



Data Management

D5.1 Data Management Plan: 1st Edition

May, 2023



Funded by
the European Union

Technical References

Project Acronym	BIOTraCes
Grand Agreement Number	101081923
Project Title	BIOdiversity and Transformative Change for plural and nature positive societies
Project Coordinator	Rosalie van Dam (WR)
Project Duration	2022 – 2026 (4 years)

Deliverable No.	D5.5
Dissemination level¹	PU
Work Package	WP5
Task	5.5
Lead beneficiary	MRU - Environmental Psychology Research Centre, Mykolas Romeris University Lead authors: Audra Balundė, Aistė Bakaitytė, Goda Kaniušonytė
Contributing beneficiary(ies)	Contributing Authors: BC3, CER, CES, UBB, UGOT, UNICT, UT & WR
Due date of deliverable	May 31 st , 2023
Actual submission date	May 31 st , 2023

- 1 PU = Public
 PP = Restricted to other programme participants (including the Commission Services)
 RE = Restricted to a group specified by the consortium (including the Commission Services)
 CO = Confidential, only for members of the consortium (including the Commission Services)

Document history

V	Date	Beneficiary	Author
V0.1	05-04-2023	MRU	Audra Balundė Aistė Bakaitytė Goda Kaniušonytė
V0.2	26-04-2023	MRU	Aistė Bakaitytė Goda Kaniušonytė
V0.3	10-05-2023	WUR	Rosalie van Dam Caroline van Oostveen
V0.4	26-05-2023	MRU	Audra Balundė Aistė Bakaitytė Goda Kaniušonytė

Summary

The Data Management Plan (DMP) offers guidelines to researchers working on the BIOTraCes project. These guidelines will inform how data should be organised and stored. The objective of this plan is to ensure that every bit of data is properly documented, thereby preventing any ethical concerns or confusion in the future, both for the researchers themselves and for others seeking to replicate the study. All researchers involved in the BIOTraCes project are obliged to document significant research steps and store their digital data, along with supporting documentation in reliable repositories. Additionally, alongside the data (both quantitative and qualitative), researchers are required to document their analysis methods and other crucial details to enable the tracing and monitoring of the methodological steps and decisions made within each case study. Some instructions indicated in the DMP (e.g. file naming instructions) are tentative and subject to change, when new more efficient instructions are established.

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.

Table of contents

Technical References 2

Document history 3

Summary 4

Disclaimer 4

1 INTRODUCTION 6

1.1 Project description and objectives..... 6

1.2 DMP description and objectives..... 6

2 DATA SUMMARY 7

3 FAIR data 10

3.1 Making data findable, including provisions for metadata 10

3.2 Making data accessible 10

3.3 Making data interoperable..... 15

3.4 Increase data re-use 16

4 Other research outputs 18

5 Allocation of resources 19

6 Data security 20

7 Ethics 21

8 Other issues 22

1 INTRODUCTION

1.1 Project description and objectives

The key objective of the BIOTraCes project is to co-produce knowledge enabling to develop approaches and strategies that contribute to transformative changes. These changes are necessary to preserve and restore biodiversity across Europe. The objective will be achieved by building upon principles of **pluralising, empowering, politicising and embedding**, developing capacities for innovation and fostering transformative (i.e., adaptive, plural and equitable) governance approaches to achieve just and nature-positive societies. The following are objectives that will specify the main objective.

- O1: Understand the role of diverse values, knowledge systems, power, and behaviour in transformative biodiversity approaches.
- O2: Demonstrate practices and key principles of transformative change for nature-positive societies.
- O3: Develop strategies to aid transformative (i.e., integrative, adaptive, inclusive and pluralistic) governance approaches.
- O4: Contribute to propelling transformative changes for biodiversity, local to global.

To achieve these objectives, various forms of outputs and knowledge will be developed, for example, datasets. In the following sections and chapters the data handling procedures will be described.

1.2 DMP description and objectives

Deliverable 5.5, the "Data Management Plan" (DMP), serves as a guide for researchers in the BIOTraCes consortium on how project data will be handled throughout the project's lifecycle and beyond. To drive transformative changes in the domain of biodiversity conservation, a variety of research methods will be employed, including participatory action research, policy analysis and surveys. The DMP identifies key strategies and steps to ensure that all data generated from these methods, whether qualitative or quantitative, are secure, sustainable, accessible (when possible), reusable and comply with quality standards and guidelines outlined by the Horizon Europe programme and are rooted in the ¹FAIR data principles. The DMP facilitates uniform and responsible treatment of data across the consortium in all stages of the project, including planning, collection, processing, organisation and storage. The document also specifies reliable tools for data treatment in these stages. The DMP ensures compliance with ethical standards for data treatment derived from renowned international regulations, as well as national data management standards. Examples of document envisioning ethical standards include the European Code of Conduct for Research Integrity issued by the European Federation of Academies of Sciences and Humanities (ALLEA) and the European General Data Protection Regulation (GDPR). In conformity with the commitments outlined in the Grant Agreement, the current version of the DMP will be updated two more times before months 24 and 45 but is subject to change when necessary.

¹ FAIR: Findable, Accessible, Interoperable, Reusable.

2 DATA SUMMARY

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

- We do not intend to re-use existing datasets, but if such need will occur current DMP will be updated to explain how reused data will be handled procedures- and ethics-wise.

What types and formats of data will the project generate or re-use?

- Multiple research methods used in this project will lead to various types of (non)digital data generation stored in open-sourced **numerical (e.g., csv), textual (e.g., txt, pdf, LaTeX), visual (e.g., jpeg, mp4, gif), audio (e.g., mp3, FLAC), and mixed media (e.g., svg, gzip) files.**

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

- In relation to specific objective indicated in the Grant Agreement (O2) - **Demonstrate practices and key principles of transformative change for nature-positive societies** – nine in-depth case studies of transformative biodiversity innovations in four high-impact sectors will be conducted. Each case study will organise at least five interactive workshops. In these case studies contextual knowledge on various variables will be gathered by the means of participatory action research and beyond. Examples of specific quantitative and qualitative methods and approaches that will be used for **pluralising, empowering, politicising, and embedding** project results are indicated in **Figure 1**. By using these methods, the data will be collected to uncover the connections between values, power lock-ins, behaviour, and leverage points that prevent or enable biodiversity conservation across Europe.

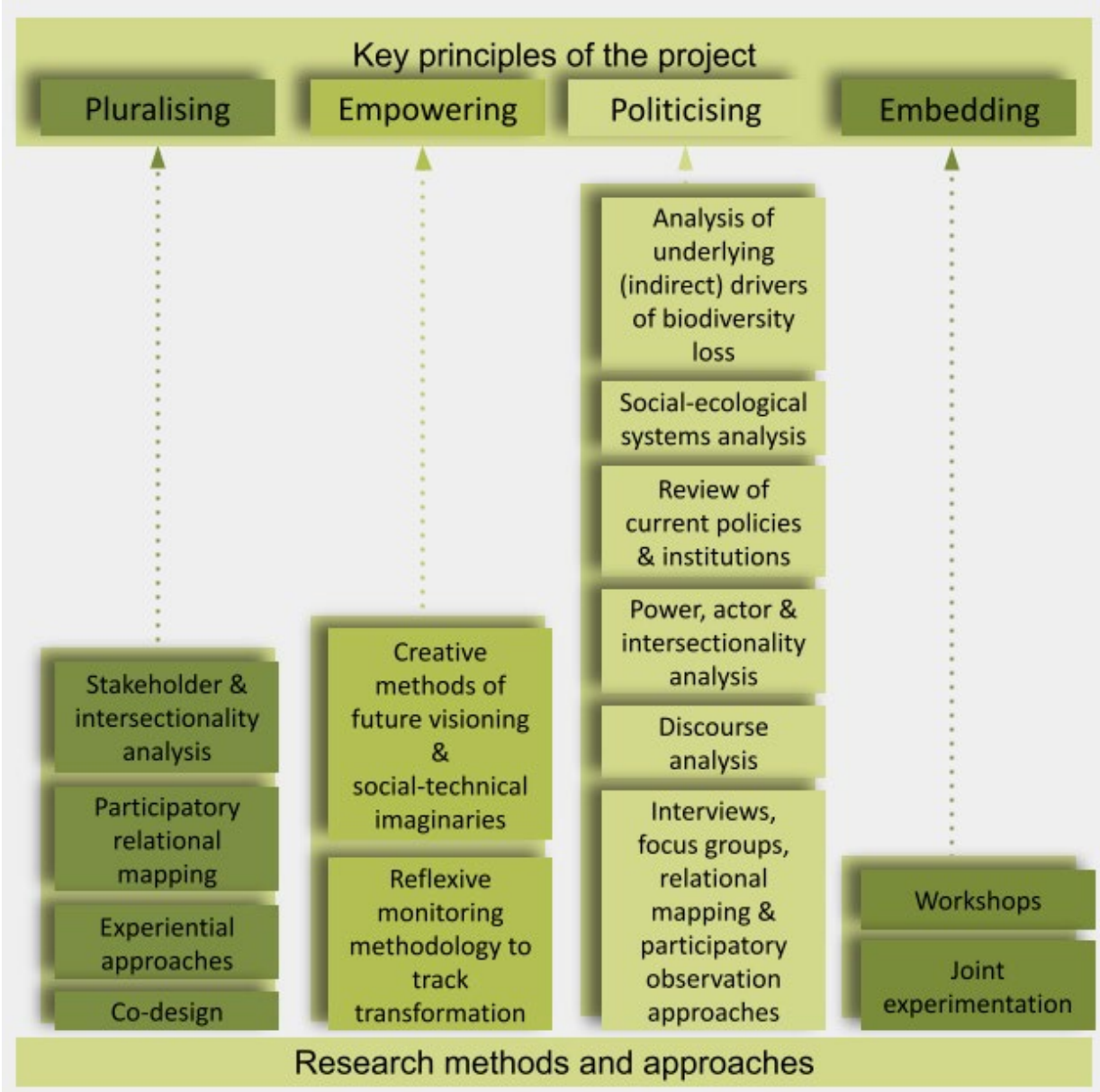


Figure 1: Examples of research methods and approaches to practise key principles of the project.

What is the expected size of the data that you intend to generate or re-use?

- At this stage there are no indicators to make such a prediction but given the multiple sources and formats of the data intended to generate during this project across nine case studies it is expected to have **rich amounts of qualitative and/or quantitative data**.

What is the origin/provenance of the data, either generated or re-used?

- The **data will derive from** interaction with various stakeholders (e.g., case study-related community members, EU and local policy makers, influencers, representatives of science-policy interface - IPBES, IPCC, CBD, IUCN), internal project team interaction and observations (e.g., from reflexivity practices), experiences (e.g., during interactions with stakeholders), content analysis (e.g., policy documents, historical artefacts, interviews, focus groups, workshops, mapping), surveys (e.g., on topics relevant to case study) and other sources relevant for each case study.

To whom might your data be useful ('data utility'), outside your project?

- We expect that data as well as other relevant outputs produced over the course of this project might be relevant but not limited to (1) scientific community, in particular transformative change and biodiversity projects, (2) science-policy interface (e.g., IPBES, IPCC, CBD, IUCN), (3) EU policy makers, (4) policy makers at the national level in member states of case studies, (5) finance and business leaders connected to the case studies, (6) members of the Stakeholder board and (7) transformative biodiversity innovators.

3 FAIR data

3.1 Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

- Each datafile prepared for publication (i.e., all sensitive personal information removed that can link to one's identity) in a repository will be given a persistent identifier, namely Digital Object Identifier (DOI). Datafiles containing raw data will contain no persistent identifier and will be handled according to ethical standards of the partner's institution who generated the data.

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

- All data that will be stored in Zenodo² will be provided with rich metadata. Metadata will include the type of data and subtype (e.g., image, video/audio, dataset), DOI, title, authors, description, and other information if needed. Metadata will be created following DataCite Metadata Schema³ requirements.

Will search keywords be provided in the metadata to optimise the possibility for discovery and then potential re-use?

- The keywords will be indicated in the metadata that reflect the nature, content and other important properties of the data.

Will metadata be offered in such a way that it can be harvested and indexed?

- The metadata of each record will be indexed and searchable directly in Zenodo's search engine immediately after publishing. As described later in the document, records will be named uniformly by creating a naming system for the datafiles.

3.2 Making data accessible

Repository:

Will the data be deposited in a trusted repository?

- **Final anonymised versions of the data** will be deposited in a trusted repository, namely Zenodo⁴. The repository is hosted by the reputable intergovernmental research institution – CERN, which is also one of the funding

² <https://zenodo.org/>

³ <https://about.zenodo.org/policies/>

⁴ <https://zenodo.org/>

bodies together with European Commission via the OpenAIRE⁵ projects, US National Institutes of Health via the GREI initiative⁶, Arcadia Fund⁷ and others. There are many indicators pointing out that data deposited in Zenodo will stay safe⁸. First, CERN as a main host of Zenodo, secured funding for their research for 20+ years; this suggests that it is highly unlikely a repository used for sharing CERN data will be unmaintained or closed. Second, CERN and Zenodo policies meticulously describe data handling and security protocols and take strict security measures⁹. Third, all data and metadata files hold multiple online and independent replicas deposited in other places, suggesting that in case of disruptions in one repository data could be retracted from another source with no risk to lose the data. Fourth, in case of an unexpected event that could lead to closing of Zenodo it is guaranteed that all data and metadata related to it will be transferred to another equivalent repository. Fifth, ZENODO provides references to relevant tools, protocols and approaches, among others, AMNESIA¹⁰, JSON schema¹¹, DataCite's Metadata Schema¹² and OpenAIRE¹³, that enables OPEN (meta)data handling and management.

- Where applicable and in compliance with institutional policies, data sets will be registered in Research Information Systems (e.g. Pure) to make the data findable and accessible in/via institutional repositories.
- **Raw data files** containing personal and sensitive information will be stored in trusted local institutional repositories in compliance with the local ethical standards and regulations. These files will be curated by the representative of each consortium institution. The representative person (**i.e. data steward**) is already assigned from each consortium institution and is indicated in the internal project's documents. Consortium partners will be obliged to ensure that the repository is trusted, to keep details of the repository and records about data handling history. Consortium coordinator and partners will collectively discuss and agree upon terms of keeping raw data and when (if at all) data will be destroyed.

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

- The properties of the Zenodo repository are fully capable of satisfying the needs of the BIOTraCes project, as well as the requirements raised by the funding body. The repository provides unlimited free and open access storage for a variety of project outputs and includes guiding tools to manage the process effectively.

⁵ <https://www.openaire.eu/>

⁶ <https://datascience.nih.gov/data-ecosystem/generalist-repository-ecosystem-initiative>

⁷ <https://www.arcadiahfund.org.uk/>

⁸ <https://about.zenodo.org/policies/>

⁹ <https://about.zenodo.org/infrastructure/>

¹⁰ <https://amnesia.openaire.eu/>

¹¹ <https://zenodo.org/schemas/records/record-v1.0.0.json>

¹² <https://schema.datacite.org/>

¹³ <https://guidelines.openaire.eu/en/latest/#>



Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

- As is indicated earlier in the document each datafile prepared for publishing in a repository will be given a persistent identifier, namely Digital Object Identifier (DOI).

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

- As is envisioned in the Grant Agreement to maximise the impact of results and to facilitate their dissemination and application the project team will comply with the latest guidelines and established practices of open science. BIODataCes will provide open access to all research outputs such as data through deposition in trusted repositories. It is not intended to close access to the (if necessary) anonymised data, yet if such need occurs the DMP will be updated in this regard accordingly before regular terms.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

- An embargo period for data is not considered in this project. Yet if such need will occur, this information will be discussed in the DMP and the DMP will be updated accordingly before regular terms.

Will the data be accessible through a free and standardized access protocol?

- All quantitative or qualitative data generated in this project will be accessible via free and standardised access protocol¹⁴, namely it will be placed in files that are readable by open-source tools. For example, numerical data will be stored in **.csv** files, text-based data in **.txt** files, visual and audio data in **.mp4** files, etc. In case other types of files readable by open-source tools will be used, the DMP will be updated accordingly.
- To facilitate the findability and management of datafiles, a tentative File Naming Convention (FNC) was developed based on OpenAIRE standards¹⁵ for openly accessible files (e.g., anonymized datasets). The file-naming system will enable consortium members and users to efficiently operate datafiles and understand, by file names, what they contain and how they are related to other datafiles. Datafile names can be generated by using the following keywords: **project acronym, date (i.e., yyyyymmdd), work package number, partner name, version of the datafile** and **edition number**. Researchers

¹⁴ <https://www.go-fair.org/fair-principles/a1-1-protocol-open-free-universally-implementable/>

¹⁵ <https://www.openaire.eu/guides/>

will be encouraged to upload final versions of the datasets. Yet it might happen that some minor changes to the final data(set) need to be applied. In that case version numbers (e.g. v1) could be indicated and updated when slight changes occur. These changes include error correction or other minor changes that do not change the content of the data itself. For example, researchers could come up with better structure/categories of the data, but content remains the same. In that case the file with new data categorization is indexed as v2, while the initial file is indexed as v1. Further, there might be cases when additional data is collected based on the same categories and structure. In this case, edition correction might be made - ed2, while the initial file is always indexed as ed1..

- Example of the naming of the initial datafile: QL_BIOTraCes_20231110_WP2_WUR_v1_ed1.txt
- Example of file naming with slight corrections that does not change the content of the data: QL_BIOTraCes_20231110_WP2_WUR_v2_ed1.txt
- The example of file naming when more data is added to the datafile: QL_BIOTraCes_20231110_WP2_WUR_v2_ed2.txt

- There is a need to add yet another identifier to the name, namely a prefix, that would help to identify the nature of the datafile and distinguish data files from other type of files that will be made openly available. For example, prefix **NM** could be used for numerical data, **QL** for qualitative data such as interviews, **VD** could be used for video files, **AD** for audio files, **WF** for workflows, etc. **Figure 2** provides example of the datafile name. QL_BIOTraCes_20231110_WP2_WUR_v1_ed1.txt.
- All file versions and editions will be stored in Zenodo. Consistent file naming will allow interested parties to follow the order and development of the files and discriminate which one is the newest version.

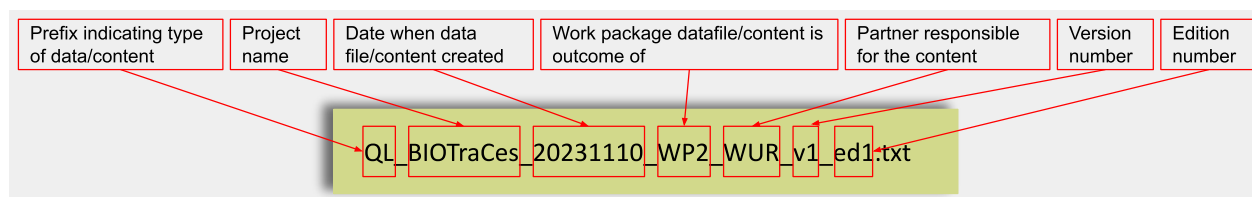


Figure 2: Example of datafile name prepared for repositing in Zenodo.

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

- During the project life-cycle and after it's end, anonymised data will be available in the Zenodo database. No restrictions are planned to be imposed on generated data.

How will the identity of the person accessing the data be ascertained?

- This project envisions two levels of data security: **anonymised** data and **raw** data containing sensitive information. Only sensitive data will be protected from public access, and access to this data will be limited to representatives from each consortium partner. No other interested parties will be able to access this data. On the other hand, anonymised data that complies with all relevant regulations will be freely available to anyone interested.

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

- Each consortium institution will be held responsible for ensuring that data collection and management is backed by proper ethical standards and procedures as well as the data collection plan is officially evaluated and approved by the institutional Ethics board. Sensitive data will be accessible for designated representative(s) in each consortium. One particular person will be held responsible for controlling who has access to sensitive data. Labelling of sensitive data will follow the same rules as earlier indicated for anonymized data.
- The need for data access committee is being discussed among consortium partners.

Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

- The anonymised data will be openly available and licensed under a public domain dedication CC BY. Additionally, the metadata will contain information

that enables interested parties to access the data. This way, anyone who is interested in the data can access it easily without any restrictions. It is important to note that the intellectual property rights are held by the data hosts. When interested parties retrieve data from Zenodo, the property rights of the data remain with the host. This means that while the data may be accessible to others, the host still retains ownership and control over the data.

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

- Data and metadata will be retained for the lifetime of the repository of Zenodo. This is the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least¹⁶.

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

- If documentation or references about the software are needed to access or read the data, they will be provided in the README file. The instructions will be thoroughly described and linked to the software sources so that interested parties can use the data without independently searching for information on how to operate the data file. Since it is planned to use open-source software if needed be, for example, JASP¹⁷ for quantitative data or RQDA¹⁸ for qualitative data that are freely available, there will be no need to upload it separately to the repository next to the data files.

3.3 Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

- During the BIOTraCes project, we will make every effort to ensure that the data we generate can be easily integrated into other data sets or systems that explore similar topics and research questions. Our aim is for interested parties to be able to merge project data with other datasets without any special efforts. To achieve this, we will (I) **use vocabularies and keywords** that are relevant to biodiversity conservation, co-creation, action research, and the theory of transformative change; (II) **use data formats and standards** that are appropriate for organising both qualitative and quantitative research outputs; and (III) **employ methodologies** that are relevant for building

¹⁶

<https://www.google.com/url?q=https://about.zenodo.org/principles/&sa=D&source=docs&ust=1682538260808905&usq=AOvVaw2Xuj6nWlvFXTkYi1MdFYLf>

¹⁷ <https://jasp-stats.org/>

¹⁸ <https://rqda.r-forge.r-project.org/>

shareable datasets¹⁹. Currently, we are exploring and discussing these vocabularies, keywords, standards, formats, and methodologies among our consortium partners. As soon as decisions are made, we will update the DMP with concrete information.

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? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

- It is important to note that, despite our best efforts, **not all data may be easily made interoperable**, particularly qualitative data such as recordings or interview transcripts, as well as data resulting from community co-creation efforts, among others. Currently, strategies are being developed to ensure that interested parties can still make use of datasets that have low compatibility with other datasets. Consortium partners will collaborate to create systems and maps on how to navigate project-specific ontologies and identify relevant vocabularies to help navigate these ontologies. As stated in the proposal, all relevant information will be made openly available to accelerate the uptake of innovations delivered during the course of this project.

Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?

- When applicable, **data published in Zenodo will be linked to other relevant research data** either developed in this project or elsewhere. Given that some datasets will be versioned we will link the newest version of the data to the previous one by providing qualified references, for example, DOI or HTTP link if DOI or other identifier is not available²⁰.

3.4 Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

- We plan to make our datasets reusable by assuring high data quality, by providing all documentation needed to support data interpretation and reuse. We will provide readme files with information methodology, codebooks (that include variables definitions, units of measurement etc.).

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

- As discussed earlier in this document and envisioned in the Grant Agreement, the data generated in this project will be made freely available on a CC BY or

¹⁹ GODAN Action. (2019, December 20). GODAN Action Online Course on Open Data Management in Agriculture and Nutrition (Version v1.0). Zenodo. <http://doi.org/10.5281/zenodo.3588148>

²⁰ <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>

equivalent license via Zenodo. Additionally, the data may also be shared on institutional platforms of consortium members, with the primary source being linked to Zenodo.

Will the data produced in the project be useable by third parties, in particular after the end of the project?

- As discussed earlier in this document, anonymised data will be stored in a trusted repository, namely Zenodo, and will be stored for an unlimited amount of time during the lifetime of the repository.

Will the provenance of the data be thoroughly documented using the appropriate standards?

- We will use metadata standards. Metadata will include the type of data and subtype (e.g., image, video/audio, dataset), DOI, title, authors, description, and other information if needed. Metadata will be created following DataCite Metadata Schema requirements.

Describe all relevant data quality assurance processes.

- Different procedures will apply for **qualitative** and **quantitative** data quality assurance processes. For example, numerical data will be quality-checked at collection/generation phase, checked for double entries, completeness, missing data, and unreasonable values. The register holder (i.e. data steward delegated by each consortium partner) assures data quality in terms of completeness and correctness of registration. Qualitative data quality, for example, interview data will be checked at all three main stages: pre-interview, during the interview, and post-interview. We will adapt data quality assurance processes for qualitative data²¹ to every data category (e.g., textual, visual, etc.) according to the guidelines and requirements to align the best practice in the field.

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

- In this project DMPs will also address other research outputs beyond data described earlier in this document. In the following section aspects related to the allocation of resources, security and ethics will be discussed.

²¹ <https://www.nice.org.uk/process/pmg4/chapter/appendix-h-quality-appraisal-checklist-qualitative-studies>

4 Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

- Possible outputs beyond research include publications (such as policy briefs, theoretical papers, and handbooks), as well as theoretical and practical models and frameworks, workflows, mappings, among others. These outputs will be publicly available under the same conditions as research data. Specifically, they will be stored in a given repository under a CC BY or equivalent license during and after the project's lifetime. The management of these outputs will be discussed in the upcoming Data Management Plans as needed.

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

- All other research outputs, if such need will occur, will follow FAIR principles.

5 Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.)? How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions).

- The costs of maintaining **sensitive data** will be budgeted under this grant's salary (i.e., person/month) expenditures line. Once properly anonymised, the sensitive data will be destroyed, and no financial or human resources will be required to maintain it beyond the project's completion. Additionally, the costs of maintaining **anonymized data** will be budgeted under the same category of expenditures as sensitive data maintenance. Repository Zenodo offers unlimited space for an unlimited period, providing storage, archiving, re-use, and security without the need for financial resources.

Who will be responsible for data management in your project?

- The Mykolas Romeris University's team members, who are the leaders of the data management task (T5.5), will draft the data management strategy and plan. However, the final strategy will always be collectively discussed with the consortium leader and partners and amended based on input from all partners. Since each consortium partner will generate a unique set of data that may be subject to data protection, as well as local ethical and data management regulations, each partner will designate a person responsible for their data management. All data management practices applied within this project will comply with the Data Management Plan.

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

- Anonymised data and their metadata will be retained for the lifetime of the Zenodo repository, which is currently the same as the lifetime of the host laboratory CERN. CERN has an experimental programme defined for at least the next 20 years. No financial resources are required to maintain the repositied data. Upon upload, it will be ensured that data files are as self-sufficient as possible to minimise the need for human assistance in understanding the data.

6 Data security

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)? Will the data be safely stored in trusted repositories for long term preservation and curation?

- As discussed earlier in this document, data security will be handled based on the following strategies: (I) **final anonymised versions of the data**, with all sensitive and personal information removed, will be deposited in a trusted repository, namely Zenodo. We rely on Zenodo's policies to ensure that deposited data remain safe and accessible to interested parties for a very long time. Zenodo policies meticulously describe data handling and security protocols and implement strict security measures. All data and metadata files are replicated multiple times and deposited in other independent repositories, ensuring that data can be retrieved from another source in the event of disruptions in one repository. In the unlikely event of Zenodo's closure, it is guaranteed that all data and metadata related to it will be transferred to another equivalent repository. The Zenodo repository properties are fully capable of satisfying the needs of the project, as well as the requirements raised by the funding body; (II) **raw data files containing personal and sensitive information** will be protected from public access. Access to this data will be limited to representatives from each consortium partner, and no other interested parties will be able to access it. The data will be stored in trusted local institutional repositories in compliance with local ethical standards and regulations, and will be curated by the representative of each consortium institution. Consortium partners will be required to ensure that the repository is trusted, keep details of the repository, and maintain records about data handling history. Each consortium institution will be responsible for ensuring that data collection and management are backed by proper ethical standards and procedures, and that the data collection plan is officially evaluated and approved by the institutional Ethics board. Designated representative(s) in each consortium will be responsible for controlling who has access to sensitive data. The consortium coordinator and partners will collectively discuss and agree upon terms for keeping raw data, and when (if at all) data will be destroyed.

7 Ethics

Are there, or could there be, any ethics 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 Description of the Action (DoA).

- Given that sensitive data will be destroyed immediately after its anonymisation process there is unlikely that legal or ethical issues will occur. The project will follow the principle of responsible research and innovation and will comply with The Brussels Declaration: Ethics and Principles for Science & Society Policy-Making²², the European Code of Conduct for Research Integrity²³ and European Union General Data Protection Regulation (GDPR)²⁴ in all stages of the project, from the research design to the dissemination of the research output. When necessary other research ethic-relevant regulations will be complied with. Study(ies) plan(s) will be submitted to the relevant Ethics Committee and will be in accordance with the national legislation that regulates research with human beings and research ethics in the relevant countries, including informed consent and data security. Archiving options of informed consent forms are still being considered among consortium partners. Information on that will be updated in the next DMP. The compliance with national legislation will be curated and assured by the representative of each consortium institution. Ethical considerations related to the research data and other relevant non-research outputs will be discussed in the **D1.3 Methodological participation guide (M6)**.

Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?

- When possible, an informed consent forms will be included and it will indicate how collected data will be handled, including the explanation that **raw** data (with personal details and sensitive information) and **anonymised** data (with no possibility to track participant's identity) will be processed following different procedures, all of which will comply with stringent ethical standards²⁵. However, the framework of Participatory action research requires a "bottom up" approach regarding consent ^{26,27,28}. Therefore, progressive forms of consent or communal consent will apply in most cases. Ethical considerations related to consents will be discussed in the **D1.3 Methodological participation guide**.

²² American Association for the Advancement of Science, 2017

²³ European Science Foundation & All European Academies, 2011

²⁴ <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

²⁵ [https://www.apa.org/news/press/releases/2014/06/informed-consent#:~:text=%E2%80%9CInformed%20Consent%2C%20psychologists%20inform%20participants,withdrawing%3B%20\(4\)%20reasonably%20foreseeable](https://www.apa.org/news/press/releases/2014/06/informed-consent#:~:text=%E2%80%9CInformed%20Consent%2C%20psychologists%20inform%20participants,withdrawing%3B%20(4)%20reasonably%20foreseeable)

²⁶ Chevalier, Jacques M., and Daniel J. Buckles. Participatory action research: Theory and methods for engaged inquiry. Routledge, 2019

²⁷ McLean, Irina. Ethics in participatory research for health and social well-being, 2019

²⁸ Cahill, Caitlin. Repositioning ethical commitments: Participatory action research as a relational praxis of social change. ACME: An International Journal for Critical Geographies, 2007

8 Other issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

- Every consortium member must comply with local institutional ethical and legal standards; thus, in every country and university there are different procedures and requirements, all of which do not contradict the current data management plan, but only help to implement it. The examples are: EU General Data Protection Regulation (the GDPR) (2018), The Guidelines of the Research Council of Lithuania on Open Access to Scientific Publications and Data (2016)