For datasets contributors (THL, UKE, UMG, QMUL , CAHMM)

**1. Data summary**

**1.1 Description of data**

• Specify the origin of data generated/collected

• Specify the types and formats of data generated/collected

• State the expected size of the data (if known)

1-What is the origin of the data?

2-What types and formats of data will the project generate/collect?

3-What is the expected size of the data?

If applicable please fill in the table:

**Table 1: Data types and its format**

**1.2 Data security levels, confidentiality of potentially disclosive personal information**

• Specify different levels of data security (for instance > open (aggregated statistics), ……..and controlled (personally identifiable).

**Table 2: Data types and its needs for long-term storage, security issues, required metadata and interoperability requirements**

For platforms (MuGVRE platform, ELIXIR-ES)

Which platform for each data type?

**2 Data management, documentation and curation**

the data workflow diagram Data Workflow figure 7 Deliverable D7.1.

• Specify the data workflow in your platform and how it differ from the project’s general workflow

**2.1 Data collection / generation**

General data collection/ deposition method is depicted in diagram..

• Specify the methodology of data deposition for user

• Define protocols for depositing new cohort raw data into the appropriate repositories and methods to provide rich metadata to foster a quality re-use of raw data for newly coming research projects.

Centralised platform for submitting. Data will be phisically storaged in the corresponding repositories (See data storage)

**2.2 Data storage**

• Specify the methodology of data storage

• Data workflow diagram

**2.3 Data sharing and access**

• Specify if existing data is being re-used (if any)

• Specify technology for data access and transfer

• Specify methodology of data re-use (analysis)

**3 FAIR data management**

**3.1 Making data findable, including provisions for metadata**

* Outline the discoverability of data (metadata provision)
* Outline the identifiability of data and refer to standard identifica on mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
* Outline naming conventions used
* Outline the approach towards search keyword
* Outline the approach for clear versioning
* Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
* Project-Describe relevant models, tools and infrastructures for (meta) data sharing and distribution within the network
* Project-Describe relevant models, tools and infrastructures for (meta) data sharing and distribution within the network

Q>

1- Are the data produced and/or used in the project discoverable with metadata, iden fiable and locatable by means of a standard iden fica on mechanism (e.g. persistent and unique iden fiers such as Digital Object Iden fiers)?

2- What naming conven ons do you follow?

3- Will search keywords be provided that op mize possibili es for re-use?

4- Do you provide clear version numbers?

5- What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

**3.2 Making data openly accessible**

* Specify which data will be made openly available? If some data is kept closed provide ra onale for doing so
* Specify how the data will be made available
* Specify what methods or so ware tools are needed to access the data? Is documenta on about the so ware needed to access the data included? Is it possible to include the relevant so ware (e.g. in open source code)?
* Specify where the data and associated metadata, documentation and code are deposited
* Specify how access will be provided in case there are any restrictions

1-Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restric ons), explain why, clearly separa ng legal and contractual reasons from voluntary restric ons.

Note that in mul -beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consor um agreement and are in line with the reasons for op ng out.

2-How will the data be made accessible (e.g. by deposi on in a repository)? What methods or so ware tools are needed to access the data?

3-Is documenta on about the so ware needed to access the data included? Is it possible to include the relevant so ware (e.g. in open source code)?

4-Where will the data and associated metadata, documenta on and code be deposited? Preference should be given to cer fied repositories which support open access where possible.

5-Have you explored appropriate arrangements with the iden fied repository?

6-If there are restrictions on use, how will access be provided?

7-Is there a need for a data access commitee?

8-Are there well described condi ons for access (i.e. a machine readable license)? How will the iden ty of the person accessing the data be ascertained?

* 1. **Making data interoperable**
* Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
* Specifywhetheryouwillbeusingstandardvocabularyforalldatatypespresentinyourdataset,toallowinter- disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
* Project-Outline methodology for cross-referencing between repositories (ELIXIR/EGA and euro-BioImaging), and the technical solutions for enabling efficient access to data using the technologies developed by ELIXIR, EGA and euro-BioImaging.
* Project-Outline harmonization methodology

1-Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) so ware applica ons, and in par cular facilitating recombina ons with different datasets from different origins)?

2-What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

3-Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

4-In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

databases within the infrastructure.

**3.4 Increase data re-use (through clarifying licenses)**

* Specify how the data will be licensed to permit the widest reuse possible
* Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
* Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
* Describe data quality assurance processes
* Specify the length of time for which the data will remain re-usable

1-How will the data be licensed to permit the widest re-use possible?

2-When will the data be made available for re-use? If an embargo is sought to give me to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

3-Are the data produced and/or used in the project useable by third par es, in par cular a er the end of the project? If the re-use of some data is restricted, explain why.

4-How long is it intended that the data remains re-usable? Are data quality assurance processes described?

**4 Allocation of resources**

• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs.

• Clearly identify responsibilities for data management in your project.

• Describe costs and potential value of long-term preservation

1- What are the costs for making data FAIR in your project?

2- How will these be covered? Note that costs related to open access to research data are eligible as 3- part of the Horizon 2020 grant (if compliant with the Grant Agreement condi ons).

4- Who will be responsible for data management in your project?

5- Are the resources for long term preserva on discussed (costs and poten al value, who decides and how what data will be kept and for how long)?

**5 Data security**

• Address data recovery as well as secure storage and transfer of sensitive data

1- What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

2- Is the data safely stored in certified repositories for long-term preservation and curation?

**6 Ethical aspects**

• To be covered in the context of the ethics review, ethics sec on of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Descrip on of the Ac on (DoA).

Is informed consent for data sharing and long term preserva on included in ques onnaires dealing with personal data?

**7 Other issues**

• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?