



SEMPRE-BIO

D7.5 Data Management Plan V2

**SEcuring doMestic PRoduction of
cost-Effective BIOMethane**

CETAQUA
WATER TECHNOLOGY CENTRE



Funded by
the European Union

PROJECT INFORMATION

GRANT AGREEMENT NUMBER	101084297
PROJECT TITLE	SEcuring doMestic PROduction of cost-Effective BIOmethane
PROJECT ACRONYM	Sempre-Bio
FUNDING SCHEME	HORIZON-IA
START DATE OF THE PROJECT	1 November 2022
DURATION	42 months
CALL IDENTIFIER	HORIZON-CL5-2021-D3-03-16
PROJECT WEBSITE	https://sempre-bio.com/

DELIVERABLE INFORMATION

DELIVERABLE N°	D7.5
DELIVERABLE TITLE	Data Management Plan v2
WP NO.	7
WP LEADER	CETAQUA
CONTRIBUTING PARTNERS	CET, AB, CRYO, DBFZ, DTU, INV, PROPULS, SINTEF, TERRA, TMB, UGE, UVIC, BIOGAS-E, INNOLAB, NAT, MASS
AUTHORS	Alejandra Córdova
REVIEWERS	DBFZ: Torsten Thalheim, Jaqueline Daniel-Gromke
CONTRACTUAL DEADLINE	31 Oct 2024
DELIVERY DATE TO EC	31 Oct 2024
DISSEMINATION LEVEL	Public

DOCUMENT LOG

VERSION 1	DATE	AUTHOR	DESCRIPTION OF CHANGE
V0.1	04-04-2023	Alejandra Córdova	Initial version
V0.2	19-04-2023	INV (Laia Mencia)	Review feedback
V1.1	21-04-2023	Alejandra Córdova	First version
V1.2	25-04-2023	INV (Laia Mencia, Estafanía González)	Review feedback
V1.3	26-04-2023	Alejandra Córdova	Final version ready for submission
VERSION 2	DATE	AUTHOR	DESCRIPTION OF CHANGE
V2.0	08-10-2024	Alejandra Córdova	Initial version ready for review
V2.1	24-10-2024	DBFZ (Torsten Thalheim, Jaqueline Daniel-Gromke)	Review feedback
V2.2	28-10-2024	Alejandra Córdova	Complete version ready for review
V2.2	29-10-2024	DBFZ (Jaqueline Daniel-Gromke)	Review feedback
V2.2	28-10-2024	Alejandra Córdova	Final version ready for submission

DISCLAIMER



**Funded by the
European Union**

"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 CINEA. Neither the European Union nor the granting authority can be held responsible for them."

CONTENT

1. Executive Summary	6
2. Data Summary	7
3. FAIR Data	8
3.1. Making data findable	9
3.1.1. Data discoverability	9
3.1.2. Data identification mechanisms	9
3.1.3. Naming conventions	10
3.1.4. Document versioning	10
3.1.5. Metadata standards	11
3.2. Making data accessible	12
3.2.1. Data availability	12
3.2.2. Data access	12
3.3. Making data interoperable	12
3.4. Increase data re-use	13
4. Allocation of resources	14
5. Data security	15
6. Ethical Aspects	16
7. Other issues	17
8. History of Changes	18

Figure index

Figure 1. Front-page of the Google Drive folder.	9
--	---

Table index

Table 1. Document template extract showing the history panel.	11
Table 2. Document template example showing the information panel on the front page.	11
Table 3. File formats for different disciplines that are either recommended or acceptable	13
Table 4. History of changes.	18

Acronym Glossary

DMP: Data Management Plan	EBIEs: European Biomethane Innovation Ecosystems
FAIR: Findable, Accessible, Interoperable and Reusable	CERIF: Common European Research Information Format

Consortium partners

	Participant organisation name	Acronym
1	CETAQUA	CET
2	AIGÜES DE BARCELONA	AB
3	CRYO INOX	CRYO
4	DEUTSCHES BIOMASSEFORSCHUNGSZENTRUM GEMEINNÜTZIGE GMBH	DBFZ
5	DANMARKS TEKNISKE UNIVERSITET	DTU
6	INVENIAM GROUP	INV
7	PROPULS	PROPULS
8	SINTEF AS	SINTEF
9	TERRAWATT	TERRA
10	TRANSPORTS METROPOLITANS DE BARCELONA	TMB
11	UNIVERSITEIT GENT	UGE
12	UNIVERSITAT DE VIC	UVIC
13	BIOGAS-E	BIOGAS-E
14	INNOLAB	INNOLAB
15	NATURGY	NAT
16	NV De Zwanebloem	MASS
17	VEOLIA WATER TECHNOLOGIES TECHNO CENTER NETHERLANDS BV	BIOTHANE

1. Executive Summary

SEMPRE-BIO (SEcuring doMestic PROduction of cost-Effective BIOmethane) is a €9.9M project financed under the Horizon Europe Cluster 5 programme running from November 2022 to April 2026. SEMPRE-BIO aims to demonstrate novel and cost-effective biomethane production solutions and pathways, deemed essential to achieve the European Green Deal and climate and energy targets for 2030 and the net zero greenhouse gas emissions by 2050, and to increase the market up-take of biomethane-related technologies.

With sites in Baix Llobregat (ES), Bourges (FR), and Adinkerke (BE), SEMPRE-BIO will establish three European Biomethane Innovation Ecosystems (EBIEs), which will be indicative of the various baseline settings for biomethane production throughout Europe. The challenge is to lower investment and operating costs, optimize feedstock supply and use, identify alternative feedstock, and reduce their costs, improve plant efficiency and operations, account for carbon savings, and increase and monetize co-benefits, such as from the commercialization of the digestate or the valorisation of residual gas streams.

The SEMPRE-BIO Data Management Plan (DMP) follows the DMP template that was designed to be applied to any Horizon Europe project that produces, collects or processes research data.

The aim of the DMP is to plan the life cycle of data within the SEMPRE-BIO project. It offers a long-term perspective by outlining how data will be generated, collected, documented, shared and preserved within the project.

The DMP is a living document to be updated as the implementation of the project progresses in the context of the periodic evaluation/assessment of the project and when significant changes occur.

2. Data Summary

The SEMPRES-BIO Data Management Plan (DMP) aims to provide a strategy for managing the data generated and collected during the project and optimise access to and re-use of research data. The DMP is intended to be a 'living' document that will outline how the SEMPRES-BIO research data will be handled during and after the project, and so it will be reviewed and updated at regular intervals. This document is structured according to the Horizon Europe template, and as stated in the Article 17 within Annex 5 of SEMPRES-BIO Grant Agreement. Hence, the DMP is based on the FAIR Data Management in Horizon Europe, that defines a Data Management Plan as a key element in ensuring that data is managed properly.

All the data sets, regardless of their categorization, will be stored in each of the participant entities databases and in the Google drive created as an internal database for all partners. In addition, the data categorised as open will be shared publicly through the library section of the project website, peer-reviewed scientific publications and Open Research Europe.

Data sets identified that will be generated in the framework of the project, which includes:

- Project deliverables.
- Articles published in Open Access scientific journals.
- Conference and workshops abstracts/articles/posters.
- Meeting presentations/agendas/minutes.
- Multimedia materials.
- Project Website, social networks, e-newsletter.

The technical, professional, and scientific publications that will be produced by the SEMPRES-BIO project will be open access in order to be compliant with the general principle of the Horizon Europe funding programmes. CETAQUA will clarify to all partners that embargo periods and article processing charges for hybrid open access publication are not allowed by Horizon Europe rule. This general rule is reinforced in section dedicated to Article 17 within Annex 5 of SEMPRES-BIO Grant Agreement (GA), which states that only publication fees in full open access venues for peer-reviewed scientific publications are eligible for reimbursement.

3. FAIR Data

The SEMPRES-BIO project will follow FAIR guiding principles for scientific data management and stewardship (mandatory for defined open data) that will be used all across the length of the project. As per the FAIR principles data should be:

1. Findable.
2. Accessible.
3. Interoperable.
4. Reusable.

Access to data will be kept as open as possible, but also as closed as necessary for commercial and exploitation purposes. Some key elements defined to ensure open access to data under this DMP are the following:

- Types of data/research outputs: reports, tools, videos, website, policy recommendations and scientific publications.
- Findability of data/research outputs: Use of Digital Object Identifier (DOI) for SEMPRES-BIO publications, use of standardised metadata frameworks, use of ORCID researchers' identifiers. Publication on Open Research Europe, encouraging partners to adhere to a fully open research publishing model.
- Accessibility of data/research outputs: Full open access will be provided to 25 out of 39 of SEMPRES-BIO deliverables. All these deliverables will be fully accessible to enable a broader impact of the project's results. The restricted deliverables will be D2, D3, D6, D8, D10, D11, D12, D13, D15, D16, D18, D19, D20, D27. They will be registered as sensitives and their access will be limited under the conditions of the GA.
- Interoperability of data/research outputs: Open-source fonts such as Barlow / Alatsi Fonts will be used as much as possible for reports since it was specifically designed to address the challenges of on-screen display.
- Reusability of data/research outputs: Open datasets will be used when possible, for the project. The results of the project will be made available in order to facilitate reproducibility. An analysis of sensitive data will be performed in order to avoid unexpected data leakage issues.
- Curation and storage/preservation costs: CETAQUA will be responsible for the storage and preservation of data.

3.1. Making data findable

The data management plan supports the effective collection and integration of SEMPRE-BIO project data. Project-internal storage, processing, and sharing of project data will mainly happen through the project's folder Google Drive "SEMPRE-BIO PROJECT", which is protected from unauthorised access and is managed by the project coordinator.

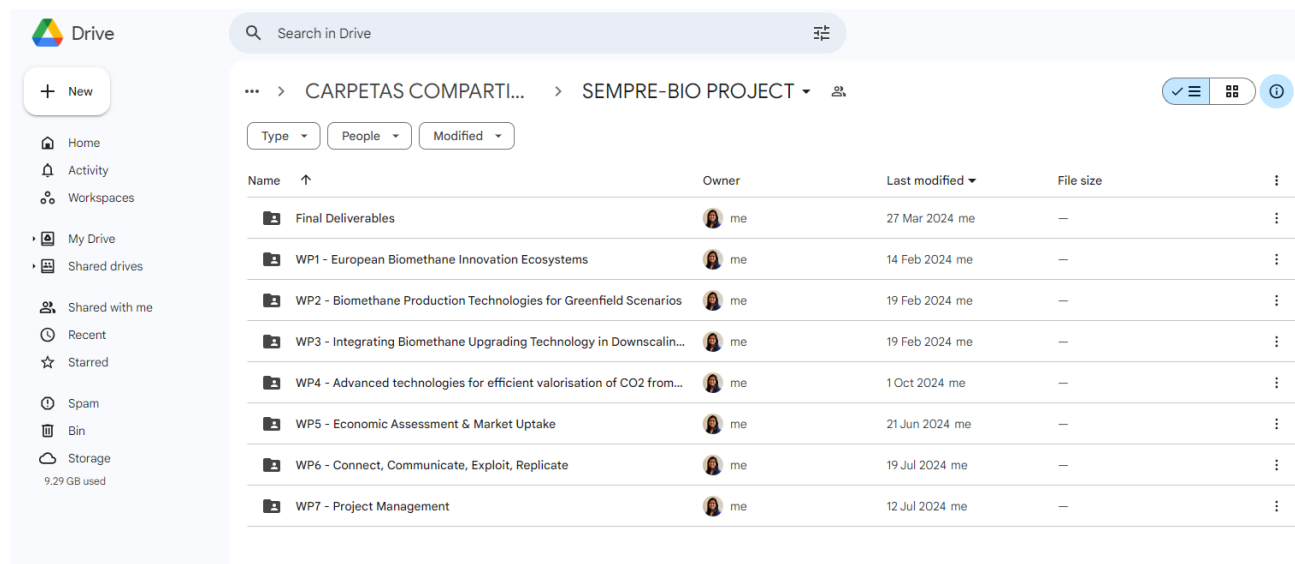


Figure 1. Front-page of the Google Drive folder.

All partners are invited to use it to exchange information and documents on project activities and any project related issues. A folder structure with major category divisions based on WPs already exists and partners can add more sub-folders as necessary or appropriate, but they must not remove any of the current folders or data. Within the folder WP7 Project Management and available to all partners, there are the Grant Agreement, Consortium Agreement, minutes for the different project meetings and an updated contact list.

The transfer of project information by email will be limited to a strict minimum. Public project data will be mainly shared through the project website, social media, community, other dissemination channels and open repositories (i.e., ZENODO. <https://zenodo.org/>). It will be ensured that public project data will not contain critical personal data or confidential information. Each data contributor, along with the designated reviewers, will be responsible for preliminary checks to exclude any sensitive information before data submission. The final verification will be conducted by the Project Coordinator, in collaboration with the Data Manager (SINTEF), who will ensure all data shared publicly complies with GDPR and Horizon Europe data security standards.

3.1.1. Data discoverability

Considering the FAIR data principles, public project data should be Findable by:

- Being assigned to a globally unique and persistent identifier.
- Containing enough metadata to fully interpret the data.
- Data and metadata being indexed in a searchable source.

By applying these principles data becomes Findable.

3.1.2. Data identification mechanisms

All documents associated with the project will be identified with a unique, and persistent, identifier that will be given at the time of the submission process.

Deliverables: SB-WPx-Dx.x-Deliverable name-Dissemination level-vx.x

- Project acronym.
- Related Work Package number.
- Deliverable number and name.
- Dissemination level [Sensitive (SEN) or Public (PU)].
- Version number [vx.x].

Example: SB-WP1-D7.10-Project Management Plan-PU-v1.0

Other documents: SB-WPx-Name-Vx.x

- Project acronym.
- Optional: Related Work Package number.
- Title of the document.
- Optional: Version number [vx.x] / Date

Example: SB-CET minutes or SB-WP7-CET minutes-v1.1

3.1.3. Naming conventions

The project document naming conventions are as per the list below developed at the University of Edinburgh¹:

1. Keep file names short, but meaningful.
2. Avoid unnecessary repetition and redundancy in file names and file paths.
3. Use capital letters to delimit words, not spaces or underscores.
4. When including a number in a file name, always give it as a two-digit number, i.e., 01-99, unless it is a year or another number with more than two digits.
5. If using a date in the file name, always state the date 'back to front', and use four-digit years, two-digit months and two-digit days: YYYYMMDD or YYYYMM or YYYY.
6. When including a personal name in a filename give the family name first followed by the initials.
7. Avoid using common words such as "draft" or "letter" at the start of file names, unless doing so will make it easier to retrieve the record.
8. Order the elements in a file name in the most appropriate way to retrieve the record.
9. The file names of records relating to recurring events should include the date and a description of the event, except where the inclusion of any of either of these elements would be incompatible with rule 2.
10. The file names of correspondence should include the name of the correspondent, an indication of the subject, the date of the correspondence and whether it is incoming or outgoing correspondence, except where the inclusion of any of these elements would be incompatible with rule 2.
11. The file name of an email attachment should include the name of the correspondent, an indication of the subject, the date of the correspondence, 'attach', and an indication of the number of attachments sent with the covering email, except where the inclusion of any of these elements would be incompatible with rule 2.
12. The version number of a record should be indicated in its file name by the inclusion of 'v' followed by the version number and, where applicable, 'DRAFT'.
13. Avoid using non-alphanumeric characters in file names.

¹ Gryzbowski, A., University of Edinburgh: (July 2007). Naming Conventions for Electronic Records. Retrieved from <https://www.ed.ac.uk/records-management/guidance/records/practical-guidance/naming-conventions> [accessed on April 2023]

3.1.4. Document versioning

Only documents created by the consortium will be versioned. To this end, document templates include a history panel (Table 1) for keeping track of the document versioning.

Table 1. Document template extract showing the history panel.

VERSION	DATE	AUTHOR	DESCRIPTION OF CHANGE

Project partners will identify different document versions by using a two-digit version number (vx.x). A document reviewed by another partner should be returned to the principal author by including the secondary author in the history panel. Only the principal author will change the version number at the document name.

3.1.5. Metadata standards

Basic metadata will be used to facilitate the efficient recall and retrieval of information by project partners. Additionally, it will contribute to facilitating the discoverability of the requested information. To this end, all documents related to the project have to include, in the second-page, an information panel with information about the project and the deliverable: deliverable number and title, WP number and leader, contributor(s), the author(s) and reviewer(s), date and dissemination level. The SEMPRES-BIO partnership will use templates and tables to display this information in the most homogenised way. For instance, Table 2 depicts the information panel of project documents.

Table 2. Document template example showing the information panel on the front page.

PROJECT INFORMATION

GRANT AGREEMENT NUMBER		101084297
PROJECT TITLE	SEcuring doMestic PROduction of cost-Effective BIOmethane	
PROJECT ACRONYM	Sempre-Bio	
FUNDING SCHEME	HORIZON-IA	
START DATE OF THE PROJECT	1 November 2022	
DURATION	42 months	
CALL IDENTIFIER	HORIZON-CL5-2021-D3-03-16	
PROJECT WEBSITE	https://sempre-bio.com/	

DELIVERABLE INFORMATION

DELIVERABLE N°	
DELIVERABLE TITLE	
WP NO.	
WP LEADER	
CONTRIBUTING PARTNERS	
AUTHORS	
REVIEWERS	
CONTRACTUAL DEADLINE	
DELIVERY DATE TO EC	
DISSEMINATION LEVEL	

When depositing a publication in a repository, the required metadata will be entered, especially the funding programme (Horizon Europe), the associated European Research Council (ERC) grant number and the grant acronym, so that the publications can be linked to this project.

To further support data discoverability and interoperability, SEMPRES-BIO will adopt standardised metadata schemes. The DataCite schema will be used for publications and datasets, ensuring consistent citation and accessibility, while DCAT (Data Catalog Vocabulary) will structure the project's data catalogues. Together, these approaches will standardise documentation, facilitating data access, integration, and reuse across different platforms.

3.2. Making data accessible

3.2.1. Data availability

Project data, materials (dissemination and communication) and deliverables that are declared as public will be made available to the public (through the project website, social media channels, community and/or open repositories). Furthermore, the pseudonymized data sets will be exploited through the creation of tables and infographics that may be used as part of the dissemination activities of the project.

3.2.2. Data access

In order to maximise the impact of SEMPRES-BIO research data, the results are shared within and beyond the consortium. Selected data and results will be shared with the scientific community and other stakeholders through publications in scientific journals and presentations at conferences, as well as through open access data repositories. The SEMPRES-BIO project results and deliverables are first stored and organised in the project's folder Google Drive, on the personal computers of project staff members, on the institutional secure server and on the project website. All data are made available for verification and re-use, unless the task leader can justify why data cannot be made openly accessible. To protect the copyright of the project knowledge, Creative Commons licence will be used in some cases. Hence, the project outcomes and public deliverables will be accessible by:

- SEMPRES-BIO project website
- Project repository in the Google Drive
- Partner database
- Open access journals
- Repositories like ZENODO, OpenAire

All data deposited in ZENODO is accessible without restriction to the public. For other data, potential users must contact the project coordinator or data owner to gain access. If necessary, appropriate Intellectual Property Rights (IPR) procedure (such as non-disclosure agreement) will be used.

3.3. Making data interoperable

The project partners will pay reasonable efforts towards storing all data in the appropriate format that will make data re-usable by external parties that might be interested in exploiting the data generated during the project. Partners will observe OpenAIRE guidelines for online interoperability, including OpenAIRE Guidelines for Literature Repositories, OpenAIRE Guidelines for Data Archives, OpenAIRE Guidelines for Current Research Information Systems (CRIS) Managers based on CERIF. These guidelines can be found at: <https://guidelines.openaire.eu/en/latest/>

The by-default language will be British English, and the vocabulary to be used will be the common one within research and manufacturing sectors, addressing the specific target audience of Small and medium-sized enterprises (SMEs), large companies, and their ecosystems. Vocabulary will not

constitute a barrier for data interoperability and reuse. A glossary has been set up and will be incremented as time goes.

The following table describes a variety of file formats for different disciplines that are either recommended or acceptable:

Table 3. File formats for different disciplines that are either recommended or acceptable

Type of data	Recommended formats	Acceptable formats
Tabular data	CSV (.csv), Tab-delimited (.tab)	Excel (.xls/.xlsx)
Geospatial data	ESRI Shapefile (.shp), GeoTIFF (.tif)	Geography Markup Language (.gml)
Textual data	PDF/A, RTF (.rtf)	DOCX (.docx), HTML (.html)
Image data	TIFF (.tif)	JPEG (.jpg), PNG (.png)
Video data	MPEG-4 (.mp4)	AVCHD video (.avchd)
Documentation and scripts	PDF/A, XHTML	DOCX (.docx), XML (.xml)

This selection is based on recommendations for data preservation from the OpenAIRE guidelines for Horizon Europe projects. For a full list of formats by data type, refer to the [OpenAIRE Guidelines for Data Archives](#)

3.4. Increase data re-use

Appropriate Licensing will be used to protect the ownership of the datasets. Depending on the need for attribution and reuse conditions, licences such as the Open Database License (ODbL) or the Public Domain Dedication and License (PDDL) will be considered for the parts of datasets for which the decision of making that part public has been made by the Consortium.

For datasets deposited on a public data repository (ZENODO) the access is unlimited. Restrictions on-use policy are applied for all protected data, whose re-use will be limited within the project partners. Specific restrictions may include the following:

- Attribution: requires users of the dataset to give appropriate credit, provide a link to the licence, and indicate if changes were made
- Limited to Non-Commercial use: prohibits the use of the dataset for commercial purposes by others.
- Share under the same terms: requires the others to use the same licence as the original on all derivative works based on the original data.

Internal process of Quality evaluation is activated throughout the entire project duration to assess both project data /products and project process. Internal peer review is performed for the project deliverables to guarantee the deliverable is developed with a high level of quality (see the D7.11 Project Management Plan).

4. Allocation of resources

Data management in SEMPRES-BIO will be done as part of the WP7 and CETAQUA, as project coordinator, will be responsible for data management in the project. CETAQUA has allocated a part of the overall WP7 budget and person-months to these activities. For the time being, the project coordinator is responsible for FAIR data management. Costs related to open access to research data are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions). In this sense, only publication fees in full open access venues for peer-reviewed scientific publications are eligible for reimbursement. Resources for long term preservation, associated costs and potential value, as well as how data will be kept beyond the project and for how long, will be discussed by the whole consortium during Project Steering Board (PSB) meetings.

5. Data security

All project deliverables and data will be stored and shared in the folder name “SEMPRE-BIO PROJECT” in Google Drive restricted to the project consortium. As an initial step, only the Consortium Partners will have access to the cloud storage where dataset and metadata are filed. Following scientific publications and articles, the dataset deliverables and the final demonstrator research results will be shared through ZENODO and other databases to promote the data making FAIR.

6. Ethical Aspects

The deliverable D7.8 Research Ethics and Data Protection monitoring Report V1 aims at ensuring that ethical requirements are met for all research undertaken in the project, in compliance with Horizon Europe standards. All partners will assure that the EU standards regarding ethics and data management are fulfilled in compliance with the ethical principles (see Article 14 and Annex 5 as set out in the GA). In addition, SEMPRE-BIO partners have to comply with:

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons regarding the processing of personal data and on the free movement of such data (GDPR).
2. Ethical principles (including the highest standards of research integrity as set out, for instance, in the European Code of Conduct for Research Integrity and including avoiding fabrication, falsification, plagiarism or other research misconduct).

7. Other issues

Researchers will ensure that all aspects of their research data management are thoroughly planned and documented, which will contribute to improving the quality, transparency, and accessibility of their research. To this end, as well as European Commission policies on open data management, Project Partners must also adhere to their own institutional policies and procedures for data management.



8. History of Changes

Table 4. History of changes.

Version Number and date	Change and justification	Section of D7.5
2.0 07/10/2024	Updated the number and name version of deliverable to correspond to the new version of the document for the previous deliverable.	Introduction (Section 1)
2.0 07/10/2024	Updated with the name of the new beneficiary to the project.	Consortium partners
2.0 07/10/2024	Removed Basecamp, as it is no longer available. Google Drive is now the platform to share documents, replacing Basecamp.	Data Summary (Section 2) FAIR Data (Section 3.1, 3.2.2) Data Security (Section 5)
2.0 07/10/2024	Replaced the Basecamp figure with Google Drive folder, as this platform is no longer used.	FAIR Data (Section 3.1)
2.0 07/10/2024	Updated the identified mechanism of the other documents, making the related Work Package number optional to the document name.	FAIR Data (Section 3.1.2)
2.0 08/10/2024	Updated the name of Project Steering Board (PSB)	Allocation of resources (Section 4)
2.1 28/10/2024	Defined roles for initial and final data checks	Making data findable (Section 3.1)
2.1 28/10/2024	Specified that Data and metadata applied for indexing in a searchable source.	Data discoverability (Section 3.1.1)
2.1 28/10/2024	Added the standardised metadata schemes.	Metadata standards (Section 3.1.5)
2.1 28/10/2024	Included recommended file formats for different disciplines.	Making data interoperable (Section 3.3)
2.1 28/10/2024	Updated licensing information to emphasize appropriate licensing options over Creative Commons.	Increase data re-use (Section 3.4)