Report

Indications

Answered (current phase)	41 / 44
Answered	46 / 49

Metrics

Metric	Score
Findability	0
Accessibility	0
Interoperability	1
Reusability	0.72

Questions

1 What data formats/types will you/your users be using?

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 "PNG" and "CSV") that help to make these data reusable. Are you using such common data formats?

You should make sure also to list the formats used in any data sets that you are re-using.

📖 Tags: *Science Europe DMP*

📖 Data Stewardship for Open Science: [njy](#)

Answers

1.b.1 Data format/type

📖 Tags: *Science Europe DMP*

✓ *macromolecular Crystallographic Information File*



<https://fairsharing.org/bsg-s000290>

1.b.2 Is this a standard data format used by others in this field?

📖 Tags: *Science Europe DMP*

✓ b. Yes

1.b.3 Does this data format enable sharing and long term archiving?

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

Tags: *Science Europe DMP*

✓ *b. Yes*

1.b.4 What volume of data of this type will you be working with?

Tags: *Science Europe DMP*

✓ *b. I can specify the total amount per year*

1.b.4.b.1 Data volume in Gigabytes per year

Specify an approximation of the expected data volume

Tags: *Science Europe DMP*

✓ *1000*

1.c.1 Data format/type

Tags: *Science Europe DMP*

✓ *Protein Data Bank Format*

 FAIRsharing

<https://fairsharing.org/bsg-s000255>

1.c.2 Is this a standard data format used by others in this field?

Tags: *Science Europe DMP*

✓ *b. Yes*

1.c.3 Does this data format enable sharing and long term archiving?

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

Tags: *Science Europe DMP*

✓ *b. Yes*

1.c.4 What volume of data of this type will you be working with?

Tags: *Science Europe DMP*

✓ *a. So small that it is not a problem*

2 Will you/your users be using new types of data?

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

3 How will you/your users be storing metadata?

For the re-usability of your data by yourself or others at a later stage, a lot of information about the data, 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.

[SEEK](#) is a webtool to store (meta)data and provenance. The public global instance [FAIRDOMHub](#) is free to users in Norway. SEEK can be integrated with the data storage and analysis platform for users in Norway [NeLS](#).

☰ Data Stewardship for Open Science: [rhm](#)

🔗 External Links: [SEEK](#)

✓ a. Explore

3.a.1 Do suitable 'Minimal Metadata About ...' (MIA...) standards exist for your experiments?

🔗 External Links: [FAIRsharing repository of standards](#)

✓ b. Yes

3.a.2 Do you know how and when you will be collecting the necessary metadata?

Often it is easiest to make sure you collect the metadata as early as possible.

🔗 External Links: [FAIRsharing repository of standards](#)

✓ a. No

3.a.3 Will you consider re-usability of your data beyond your original purpose?

Adding more than the strict minimum metadata about your experiment will possibly allow more wide re-use of your data, with associated higher data citation rates. Please note that it is not easy for yourself to see all other ways in which others could be reusing your data.

✓ b. Yes, I will document more metadata than needed for reproducibility

3.a.3.b.1 How will you balance the extra efforts with the potential for added reusability?

✓ b. I will use preselected additional standard modules of metadata

3.a.3.b.2 Do you need to exchange your data with others?

✓ b. Yes

3.a.4 Did you consider how to monitor data integrity?

Working with large amounts of heterogenous data in a larger research group has 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: [spg](#)

✓ a. Explore

3.a.4.a.1 Will you be keeping a master list with checksums of certified/correct/canonical/verified data?

Data corruption or mistakes can happen with large amounts of files or large files. Keeping a master list with data checksums can be helpful to prevent expensive mistakes. It can also be helpful to keep the sample list

under version control forcing that all changes are well documented.

✓ *a. No*

3.a.4.a.2 Will you define a way to detect file or sample swaps, e.g. by measuring something independently?

This will dependent on the applied methods. Examples could include e.g. verifyBamID for known genotypes

✓ *a. No*

3.a.5 Do all datasets you work with have a license?

It is not always clear to everyone in the project (ad outside) what can and can not be done with a data set. It is helpful to associate each data set with a license as early as possible in the project. A data license should ideally be as free as possible: any restriction like 'only for non-commercial use' or 'attribution required' may reduce the reusability and thereby the number of citations. If possible, use a computer-readable and computer actionable license.

✓ *b. Yes*

3.a.5.b.1 Will you store the licenses with the data at all time?

It is very likely that data will be moved and copied. At some point people may lose track of the origins. It can be helpful to have the licenses (of coarse as open as possible) stored in close association with the data.

📖 Data Stewardship for Open Science: [atw](#)

✗ **This question has not been answered yet!**

3.a.6 How will you keep provenance?

To make your experiments reproducible, all steps in the data processing must be documented in detail. The software you used, including version number, all options and parameters. This information together for every step of the analysis is part of the so-called data provenance. There are more questions regarding this in the chapter on data processing and curation.

✓ *a. All steps will be documented in an (electronic) lab notebook*

3.a.7 How will you do file naming and file organization?

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

✓ *a. Explore*

3.a.7.a.1 Did you make a SOP (Standard Operating Procedure) for file naming?

It can help if everyone in the project uses the same naming scheme.

✓ *a. No*

3.a.7.a.2 Will you be keeping the relationships between data clear in the file names?

Advice: Use the same identifiers for sample IDs etc throughout the entire project.

✓ *a. No*

3.a.7.a.3 Will all the metadata in the file names also be available in the proper metadata?

The file names are very useful as metadata for people involved in the project, but to computers they are just identifiers. To prevent accidents with e.g. renamed files metadata information should always also be available elsewhere and not only through the file name.

✘ This question has not been answered yet!

4 Please specify what data you will acquire using measurement equipment

You can use any name for the data set, make sure that it identifies the data set to yourself.

Tags: *Science Europe DMP*

Answers

4.b.1 Who will do the measurements? And where?

Tags: *Science Europe DMP*

✘ This question has not been answered yet!

4.b.2 Instruments used for data collection

Specify what technical instruments you are using to collect the data.

Tags: *Science Europe DMP*

Answers

4.b.2.b.1 Instrument name

Tags: *Science Europe DMP*

✓ *Max IV, BESSY, ESRF*

4.b.2.b.2 Instrument description

Tags: *Science Europe DMP*

✓ *Synchrotron*

4.b.3 Is the equipment completely standard and well described?

If the technology is very much under development, you may want to come back later to understand exactly how the measurements have been made. Is the measurement equipment and protocol sufficiently standard that you will be able to explain how it is done or refer to a standard explanation?

Tags: *Science Europe DMP*

✓ *a. Very well described and known*

4.b.4 Is special care needed to get the raw data ready for processing?

Where does the data come from? And who will need it? Sometimes the raw data is measured somewhere else than where the primary processing is taking place. In such cases the ingestion or transport of the primary data may take special planning. You also need to make sure that data is secure and that data integrity is guaranteed.

✓ *a. No, this is all fine*

4.b.5 Will you be using quality processes?

Tags: *Science Europe DMP*

✓ *b. Yes*

4.b.5.b.1 Are you calibrating measurements?

Tags: *Science Europe DMP*

b. Yes

4.b.5.b.2 Are you running repeat samples or are you repeating measurements?

Tags: *Science Europe DMP*

b. Yes

4.b.5.b.3 Are you running standardized data capture or recording?

Tags: *Science Europe DMP*

b. Yes

4.b.5.b.4 Are you doing Data Entry validation?

Tags: *Science Europe DMP*

b. Yes

4.b.5.b.5 Are you using data peer review?

Tags: *Science Europe DMP*

a. No

4.b.5.b.6 Are you using controlled vocabularies?

Tags: *Science Europe DMP*

a. No

4.b.5.b.7 Are you using any other quality processes?

Tags: *Science Europe DMP*

b. Yes

4.b.5.b.7.b.1 What other quality processes do you use?

PDB validation server

5 Do you have any non-equipment data capture?

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

Tags: *Science Europe DMP*

Data Stewardship for Open Science: [ybw](#)

a. No

6 Is there a data integration tool that can handle and combine all the data types you are dealing with

in your RI?

✓ b. Yes

6.b.1 What software will you be using to collect all data?

✓ d. Software not listed here

7 Will you be storing physical samples?

📖 Data Stewardship for Open Science: [kuz](#)

✓ a. No

8 Will you need consent for any newly collected personal data?

🔖 Tags: *maDMP, Science Europe DMP*

🔗 External Links: [NSD Information and consent](#), [REC Informed consent](#)

✓ a. No, We do not collect any new personal data

9 How is the ownership of the collected data arranged?

🔖 Tags: *Science Europe DMP*

✓ b. All data will be owned by the Principle Investigator/user

IV. Data sensitivity

Ethical and legal issues

adapted from 2019 version of [NSD DMP tool](#) and [Tryggve Checklist on ELSI issues and GDPR compliance](#)

Report

Indications

Answered (current phase)	3 / 3
Answered	3 / 4

Metrics

No metrics for this chapter.

Questions

1 Will you collect or generate data about people?

✓ b. No

2 Will the RI follow any institutional policies, codes of conducts or other ethical guidelines?

Each researcher has an independent responsibility for making sure that the research is being carried out in accordance with general scientific and ethical principles and guidelines. For an overview of general and subject-specific research ethics guidelines, see the [Norwegian National Research Ethics Committees](#). Note that in multidisciplinary projects it may be relevant to look to guidelines for several subject areas. In addition, the [Research Ethics Act](#) applies to all research in Norway. Also, check which guidelines apply to your institution.

✓ a. Yes

2.a.1 Provide names and links below.

✓ *general research ethic guidelines at national level and the involved institutions*

3 **Other ethical / legal issues.**

✗ **This question has not been answered yet!**

V. Processing data

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

Report

Indications

Answered (current phase)	24 / 27
Answered	24 / 29

Metrics

Metric	Score
Accessibility	1
Good DMP Practice	0

Questions

1 **Will you be providing the data to the user through a shared working space ?**

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

Tags: *Science Europe DMP*

✓ *b. Yes*

ELIXIR Norway offers [NeLS](#) a multi tiered shared storage for collaborating on data sets

1.b.1 **Will this work space be run by dedicated specialists?**

If your work space is run and maintained by specialists, e.g. the ICT department of one of the institutes involved in the projects, this means that backup and restore as well as access management is properly addressed.

✓ *b. Yes*

1.b.2 **How will you/your users work with the data?**

There are several questions regarding the dynamics of the data in the working area, who works with it, the software that is run on it, etc.

Tags: *Science Europe DMP*

✓ *a. Explore*

1.b.2.a.1 **What kind of data will you/your users have in the work space?**

When making the work space, it helps to know whether you expect to work with very many small files, a few very large files, whether you will use a (SQL) database to store most of the data. Maybe your data is suitable for a system like Hadoop? Such information can be collected here.

✓ *diffraction images, structure factor files, coordinate files.*

1.b.2.a.2 Do you/your users need the work space to be close to the compute capacity?

If you have large volumes of data that are intensely and repeatedly used by the computing work flow, it may be needed to keep the storage in the same place as where the computing takes place.

 Data Stewardship for Open Science: [wia](#)

✓ *a. No*

1.b.2.a.3 Will you/your users be working with your data in another form than the way it will be archived?

Archival and working with data have different requirements. You want archived information to be in a form that others could read and in a format that is also understandable in a number of years. When working with the data, you need to be able to address it efficiently. If the two differ, you need to plan for conversions.

✓ *a. No, data format will be archived in the same way we work with it*

1.b.2.a.4 How does the storage need change over time?

To perform capacity planning, it is important to know what the need for storage capacity at the beginning and the end of the project will be.

 Tags: *Science Europe DMP*

✓ *a. Storage needs will be the same during the whole RI runtime*

1.b.2.a.5 Will you need to temporarily archive data sets (e.g. to tape)?

Usually, data sets will be archived if it is unlikely you need them in the short term, but it would be hard to create them again, and/or they are essential for reproducing your work. Archival storage of large volumes can be significantly cheaper than keeping it in the working area for an extensive period.

✓ *b. Yes*

1.b.2.a.6 How will your first data come in?

✓ *d. Initial data will arrive on separate media and will need to be copied to the work space*

1.b.2.a.7 How will the RI partners/ the users access the work space?

✓ *a. Explore*

1.b.2.a.7.a.1 Who will arrange access control?

✓ *b. RI management will need to be able to give people access*

1.b.2.a.7.a.2 Will the work space storage need to be remote mounted?

✓ *b. No, data will be copied if it needs to be processed elsewhere*

1.b.2.a.7.a.3 Will data be copied out and in to the workspace storage by remote users?

✓ *c. Yes, occasionally*

1.b.3 How available/reliable should must the work space be?

There are a number of questions that can help you to decide whether your work space will be reliable enough for

your project.

Tags: *Science Europe DMP*

✓ *a. Explore*

1.b.3.a.1 How do you prevent a total loss of data in the work space?

Tags: *Science Europe DMP*

✓ *b. All essential data is also stored elsewhere*

1.b.3.a.1.b.1 Is there software in the work space? Can it also be restored quickly?

Data Stewardship for Open Science: [cbq](#)

✓ *b. Software in the work space is only a copy*

1.b.3.a.2 Can you/your users handle it when the work space is off line for a while?

✓ *a. We could handle a few days of offline time per year*

1.b.3.a.3 How long can you/your users wait for a restore if the storage fails?

✓ *a. We can wait for repair and a restore*

1.b.3.a.4 How long can you wait for a restore if you accidentally damage a file?

✓ *a. Days*

1.b.3.a.5 Will you make backup copies of your/your users data that is not in the work space?

Are there any data files e.g. on laptops of project members? Also: supercomputing centers and other high performance computer centers often write in their terms of use that you need to take care of your own backups

Tags: *Science Europe DMP*

✓ *a. This is left to the project participants*

1.b.4 How will access control to the work space be controlled?

Tags: *Science Europe DMP*

✗ **This question has not been answered yet!**

2 Data storage systems and file naming conventions

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.

Tags: *Science Europe DMP*

✗ **This question has not been answered yet!**

3 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*

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

✗ This question has not been answered yet!

5 How will you validate the integrity of the results?

✗ This question has not been answered yet!

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 available, some planning is necessary. Do you think you need to do compute capacity planning?

✓ *a. No*

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

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.

Tags: *Science Europe DMP*

✗ This question has not been answered yet!

8 Do you have a contingency plan?

What will you do if the compute facility is down?

✓ *b. We have an alternative*

9 Will you version datasets?

[SEEK](#) which is used in [FAIRDOMHub](#) and can be used together with [NeLS](#) supports versioning by default.

[NeLS](#) can also be used with [Git Large File Storage \(LFS\)](#)

External Links: [FAIRDOMHub](#), [SEEK](#), [NeLS](#), [Git Large File Storage \(LFS\)](#)

✓ *b. No*

VI. 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.

Report

Indications

Answered (current phase)	7 / 16
Answered	7 / 16

Metrics

Questions

1 How will you be doing the integration of different data sources?

✓ a. Explore

1.a.1 List the data formats you will be using for data integration

Answer some questions for each

Answers

1.a.1.b.1 Data format:

✓ macromolecular Crystallographic Information File



<https://fairsharing.org/bsg-s000290>

1.a.1.b.2 How is the data structured in general?

✓ a. (meta)data fields in a domain specific file

1.a.1.b.2.a.1 Can all of the data you want to couple be captured in that format?

✓ b. Yes

1.a.1.b.2.a.1.b.1 Will you be doing it that way?

✓ b. Yes

1.a.1.b.2.a.2 Does the domain specific format come with its own suite of integration tools that you will use?

✓ a. No

1.a.2 Will you/your users be using a workflow for data integration, e.g. with tools for database access or conversion?

📖 Data Stewardship for Open Science: [qqb](#)

✗ This question has not been answered yet!

1.a.3 Will you/your users use a 'linked data' approach?

🔗 External Links: [Linked data \(wikipedia\)](#)

✗ This question has not been answered yet!

2 Will you/your users be using common or exchangeable units?

✗ This question has not been answered yet!

3 Will you/your users be using common ontologies?

✗ This question has not been answered yet!

4 Will there be potential issues with statistical normalization?

✗ This question has not been answered yet!

5 Will you/your users be integrating different data sources to get more samples or more data points?

✗ This question has not been answered yet!

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

✗ This question has not been answered yet!

7 Do you/your users have all tools to couple the necessary data types?

✗ This question has not been answered yet!

8 Will you/your users be doing (automated) knowledge discovery?

📖 Data Stewardship for Open Science: [bzu](#)

✗ This question has not been answered yet!

VII. Preserving data

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

Report

Indications

Answered (current phase)	23 / 25
Answered	31 / 34

Metrics

Metric	Score
Findability	1
Accessibility	0.29
Reusability	0.75
Good DMP Practice	0.8

Questions

1 Will you /your users be archiving data (using so-called 'cold storage') for long term preservation already during the RI runtime/project?

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](#)

✓ *b. Yes*

1.b.1 Is the archived data changing over time, needing re-archival?

☰ Data Stewardship for Open Science: [tgk](#)

✓ *a. No*

1.b.2 Will the archive be stored on disk or on tape?

Data stored though [StoreBioinfo](#) and [NIRD](#) is backedup on disk.

☞ External Links: [NeLS](#), [NIRD](#)

✓ *a. Disk*

1.b.3 Will the archive be stored in a remote location, protecting the data against disasters?

Data stored though [StoreBioinfo](#) or [NIRD](#) is geo replicated.

☞ External Links: [NeLS](#), [NIRD](#)

✓ *b. Yes*

1.b.4 Will the archive need to be protected against loss or theft?

✓ *b. Yes*

1.b.4.b.1 Will the archive be encrypted?

✓ *a. No*

1.b.4.b.2 Is it clear who has physical access to the archives?

✓ *b. Yes*

1.b.5 Will your project require the archives to be available on-line?

☰ Data Stewardship for Open Science: [ybd](#)

✓ *a. No*

1.b.6 Has it been established who has access to the archive, and how fast?

✓ *a. No*

1.b.7 Has it been established how long the archived data need to be kept? For each of the different parts of the archive (raw data / results)?

Deposition repositories can be an option for storage of these.

☰ Data Stewardship for Open Science: [kdp](#)

✓ *a. No*

1.b.8 Will the data still be understandable after a long time?

See also all questions about keeping metadata and data formats. Make sure the metadata is kept close to the data in the archive, and that community supported data formats are used for all long term archiving.

☰ Data Stewardship for Open Science: [zmu](#)

✓ *b. Yes*

2 Specify details of data types which will be produced at your RI

It is useful to think about a data types as some collection of data that will be ending up in the same place.

Tags: *maDMP, Science Europe DMP*

✗ **This question has not been answered yet!**

3 Will any of the repositories you use charge you/your users for their services?

Tags: *Science Europe DMP*

✗ **This question has not been answered yet!**

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

Tags: *Science Europe DMP*

✓ *b. Yes*

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

✓ *b. Yes, all data sets will have links to the related data*

6 Are there any recurring fees to keep data or documents available?

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

✓ *a. No*

7 Will you be archiving your data after the RI runtime in 'cold storage'?

Will you be storing (in cold storage) copies of your own data for a longer period after the project has ended? Possibly as a continuation of archival as part of data storage strategy during the project? Data archival is distinct from data publishing, an archive is usually limited in who can access the data.

Data Stewardship for Open Science: [fxe](#)

✓ *b. Yes*

7.b.1 Will data formats of data in cold storage be upgraded if they become obsolete?

✓ *a. No*

7.b.2 Will data be migrated regularly to more modern storage media (e.g. newer tapes)?

✓ *a. No*

8 Will you also publish data if the results of your study are negative/inconclusive or unpublishable?

Even if you do not obtain the results you had foreseen from your own study, the data can still be valuable for reuse in another context. Also, publishing the data can avoid that someone else collects a similar data set with a similar negative result.

✓ *b. Yes*

9 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!**

10 How will you be making sure there is good provenance of the data (and analysis)?

Data analysis is normally done manually on a step-by-step basis. It is essential to make sure all steps are properly documented, otherwise results will not be reproducible.

Tags: *Science Europe DMP*

✓ *a. We use lab notebooks*

Make sure to make the notes available in electronic form along with your data

11 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)?

Data Stewardship for Open Science: *rbz*

✓ *b. Yes*

11.b.1 What will the Intellectual Property be like?

Who will own the rights to the reference data set? Who will be able to use it?

Data Stewardship for Open Science: *hct*

✓ *CC0*

11.b.2 How will you maintain it?

How will maintenance be paid for in the long run? Will you host it yourself or deposit it with a repository? How will you deal with requests for help? And with requests for adding data?

Data Stewardship for Open Science: *usx*

✓ *deposited to PDB*

11.b.3 Will you be updating the reference data at regular intervals?

✓ *b. No*

12 How will you document your/the user data?

For reusability, the data should be well documented. In this section of the questionnaire you can specify what kinds of documentation you will be providing.

Tags: *Science Europe DMP*

✓ *a. Explore*

12.a.1 Will you be documenting the data with Dublin Core metadata?

Dublin Core is a standard documenting domain independent aspects of a resource; including who has created it, audience, function, formatting and licensing. Does your documentation follow the Dublin Core standard?

Tags: *Science Europe DMP*

External Links: [Dublin Core Metadata Terms](#), [Dublin Core Initiative](#)

✓ a. No

12.a.2 Will you be documenting the data with W3C PROV provenance?

The W3C Prov standard documents processes (workflow) that were used to produce a resource. This can be used to document e.g. the software (including version) and parameters you use to analyze the data. Will your documentation follow the W3C Prov standard?

Tags: *Science Europe DMP*

External Links: [W3C Prov primer](#)

✓ b. Yes

13 Will you do systems biology modeling (for users)?

✓ a. No

14 Will you do structural modeling?

✓ b. Yes

14.b.1 Will you use model building by homology?

✓ b. Yes

14.b.2 Will you use enzyme activity models?

✓ b. Yes

14.b.3 Will you use molecular dynamics?

✓ b. Yes

VIII. 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.

Report

Indications

Answered (current phase)	20 / 20
Answered	20 / 20

Metrics

Metric	Score
Accessibility	1
Good DMP Practice	1
Openness	0.46

Questions

1 Will you be working with the philosophy 'as open as possible' for your data/your users data?

Tags: *Science Europe DMP*

Data Stewardship for Open Science: [jvm](#)

✓ a. No

You will need to explain!

2 Are there potential copyright and Intellectual Property Rights (IPR) issues?

✓ a. Yes

2.a.1 How will you manage copyright and Intellectual Property Rights (IPR) issues?

✓ /

3 Can all of your data at your RI become completely open immediately?

Tags: *maDMP, Science Europe DMP*

✓ a. No

3.a.1 Are there legal reasons why (some of your) data can not be completely open?

Tags: *maDMP, Science Europe DMP*

✓ b. Yes

3.a.1.b.1 Are there privacy reasons why data can not be open?

Tags: *maDMP*

✓ b. Yes

3.a.1.b.1.b.1 Are there restrictions on where the data need to be stored?

Tags: *maDMP*

✓ a. No

3.a.1.b.1.b.2 Could pseudonymization be used to make the data more openly available?

Legally, pseudonymous data (which means that someone has the key to reverse the process) is still considered privacy sensitive information. However, the EU is working on special cases where the data can still be opened as long as the key availability is sufficiently limited.

Tags: *maDMP*

✓ a. No

3.a.1.b.1.b.3 Could anonymization be used to make the data more openly available?

Different anonymization techniques exist. Disadvantage of anonymization is that data integration becomes virtually impossible, but it may be the only way to open up your data for other research

Tags: *maDMP*

✓ a. No

3.a.1.b.1.b.4 Could you use data aggregation to make the data openly available?

Aggregated data, where typically at least 15 individuals are in any data point, are considered sufficiently anonymous. This is an alternative way of making data openly available for future research

Tags: *maDMP*

✓ *a. No*

3.a.1.b.2 Are there IP reasons why data can not be open?

✓ *b. Yes*

3.a.1.b.2.b.1 Is it clear who owns data and documents?

✓ *b. Yes*

3.a.1.b.2.b.1.b.1 Who will own the intellectual property rights (copyrights) of the data that you will collect or create?

✓ *I*

3.a.1.b.2.b.2 Will someone be given decision power to move documents or data to a new place after the project has finished?

In one case in the past, all documents that had been assembled by a project in a documentation system had to be deleted because not a single person could decide to move them to a new platform when the documentation system was going off-line.

✓ *b. Yes*

3.a.1.b.3 Will you/your users be allowing authenticated access to the data?

Tags: *Science Europe DMP*

✓ *a. No*

3.a.2 Are there business reasons why (some of) the data at your RI can not be completely open?

Tags: *Science Europe DMP*

✓ *b. Yes, patents still need to be applied for*

3.a.3 Are there other reasons why (some of) the data at your RI can not be completely open?

Tags: *Science Europe DMP*

✓ *b. Yes, papers need to be submitted first*

3.a.4 Will you use a limited embargo?

Tags: *Science Europe DMP*

✓ *c. Yes, data that is not legally restrained will be released after a fixed time period, unconditionally*

3.a.4.c.1 What embargo period are you using?

After what period will restrictions on the reuse of data (except ethical and legal restrictions) fall away?

Tags: *Science Europe DMP*

4 Will there be valorization or translational returns of the data generated at your RI?

✓ a. No