

# Data Management Plan

## 2024-2029

**Version:** 1.0

**Date of version:** 29.02.2024

**Created by:** Susana Pascoal

**Status:** In progress

**Contact:** [nimsb@unl.pt](mailto:nimsb@unl.pt)

## 1. Data Management Plan

NIMSB will strive to ensure that the very large and complex datasets that will be generated will be as FAIR (Findable, Accessible, Interoperable, Reusable) and open as possible, within legal boundaries, to ensure greater scientific and societal value.

Teaming synergies: NIMSB will create a Data Management Plan (DMP) (WP7) based on previous experience and policies in place at the MDC and NOVA University, which will ensure the most effective use of resources, transparency, re-use, participation, accountability, and reproducibility. The DMP will address all aspects of the research data life cycle (planning, data collection, management, and analysis, preservation and sharing) and follow international guidelines and legal requirements, including the EU's General Data Protection Regulation (GDPR) (EU 2016/679). There will be NO work performed on embryonic stem cells, human embryos, non-human primates, cloned farm animals, or endangered species. Personal data will be protected as samples from patients will be anonymized. There will be NO environmental damage, stigmatization, or social groups or other adverse consequences.

<b>1.1. Planning</b>	<b>1.2. Managing</b>	<b>1.3. Sharing</b>	<b>1.4. Policy</b>
1.1.1. Police and Guidelines for funders	1.2.1. Organizing data 1.2.2. Storage and Backup 1.2.3. Electronic Lab Notebook	1.3.1. Publishing data 1.3.2. Collection	1.4.1. Research Data management policy 1.4.2. Collection of funders policy and guidelines

### 1.1. Planning

A data management plan (DMP) is a document that outlines the strategies and procedures for managing research data throughout the entire life cycle of a project. It is a structured document that provides information on how data will be collected, processed, analysed, stored, shared, and preserved. The primary goals of a data management plan are to ensure the integrity, accessibility, and long-term preservation of research data, as well as to comply with any relevant policies, regulations, or funding agency requirements.

Proper data management planning was shown to save time for scientists, increase research efficiency, visibility and prevent duplication of efforts.

FAIR wizard plan (<https://fair-wizard.com/get-started>) and OpenAIRE Argos (<https://argos.openaire.eu/splash/>) are being tested to evaluate which application cope with the demands of NIMSB.

### **1.1.1. Policies and guidelines of the funders:**

1. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon (April 2017). Horizon 2020 Programme, Retrieved April 19, 2022 from [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

2. Guidelines on Implementation of Open Access to Scientific Publications and Research Data (April 2017). European Research Council (ERC). Retrieved April 19, 2022, from [https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf)

3. Sharing Research Resources (October 1, 2021). National Institutes of Health (NIH). Retrieved April 19, 2022, from [https://grants.nih.gov/grants/policy/nihgps/html5/section\\_8/8.2.3\\_sharing\\_research\\_resources.htm](https://grants.nih.gov/grants/policy/nihgps/html5/section_8/8.2.3_sharing_research_resources.htm)

Research Tools Policy. National Institutes of Health (NIH). Retrieved April 19, 2022, from <https://sharing.nih.gov/other-sharing-policies/research-tools-policy>

Where to Submit Genomic Data. National Institutes of Health (NIH). Retrieved April 19, 2022, from <https://sharing.nih.gov/genomic-data-sharing-policy/submitting-genomic-data/where-to-submit-genomic-data>

4. Research Data Management and Sharing. Fundação para a Ciência e Tecnologia (FCT). Retrieved May 2014, from [https://www.fct.pt/wp-content/uploads/2023/01/PoliticaAcessoAberto\\_Dados.pdf](https://www.fct.pt/wp-content/uploads/2023/01/PoliticaAcessoAberto_Dados.pdf)

Open Access to Scientific Publications. Fundação para a Ciência e Tecnologia (FCT). Retrieved 2014, from [https://www.fct.pt/wp-content/uploads/2023/01/PoliticaAcessoAberto\\_Publicacoes.pdf](https://www.fct.pt/wp-content/uploads/2023/01/PoliticaAcessoAberto_Publicacoes.pdf)

Política Nacional de Ciência Aberta. Resolução do Conselho de Ministros n.º 21/2016, de 11 de abril. Retrieved April 2016, from <https://diariodarepublica.pt/dr/detalhe/resolucao-conselho-ministros/21-2016-74094659>

5. Horizon Europe (HORIZON) Euratom Research and Training Programme (EURATOM) General Model Grant Agreement EIC Accelerator Contract (15 Dec 2021), Retrieved April 19, 2022 from [https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/agr-contr/general-mga\\_horizon-euratom\\_en.pdf](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/agr-contr/general-mga_horizon-euratom_en.pdf)

NIH defined that: "Scientific data does not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects such as laboratory specimens." From <https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-and-sharing-policies/research-covered-under-the-data-management-sharing-policy#after>

Other resources:

EC/Horizon 2020

Guidelines on FAIR Data Management in Horizon 2020:

[https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

Open Science: [https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science\\_en#the-eus-open-science-policy](https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en#the-eus-open-science-policy)

UNESCO

UNESCO Recommendation on Open Science:

<https://www.unesco.org/en/openscience/about?hub=686>

Overview of funder requirements for the retention and preservation of research data

<http://www.open.ac.uk/library-research-support/sites/www.open.ac.uk.library-research-support/files/files/RDM-Overview-of-Funders-Policies-updated-April-2014.pdf>

## **1.2. Managing**

### **1.2.1. Organizing data**

#### 1.2.1.1. Directory structure

A clear directory structure will make it easier to locate files and versions and this is particularly important when collaborating with others. A hierarchical and systematic file structure will be considered, starting from broad topics to more specific ones. The structure typically starts with a root directory and branches out into subdirectories.

NIMSB will have a directory structure accordingly with the UK data services guidelines:

<https://ukdataservice.ac.uk/learning-hub/research-data-management/>  
[https://datadryad.org/stash/best\\_practices](https://datadryad.org/stash/best_practices)

#### 1.2.1.2. File naming

File naming is a crucial aspect of organizing and managing digital files. A well-thought-out naming

convention can make it easier to locate, understand, and organize files. Below some best practices for file naming:

- **Be Descriptive:**

Use descriptive and meaningful names that reflect the content or purpose of the file. Avoid generic names like "Untitled" or "Document."

- **Include Dates:**

For files associated with specific dates, consider including a date in the file name. Use a consistent date format, such as YYYY-MM-DD, to ensure proper sorting.

- **Use Underscores or Hyphens:**

Use underscores (\_) or hyphens (-) to separate words in a file name. This enhances readability and can help with searchability.

- **Avoid Special Characters:**

Avoid using special characters, spaces, or symbols that may cause issues on different operating systems or in web environments. Stick to alphanumeric characters, underscores, and hyphens.

- **Consistent Capitalization:**

Choose a consistent capitalization style (e.g., CamelCase, Title Case, lowercase) and apply it consistently across your files.

- **Include Version Numbers:**

If your files undergo multiple versions, consider including a version number in the file name. This can be especially helpful for documents or projects that undergo frequent updates.

- **Short and Concise:**

Keep file names short and to the point while still being descriptive. Avoid overly long names that may become cumbersome.

- **Use Leading Zeros for Numbers:**

When including numbers in file names, use leading zeros to ensure proper sorting. For example, "001", "002", "010" instead of "1", "2", "10".

- **Organize by Project or Category:**

If your files belong to specific projects or categories, consider including a project or category identifier in the file name.

- **Avoid Special Keywords:**

Avoid using special keywords that may conflict with the operating system or software you're using.

- **Consider File Type Extensions:**

If applicable, include the file type extension in the name to quickly identify the file format (e.g., .docx, .pdf, .jpg).

Examples of good file names based on these practices:

- 2024-02-26\_Project\_Report.docx
- NIMSB\_Presentation\_v002.pdf

Establishing and adhering to a consistent file naming convention will help to easily manage and navigate

through digital files.

- Maximum number of characters recommended is 30 per file name.
- Don't use special characters `^!@%$#\"&/()=?»}][{£`

#### 1.2.1.3. File format

File format refers to the structure or arrangement of data within a computer file. It defines how the data is encoded, organized, and stored, allowing software applications to interpret and process the information correctly. The choice of file formats plays an essential role for long term data storage and archiving, data sharing, searchability, accessibility, and has a significant impact on data reusability. It is advisable to consider open files format that allows the data to be imported and accessed by different tools and not vendor locked in case a tool is no longer supported.

Consider the following:

- Choose standard file formats most commonly used in your field.
- Convert data to a standard format.
- Choose a format which is required for data deposition i.e. repository requirements, archival compression.
- Consider exporting or converting from original format to a more open/preferred format but keep in mind that some data might be lost or altered during the process e.g., text formatting in documents, decimal point formatting, date and time values.
- Keep in mind there are no standard preferred file formats, and none are perfect, but consider choosing open formats that are most applicable for your use and field.
- When archiving data, combine the whole project (i.e., raw data, analysis, documentation, code and software) in one package.
- For software consider the use of containers to enable interoperability and long-term re-use.

Recommended file formats by the UK data archive:

<https://ukdataservice.ac.uk/learning-hub/research-data-management/>

Dryad guide: [https://datadryad.org/stash/best\\_practices#accessible](https://datadryad.org/stash/best_practices#accessible)

#### 1.2.1.4. Quality Control

Quality control is fundamental to ensure the reliability, validity, and integrity of research findings. It involves systematically monitoring and assessing various aspects of the research process to identify and address any potential issues or errors. Consider the following quality control measures:

- Data collection
  - Outline the number of measurements/samples/procedures repeated.
  - Outline instrument calibration tests & data set or samples used for calibration.
  - Outline standardized controls (e.g., sample controls).
  - Use of standardized protocols and methods with clear instructions and documentation.
- Data entry

- Decide a method for documentation i.e., Electronic lab notebooks vs paper.
- Outline the non-digital data structure and strategy for digitization.
- Collect and create metadata throughout the data collection and handling process.
- Use controlled vocabularies.
- Outline how the data/samples/variables are labelled.
- Document terminology used.
- Describe how to flag/tag questionable data.
- Ensure data and time is represented in a machine-readable format and valid.
- Set up validation rules or input masks in data entry software.
- Data Analysis and checking
  - Outline software/code used for analysis.
  - Outline strategy for data transfer and controls.
  - Outline how the data will be cross-checked and validated.
  - Assign person/expert for quality assurance and data checks and/or peer review.
  - Outline database structure to organize data and data files.
  - Document any modifications and outline versioning strategy to avoid duplicate error checking.
  - Check and flag questionable data.
  - Verify your analysis by using a random data set/samples compare to original data.
  - Double-check the code for any errors and ensure appropriate documentation.
  - Use statistical analysis to detect erroneous and/or anomalous values.
- Qualitative data

Qualitative data can be collected through various methods such as interviews, focus groups, observations, open-ended surveys, and content analysis.

- For qualitative data such as interviews:
  - Outline guided interview questions.
  - Make use of software tools such as text to speech.
  - Control the quality of audio/video/transcripts files.

Data on quality control (UK guidelines):

<https://www.kcl.ac.uk/researchsupport/managing/organise>  
<https://ukdataservice.ac.uk/learning-hub/research-data-management/>  
<https://rd-alliance.github.io/metadata-directory/standards/>

#### 1.2.1.5. Versioning

Versioning refers to the practice of keeping track of different versions or revisions of a document, file, software, or any other type of information. This is particularly important in collaborative environments or when managing the development and evolution of a product or project over time. Version control helps prevent data loss, allows collaboration, and facilitates the tracking of changes.

Consider the following practices:

- Outline the master file and identify major files for instance; original, pre-review, 1st revision, 2nd revision, final revision, submitted.
- Outline strategy for archiving and storing: Where to store the minor and major versions, how long will you retain them accordingly.
- Maintain a record of file locations, a good place is in the README files.
- Record any related files and documents and any updates/changes made to them.
- Use a systemic and unique naming system to identify the different versions, e.g., numbers and/or dates.
- Include a version control table that outlines the file history, which version, where the other versions are located, list all associated files and their versions and modifications, add dates, authors, access rights, licensing, and details of changes made since the last version.

#### 1.2.1.6. README files

A README file is a document, usually in plain text format, that provides essential information about a project, directory, or code repository. README stands for "Read Me" and serves as a guide for users or developers to understand and navigate the content of a project. This can be very helpful in case to sharing data with others, or to keep track of content and edits or changes made in multiple projects.

A README file is better suited for a collection of data such as a directory for a specific project or experiment, software tool, or any data that is related to each other "logically".

Place the README file in a parent directory associated with the content described.

Use plain markdown or a simple text editor to create the README file in either .md or .txt file format.

#### 1.2.2. Storage and Backup

To store all research data at NIMSB, storage solutions will be provided accordingly with the data produced preserving digital information in a structured manner. Data backup involves creating duplicate copies of data to protect against loss due to hardware failure, human error, cyber attacks, natural disasters, or other unforeseen events.

For data storage, data can be stored on various types of storage media, including hard disk drives (HDDs), solid-state drives (SSDs), magnetic tapes, cloud storage, network-attached storage (NAS) and storage area network (SAN).

The research data management office from NIMSB will select appropriate storage solutions based on their specific requirements, such as data volume, access speed, durability, and budget constraints.

Guidelines documents will be produced for:

- storage, backup, archival and disposal;
- archival, destruction and retention periods;
- and in case of sensitive data.

### **1.2.3. Electronic Lab Notebook (ELN)**

An electronic lab notebook (ELN) is a digital tool used by researchers, scientists, and other professionals to document and manage their laboratory work, experiments, observations, and research findings in an electronic format, similar to traditional paper lab notebooks. ELNs serve as a central repository for recording experimental data, procedures, protocols, observations, and analysis results. However, ELNs offer several advantages over traditional paper notebooks, including:

- **Digital Format:** ELNs are digital, allowing users to create, store, search, and access their laboratory records electronically. This enables easier organization, retrieval, and sharing of data compared to paper-based notebooks.
- **Collaboration:** ELNs facilitate collaboration among team members by providing a centralized platform for sharing research data, protocols, and findings. Users can collaborate in real-time, share notes, and provide feedback, regardless of their physical location.
- **Data Integrity:** ELNs often include features such as electronic signatures, timestamps, and version control to ensure the integrity and authenticity of laboratory records. This helps meet regulatory compliance requirements and maintain data traceability and auditability.
- **Search and Retrieval:** ELNs offer powerful search capabilities, allowing users to quickly find and retrieve specific experiments, data sets, or observations using keywords, tags, or metadata. This saves time and reduces the risk of data loss or misplacement.
- **Integration:** ELNs can integrate with other laboratory software and tools, such as laboratory information management systems (LIMS), data analysis software, instrumentation, and databases. This seamless integration streamlines data capture, analysis, and reporting workflows.
- **Security:** ELNs include security features to protect sensitive research data from unauthorized access, manipulation, or loss. This may include user authentication, access controls, encryption, and backup and disaster recovery mechanisms.
- **Compliance:** ELNs help researchers adhere to regulatory requirements, such as Good Laboratory Practice (GLP), Good Manufacturing Practice (GMP), or data integrity guidelines. ELNs may offer compliance features, such as electronic signatures, audit trails, and data validation checks.

Having an institutional ELN directly responds to the increasing demands of multiple funding agencies and research journals. NIMBS will do a subscription to an electronic lab notebook tool. At this moment we are considering several ones that are in the market, as RSpace (<https://www.researchspace.com/>) and Elab FTW (<https://www.elabftw.net/>). Once implemented, training and FAQs will be provided by the research data management office.

## **1.3. Sharing**

### **1.3.1. Publishing data**

When discussing the publishing of data, it's important to emphasize the significance of making research

data openly accessible to the scientific community. Open science policies will be described in the Delivery D7.2.

For that some points should be considered:

- **Transparency and Reproducibility:** Publishing data promotes transparency and reproducibility in research. By sharing data sets, researchers allow others to verify and build upon their findings, contributing to the credibility and reliability of scientific research.
- **Maximizing Impact:** Making data publicly available can maximize the impact of research by enabling broader access to valuable resources. This facilitates collaboration, fosters interdisciplinary research, and accelerates scientific progress.
- **Compliance with Funding Agency Policies:** Many funding agencies and research institutions require researchers to publish data resulting from publicly funded projects. Compliance with these policies ensures accountability and transparency in the use of public funds.
- **Citation and Recognition:** Publishing data allows researchers to receive proper citation and recognition for their work. Data sets are assigned unique identifiers (e.g., DOIs) that can be referenced in publications, increasing visibility and facilitating attribution.
- **Data Sharing Platforms:** Researchers can publish data on various data sharing platforms and repositories tailored to specific disciplines or types of data.
- **Data Management and Documentation:** Prior to publication, researchers should ensure that their data sets are well-managed, documented, and annotated to facilitate understanding and reuse by others. This includes providing detailed metadata, descriptions of experimental protocols, and explanations of data formats.
- **Data Access and Licensing:** Researchers should consider licensing options for their data sets to specify usage rights and restrictions. Common licenses for research data include Creative Commons licenses, which allow users to reuse data with proper attribution under specified conditions.
- **Ethical Considerations:** Researchers should adhere to ethical guidelines and regulations governing data sharing, including data privacy, confidentiality, and informed consent. Sensitive or personally identifiable information should be anonymized or appropriately managed to protect participants' rights and privacy.
- **Long-Term Preservation:** Consideration should be given to the long-term preservation and stewardship of published data sets. Data repositories often provide mechanisms for ensuring data integrity, accessibility, and preservation over time.
- **Community Engagement:** Engaging with the research community through conferences, workshops, and online forums can help raise awareness of published data sets and foster collaboration and feedback from fellow researchers.

Researchers can meet this condition by submitting their data to the repository that assigns persistent identifiers (DOIs and accession numbers) to the data. Such data is properly citable and can be referenced. Information of how to choose a repository will be given, to identify whether commonly used

discipline-specific, community-recognized repository exists to submit the data there and if not to deposit in a more generalist repository. Funders and publishers usually list eligible repositories on their websites, a summary of the repositories will be made to be available to the NIMSB community.

### **1.3.2. Collection of requirements from publishers for storing data into repositories**

A summary/table of the requirements/recommendation from publishers for storing data into repositories with an appropriate license will be created focusing on the following points:

- if is a discipline-specific repository or
- a generalist repository,
- the license and
- how data will be accessible

## **1.4. Policy**

### **1.4.1. Research data management policy**

A policy framework for research data management (RDM) will be produced and submitted to the NIMSB Scientific Advisory Board with focus on:

A) Jurisdiction

B) Intellectual property rights

C) Handling research data

- Data storage
- Data access and re-use
- Deletion of data

D) Responsibilities, rights and duties

- Principal Investigators and Researchers are responsible for
- The NIMSB is responsible for

### **1.4.2. Collection of funders policy and guidelines**

The requirements for the retention and preservation of research of the top funders will be collected and summarized to be accessible for the NIMSB community.

Guidelines to be followed: see point **1.1.1**

### **Preliminary Template:**

Annex - NIMSB Data Management Plan template.

## NIMSB Data Management Plan Template

Project information	
Contact information	
Title of project	
Principal Investigator	
Project duration	
Funder/Funding scheme (if applicable)	

Collecting data and study materials	
Questions	Details
<ul style="list-style-type: none"> <li>• <b>Are these new or existing data/materials?</b></li> <li>• <b>Type(s)</b></li> <li>• <b>Format(s)</b></li> <li>• <b>Methods of data/materials collection</b></li> <li>• <b>Approaches to ensuring quality</b></li> <li>• <b>Expected volume</b></li> <li>• <b>Indicate what documentation will accompany the data</b></li> <li>• <b>Indicate which metadata will be provided to help others identify and discover the data.</b></li> </ul>	<p>Explain which existing data you will re-use and state any constraints on re-use of existing data if there are any.</p> <p>Explain which methodologies or software will be used if new data are collected or produced.</p> <p>Give details on the kind of data: for example numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed media.</p> <p>Give details on the data format: the way in which the data is encoded for storage, often reflected by the file name extension (for example pdf, xls, doc, txt, or rdf).</p> <p>Justify the use of certain formats. For example, decisions may be based on staff expertise within the host organisation, a preference for open formats, standards accepted by data repositories, widespread usage within the research community, or on the software or equipment that will be used.</p> <p>Give preference to open and standard formats as they facilitate sharing and long-term re-use of data. Several repositories provide lists of such 'preferred formats'.</p> <p>How much storage will your project require in total: (10 GB; 10 – 100 GB; 100 – 1000 GB; &gt;1000 GB)</p> <p>Consider what other documentation is needed to enable re-use. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, and so on. Consider how this information will be captured and where it will be recorded, for example in a database with links to each item, a 'readme' text file, file headers, code books, or lab notebooks.</p> <p>Indicate how the data will be organised during the project, mentioning for example conventions, version control, and folder structures. Consistent, well-ordered research data will be easier to find, understand, and re-use.</p>

	<p>To be findable, accessible, interoperable and reusable, data must be accompanied with descriptive information in the form of metadata.</p> <ul style="list-style-type: none"> <li>- Where these are in place, researchers are advised to use community metadata standards.</li> <li>- Depositing data in a certified or trustworthy repository will typically involve providing information about the data according to a metadata standard scheme and specify here which one are used.</li> </ul>
--	---

<b>Storage, security and back up</b>	
--------------------------------------	--

<b>Questions</b>	<b>Details</b>
------------------	----------------

- **Where will you store your data/study materials?**
- **Approaches to securing data and study materials**
- **What is your backup strategy?**
- **Are you using any personal, identifiable or pseudonymised data? (If so, which data protection legislation apply to your research?)**
- **How will you handle issues regarding the processing of personal information and intellectual property rights and ownership?**

Describe where the data and metadata will be stored and backed up during the project:  
(Institution networked research storage or Other (please specify))

Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of NIMSB. Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended.

How will data security and protection of sensitive data be taken care of during the research?  
(Not applicable (no sensitive data); Default security measures of NIMSB networked research storage; Additional security measures (please specify))

Consider data protection, particularly if your data is sensitive – for example, containing personal data, politically sensitive information or information relating to religion and health, trade secrets or national security information. Describe the main risks and how these will be managed. Explain how the data will be recovered in the event of an incident. Explain who will have access to the data during the research and how access to data is controlled, especially in collaborative partnerships.

Explain which institutional data protection policies are in place.

Will you process and/or store personal data during your project? If yes, ensure that when dealing with personal data, data protection laws (for example GDPR) are complied with: ~

- Gain informed consent for preservation and/or sharing of personal data.
- Consider anonymisation of personal data for preservation and/or sharing (truly anonymous data are no longer considered personal data).
- Consider pseudonymisation of personal data (the main difference with anonymisation is that pseudonymisation is reversible).
- Consider encryption which is seen as a special case of pseudonymisation (the encryption key must be stored separately from the data, for instance by a trusted third party).
- Explain whether there is a managed access procedure in place for authorised users of personal data.

Explain who will be the owner of the data, meaning who will have the rights to control access. Make sure to cover the ownership of the data and intellectual property rights to control access for multi-partner projects and multiple data owners in the consortium

agreement. Indicate whether there are any restrictions on the re-use of third-party data.

Archiving, preservation and curation	
Questions	Details
<ul style="list-style-type: none"> <li>• <b>What data / physical items do you plan on archiving and preserving?</b></li> <li>• <b>How long for will you preserve the data/output?</b></li> <li>• <b>How and where will you store these items?</b></li> <li>• <b>Which unique identifiers do you plan on using?</b></li> <li>• <b>How will you curate the data/research materials on a medium to longer-term basis?</b></li> <li>• <b>If applicable, how will you dispose of any research materials?</b></li> </ul>	<p>All data resulting from the project will be preserved for at least ___years (for example 10 years) or Other (please specify)</p> <p>Indicate what data must be retained or destroyed for contractual, legal, or regulatory purposes. Indicate how it will be decided what data to keep. Describe the data to be preserved long-term.</p> <p>For storage, have you considered the formats of these items? Have you considered digitising physical items? And about software code, simulated data, databases structures?</p> <p>Describe which identifiers will be used like ORCID for researchers, DOIs for data, publications, metadata records, study protocols, posters, presentations, etc.</p> <p>Once the retention period has passed and you feel that the data is no longer of value or to meet ethical requirements it can be disposed.</p> <p>When data is destroyed it must be irreversible with no chance of recovery. Paper can be shredded using an office shredder. Extra care should be taken with sensitive or confidential information and a secure paper destruction service bin used.</p> <p>Digital data may be destroyed by deleting or overwriting information, purging magnetic media through degaussing (exposure to a strong magnetic field), or destroying the physical media (e.g. CD-ROMS, DVDs).</p>

Discovery, access and sharing	
Questions	Details
<ul style="list-style-type: none"> <li>• <b>How do you plan on making your research discoverable?</b></li> <li>• <b>Are there any (potential) limits on data access and sharing?</b></li> <li>• <b>Which usage license(s) do you plan on applying to items you plan on archiving?</b></li> </ul>	<p>As much as possible, research data should be made publicly available for re-use.</p> <p>Indicate what data will be made available for re-use. This may differ from the data that is preserved, when the data are so large that it is unfeasible to deposit the data in a repository in its entirety, or if there are reasons that prohibit making data available for re-use. If there are any restrictions on the re-use of the data, or if it is necessary to restrict access to certain parts of the data or to apply a data sharing agreement, explain how and why. Explain what actions will be taken to overcome or to minimize restrictions. Check the guidelines of the funders, in case they exist.</p> <p>Indicate where the data will be deposited and made available for re-use. Repository Finder can help you find an appropriate repository to deposit your research data.</p> <p>Indicate whether a persistent identifier will be pursued. Typically, a trustworthy, long-term repository will provide a persistent identifier.</p> <p>Indicate under which license the data may be re-used. Check the commonly used Creative Commons licenses (<a href="https://ufal.github.io/public-license-selector/">https://ufal.github.io/public-license-selector/</a>).</p> <p>Indicate whether the repository is certified. In case no such repositories can be found or are suitable, under adherence to the</p>

	<p>following minimum selection criteria: provision of persistent and unique identifiers; use of metadata standards that are broadly accepted by the scientific community; provision of information that is publicly available; enabling access to data under well-specified conditions and following open and standard access protocols; provision of information about licenses and permissions; ensuring persistence of data and metadata. Indicate whether potential users need specific tools or software (e.g. specific scripts, codes or algorithms developed during the project) to access, interpret and (re-)use the data. Indicate how these items will be made available.</p>
--	--

<b>Metadata</b>	
<b>Questions</b>	<b>Details</b>
<ul style="list-style-type: none"> <li>• <b>What kinds of metadata do you plan on providing?</b></li> <li>• <b>In which formats will metadata be available?</b></li> <li>• <b>What types of metadata will you provide?</b></li> <li>• <b>Will you apply a metadata standard?</b></li> <li>• <b>Will there be any restrictions on the availability of the metadata? E.g. embargo?</b></li> </ul>	<p>Describe the types and format of metadata that will be provided, e.g. data dictionaries, codebooks, README files, technical descriptions, commentary for coding, audit trails, etc. Format in XML, PDF, .csv, etc.</p> <p>A list of metadata standards and other research-related standards and policies can be found here: <a href="https://fairsharing.org/">https://fairsharing.org/</a></p>

<b>Rights and responsibilities</b>	
<b>Questions</b>	<b>Details</b>
<ul style="list-style-type: none"> <li>• <b>Who is/are the copyright owner(s)?</b></li> <li>• <b>Who will create and maintain the metadata?</b></li> <li>• <b>Who will manage the research data/study materials: a) during the project; b) once the project is complete?</b></li> <li>• <b>Who will oversee compliance with relevant applicable laws?</b></li> </ul>	