NALMIN_GapAnalysis

Core DS Knowledge Model

Organization	ELIXIR Norway
Created by	xian hu (huxian@gmail.com)
Based on	RI gap analysis, 0.0.1 (elixir.no:ri-elixir-norway:0.0.1)
Project Phase	Before Submitting the Proposal
Created at	11 Mar 2021

I. Administrative details

Report

Indications

Answered (current phase)	4 / 4
Answered	13 / 18

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*

Answers

1.b.1 **Name**

🏷 Tags: *maDMP, Science Europe DMP*

✔ *Xian Hu*

1.b.2 **E-mail address**

🏷 Tags: *maDMP, Science Europe DMP*

✔ *xianh@ncmm.uio.no*

1.b.3 **ORCID Identifier**

🏷 Tags: *maDMP, Science Europe DMP*

✔ *https://orcid.org/0000-0002-3381-7514*

1.b.4 **Affiliation**

🏷 Tags: *Science Europe DMP*

✔ *University of Oslo*

1.b.5　Role

Roles in a project should be given as they are defined by [datacite](#).

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".

🏷 Tags: *maDMP, Science Europe DMP*

✔ *d. Data Manager*

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*

✔ *NALMIN*

2.b.2　Project short discription

🏷 Tags: *maDMP, Science Europe DMP*

✔ *NALMIN's principal goal is to establish a network at international competence level for advanced light microscopy. The network consists of 6 national nodes located in Oslo (2 nodes), Bergen, Trondheim (2 nodes) and Tromsø, each with specialized competence. All nodes are members of the national NorBioimaging infrastructure organization, and three of the nodes have already qualified, pending financing, for membership in the ESFRI network EuroBioimaging. Three of the nodes are associated with Norwegian Centres of Excellence (CCB, CIR and CEMIR). Research groups in Norway will have access to both the national centres and the international EuroBioimaging nodes for their specific research, providing cutting-edge imaging techniques to a broad community.*

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*

✔ *Norges Forskningsråd*

 http://dx.doi.org/10.13039/501100005416

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*

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

---

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?

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

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)	2 / 2
Answered	2 / 2

Metrics
No metrics for this chapter.

Questions

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

✔ *Bioimaging Data from cellular and tissue imaging. Might be useful to other research who uses bioimaging as tool of research.*

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*

✔ *a. No*

You know that this is very unlikely? 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

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)	14 / 14
Answered	32 / 33

Metrics

Metric	Score
Findability	0.33
Interoperability	1
Reusability	0.67

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*

✔ *Open Microscopy Environment - Tagged Image File Format*

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*

✔ *d. The user will have to take care of data storage immediately*

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 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*

✔ *b. Yes*

`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?**
>
> ✔ *a. I will see what I can do*
>
> `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*

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

`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.

✔ *a. No*

`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.

✔️ *b. Yes*

`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.

✔️ *a. No, the file names in the project are an essential part of the metadata*

`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*

✔️ *a. Users using our equipment*

`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*

✔️ *Optical Microscopes*

`4.b.2.b.2` **Instrument description**

🏷️ Tags: *Science Europe DMP*

✔️ *Various optical microscopes with different imaging modality.*

`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*

✔ *a. No*

---

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?**

✔ *a. No*

6.a.1 **Can all data be brought into the same format?**

✔ *b. Yes*

---

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.**
>
> ✔ *https://www.uio.no/english/for-employees/support/research/research-data-management/*

`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)	28 / 28
Answered	68 / 68

Metrics

Metric	Score
Accessibility	0.6
Reusability	0.48
Good DMP Practice	0.51

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.

✔ *Image data, metadata, image processing software/workflow/source code Tentatively a modest number of very large files and about twice as many small ones*

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*

✔ *b. Yes*

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.

✔ *b. Yes, archival will require a conversion step*

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*

✔ *c. Storage needs are small at the beginning of the RI runtime and will grow later*

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?**

✔ *c. We will need a high-speed network connection to copy the initial data*

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

✔ *a. Explore*

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

    ✔ *c. The work space should be connected to a single-sign-on system*

    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*

    ✔ *a. The work space provides sufficient guarantees*

    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?**

    ✔ *b. Hours*

    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*

✔️ *f. Only user of specific user projects will have read access; only selected RI members will be able to write data*

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*

✔️ *a. Explore*

2.a.1 Are you using a filesystem with files and folders?

Are some of the data in the project stored in a filesystem with files and folders?

🏷 Tags: *Science Europe DMP*

✔️ *b. Yes*

2.a.1.b.1 Will you use a folder for each sample/subject?

🏷 Tags: *Science Europe DMP*

✔️ *a. No*

2.a.1.b.2 Will you use a (sub)folder for each (repeated) analysis?

🏷 Tags: *Science Europe DMP*

✔️ *b. Yes*

2.a.1.b.2.b.1 What are the naming conventions for the analysis folders?

What appointment have you made for the naming of the folders? Make sure names are relatively short, and avoid spaces and special characters.

✔️ *In the end up to the user, this is why we want to have meetings on the handling of metadata. Generally, one includes the type of experiment, run number. and parameters.*

2.a.1.b.3 Will you use a (sub)folder for each step in the analysis workflow?

🏷 Tags: *Science Europe DMP*

✔️ *a. No*

2.a.1.b.4 What appointments have you made about the naming of files?

Make sure names are relatively short, and avoid spaces and special characters. You can use underscore characters, and consider using unique identifiers for the samples/experiments. You can consider to add versioning using the date in YYYYMMDD format.

🏷 Tags: *Science Europe DMP*

✔️ *If we are given a tool to prohibit or suppressed the use of spaces and special or invisible characters, we'd hail that. But naming conventions are of little interest to us as the intricacies are encoded in the metadata and its keywords. We see no value in obsessing about the naming convention.*

2.a.2 Will you be storing data in an "object store" system?

🏷 Tags: *Science Europe DMP*

✔️ *a. No*

Will you use a relational database system to store project data?

🏷 Tags: *Science Europe DMP*

✔ *b. Yes*

> 2.a.3.b.1 **How will you handle changes in the data?**
>
> Database systems can be configured to keep all data, so that it is possible to reconstruct any past state of the data. How are changes in the data handled by your database?
>
> 🏷 Tags: *Science Europe DMP*
>
> ✔ *b. We will be allowing Create, Update and Delete*

2.a.4 **Will you use a graph database for data in the project?**

🏷 Tags: *Science Europe DMP*

✔ *a. No*

2.a.5 **Will you be storing data in a triple store?**

🏷 Tags: *Science Europe DMP*

✔ *a. No*

---

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?**

✔ *a. Explore*

4.a.1 **Will you keep results together with all processing scripts or workflows including documentation of the versions of the tools that have been run?**

✔ *b. Yes*

4.a.2 **Will you make use of the metadata fields in your output data files to register how the data was obtained?**

File formats like VCF (for genetics) and TIFF (for images) have possibilities to document metadata in the file header. It is a good idea to use work flow tools that use these fields to document what was done to obtain the data.

✔ *b. Yes*

4.a.3 **Will you use a central repository for all tools and their versions as used in your RI/for each user project?**

Especially if analysis and processing of data in the project is done on multiple different computers by different people, it is a good idea to have your own repository of tools and their blessed versions.

📄 Data Stewardship for Open Science: *pzq*

✔️ *a. No*

4.a.4 **Will you use a central repository for reference data used at your RI?**

Especially if analysis and processing of data in the project is done on multiple different computers by different people, it is a good idea to have your own repository of reference data versions.

📄 Data Stewardship for Open Science: *pzq*

✔️ *b. Yes*

4.a.5 **Will you make use of standard workflow engines and automatic workflows for all data analysis at your RI?**

It is much easier to guarantee consistency and reproducibility if all data processing is done using automated work flows, especially if the workflow engine automatically keeps adequate provenance data.

✔️ *b. Yes*

4.a.6 **Are all software tools in the workflow professionally maintained, with version control?**

Will you be able to find and reproduce exactly which version was used for any analysis? Not only for the major tools in the workflows, but also for all 'glue' code and small tools you created especially for the project?

✔️ *a. No*

5 **How will you validate the integrity of the results?**

✔️ *a. Explore*

5.a.1 **Will you run a subset of your jobs several times across the different compute infrastructures you are using?**

There are surprisingly many complications that can cause (slight) inconsistencies between results when workflows are run on different compute infrastructures. A good way to make sure this does not bite you is to run a subset of all jobs on all different infrastructure to check the consistency.

✔️ *a. No*

5.a.2 **Will you be instrumenting the tools into pipelines and workflows using automated tools?**

Surrounding all tools in your data processing and analysis workflows with the 'boilerplate' code necessary on the computer system you are using is tedious and error prone. Especially if you are using the same tools in multiple different work flows and/or on multiple different computer architectures. Automated instrumentation, e.g. by using a workflow management system, can prevent many mistakes.

✔️ *a. No*

5.a.3 **Will you use independently developed duplicate tools or workflows for critical steps to reduce or eliminate human errors?**

Validation of results without a golden standard is very hard. One way of doing it is to develop two solutions for a problem (two independent workflows or two independently developed tools) to check whether the results are identical or comparable.

✔️ *a. No*

5.a.4 **Will you run part of data sets repeatedly to catch unexpected changes in results?**

Running a small subset of the data repeatedly can be useful to catch unexpected problems that would otherwise be very hard to detect.

📖 Data Stewardship for Open Science: *egv*

✔ *b. Yes*

---

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?

✔ *b. Yes*

Standardized computing workflows and multiple tier storage is provided through NeLS

You can directly apply for big scale e-infrastructure resources in Norway at *Sigma2*

6.b.1 **Do you know how much CPU power, memory and I/O band width a typical analysis will take?**

Did you run pilot jobs? Do you know this information from comparable projects? Did you test whether the work scales up as you expected if you run more than one job?

✔ *b. Yes*

6.b.2 **What type of compute architecture is most suitable for your work? Will you have that available?**

✔ *d. We will use a mix of computing architectures for different parts of the work*

6.b.3 **Is there sufficient experience with the chosen computer in the RI, and sufficient time to support the RI partners/users?**

ELIXIR-Norway provides support through its helpdesk and the option for collaborations on bigger projects

✔ *d. We have sufficient knowledge in the project*

6.b.4 **Have you arranged with the provider when will you need the compute capacity?**

Do you need the compute capacity also for development? Can you start developing locally and start with a deployment test later?

🔗 External Links: *NeLS, Sigma2*

✔ *b. Yes*

6.b.5 **Is all required compute capacity available close to the project working storage area?**

✔ *b. Yes*

6.b.6 **Will different groups work on different parts of the workflows at your RI? Will parts of the computing be done on 'local' infrastructure to the RI?**

✔ *b. Some steps may be performed at local computing locations*

6.b.6.b.1 **Is there sufficient network capacity to the other computing locations?**

✔ *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*

✔ *a. Explore*

`7.a.1` **Do RI members store data or software on computers in the lab or external hard drives connected to those computers?**

When assessing the risk, take into account who has access to the lab, who has (physical) access to the computer hardware itself. Also consider whether data on those systems is properly backed up

🏷 Tags: *Science Europe DMP*

✔ *a. No*

`7.a.2` **Do RI members carry data with them?**

Does anyone carry project data on laptops, USB sticks or other external media?

🏷 Tags: *Science Europe DMP*

✔ *b. Yes*

> `7.a.2.b.1` **Are all data carriers encrypted? Are accounts on the laptop password protected?**
>
> 🏷 Tags: *Science Europe DMP*
>
> ✔ *a. No*

`7.a.3` **Do RI members store project data in cloud accounts?**

Think about services like Dropbox, but also about Google Drive, Apple iCloud accounts, or Microsoft's Office365

✔ *b. Yes*

Make sure your users are aware of the risks of cloud storage (not so much that the cloud is unreliable, but there is no protection against "accidentally" sharing a cloud folder with people outside the project)

`7.a.4` **Do RI members send project data or reports per e-mail or other messaging services?**

✔ *b. Yes*

`7.a.5` **Do all data centers where RI data is stored carry sufficient certifications?**

🏷 Tags: *Science Europe DMP*

✔ *a. No*

`7.a.6` **Are all RI web services addressed via secure http (https://)?**

🏷 Tags: *Science Europe DMP*

✔ *b. Yes*

`7.a.7` **Have RI members been instructed about the risks (generic and specific to the project)?**

RI members may need to know about passwords (not sharing accounts, using different passwords for each service, and two factor authentication), about security for data they carry (encryption, backups), data stored in their own labs and in personal cloud accounts, and about the use of open WiFi and https

✔ *b. Yes*

7.a.8 **Did you consider the possible impact to the RI or organization if information is lost?**

✔ *c. Yes; the risk is acceptably low*

7.a.9 **Did you consider the possible impact to the RI or organization if information leaks?**

✔ *d. Yes; we will need to work on this.*

7.a.10 **Did you consider the possible impact to the RI or organization if information is vandalized?**

✔ *c. Yes; the risk is acceptably low*

7.a.11 **Are personal data sufficiently protected?**

✔ *d. Yes, all data will be anonymized as early as possible*

Please note that GDPR law in Europe specifies that data is only anonymous as long as nobody in the world has enough information to re-identify the subject.

---

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)*

✔ *a. Yes*

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)	12 / 12
Answered	25 / 25

Metrics

Metric	Score
Interoperability	0.25
Reusability	1

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:**

✔ *Nikon, Zeiss, Olympus, Leica formats*

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

✔ *c. Complex data, like a graph*

1.a.1.b.2.c.1 **Are you representing the graph in RDF?**

RDF is the graph representation used in the 'semantic web'

✔ *a. No*

1.a.1.c.1 **Data format:**

✔ *Imaris formatt*

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

✔ *a. (meta)data fields in a domain specific file*

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

✔ *a. No*

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

✔ *b. Yes*

1.a.1.d.1 **Data format:**

✔ *Metamorph*

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

✔ *a. (meta)data fields in a domain specific file*

> **1.a.1.d.2.a.1** **Can all of the data you want to couple be captured in that format?**
>
> ✔ *a. No*

> **1.a.1.d.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*

✔ *b. Yes*

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

↗ External Links: *Linked data (wikipedia)*

✔ *b. Yes*

> **1.a.3.b.1** **Are your data sources using linked data?**
>
> ✔ *b. Partly*

> **1.a.3.b.2** **Will you provide your results as semantically interoperable linked data?**
>
> 📘 Data Stewardship for Open Science: *fxm*
>
> ✔ *b. Yes*

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

✔ *a. No*

3 **Will you/your users be using common ontologies?**

✔ *a. No*

4 **Will there be potential issues with statistical normalization?**

✔ *a. No*

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

✔ *b. Yes*

> **5.b.1** **Have these been collected with sufficiently identical protocols?**
>
> ✔ *a. No*

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

✔ *a. No*

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

✔ *b. Yes*

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

▤ Data Stewardship for Open Science: *bzu*

✔ *b. Yes*

VII. Preserving data

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

Report

Indications

Answered (current phase)	14 / 14
Answered	76 / 77

Metrics

Metric	Score
Findability	0.54
Accessibility	0.77
Reusability	0.28
Good DMP Practice	0.78

Questions

☐ 1 **Will you /your usersbe 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*

✔ *b. Yes*

☐ 1.b.1.b.1 **Do you/your users need frequent backups?**

The general term 'backup' is used for protection against two different kinds of problems: equipment failure and human error. Protections against these two may need different solutions. Both are considered backup for this and subsequent questions.

✔ *b. Yes, data changes frequently*

☐ 1.b.1.b.2 **Will you be relying on these backups to recover from human error (accidental changes**

or deletions)?

✔ *b. Yes*

This puts other demands on the possibility to restore files. Make sure this is covered

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

Data stored though StoreBioinfo and NIRD is backuped 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*

✔ *a. No*

`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?**

   ✔ *c. Yes*

      `1.b.4.b.1.c.1` **Is it clear who has access to the key? Also in case of a required data restore?**

      ✔ *b. Yes*

   `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?**

✔ *b. Yes*

   `1.b.6.b.1` **Has it been established who can ask for a restore?**

   ✔ *b. Yes*

   `1.b.6.b.2` **Is the data volumnious?**

   ✔ *a. Yes*

      `1.b.6.b.2.a.1` **If the data is voluminous, will your RI/the users be able to cope with the time needed for a restore?**

✔ *a. Yes*

1.b.6.b.3 **Has authority over the data been arranged for when the RI/or a user project is finished (potentially long ago)? Is there a data access committee?**

Consider who would decide on this when people or the PI leaves the project/institute.

✔ *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*

✔ *a. No*

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*

Answers

2.b.1 **Data type:**

Consider one data set as a collection of data from one set of samples.

◈ Tags: *maDMP, Science Europe DMP*

✔ *Imaging*

2.b.2 **Description of the data type**

Examples could be "Field observations", "raw instrument data", "genomic variants".

◈ Tags: *Science Europe DMP, maDMP*

✔ *Recored high dimensional, multispectral, high dynamic range, high resolution datasets with many time points.*

2.b.3 **Identifier of the data type**

Please add all "formal" identifiers you have for this data set: these can be handles or DOIs or any other type. One important purpose of these identifiers is to be able to find the dataset back.

A good identifier is *persistent* (i.e. it does not change, and also the same identifier will never be used for another data set), *globally unique* (nobody else uses the same identifier for a different data set) and *resolvable* (you can actually locate the data set if you only know the identifier).

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

2.b.4 **Will this data types be published?**

Will you publish the data set somewhere? Note that this does not necessarily mean that the data set becomes openly available, conditions for access and use may apply.

✔ *b. Yes*

2.b.4.b.1 **Specify where you will distribute this data from**

Add each of the locations where this data set will be made available. Give each of these a short name that identifies it to yourself.

Answers

2.b.4.b.1.b.1 **What repository will this data be stored in?**

Domain repositories often have the best functionality to make the data findable and reusable. Many of them are listed in https://fairsharing.org/

If a repository offers to give your data set a DOI it is a good idea to use that option.

✔ *c. Our institutional repository*

Disadvantage of a general purpose repository is the lack of data-specific features (e.g. 'play' instead of 'download' for an audio file) and limited findability

Contact your local library for more information.

2.b.4.b.1.b.2 **Who will the data in this place be shared with?**

✔ *c. Closed: This is an archive, it is not meant to share the data with anyone*

2.b.4.b.1.b.3 **Licenses under which this distribution of the data set will be available**

Please add each license that this data set will be available as. For each license you will be able to specify when it starts being applicable, so that you can e.g. specify that the data is restricted for a few months and open afterwards.

Answers

2.b.4.b.1.b.3.b.1 **Under what license will the data set be made available?**

✔ *c. They will be available under some restrictions*

2.b.4.b.1.b.3.b.1.c.1 **What conditions?**

Please summarize the restrictions

✔ *Set out by the owner during cooperation, the partners after publication, the journal's TOS*

2.b.4.b.1.b.3.b.1.c.2 **Link to the license text**

If the license is available as a text on the internet, please specify its URL here.

✔ *no unified license available*

2.b.4.b.1.b.3.b.1.c.3 **How will re-users be able to get access?**

✔ *c. Through another process*

2.b.4.b.1.b.3.b.1.c.3.c.1 **What process?**

✔ *They will be granted access to shared data. With project termination, this data will be made unavailable.*

2.b.4.b.1.b.3.b.1.c.4 **Will the conditions be published as part of open metadata?**

✔ *a. No*

2.b.4.b.1.b.3.b.2 **Starting date**

From which date will data be available under this license?

✔ *ab initio*

2.b.4.b.2 **Will you be adding a reference to the published data to at least one data catalogue?**

Data is sometimes difficult to locate, especially if it is not in a domain-specific repository. Data catalogues may increase findability.

✔ *a. No*

2.b.5 **How long will this data set be kept?**

For optimum reusability data needs to be available for as long as possible. There may be financial reasons why you can't keep the data any longer; there may be legal reasons requiring you to delete the data.

✔ *a. As long as technically possible*

2.b.6 **Will the metadata be available even when the data no longer exists?**

This is a one of the FAIR principles.

🏷 Tags: *Science Europe DMP*

✔️ *b. Yes*

2.b.7 **Does the data usually contain personal data?**

Is there anything in this dataset that could be tied to a person? This could be a physical characteristic, but also behavior of a person, movements, communications. Note that e.g. readouts about the performance of an airplane are considered to contain personal data of the pilot!

🏷 Tags: *Science Europe DMP, maDMP*

✔️ *a. No*

2.b.8 **Does this data contain sensitive information?**

Personal information can be sensitive if it is for instance about the health, sexual orientation, religion of a person. But there are also other classes of sensitive information: e.g. locations of rare species in biodiversity could be sensitive and should not leak to poachers.

🏷 Tags: *Science Europe DMP, maDMP*

✔️ *b. Yes*

2.b.9 **Do you make use of persistent and unique identifiers such as Repository specific Identifiers or Digital Object Identifiers for this ?**

✔️ *a. Yes*

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

🏷 Tags: *Science Europe DMP*

✔️ *a. No*

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*

✔️ *a. No*

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?**

✔️ *d. Yes, we have made other arrangements*

5.d.1 **What other arrangements?**

✔️ *All original data will be held locally. Collaborations between research groups are not our immediate concern.*

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

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.

✔ *a. No*

9 **Specify a list of software packages you will be publishing**

Specify a short name for each software package.

Answers

> **9.b.1** **Software package:**
>
> ✔ *Imaris*
>
> **9.b.2** **Will you be adding a proper open-source license?**
>
> ⧉ External Links: *Choose an open source license,* *Open Source Initiative: Licenses*
>
> ✔ *a. No*
>
> **9.b.3** **Where will the software package be available?**
>
> ✔ *at some or all of the local nodes*
>
> **9.b.4** **Will this software be listed in a catalogue?**
>
> ✔ *b. Yes*
>
> > **9.b.4.b.1** **What catalogue?**
> >
> > ✔ *TBD certaily our website*
>
> **9.c.1** **Software package:**

✔ *SVI Huygens*

9.c.2 **Will you be adding a proper open-source license?**

⬁ External Links: *Choose an open source license, Open Source Initiative: Licenses*

✔ *a. No*

9.c.3 **Where will the software package be available?**

✔ *floating license local to our nodes*

9.c.4 **Will this software be listed in a catalogue?**

✔ *b. Yes*

9.c.4.b.1 **What catalogue?**

✔ *website*

9.d.1 **Software package:**

✔ *FIJI*

9.d.2 **Will you be adding a proper open-source license?**

⬁ External Links: *Choose an open source license, Open Source Initiative: Licenses*

✔ *b. Yes, we have decided on an open source license*

9.d.2.b.1 **What is the license?**

✔ *generic ImageJ FIJI license*

9.d.3 **Where will the software package be available?**

✔ *all local nodes*

9.d.4 **Will this software be listed in a catalogue?**

✔ *b. Yes*

9.d.4.b.1 **What catalogue?**

✔ *website*

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*

✔️ *PI owner of the data will decide data will be anonymized and may be reduced*

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*

✔️ *Same as all our bulk data Full res version may even be migrated to archieve*

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

✔️ *a. Yes*

11.b.3.a.1 | **What will be the release interval?**

✔️ *Upon imminent obsolescence of the underlying dataformat Likely around once every 5 years*

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*

✔️ *a. No*

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

✔ *a. No*

14 **Will you do structural modeling?**

✔ *a. No*

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)	15 / 15
Answered	29 / 29

Metrics

Metric	Score
Findability	0
Accessibility	0.8
Good DMP Practice	1
Openness	0.42

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?**

✔ *PI owner decide on all privacy and sharing rights*

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*

✔ *d. Yes, they must stay in the same institute*

`3.a.1.b.1.b.1.d.1` **Are you going to use a dedicated platform at your intituion?**

✔ *a. Yes*

`3.a.1.b.1.b.1.d.1.a.1` **Which service are you going to use?**

↗ External Links: *HUNT DB, EUTRO*

✔ *Locally: a high performance multi-tiered disk subsystem. On platform encrypttion. Mirroring of encryoted data on nearby backup.*

`3.a.1.b.1.b.1.d.1.a.2` **Have you made sure that the platform fullfills the requirments for your data?**

✔ *a. Yes*

`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*

✔ *b. Yes*

`3.a.1.b.1.b.2.b.1` **Can you make use of an existing 'trusted third party' for pseudonymization?**

Making use of the same pseudonymization for different studies makes it possible to integrate information later. Obviously it also raises the risk of re-identification

✔ *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*

✔ *b. Yes*

`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*

✔ *b. Yes*

`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?**

✔ *PI of the group generating the data+PI of the group that did the processing*

`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*

✔ *b. Yes*

`3.a.1.b.3.b.1` **Where will the data be stored?**

↗ External Links: *The European Genome-phenome Archive (EGA)*

✔ *c. In a dedicated repository*

make sure you can guarantee longevity of this repository!

`3.a.1.b.3.b.2` **Who will take care of authentication of potential users?**

✔ *b. We will use a single sign-on system such as FEIDE or ELIXIR-AAI*

`3.a.1.b.3.b.3` **Who will take care of authorization of potential users?**

🏷 Tags: *Science Europe DMP*

✔ *d. We will make other arrangements*

`3.a.1.b.3.b.3.d.1` **What other arrangements?**

🏷 Tags: *Science Europe DMP*

✔ *Owner decides on level of sharing and group to share with. ELIXIR will provide the authentication for authorized groups accessing.*

`3.a.1.b.3.b.4` **Are the criteria for application to access the data openly available (e.g . are there well described conditions for access (i.e. a machine readable license)?**

✔ *b. Yes*

3.a.1.b.3.b.5 **Has auditing for the re-use been arranged?**

✔ *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*

✔ *a. No*

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

🏷 Tags: *Science Europe DMP*

✔ *a. No*

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

🏷 Tags: *Science Europe DMP*

✔ *b. No, data will be released only as soon as restrictions are falling away*

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

✔ *b. Yes*