UNIVERSITÀ DEGLI STUDI DI MILANO

DATA MANAGEMENT PLAN

The DMP is a document that provides details regarding all the research data collected and generated within the PhD project. In particular, it explains the way research data are handled, organized, licensed, and made openly available to the public, and how they will be preserved after the project is completed. The DMP also provides motivations when versions or parts of the project research data cannot be openly shared on account of third-party copyright issues, confidentiality, or personal data protection requirements or when open dissemination could jeopardize the project's achievements.

Project Title	Designing foundational strategies for developing a family of food systems sustainability ontologies
PhD student	Giorgio A. Ubbiali (ORCID: 0000-0001-7872-1770)
Tutor	Prof. Andrea Borghini (ORCID: 0000-0002-2239-1482) and Dr. Matthew Lange (ORCID: 0000-0002-6148-7962)
PhD School	Human Mind and its Explanations HUME; Department of Philosophy "Piero Martinetti"
Partner Company	International Center for Food Ontology Operability Data & Semantics (IC-FOODS)
Date of Issue	2025-01-09
File name	Ubbiali PhD project DMP. V.1.0
Storage location	https://github.com/gioUbbiali/Data-Management-Plan-Giorgio-A.-Ubbiali-s-PhD-project/
PhD Thesis	https://air.unimi.it/handle/2434/1161515

Revision n. 1	
Date of Revision	
Modified points:	
Note:	
Reason for modifications:	
File name	
Storage location	

This document was drawn up on the basis of the DMP Model Horizon 2020.

Table of Contents

1. PROJECT ABSTRACT

This document is the Data Management Plan (DMP) for the doctoral project entitled "Designing foundational strategies for developing a family of food systems sustainability ontologies". This project is co-funded by the Department of Philosophy "Piero Martinetti" University of Milan and the International Center for Food Ontology Operability Data & Semantics (IC-FOODS), as part of the "Departments of Excellence 2023-2027" Project granted by the Italian Ministry of University and Research (MUR), Italy. The main goal of this project is to deliver strategies to support ontologists in developing ontologies that are able to consistently address sustainability, especially food systems sustainability. Developed strategies are documented in the PhD thesis(each thesis chapter corresponds to one strategy).

This project covers three strategies, The first strategy corresponds to a literature review of available ontologies dealing with sustainability. Strategy 1 also proposes to establish a family of interoperable sustainability ontologies and presents a Sustainability Core Ontology (SCO) to consistently structure the family. The second strategy provides a meta-framework, named Food Systems Sustainability Meta-Framework (FSSMF), to support the development of ontologies that effectively address food systems sustainability. The third strategy highlights guidelines for employing the proposed meta-framework and provides a use-case core ontological module representing meat systems sustainability (Sustainable Meat System Ontology, SuMSO), accordingly developed.

This DMP presents a comprehensive overview of the sets of data that have been collected and generated as part of these three strategies. It equally clarifies our research approach and methodologies employed to make these data reusable, interoperable, and accurately stored and preserved. The DMP identifies nine sets of data:

- Set 1 covers the lists of ontologies and terms reviewed during the ontology assessment performed in Strategy 1. Lists are provided as an annex to the thesis manuscript (Chapter 1, Annex 1) and are stored on a dedicated GitHub repository, in both CSV and xlsx formats. Link to the GitHub repository; https://github.com/gioUbbiali/The-List-of-Sustainability-Ontologies.git.
- Set 2 covers the Sustainability Core Ontology (SCO) that we developed in Strategy 1. The SCO file is provided as an annex to the thesis manuscript (Chapter 1, Annex 2) and is stored on a dedicated GitHub repository in OWL, RDF, and TTL formats. Link to the GitHub repository: https://github.com/gioUbbiali/Sustainability-Core-Ontology.git.
- Set 3 covers an initial structured vocabulary we created as the starting basis for developing SCO. A related graphical representation complements this structured vocabulary. The structured vocabulary is provided as an annex to the thesis manuscript (Chapter 1, Annex 2; this and the previous set compose a unique annex) and is stored on the SCO GitHub repository, in both CSV and xlsx formats. The graphical representation is stored in JPG format. Link to the GitHub repository: https://github.com/gioUbbiali/Sustainability-Core-Ontology.git.
- Set 4 covers the protocol documenting the SCO design. This set is a complement to Chapter 1, Annex 2 and is stored on the SCO GitHub repository, in both docx and PDF formats. Link to the GitHub repository: https://github.com/gioUbbiali/Sustainability-Core-Ontology.git.
- Set 5 covers a series of charts that graphically represent major terms and relations belonging to the meta-framework that we propose in Strategy 2. Charts are provided as an annex to the thesis manuscript (Chapter 2, Annex 1) and are stored on a dedicated GitHub repository in CXL and PDF formats. Link to the GitHub repository: https://github.com/gioUbbiali/Food-Systems-Sustainability-Meta-Framework-FSSMF.git.
- Set 6 covers a supplementary text presenting details on the meta-framework. This supplementary text is provided as an annex (Chapter 2, Annex 2) to and included in the thesis manuscript.
- Set 7 covers the SuMSO ontological module we propose in Strategy 3. The file is provided as an annex to the thesis manuscript (Chapter 3, Annex 1) and is stored on a dedicated GitHub repository in OWL,

RDF, and TTL formats. Link to the GitHub repository: https://github.com/gioUbbiali/Sustainable-Meat-Systems-Ontology

- Set 8 covers materials for SuMSO implementation, including a list of ontologies to reuse (docx and PDF formats), additional terminologies to cover and related file guidelines (xlsx and PDF formats), lists of Competency Questions (CQs) and topics for implementation (docx and PDF formats), and reference to an RDF file that contains terms from FoodOn to reuse (RDF format). These materials are a complement to Chapter 3, Annex 1, and are stored on the SuMSO GitHub repository. Link to the GitHub repository: https://github.com/gioUbbiali/Sustainable-Meat-Systems-Ontology
- Set 9 covers a supplementary text detailing additional information regarding the SuMSO ontological module and its development process. This supplementary text is provided as an annex (Chapter 3, Annex 2) to and included in the thesis manuscript.

2. TYPE OF DATA

2.1 DATA GENERATING DURING THE PROJECT

Table 1 – Summary of data generating

Number of Dataset	Type of Data	Formats	Size	Generated by
1	Tabular data	xlsx + CSV	128 KB	Ubbiali Giorgio A.
2	Ontology data	OWL + RDF + TTL;	921 KB + 626 KB + 517 KB	Ubbiali Giorgio A.
3	Tabular data and image data	xlsx + CSV and JPG	513 KB + 7 KB and 362 KB	Ubbiali Giorgio A.
4	Text data	Docx + PDF	293 KB + 245 KB	Ubbiali Giorgio A.
5	Graphical representation data (concept maps)	Cmaps (cxl) + PDF	257 KB + 1370 KB; 468 KB + 3262 KB; 439 KB + 2312 KB; 127 KB + 1489 KB	Ubbiali Giorgio A.
6	Text data	PDF (initially docx)	305 KB	Ubbiali Giorgio A.
7	Ontology data	OWL + RDF + TTL	561 KB + 509 KB + 627 KB	Ubbiali Giorgio A.
8	Text, tabular, and ontology data	Docx + Xlsx + PDF + RDF	CQs: 18 KB + 91 KB; List of ontologies to reuse: 24 KB + 165 KB; Vocabulary: 76	Ubbiali Giorgio A.

			KB + 129 KB + 31 KB; Topics for implementation: 19 KB +100 KB	
9	Text data	PDF (initially docx)	346 KB	Ubbiali Giorgio A.

2.2 EXISTING DATA USED IN THE PROJECT

Table 2 – Summary of existing data

Type of Data	**Formats**	**Size**	**Owner**	**Free to be used/ Third Party Rights**
Public data (links to documentation of explored ontologies and ontology terms) Reference dataset: 1	URIs	Not applicable	Developers (and related institutions) of explored ontologies	Free to be used
Public data (links to object properties and terms belonging to OBO-Foundry and compliant ontologies that SCO and SCO SV reuse) Reference dataset: 2 and 3	URIs	Not applicable	Developers (and related institutions) of explored OBO-Foundry and compliant ontologies	Free to be used
Public data (links to OBO-Foundry and compliant ontologies	URIs	Not applicable	Developers (and related institutions) of explored OBO-Foundry and	Free to be used

reused to construct SCO) Reference dataset: 4			compliant ontologies	
Public data (links to object properties belonging to OBO-Foundry, compliant ontologies, and other ontological resources that we associated with the series of charts) Reference dataset: 5	URIs	Not applicable	Developers (and related institutions) of explored OBO-Foundry and compliant ontologies	Free to be used
Public data (DOI of publications and links to webpages consulted to develop the meta-framework that we associated with the series of charts) Reference dataset: 5	URIs	Not applicable	Authors (and related institutions) of the papers and webpages	Free to be used
Public data (links to object properties and classes belonging to OBO-Foundry ontologies, OBO-Foundry compliant ontologies, and other resources that SuMSO reuses)	URIs	Not applicable	Developers (and related institutions)	Free to be used

Reference dataset: 7				
Public data (links to OBO-Foundry ontologies, OBO-Foundry compliant ontologies, and other resources that we reused or pointed to for designing SuMSO) Reference dataset: 7,8,9	URIs	Not applicable	Developers (and related institutions)	Free to be used
Information gathered from discussions with SuMSO stakeholders and the working documents that they shared with us Reference dataset: 7,8, and 9	Docx	Not applicable	SuMSO Stakeholders	Permission to reuse this information has been given by stakeholders

3. FAIR DATA

This part has been completed for all datasets indicated in Table 1.

3.1 Make your data Findable and Interoperable

DATASET 1

Dataset 1 provides the list of sustainability ontologies. This list covers guidelines and two sections. The first section documents the ontologies assessed in Strategy 1 (List A) and the second the terms related to sustainability that are covered by those ontologies (List B). For each ontology, List A presents the name, the acronym, a reference, and the modality through which we retrieved it (ontology portal exploration or paper review). For each term, List B presents the name, IRI, and the parental ontology

(acronym). The list of sustainability ontologies is stored on GitHub (guidelines and sections are stored separately): https://github.com/gioUbbiali/The-List-of-Sustainability-Ontologies.git.

- Metadata provision.

This dataset is described using the Dublin Core Metadata Standard on GitHub.

- Naming conventions.

The DMP identifies the following naming for dataset 1 on GitHub:

< number of the annex> <name of the dataset> (no spaces)

Example:

Annex1SustainabilityOntologyListA

Initially, before uploading this dataset on GitHub, we ensured file versioning via the Google Drive doc versioning system. During the final revisions, we included the date in the file name. The DMP identifies the following naming for dataset 1:

<Name of the chapter> < number of the annex> <name of the dataset> <date (yy-mm-dd)>.

Example:

Chapter 1 Annex 1 The list of sustainability ontology 2024- 02-25

- Raw data storage location

During the ontology review assessment, data was collected into an xlsx file and stored in a Google Drive folder shared among researchers involved in the project and potentially interested colleagues. The versions have always been kept aligned. Updates in one version were also included in the others.

- Specify tools, hardware, and software needed to access the data

Dataset 1 is provided in xlsx and CSV file formats. A Microsoft Office Suite license is required to access the xlsx file.

DATASET 2

Dataset 2 provides the Sustainability Core Ontology. SCO is stored on GitHub: https://github.com/gioUbbiali/Sustainability-Core-Ontology.git.

- Metadata provision

This dataset is described using the Dublin Core Metadata Standard on GitHub.

- Naming conventions

The DMP identifies the following naming for dataset 2 on GitHub:

<Name of the ontology>

Example:

SCO

The SCO version before release included the naming "-edit".

Example:

SCO-edit

Initially, before uploading SCO to GitHub, we ensured file versioning including the date in the file name on Giorgio A. Ubbiali's personal laptop. The DMP identifies the following naming for dataset 2:

<Name of the ontology> <date (yy-mm-dd)>

Example:

SCO 24-09-13

- Raw data storage location

During the design of SCO, data was stored in a folder on Giorgio A. Ubbiali's personal devices (see Data Security section). A few final revisions were made once SCO had been uploaded to GitHub. Updates were equally included in the other versions.

- Specify tools, hardware, and software needed to access the data

Dataset 2 is provided in OWL, RDF, and TTL formats.

DATASET 3

Dataset 3 provides "the structured vocabulary of SCO". This structured vocabulary covers guidelines and a table documenting an initial set of terms to use for developing SCO, related definitions, synonyms, links to ontologies from which we reuse some terms (and term URIs) if applicable, the is_a hierarchy, and notes. A nodes-and-edges graphical representation (chart) complements the table, showing relations occurring among entities these terms refer to. The representation was created using Cmaps Tools software and provided as a JPG/JPEG file. The structured vocabulary of SCO is stored on GitHub: https://github.com/gioUbbiali/Sustainability-Core-Ontology.git.

- Metadata provision

This dataset is described using the Dublin Core Metadata Standard on GitHub.

- Naming conventions

The DMP identifies the following naming for dataset 3 on GitHub.

<Name of the chapter> < number of the annex> <name of the dataset> (no spaces)

Example:

Annex2SCOstructuredvocabulary

The graphical representation presents the following file name.

Supplementary Fig. S1 The structured vocabulary of SCO

Initially, before uploading the structured vocabulary of SCO to GitHub, we ensured file versioning included the date in the file name. The DMP identifies the following naming for dataset 3:

<Name of the chapter> < number of the annex> <name of the dataset> <date (yy-mm-dd)>

Example:

Chapter 1 Annex 2 SCO structured vocabulary 2024- 05-11

- Raw data storage location

During the design of the SCO structured vocabulary, data was stored in a folder on Giorgio A. Ubbiali's personal devices (see Data Security section). and in folders on Google Drive and Microsoft Teams shared among researchers involved in the project. The graphical representation was stored in a CmapsTool folder on Giorgio A. Ubbiali's personal devices and in a folder on CmapsTool Cloud shared among researchers involved in the project. The versions have always been kept aligned. Updates in one version were also included in the others.

- Specify tools, hardware, and software needed to access the data

Dataset 3 is provided in xlsx, CSV, and JPG (only the chart) file formats. A Microsoft Office Suite license is required to access the xlsx file.

DATASET 4

Dataset 4 provides the protocol detailing the development of SCO. The SCO development protocol is stored on GitHub: https://github.com/gioUbbiali/Sustainability-Core-Ontology.git.

- Metadata provision

This dataset is described using the Dublin Core Metadata Standard on GitHub.

- Naming conventions

The DMP identifies the following naming for dataset 4 on GitHub:

<Name of the dataset>

Example:

SCO Development Protocol

We ensured file versioning via the GitHub versioning system.

- Raw data storage location

During the design of the SCO development protocol, data was stored on GitHub and in a folder on Giorgio A. Ubbiali's personal devices (see Data Security section). When the protocol was finalized, we also stored a version in a folder on Google Drive shared among researchers involved in the project. The versions have always been kept aligned. Updates in one version were also included in the others.

- Specify tools, hardware, and software needed to access the data

Dataset 4 is provided in docx and PDF file formats. A Microsoft Office Suite license is required to access the docx file.

DATASET 5

Dataset 5 provides the FSSMF Atlas. FSSMF Atlas covers four charts: a main chart named FSSMF-MC, and three auxiliary ones named FSSMF-AC1, FSSMF-AC2, and FSSMF-AC3. We created these charts using Cmaps Tools software. FSSMF Atlas is stored on GitHub: https://github.com/gioUbbiali/Food-Systems-Sustainability-Meta-Framework-FSSMF.git.

- Metadata provision

This dataset is described using the Dublin Core Metadata Standard on GitHub.

- Naming conventions

The DMP identifies the following naming for dataset 5 on GitHub:

<name of the dataset>

Example:

FSSMF-MC

Initially, before uploading FSSMF Atlas to GitHub, we ensured file versioning including the date in the file name. The DMP identifies the following naming for dataset 5:

<name of the dataset> <date (yy-mm-dd)>.

Example:

FSSMF-MC extending SCO 2024- 05-12*

*FSSMF-MC extends the graphical representation of the SCO structured vocabulary. So, we include this clarification in the file name.

- Raw data storage location

FSSMF Atlas charts were stored in both a CmapsTool folder on Giorgio A. Ubbiali's personal laptop and in a folder on CmapsTool Cloud shared among researchers involved in the project. The versions have always been kept aligned. Updates in one version were also included in the others. A few final revisions were made once the charts had been uploaded to GitHub. Updates were equally included in the other versions.

- Specify tools, hardware, and software needed to access the data

Dataset 5 is provided in CXL and PDF file formats. We suggest using the CmapsTool software (freely downloadable here https://cmap.ihmc.us/) to optimally visualize FSSMF Atlas charts.

DATASET 6

Dataset 6 provides a supplementary text presenting details on FSSMF. This supplementary text is included in the thesis manuscript.

- Metadata provision

This dataset is described using the Extensible Metadata Platform.

- Naming conventions

The DMP identifies the following naming for dataset 6:

< Number of the annex> <name of the chapter> <date (yy-mm-dd)>

Example:

Annex 2 Chapter 2 24-12-18

In the initial versions, this dataset was numbered as Annex 1. For reasons of chapter consistency, we further changed it in Annex 2.

We ensured file versioning via the Google Drive and Microsoft Teams doc versioning systems and by including the date in the file name.

- Raw data storage location

During the design of this dataset, data was stored in a folder on Giorgio A. Ubbiali's personal devices (see Data Security section) and in folders on Google Drive and Microsoft Teams, shared among researchers involved in the project. The versions have always been kept aligned. Updates in one version were also included in the others.

- Specify tools, hardware, and software needed to access the data

Dataset 6 is provided in PDF file format.

DATASET 7

Dataset 7 provides the core ontological module of SuMSO. We constructed this resource by engaging with food systems stakeholders. The insights from discussions and materials stakeholders shared with us served as the foundational basis for the ontology development. The core ontological module of SuMSO is stored on GitHub: https://github.com/gioUbbiali/Sustainable-Meat-Systems-Ontology

- Metadata provision

This dataset is described using the Dublin Core Metadata Standard on GitHub.

- Naming conventions

The DMP identifies the following naming for dataset 7 on GitHub:

<Name of the ontology>

Example:

SuMSO

SCO version before release included the naming "-edit".

Example:

SuMSO-edit

We ensured file versioning via the GitHub versioning system.

- Raw data storage location

During the design of SuMSO, data was stored on GitHub and in a folder on Giorgio A. Ubbiali's personal devices (see Data Security section). The versions have always been kept aligned. Updates in one version were also included in the others.

- Specify tools, hardware, and software needed to access the data

Dataset 7 is provided in OWL, RDF, and TTL formats.

DATASET 8

Dataset 8 is a collection of materials to complement the process of development of the SuMSO ontological module. Dataset information includes a list of ontologies to reuse and that have been reused, a list of covered terms and additional ones to cover (and related file guidelines), lists of Competency Questions (CQs) and topics for implementation, and reference to an RDF file that contains terms from FoodOn to reuse. These materials are stored on GitHub: https://github.com/gioUbbiali/Sustainable-Meat-Systems-Ontology

- Metadata provision

This dataset is described using the Dublin Core Metadata Standard on GitHub.

- Naming conventions

The DMP identifies the following naming for dataset 8 on GiHub:

<Name of the dataset> <date (yy-mm-dd)>

Example:

SuMSO CQs 24-12-12

We ensured file versioning via the GitHub versioning system and including the date in the file name.

- Raw data storage location

During the design of these implementation materials, data was initially stored in a folder on Giorgio A. Ubbiali's personal devices (see Data Security section). Once materials were substantially structured, we uploaded them to GitHub. The versions have been kept aligned. Updates in one version were also included in the others.

- Specify tools, hardware, and software needed to access the data

Dataset 8 materials are provided in docx, xlsx, PDF, and RDF formats. A Microsoft Office Suite license is required to access the docx and xlsx files.

DATASET 9

Dataset 6 provides a supplementary text presenting details on the SuMSO ontological module and its design process. This supplementary text is included in the thesis manuscript.

- Metadata provision

This dataset is described using the Extensible Metadata Platform.

- Naming conventions

The DMP identifies the following naming for dataset 9:

< Number of the annex> <name of the chapter> <date (yy-mm-dd)>

Example:

Annex 2 Chapter 3 24-12-18

In the initial versions, this dataset was originally named Annex 1. For reasons of chapter consistency, we further changed it in Annex 2.

We ensured file versioning via the Google Drive and Microsoft Teams doc versioning systems and by including the date in the file name.

- Raw data storage location

During the design of this dataset, data was stored in a folder on Giorgio A. Ubbiali's personal devices (see Data Security section) and in folders on Google Drive and Microsoft Teams, shared among researchers involved in the project. The versions have always been kept aligned. Updates in one version were also included in the others.

- Specify tools, hardware, and software needed to access the data

Dataset 9 is provided in PDF file format.

3.2 Specify Access conditions to your data

As a guiding principle, this project seeks to make research data openly available, whenever possible, in order to allow dissemination, validation, and re-use of research results. When restrictions to access need to be applied, they are motivated, and it is indicated how and who can grant access rights.

This part will also support Data sharing and Reusability.

Table 3 – Summary of access to datasets

Dataset	**Open from**	**How**	**Restrictions Yes/Not***
1	08/2024	Associated with thesis manuscript and related publications. We created dedicated GitHub repositories to store data generated within this	No
2	09/2024		No
3	08/2024		No
4	09/2024		No

5	01/2025	project. A link to this repository will be available from the thesis manuscript and related publications.	No
6	05/2025		No
7	11/2024		No
8	11/2024		No
9	05/2025		No

* Throughout the duration of the project and until the public release, we restricted access to data to ensure confidentiality for publication purposes. Prof. Andrea Borghini, Dr. Matthew Lange, and Dr. Giorgio A. Ubbiali grant access rights to these datasets.

4. DATA QUALITY AND SAFETY STRATEGIES

This part has been completed for all datasets indicated in Table 1.

4.1 Data quality

Regarding dataset 1, ontologies have been retrieved via a keyword search on Google Scholar and PubMed search engines and on the ontology portals OLS, Ontobee, AgroPortal, and BioPortal (keywords are provided in the thesis manuscript). Terms have been retrieved via the keyword search on these four ontology portals. We performed both searches twice. List A provides links to papers we analyzed or cited in related bibliographies regarding the ontologies retrieved via browser exploration. List a provides links to documentation cited by portals regarding the ontologies retrieved via portals' exploration. If no documentation was available, we explored ontology repositories and/or homepages, and we queried for ontology names, author names, and names of the belonging projects in Google Scholar and PubMed browsers and Google search engine to identify a related publication to include in List A. We noted down the development process of the list while happening. When we converted the file from xlsx format into CSV format, we ensured no data loss and/or alteration occurred.

Regarding dataset 2, we constructed SCO on the basis of SCO structured vocabulary and extending BFO. Terms and relations have been chosen according to the themes addressed by the sustainability papers we explored in Strategy 1. Relations have been chosen among those provided by OBO-Foundry and compliant ontologies. We followed the OBO-Foundry community best practices and guidelines reported in Arp et al., (2015)**. We employed Protégé, one of the most widely used software for ontology development today, to formalize SCO. We validated SCO using Protégé automatic reasoners ELK 0.5.0 and HermiT 1.4.3.456. We noted down the SCO development process while happening.

Regarding dataset 3, terms and relations have been chosen according to the themes addressed by papers regarding sustainability we explored. Relations have been chosen among those provided by OBO-Foundry and compliant ontologies. We followed the OBO-Foundry community best practices and guidelines reported in Arp et al., (2015)** to develop the SCO structured vocabulary. We noted down the development process of SCO structured vocabulary while happening. When we converted the file from xlsx format into CSV format, we ensured no data loss and/or alteration occurred.

Regarding dataset 4, we noted down and documented the development process of SCO while happening. When we converted the file from docx format into PDF format, we ensured no data loss and/or alteration occurred.

Regarding dataset 5, terms and relations have been chosen according to the themes addressed by papers we reviewed in developing the meta-framework. Relations have been chosen among those provided by OBO-Foundry and compliant ontologies. FSSMF links to publications and webpages via related DOI and URIs. Linked resources count among those we reviewed to develop FSSMF. We followed the OBO-Foundry

community best practices and guidelines reported in Arp et al., (2015)** to construct the FSSMF Atlas. We noted down the FSSMF Atlas development process while happening. When we converted the file from CXL format into PDF format and we cropped the files, we ensured no data loss and/or alteration occurred.

Regarding dataset 6, we noted down changes made to the document while writing it. This dataset has been produced in docx format. We further converted it into PDF format. When we converted the file, we ensured no data loss and/or alteration occurred.

Regarding dataset 7, we constructed SuMSO extending SCO. Terms and relations have been chosen according to the themes addressed by meat systems and sustainability papers we explored in Strategy 3 and according to suggestions from the stakeholders we interviewed. Several terms have been imported from existing ontologies, primarily OBO-Foundry and compliant ontologies. Relations have been chosen among those provided by OBO-Foundry and compliant ontologies. We followed the OBO-Foundry community best practices and guidelines reported in Arp et al., (2015)** and we utilized FSSMF and related methodology (Strategy 2 and 3). We employed Protégé to formalize SuMSO. We validated SuMSO using Protégé automatic reasoners ELK 0.5.0 and HermiT 1.4.3.456. Stakeholders validated SuMSO as well. We noted down the SuMSO development process while happening.

Regarding dataset 8, materials for implementing SuMSO have been prepared, noted down, and documented during the ontology development process. When we converted the file from docx or xlsx format into PDF format, we ensured no data loss and/or alteration occurred.

Regarding dataset 9, we noted down changes made to the document while writing it. This dataset has been produced in docx format. We further converted it into PDF format. When we converted the file, we ensured no data loss and/or alteration occurred.

For additional details, we refer readers to the datasets' guidelines and the thesis manuscript.

** Arp, R., Smith, B., & Spear, A. D. (2015). *Building ontologies with Basic Formal Ontology*. Cambridge, Massachusetts: Massachusetts Institute of Technology.

4.2 Data security

To ensure data recovery, storage, backups, and possible transfer, we implemented both a short-term approach and a long-term approach. The short-term approach includes mixed storage on Giorgio A. Ubbiali's personal devices, in clouds, and on GitHub, according to the necessities and the development of the PhD project. The long-term approach involves storage on GitHub repositories. Giorgio A. Ubbiali's devices are his personal laptop and an external hard drive. Clouds used are Google Drive, Microsoft Teams, and Cmaps Cloud.

Table 4 – Term of storage

Dataset	Short-term	Repository for long-term storage
1	In Cloud storing + Storing on Giorgio A. Ubbiali's personal devices.	GitHub repository
2	Storing on Giorgio A. Ubbiali's personal devices.	GitHub repository
3	In Cloud storing + Storing on Giorgio A. Ubbiali's personal devices.	GitHub repository

4	On GitHub repository + In Cloud storing + Storing on Giorgio A. Ubbiali's personal devices.	GitHub repository
5	In Cloud storing + Storing on Giorgio A. Ubbiali's personal devices.	GitHub repository
6	In Cloud storing + Storing on Giorgio A. Ubbiali's personal devices.	As part of the thesis manuscript on the UNIMI IrisAIR archive (the expected storage location of the PhD thesis).
7	On GitHub repository + Storing on Giorgio A. Ubbiali's personal devices.	GitHub repository
8	On GitHub repository + Storing on Giorgio A. Ubbiali's personal devices.	GitHub repository
9	In Cloud storing + Storing on Giorgio A. Ubbiali's personal devices.	As part of the thesis manuscript on the UNIMI IrisAIR archive (the expected storage location of the PhD thesis).