D5.2 Data Management Plan
Version 1.1

Document Information

Contract Number	828947
Project Website	www.enerxico-project.eu
Contractual Deadline	M6
Dissemination Level	Public
Nature	Report
Author	Josep de la Puente (BSC), Marta Rosello (BSC), Claudia Rosas (BSC)
Contributor(s)	All partners
Reviewer	
Keywords	data description, data collection, FAIR

Notice:
The research leading to these results has received funding from the European Union's Horizon 2020 Programme under the ENERXICO Project (www.enerxico-project.eu), grant agreement no 828947 and under the Mexican CONACYT-SENER-Hidrocarburos grant agreement B-S-69926.

© 2019 ENERXICO Consortium Partners. All rights reserved.

Change Log

Version	Author	Description of Change
1.0	Claudia Rosas	Initial draft
1.1	Marta Rosselló	Final draft

Table of Contents

Executive Summary

This deliverable presents the data management plan (DMP) of the ENERXICO project, which describes the data management life-cycle for all codes and datasets to be collected, processed and/or generated during the lifetime of the project.

The document describes the contents and organization of the DMP, considering that the actual data (and metadata) description will be furnished as a "live" Annex that will be periodically updated during the development of the project. Specifically, this deliverable describes:

- How the different types of datasets that will be generated, collected, and processed during the project will be managed during and after it. This affects mainly the ENERXICO scientific codes and, to a lesser extent, other research activities.
- Which methodologies and which standards (if any) will be applied to manage each of the ENERXICO datasets.
- How the datasets will be stored and handled during the lifetime of the project and after its conclusion, as well as how the datasets will be made (openly) accessible.

1. Introduction

The ENERXICO project is part of the H2020 FETHPC Open Research Data Pilot (ORD pilot), aimed at improving and maximizing access and re-use of research data, while taking into account the need to balance openness and protection of scientific information, commercialization and Intellectual Property Rights (IPR), privacy concerns, security, and related data management and preservation questions.

This DMP describes how data (in its broadest sense) will be managed during the project, but it is not a final closed document. The DMP described in this report (D5.6) will be updated over the course of the project to account for the generation/acquisition of new datasets, implementation of consortium policies, and/or other external factors.

This document follows the Horizon 2020 "FAIR" DMP template and follows the FAIR data guiding principles; i.e. that data must be Findable, Accessible, Interoperable, and Re-usable.

2. Data Summary and Structure of the ENERXICO DMP Annex

As stated in the Grant Agreement (GA), ENERXICO will promote optimization and performance and efficiency enhancements in European and Mexican parallel codes, related to the fields of Geophysics, Combustion and Wind simulation, for the upcoming pre-Exascale and Exascale supercomputers, and will further improve the solution of scientific problems requiring of Exascale computing. Some of the final demonstrators can be considered as small-scale demonstrators for optimizing and testing codes on

Exascale hardware prototypes and for addressing the Exascale challenges.

The reference parallel applications will be the main sources of project data, which will include:

- The source code of the parallel applications themselves.
- Datasets generated to evaluate code performance.
- Experimental datasets to be used for the performance audits.
- Datasets generated from the execution of models.

For this reason, the ENERXICO DMP has been organized around a (dynamic) spreadsheet annex document, with one dedicated sheet for each of the parallel scientific codes (i.e. 8 sheets in total).

2.1 Dataset sheet

For each dataset in a parallel scientific code, a dataset sheet exists with information on the following data items:

Item	Comments/explanation
Name	Descriptive name to identify the dataset
Description	Short description of the contents
Data category	Data category code (see Table 2 for the corresponding codes)
License	Chosen among the most appropriated and most open ones
Repository location	Institutional or public repository name
Author	Data author(s) name(s)
Naming Conventions	File names structure and conventions
Versioning	How and where the version of the dataset can be found
Format	Standards, definitions, ontologies, etc.
Size	Total estimated size, or single file size and number of expected files
Storage	Physical support selected, dependent on availability needs
Archive path	Folders structure
Associated metadata	Selected metadata standards, and to metadata set
Provenance	Structured dataset origin information

Backup needs	Periodicity, subsets backup needs analysis, etc.
Access permissions	Lifecycle dependency: selected groups, or public
Legal/ethical restrictions	Privacy and security issues
Reproducibility	If yes: connection to code and environment
Data transfer needs	Replicas and periodic transfers to/from other repositories
Long term preservation	Needs at 3-5-7-10 years (if any)
Metadata management	Way to access metadata when data are not available
Resources need	Analysis of resources needs at each step of data lifecycle

Table 1. List of items in a dataset sheet and their definition.

Data category	Code	Name	Comments
Scientific data	1.1	Models	
	1.2	Experimental	Data coming from observation, measurements or produced by detectors/sensors or by any other experimental device and or activity.
	1.3	Synthetic	Data generated by a simulation and/or are not obtained by direct measurement
	1.4	Test	Datasets (experimental or synthetic) used to validate models
Software	2.1	Libraries	
	2.2	Applications	
	2.3	Services	
	2.4	API	
Administrative	3.1	Documents	Any documentation, either public

docs			or private, such as code documentation, technical notes, etc., not directly mentioned in the project deliverables list.
	3.2	Internal reports	Meetings minutes, internal notes to document the evolution of the project, such as calendar, resources management, mailing lists, etc.
	3.3	Deliverables	Project output documents
Other	4.1	Metadata	Any data describing data properties. If they contain scientific information, they can also be classified as scientific data.

Table 2. Summary of the different data categories

2.2 Software sheet

Similar to the dataset sheets, one software sheet per code (or related software component) will contain information on the following aspects:

Item	Comments/explanation
Reference name of the program or workflow	Name of the code
Description	Brief description
Author	Main developers
Programming language	Specify
Rules and best coding practices	Conventions for filenames, link to an external manual (if exists), e.g. PEP8.
Access permissions and license	Lifecycle dependency: only specific groups of collaborators, all partners, whole community, etc.
Code size if relevant (to be updated)	
Repository type	GitHub, GitLab, Bitbucket, SourceForge
Repository structure	Branches, tags, etc.
Provenance information	Containers, virtual environments

Backup and Archiving needs	If any
Legal/ethical restrictions	If any
Versioning control and rules/workflows managing	Specify the repository
Code transfer needs and security	If any
Long term preservation needs	Only if applies to a given official release version
Documentation and inline comments rules	Specify
Metadata management	(available even when the software is not)
Resources need	At each step of the life cycle

Table 3. List of items in a software sheet.

3. FAIR data

The FAIR Guiding Principles (Wilkinson et al.; 2016: DOI: 10.1038/sdata.2016.18) describes distinct considerations for contemporary data publishing environments with respect to supporting both manual and automated deposition, exploration, sharing, and reuse. A metric to quantify the degree of "FAIRness" of each dataset in ENERXICO has been defined. It results on a normalized value (between 0 and 1) for each of the 4 FAIR components. In turn, this (0,1) value results from assigning a flag value 0/1 to each of the FAIR subcomponents defined by Wilkinson et al. (2016) and listed in Table 4.

F	**FINDABLE**		
F.1	Persistent Identifiers (PDI)	(meta)data are assigned a globally unique and persistent identifier	0/1
F.2	Rich metadata	data are described with rich metadata (defined by subcomponent R.1 below)	0/1
F.3	Data registered in searchable resources	(meta)data are registered or indexed in a searchable resource	0/1
F.4	Metadata specifies the PDI	metadata clearly and explicitly include the identifier of the data it describes	0/1
A	**ACCESSIBLE**		

A.1	Retrievable by the PDI with a standardized protocol	(meta)data are retrievable by their identifier using a standardized communications protocol	0/1
A.1.1	Open, free protocol	The protocol is open, free, and universally implementable	0/1
A.1.2	Authentication and Authorization	The protocol allows for an authentication and authorization procedure, where necessary	0/1
A.2	Metadata availability	Metadata are accessible beyond the data availability	0/1
I	**INTEROPERABLE**		
I.1	Formal, accessible, shared and applicable language	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	0/1
I.2	FAIR vocabulary	(meta)data use vocabulary that follow FAIR principles	0/1
I.3	Metadata references	Metadata includes qualified references to other metadata	0/1
R	**REUSABLE**		
R.1	Relevant metadata	(meta)data have plurality of accurate and relevant attributes	0/1
R.1.1	Usage license	(meta)data are released with clear and accessible data usage license	0/1
R.1.2	Provenance	(meta)data are associated with detailed provenance	0/1
R.1.3	Community standards	(meta)data meet domain-relevant community standards	0/1

Table 4. Definition of the different FAIR components and flag value (0/1) used to quantify the degree of fairness of each dataset.

3.1 Making ENERXICO data Findable

ENERXICO datasets suited for publication will be easily citable and easily findable with the assignation of Persistent Identifiers.

- The codes will be stored in repositories which permit versioning and tags for the identification of official releases and the connection with their outputs.
- Whenever possible, a rich metadata model and the register in disciplinary

repositories will be used to allow other scientists to find the datasets produced by the project.

- Given the variety of the data of the project, the specific solutions and data models adopted for each dataset and software will be found in the corresponding sheet of the DMP.

3.2 Making ENERXICO data openly Accessible

The open-data will be made accessible as follows:

- The source-codes of the parallel scientific codes in ENERXICO and related software components licensed as open-source can be included in a web repository for codes and toolkits at the end of the project. It will be useful for archiving project results.
- Datasets access will depend on the different case and it will be described in the corresponding dataset sheet. Restriction of access will be guaranteed in cases ethical or proprietary issues arise. Metadata will be made available as soon and as long as possible, independently on the accessibility of data.

3.3 Making ENERXICO data Interoperable

The choice of metadata standards and of the way to access the data is still under discussion between the consortium members. Whenever possible, data coming from other resources will also be described in the DMP. Metadata standards will be chosen to guarantee the maximum interoperability.

3.4 Increase ENERXICO data Re-use

The ENERXICO open-datasets will be licensed under Creative Commons data licensing (see Table 5) to let the widest reuse possible of it, since this license allows both commercial and non-commercial use of the data without any restriction. If necessary, an embargo in the data may exist to guarantee publication of results for a maximum of 1 year after the conclusion of the project. In any case, this will be specified in the corresponding dataset sheet.

		Allowed			
Creative Commons	**Description**	**Modification of the content**	**Commercial use**	**Free cultural works**	**Open definition**
CC0	Free content, no restrictions	yes	yes	yes	yes
BY	Attribution	yes	yes	yes	yes
BY-SA	Attribution + Share Alike	yes	yes	yes	yes
BY-NC	Non Commercial	yes	no	no	no

BY-ND	No Derivatives	no	yes	no	no
BY-NC-SA		yes	no	no	no
BY-NC-ND		no	no	no	no

Table 5. Data licensing options.

4. Allocation of Resource

There is no additional cost for making the ENERXICO datasets identified in Section 2 FAIR:

- The code performance evaluation datasets will be maintained at BSC facilities and included in publications.
- The rest of the open-data will be stored at the project site for at least three years after the end of the project. The infrastructure and personnel funds granted from the European Community will cover the storage, hardware and staff time to manage the servers on which the data will be stored.

5. Data security

Each dataset will be evaluated separately and excepcional security measures will be identified and applied. Regular backups for preventing loss of information will be used.

6. Ethical aspects

Early warning and hazard assessment can potentially have ethical implications: the diffusion of hazard results or a warning message can be risky for public order, and have social and economic impact, the project will distribute the ultimate simulations results (application to real cases) only under specific conditions and to the appropriate stakeholders, while the scientific results and products, like the models, etc., will be openly accessible. The limitations and conditions of distribution of each dataset will be indicated in the corresponding dataset sheet.

7. Annex I

Dataset Sheets

DataSet Sheet (WRF)

Valid values: 1(totally compliant), 0.5(partially/ongoing), 0(not compliant)

		Value		
Name	Data simulated with and from the WRF code	0,5	F.1	Persistent identifiers
Description	A set of problems on weather and climate that can be solved where the topography is a key component	0,2	F.2	Rich metadata
Data Category	1.01, 1.02, 1.03, and 1.04	0	F.3	Data registered in
Licence	TBD	0	F.4	Metadata specify the
Repository location	ERAS for input data; To be implemented at ciemat.es domain for outputdata		0,175	
Author	Several (Jorge Navarro, Pablo Garcia Müller)			
Naming Conventions	Just structured in a test per folder basis	Value	A	ACCESSIBLE
Versioning	No versioning established	0	A.1	Retrievable by the PDI
Format	To be decided, each dataset has its own format	0,5	A.2	Protocol is open, free
Size	Several files (less than 100 files, up to 2 GB per year), under construction	1	A.3	Protocol allows
Storage	TBD	0,3	A.4	Metadata accessible
Archive path	TBD		0,45	
Associated metadata	none, only README files			
Provenance	not structured, different datasets have different origins	Value	I	INTEROPERABLE
Backups needs	None (under Netapp protocol)	0,8	I.1	Language are formal,
access permissions	Specific permission given by CIEMAT, might be open upon agreement by the authors	0,8	I.2	Vocabulary is FAIR
Legal/ethical restrictions	None foreseen	0,2	I.3	Metadata includes
Reproducibility	Possible		0,6	
Data transfer needs	No need, everything is stored together			
Long term preservation	Initially, until the project ends	Value	R	REUSABLE
Metadata management	NA	0,5	R.1	(Meta)data have
Resources need	NA	0,5	R.2	Released with a clear
References to other datasets	NA	0,5	R.3	Provenance information
		0,5	R.4	Domain-relevant
			0,5	

DataSet Sheet Topography suite

Valid values: 1(totally compliant), 0.5(partially/ongoing), 0(not compliant)

		Value		
Name	Topography Test Suite	0,5	F.1	Persistent identifiers
Description	A set of problems that can be solved where the topography is a key component	0,2	F.2	Rich metadata
Data Category	Models and synthesized data (1.01 and 1.03)	0	F.3	Data registered in
Licence	TBD	0	F.4	Metadata specify the
Repository location	https://gitlab.lrz.de/sebastian-wolf/topography-benchmarks		0,175	
Author	Many (Sebastian Wolf, Otilio Rojas, Josep de la Puente, ...)			
Naming Conventions	Just structured in a test per folder basis	Value	A	ACCESSIBLE
Versioning	No versioning established	0	A.1	Retrievable by the PDI
Format	To be decided, each dataset has its own format	0,5	A.2	Protocol is open, free
Size	Several files (less than 100 files, 1.1 GB currently), under construction	1	A.3	Protocol allows
Storage	gitlab	0,3	A.4	Metadata accessible
Archive path	gitlab		0,45	
Associated metadata	none, only README files			
Provenance	not structured, different datasets have different origins	Value	I	INTEROPERABLE
Backups needs	None (under git)	0,8	I.1	Language are formal,
access permissions	Specific permission given by LRZ, might be open upon agreement by the authors	0,8	I.2	Vocabulary is FAIR
Legal/ethical restrictions	None foreseen	0,2	I.3	Metadata includes
Reproducibility	In principle possible		0,6	
Data transfer needs	No need, everything is stored together			
Long term preservation	Initially, until the project ends	Value	R	REUSABLE
Metadata management	NA	0,5	R.1	(Meta)data have
Resources need	NA	0,5	R.2	Released with a clear
References to other datasets	NA	0,5	R.3	Provenance information
		0,5	R.4	Domain-relevant
			0,5	

Software Sheet

Software Sheet - BSIT

Reference name of the program or	Barcelona Subsurface Imaging Tools (BSIT)
Description	An HPC geophysical imaging tools that includes seismic and EM capabilities for modelling, migration and inversion of 2D or 3D datasets
Author	BSC (several authors inc. Mauricio Hanzich, Josep de la Puente, J.E. Rodriguez, N. Gutierrez, J. Kormann, M. Ferrer, A. Farrés)
Programming language	C
Rules and best coding practices	Internal best practices and documentation rules
Access permissions and license	Intellectual property of Repsol. Pending academic usage license at BSC
Code size	100,000 lines
Repository type	Internal svn repository at BSC, unknown repositories at Repsol
Repository structure	One yearly deliverable in 2010-2018, internally several development branches
Provenance information	Build information on executables
Backup and Archiving needs	None
Legal/ethical restrictions	None
Versioning control and	Not available for public usage
Code transfer needs and security	Private code, under industrial secret
Long term preservation needs	NA
Documentation and inline comments	Internal documentation rules, documentation compiled with doxygen
Metadata management	NA
Resources need	Requirements for sw at each step of the lifecycle (access to repository, computational needs, accessibilities, permissions....)

Software Sheet - WRF	
Reference name of the program or	Weather Research Forecast (WRF)
Description	A mesoscale numerical weather prediction system designed for both atmospheric research and operational forecasting applications
Author	NCAR (A couple of modules designed and implemented by CIEMAT staff: PBL module with YSU and SL
Programming language	F90, F95, C
Rules and best coding practices	Internal best practices and documentation rules
Access permissions and license	Open source
Code size	100,000 lines
Repository type	https://www2.mmm.ucar.edu/wrf/users/downloads.html
Repository structure	To be implemented at ciemat.es domain
Provenance information	Build information on executables
Backup and Archiving needs	None
Legal/ethical restrictions	None
Versioning control and	wrfv4.1
Code transfer needs and security	Open source
Long term preservation needs	NA
Documentation and inline comments	https://www2.mmm.ucar.edu/wrf/users/pub-doc.html
Metadata management	NA
Resources need	FORTRAN 90 or 95 and C, perl, OpenMP, MPI RSL-LITE, C-shell and Bourne shell, make, M4, sed, awk, netCDF-4. PHD5. GriB-1. NCL. RIB4. ARWpost

Software Sheet - Alya	
Reference name of the program or	Alya
Description	A multi-physics software developed in CASE department designed to work in massive parallel environments.
Author	The main architect is Guillaume Houzeaux, in the context of ENERXICO main contributions are from Matias Avila, Daniel Mira, Ambrus Both and Oriol Lehmkuhl. NCAR (A couple of modules designed and implemented by CIEMAT staff: PBL module with YSU and SL implementations)
Programming language	F90
Rules and best coding practices	Internal best practices and documentation rules
Access permissions and license	Via LUL agremment between BSC and the user
Code size	700,000 lines
Repository type	Git Lab
Repository structure	
Provenance information	Build information on executables
Backup and Archiving needs	None
Legal/ethical restrictions	None
Versioning control and	Following git format
Code transfer needs and security	Open source
Long term preservation needs	NA
Documentation and inline comments	https://gitlab.bsc.es/alya/alya/-/wikis/home
Metadata management	NA
Resources need	FORTRAN 90 or 95 and C, MPI, make, python, paraview

Software Sheet - SeisSol	
Reference name of the program or	SeisSol
Description	SeisSol solves seismic wave propagation (elastic, viscoelastic) and dynamic rupture problems on heterogeneous 3D models.
Author	LMU and TUM
Programming language	C++, Fortran, Python
Rules and best coding practices	Internal best practices and documentation rules
Access permissions and license	Open source (BSD-3)
Code size	950000
Repository type	https://github.com/SeisSol/SeisSol
Repository structure	Rolling release, snapshots for verification, development/feature branches
Provenance information	Build information on executables
Backup and Archiving needs	None
Legal/ethical restrictions	None
Versioning control and	feature branches with pull requests/code review
Code transfer needs and security	Open source
Long term preservation needs	NA
Documentation and inline comments	https://seissol.readthedocs.io/en/latest/
Metadata management	NA
Resources need	gcc or intel, numpy, parmetis, libxsmm, pspamm, mpi, netcdf, hdf5, external meshing tool (gmsh, simmetrix)

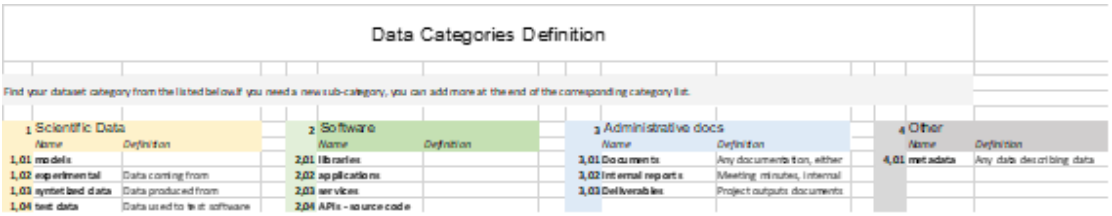
Software Sheet - ExaHyPE	
Reference name of the program or	ExaHyPE Engine
Description	An engine for the simulation of systems of hyperbolic PDEs, as stemming from conservation laws
Author	The ExaHyPE Consortium
Programming language	C++, Fortran, Python
Rules and best coding practices	Internal best practices and documentation rules
Access permissions and license	Open source
Code size	500,000 lines of code
Repository type	https://gitlab.lrz.de/exahype/ExaHyPE-Engine
Repository structure	Running release mode, stable versions available on branches, several development branches
Provenance information	Build information on executables
Backup and Archiving needs	None
Legal/ethical restrictions	None
Versioning control and	Running release mode
Code transfer needs and security	Open source
Long term preservation needs	NA
Documentation and inline comments	http://www.peano-framework.org/exahype/guidebook.pdf
Metadata management	NA
Resources need	Intel or GNU compilers, MPI, TBB, Python3

Software Sheet - SEM46	
Reference name of the program or workflow	Barcelona Subsurface Imaging Tools (BSIT)
Description	3D seismic modeling and inversion code, developed mainly in the frame of the SEISCOPE project (https://seiscope2.osug.fr/) for tackling modeling and full waveform inversion topics from the near surface to the deep crustal scale.
Author	Romain Brossier romain.brossier@univ-grenoble-alpes.fr
Programming language	
Rules and best coding practices	Internal best practices and documentation rules
Access permissions and license	SEM46 is available upon request to Romain Brossier romain.brossier@univ-grenoble-alpes.fr SEM46 relies on a BSD-like license but include restrictions for the diffusion, imposed by the funding partners of the SEISCOPE project
Code size	
Repository type	
Repository structure	
Provenance information	
Backup and Archiving needs	None
Legal/ethical restrictions	None
Versioning control and rules/workflows managing	
Code transfer needs and security	
Long term preservation needs	NA
Documentation and inline comments rules	
Metadata management	NA
Resources need	

Software Sheet - DualSPHysics and Black Hole codes	
Reference name of the program or	DualSPHysics and Black Hole codes
Description	DualSPHysics is a hardware accelerated Smoothed Particle Hydrodynamics code developed to solve free-surface flow problems. The extension to multiphase flow has been developed through the ENERXICO project and is called the Black Hole code.
Author	The DualSPHysics and Black Hole codes developers
Programming language	C++, CUDA, Java
Rules and best coding practices	Internal best practices and documentation rules
Access permissions and license	Open source (DualSPHysics), non-open source (BH code)
Code size	100,000 lines of code
Repository type	https://github.com/DualSPHysics/DualSPHysics/tree/develop
Repository structure	Running release mode, stable versions available on branches, several development branches
Provenance information	Build information on executables
Backup and Archiving needs	None
Legal/ethical restrictions	None
Versioning control and rules/workflows	Running release mode
Code transfer needs and security	Open source
Long term preservation needs	NA
Documentation and inline comments	https://dual.sphysics.org/doxygen/
Metadata management	NA
Resources need	C++ and CUDA compilers, MPI, Java

Formats definition

Data Categories definition

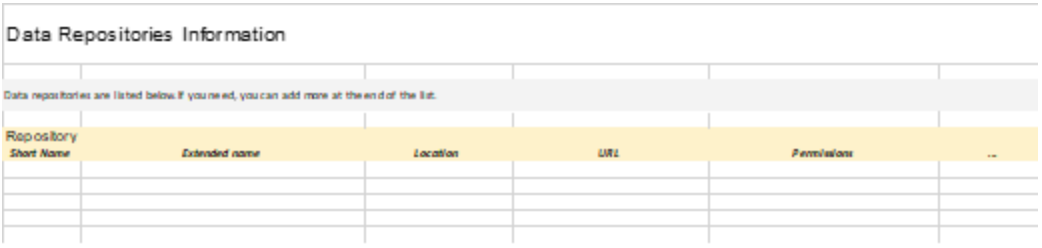


Data Categories Definition

Find your dataset category from the listed below.If you need a new sub-category, you can add more at the end of the corresponding category list.

1 Scientific Data		2 Software		3 Administrative docs		4 Other	
Name	*Definition*	*Name*	*Definition*	*Name*	*Definition*	*Name*	*Definition*
1,01 models		2,01 libraries		3,01 Documents	Any documentation, either	4,01 metadata	Any data describing data
1,02 experimental	Data coming from	2,02 applications		3,02 Internal reports	Meeting minutes, internal		
1,03 synthetized data	Data produced from	2,03 services		3,03 Deliverables	Project outputs documents		
1,04 test data	Data used to test software	2,04 APIs - source code					

Data repositories Information

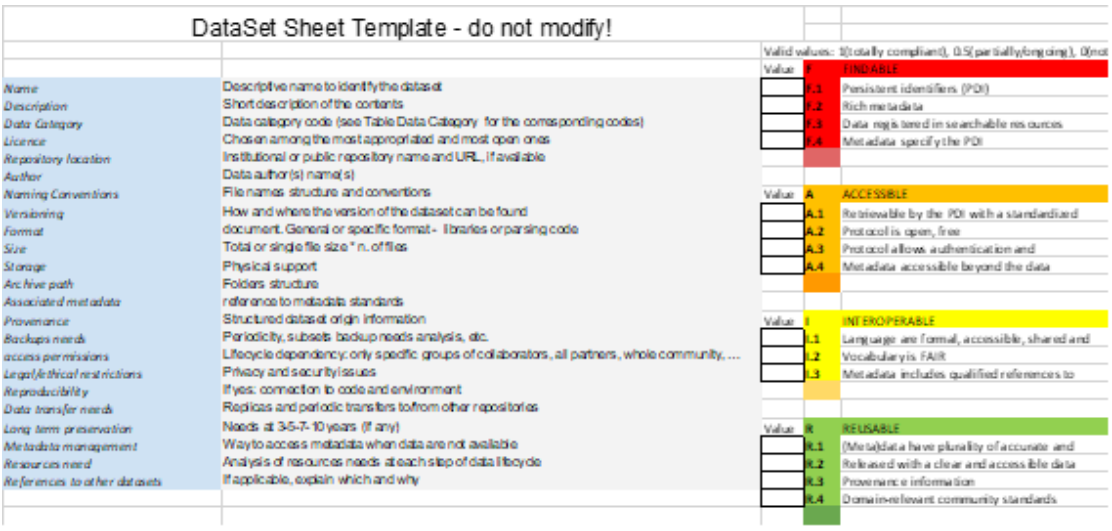


Data Repositories Information

Data repositories are listed below. If you need, you can add more at the end of the list.

Repository					
Short Name	*Extended name*	*Location*	*URL*	*Permissions*	*...*

DataSet Sheet Template - do not modify



DataSet Sheet Template - do not modify!

Valid values: 1(totally compliant), 0.5(partially/ongoing), 0(not

		Value		
			F	FINDABLE
Name	Descriptive name to identify the dataset		F.1	Persistent identifiers (PDI)
Description	Short description of the contents		F.2	Rich metadata
Data Category	Data category code (see Table Data Category for the corresponding codes)		F.3	Data registered in searchable resources
Licence	Chosen among the most appropriated and most open ones		F.4	Metadata specify the PDI
Repository location	Institutional or public repository name and URL, if available			
Author	Data author(s) name(s)			
Naming Conventions	File names structure and conventions	Value	A	ACCESSIBLE
Versioning	How and where the version of the dataset can be found		A.1	Retrievable by the PDI with a standardized
Format	document. General or specific format - libraries or parsing code		A.2	Protocol is open, free
Size	Total or single file size * n. of files		A.3	Protocol allows authentication and
Storage	Physical support		A.4	Metadata accessible beyond the data
Archive path	Folders structure			
Associated metadata	reference to metadata standards			
Provenance	Structured dataset origin information	Value	I	INTEROPERABLE
Backups needs	Periodicity, subsets backup needs analysis, etc.		I.1	Language are formal, accessible, shared and
access permissions	Lifecycle dependency: only specific groups of collaborators, all partners, whole community, ...		I.2	Vocabulary is FAIR
Legal/ethical restrictions	Privacy and security issues		I.3	Metadata includes qualified references to
Reproducibility	If yes: connection to code and environment			
Data transfer needs	Replicas and periodic transfers to/from other repositories			
Long term preservation	Needs at 3-5-7-10 years (if any)	Value	R	REUSABLE
Metadata management	Way to access metadata when data are not available		R.1	(Meta)data have plurality of accurate and
Resources need	Analysis of resources needs at each step of data lifecycle		R.2	Released with a clear and accessible data
References to other datasets	If applicable, explain which and why		R.3	Provenance information
			R.4	Domain-relevant community standards

Software Sheet - do not modify

Software Sheet - do not modify	
Reference name of the program or	Name of the code
Description	Brief description of the functionality and applicability of the software
Author	specify
Programming language	specify the programming language(s)
Rules and best coding practices	conventions for filenames, link to an external manual, if exists (ex: PEP8, etc.)
Access permissions and license	lifecycle dependency: only specific groups of collaborators, all partners, whole community, etc.
Code size	if relevant (to be updated)
Repository type	GitHub, GitLab, Bitbucket, SourceForge
Repository structure	Branches, tags, etc.
Provenance information	Containers, virtual environments
Backup and Archiving needs	If any
Legal/ethical restrictions	If any
Versioning control and	Specify the repository
Code transfer needs and security	If any
Long term preservation needs	Only if applies to a given official release version
Documentation and inline comments	specify
Metadata management	(available even when the software is not)
Resources need	Requirements for sw at each step of the lifecycle (access to repository, computational needs, accessibilities, permissions,

FAIR DATA definitions - do not modify

FAIR DATA definitions - do not modify

GoFAIR - principles definition link

To evaluate the FAIRness of a dataset, a punctuation should be given to each of the following points (elaborated by www.force11.org)Valid values: 1 (totally compliant), 0.5 (parti

F	FINDABLE	A	ACCESSIBLE	I	INTEROPERABLE	R	REUSABLE
F.1	(Meta)data are assigned a globally	A.1	(Meta)data are retrievable by their	I.1	(Meta)data use a formal,	R.1	Meta(data) are richly described
F.2	Data are described with rich metadata	A.2	The protocol is open, free, and	I.2	Metadata use vocabularies that	R.2	(Meta)data are released with a
F.3	(Meta)data are registered or indexed	A.3	The protocol allows for an	I.3	(Meta)data include qualified	R.3	(Meta)data are associated with
F.4	Metadata clearly and explicitly include	A.4	Metadata are accessible, even when			R.4	(Meta)data meet domain-

Data licensing

Creative	Description	Modification of the	Commercial Use	Free cultural works	Open definition
		Allowed			
CC0	Free content, no restrictions	yes	yes	yes	yes
BY	Attribution	yes	yes	yes	yes
BY-SA	Attribution+ShareAlike	yes	yes	yes	yes
BY-NC	NonCommercial	yes	no	no	no
BY-ND	NoDerivatives	no	yes	no	no
BY-NC-SA		yes	no	no	no
BY-NC-ND		no	no	no	no

Definitions of common license types

data.world: Common license types for datasets

For details, click here

Public Domain	The work has been dedicated to the public domain by waiving all rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law.
Attribution	You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
Share-alike	If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.
Non-commercial	You may not use the material for commercial purposes.
Database Only	License applies to the database only and not its contents or data.
No Derivatives	No Derivative Works. You may not alter, transform, or build upon this work.