D1.2 Project Data Management Plan

Document Information

Contract Number	821418
Project Website	https://estimate.bsc.es
Contractual Deadline	M6
Dissemination Level	Public
Nature	R
Author	Eduardo J. Pérez Sánchez (BSC)
Contributor(s)	All the members from the consortium
Reviewer	Daniel Mira (BSC), Oriol Lehmkuhl (BSC)
Keywords	Management, repositories, FAIR metrics

Notice
The research leading to these results has received funding from the European Union's Horizon 2020 Programme under the ESTiMatE project (www.estimate-project.eu), grant agreement nr. 821418.

Change Log

Version	Author	Description of Change
V1.1	Eduardo J. Pérez Sánchez	Initial Draft
V1.2	All members from the consortium	Final version

Table of Contents

Executive Summary

The Data Management Plan (DMP) of the ESTiMatE project describes the management of datasets that will be generated as well as the software that will be used during the lifetime of this project. This document is deliverable D1.2 from the project and gathers such information. To this purpose the following information is put forward:

- The datasets generated during the project and their management during and after it.
- The methodologies and standards (if any) that will be applied to manage each of the datasets.
- The datasets storage during and after the project and their accessibility after the conclusion of the project.

Some of the datasets generated in the project are expected to be confidential and, in consequence, not distributable. The selection of which of them will be public or not has still to be discussed with the Topic Manager of the project. Any relevant change with regards to the current DMP contained in this document will be submitted to the Commission.

Introduction

The ESTiMatE is a Clean Sky H2020 project aimed at developing a modelling strategy using CFD simulations for the prediction of soot in terms of chemical evolution and particle formation in conditions relevant to aero engine operation. This DMP describes how data generated during the project will be managed during and after it.

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

Structure of the ESTiMatE annex

As stated in the Grant Agreement (GA), in the ESTiMatE project several flame configurations as well as the atomization process for an air-blast atomizer will be measured and simulated, each one of them referred as a case configuration. On one hand, experimental detailed information about velocity, species, etc. spatial fields and soot measurements or particle size distributions will be obtained depending on the experiment. On the other hand,

such measurements will be compared with simulations that will require High Performance Computing (HPC) and the application of advanced combustion and soot models.

In this way, for each case configuration several databases will be created according to the following general structure (some of the following database may be omitted depending on the configuration case):

- Boundary conditions for the configuration.
- Experimental measurements of the configuration.
- Simulation set-up for the configuration (constant models, meshes, etc.).
- Simulation results for the configuration.

Each configuration has a summary sheet where the main information about the configuration is included and a second part where sheets with detailed information about the repositories are given together with the FAIR metrics. In addition, each code used in the project has a descriptive sheet with its main characteristics. This information is given in the annex of this document.

In the following, the items included in each of the different sheets are described.

Data summary

In this sheet a summary of the dataset related to one configuration case is given with the following entries:

Item	Comments/explanation
Project	Name of the configuration case
Relevant aspects	Aspects to be emphasized about the case configuration
Codes	Codes used for calculations
WPs involved	Project work packages involved in the configuration case
Description	Description of the activities carried out in the configuration case

Table 1: list of items that describe the main characteristics of each case configuration.

Item	Comments/explanation
Name	Name of the datasets related to the configuration case
Description	Description of the datasets related to the configuration case
Data category	Data category according to table 4
Repository location	Name of the repository where datasets are located
FAIR code	Average mark for each category of the FAIR metrics
References to other datasets/software	Name of other referenced datasets/software

Table 2: list of items that describe the datasets for each configuration.

Dataset sheet

The following information is included for each dataset sheet related to a configuration case:

Item	Comments/explanation
Name	Descriptive name to identify the dataset
Data category	Data category code (see Table Data Category for the corresponding codes)
Licence	Chosen among the most appropriate ones
Repository location	Institutional or public repository name and URL, if available
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	Standard formats and content standards, definitions, ontologies, etc. Link to description of format document. General or specific format - libraries or parsing code
Size	Estimation of total files size
Storage	Physical support
Archive path	Folders structure
Associated metadata	Reference to metadata standards
Provenance	Structured dataset origin information
Backups needs	Periodicity, subsets backup needs analysis, etc.
Access permissions	Lifecycle dependency: only specific groups of collaborators, all partners, whole community…
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
References to other datasets	Name of other referenced datasets

Table 3: list of items in the dataset sheet and their definition.

The list of the data category is given here.

Data category	Code	Name	Comments
Scientific data	1.1	Models	Data generated by the application of 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 synthetical) used to validate models
Software	2.1	Libraries	Implementation of libraries
	2.2	Applications	Development of applications
	2.3	Services	Services provided
	2.4	APIs	Creation of application programming interfaces
Administrative documents	3.1	Documents	Any documentation, either public or private, such as code documentation, technical notes, etc., not directly mentioned in the project deliverable list.
	3.2	Internal reports	Meeting 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 4: summary of the different data categories.

Software sheet

In a similar way to the dataset sheet, the software sheet contains a detailed description of the codes used in the simulations according to the following table:

Item	Comments
Reference name of the program or workflow	Name of the code
Description	Brief description of the functionality and applicability of the software
Author	Authors of the software
Programming language	Programming language(s) used for code implementation
Rules and best coding practices	Conventions for filenames, link to an external manual, if exists (ex: PEP8, etc.)
Access permissions and license	Lifecycle dependency: groups of collaborators, all partners, whole community, etc.
Code size	Code size
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	If any
Metadata management	Available even when the software is not
Resources need	Requirements for software at each step of the life cycle (access to repository, computational needs, accessibilities, permissions, ...)

Table 5: list of items in the software sheet.

FAIR data

The FAIR Guiding Principles (Wilkinson et al.; 2016; DOI: 10.1038/sdata.2016.18) describe 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 ESTiMatE 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 again between 0 and 1 to each of the FAIR subcomponents defined by Wilkinson et al. (2016) and listed in Table 6.

F	**FINDABLE**	
F.1	Persistent Identifiers (PDI)	(Meta)data are assigned a globally unique and persistent identifier
F.2	Rich metadata	Data are described with rich metadata (defined by subcomponent R.1 below)
F.3	Metadata specifies the PDI	Metadata clearly and explicitly include the identifier of the data it describes
F.4	Data registered in searchable resources	(Meta)data are registered or indexed in a searchable resource
A	**ACCESSIBLE**	
A.1	Retrievable by the PDI with a standardized protocol	(Meta)data are retrievable by their identifier using a standardized communications protocol.
A.1.2	Open, free protocol	The protocol is open, free and universally implementable
A.1.3	Authentication and authorization	The protocol allows for an authentication and authorization procedure, where necessary
A.2	Metadata availability	Metadata are accessible beyond the data availability
I	**INTEROPERABLE**	
I.1	Formal, accessible, shared and applicable language	(Meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation
I.2	FAIR vocabulary	(Meta)data use vocabularies that follow FAIR principles
I.3	Metadata references	Metadata includes qualified references to other metadata
R	**REUSABLE**	
R.1	Relevant metadata	(Meta)data have plurality of accurate and relevant attributes

R.1.1	Usage license	(Meta)data are released with a clear and accessible data usage license
R.1.2	Provenance	(Meta)data are associated with detailed provenance
R.1.3	Community standards	(Meta)data meet domain-relevant community standards

Table 6: definition of the different FAIR components used to quantify the degree of fairness of each dataset.

Making ESTiMatE data Findable

ESTiMatE 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 this DMP.

Making ESTiMatE data openly Accessible

Datasets access will depend on the different case and will be described in the corresponding dataset sheet. Restriction of access will be guaranteed in cases confidential data from the Topic Manager is used or generated. Metadata will be made available in the web, independently on the accessibility of data.

Making ESTiMatE data Interoperable

The choice of metadata standards and the way to access the data is still under discussion between the consortium members. Metadata standards will be chosen to guarantee the maximum interoperability.

Increase ESTiMatE data Re-use

The ESTiMatE open-datasets will be licensed under some Creative Commons data licensing (see Table 7).

		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+ 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

Table 7: data licensing options.

Allocation of resources

There is no additional cost for making the ESTiMatE datasets FAIR:

- The code performance evaluation datasets of the open source codes of the project will be maintained at BSC facilities and could be 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.

Data security

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

Engagement with EUDAT

Solutions for data management and movement will be provided. In particular, the use of EUDAT services to store and publish research data (B2SHARE), distribute and store large volumes of data based on data policies (B2SAFE) and transfer data between data resources and external computational facilities (B2STAGE), exploiting data citation (B2HANDLE), that for EUDAT hosted data is managed through Persistent Identifiers (PIDs), and metadata enrichment (B2NOTE), will be fostered.

Annex

The management of data information is going to be carried out to permit researchers to keep track of the stored data and identify issues, common requirements and solutions. The structure and the final format of the DMP may change in the future to fit the needs of the project so, in such case, an updated version of DMP will be submitted to the Commission.

In the following, the summary sheet and dataset sheets are given for the following configuration cases:

- Case configuration I: results of soot formation and pollutant emissions on laminar counterflow diffusion flame rig at elevated pressure.
- Case configuration II: results of soot formation and pollutant emissions on turbulent flames at low TRL test rigs.
- Case configuration III: results of soot formation and pollutant emissions on a single sector rig.
- Case configuration IV: results of soot formation and pollutant emissions on a full annular configuration.
- Case configuration V: results of primary atomization in air-blast atomizers.

As the configuration cases I to IV follow the same repository structure, the database sheets expected for each case have the same nature, so only the dataset sheets for one configuration case are shown for cases I to IV. Case V is essentially different and it is treated separately. Subsequently, the description of FAIR metrics are provided. At this stage of the project, we foresee the FAIR metrics of the case configurations are essentially the same and will be treated in the same way, so only one header for FAIR is described here that should be representative of all the data sheets. As said before, if there are changes on the management of the data, a new DMP plan will be submitted. Then the software sheet is given for each code. Finally, the formats definition and the data repositories information are provided.

Summary sheets

Case configuration I

Project	Results of soot formation and pollutant emissions on a laminar counter flow diffusion flame rig at elevated pressure
Relevant aspects	-
Codes	Alya
WPs involved	2, 3, 4 & 5
Description	Experiments and numerical simulations in laminar conditions at elevated pressure will be conducted to validate the kinetics and soot models. The data sets included in this configuration are a first and relevant feedback/optimization loop for cross-validating the kinetics and soot models. Regarding the experiments the datasets include major species (e.g. H2, CO, CO2) and intermediate/minor species, focusing at those that act as soot precursors (e.g. acetylene, benzene, naphthalene, acenaphthalene, cyclohexane etc.), temperature profiles, soot volume fractions measured with laser light extinction and particle size distributions determined with Laser-Induced Incandescence (LII) for selected conditions.

Dataset name	***Description***	***Data Category***	***Repository location***	***F***	***A***	***I***	***R***	***References to other datasets/ software***
Case configuration	Description of the geometry and nominal conditions of the experiment	Scientific data	EUDAT, BSC, TUD, IVLR, KIT	0.75	0.875	0.67	0.575	No
Experimental data	Experiments measured data including images and other relevant information	Scientific data	EUDAT, BSC, TUD, IVLR, KIT	0.75	0.875	0.67	0.575	No
Model set-up	Set-up of the simulations including models and their parameters, etc.	Scientific data	EUDAT, BSC, TUD, IVLR, KIT	0.75	0.875	0.67	0.575	No

Simulation results	Fields and variables obtained from the simulations including velocity, temperature, etc. fields, soot mass, etc.	Scientific data	EUDAT, BSC, TUD, IVLR, KIT	0.75	0.875	0.67	0.575	No

Case configuration II

Project	Results of soot formation and pollutant emissions on turbulent flames at low TRL test rigs
Relevant aspects	-
Codes	Alya
WPs involved	2, 3, 4 & 5
Description	This DMP is dedicated to numerical simulations using LES models for the assessment of the prediction of soot in low TRL test rigs. The test rigs correspond to a turbulent diffusion flame on a counter flow rig and a swirling configuration and are the baseline configurations for modelling validation. Regarding the experiments the velocity fields are measured by PIV and normalized density distribution of the flame are estimated together with local temperature measurements obtained by suction pyrometry. Furthermore, OH* chemiluminescence images of the flame and species concentration measurements (OH and CH2O PLIF) as well as soot measurements (LII) will be provided.

Dataset name	***Description***	***Data Category***	***Repository location***	***F***	***A***	***I***	***R***	***References to other datasets/ software***
Case configuration	Description of the geometry and nominal conditions of the experiment	Scientific data	EUDAT, BSC, UPV, TUD	0.75	0.875	0.67	0.575	No
Experimental data	Experiments measured data including images and other relevant information	Scientific data	EUDAT, BSC, UPV, TUD	0.75	0.875	0.67	0.575	No

Model set-up	Set-up of the simulations including models and their parameters, etc.	Scientific data	EUDAT, BSC, UPV, TUD	0.75	0.875	0.67	0.575	No
Simulation results	Fields and variables obtained from the simulations including velocity, temperature, etc. fields, soot mass, etc.	Scientific data	EUDAT, BSC, UPV, TUE, TUD	0.75	0.875	0.67	0.575	No

Case configuration III

Project	Results of soot formation and pollutant emissions on a single sector rig
Relevant aspects	-
Codes	Alya
WPs involved	2, 3, 4 & 5
Description	This DMP is dedicated to run numerical simulations with the models for the assessment of the prediction of soot in realistic configurations. This DMP provides relevant information about the performance of different soot models.

Dataset name	***Description***	***Data Category***	***Repository location***	***F***	***A***	***I***	***R***	***References to other datasets/ software***
Case configuration	Description of the geometry and nominal conditions of the experiment	Scientific data	EUDAT, BSC, UPV, TUE, TUD	0.75	0.875	0.67	0.575	No
Experimental data	Experiments measured data including images and other relevant information	Scientific data	EUDAT, BSC, UPV, TUE, TUD	0.75	0.875	0.67	0.575	No

Model set-up	Set-up of the simulations including models and their parameters, etc.	Scientific data	EUDAT, BSC, UPV, TUE, TUD	0.75	0.875	0.67	0.575	No
Simulation results	Fields and variables obtained from the simulations including velocity, temperature, etc. fields, soot mass, etc.	Scientific data	EUDAT, BSC, UPV, TUE, TUD	0.75	0.875	0.67	0.575	No

Case configuration IV

Project	Results of soot formation and pollutant emissions on a full annular configuration
Relevant aspects	-
Codes	Alya
WPs involved	2, 3, 4 & 5
Description	This DMP is dedicated to run numerical simulations for the assessment of the prediction of soot in a real configuration. It corresponds to large-scale simulations of a realistic aeroengine configuration.

Dataset name	***Description***	***Data Category***	***Repository location***	***F***	***A***	***I***	***R***	***References to other datasets/ software***
Case configuration	Description of the geometry and nominal conditions of the experiment	Scientific data	EUDAT, BSC, UPV, TUE, TUD	0.75	0.875	0.67	0.575	No
Experimental data	Experiments measured data including images and other relevant information	Scientific data	EUDAT, BSC, UPV, TUE, TUD	0.75	0.875	0.67	0.575	No

Model set-up	Set-up of the simulations including models and their parameters, etc.	Scientific data	EUDAT, BSC, UPV, TUE, TUD	0.75	0.875	0.67	0.575	No
Simulation results	Fields and variables obtained from the simulations including velocity, temperature, etc. fields, soot mass, etc.	Scientific data	EUDAT, BSC, UPV, TUE, TUD	0.75	0.875	0.67	0.575	No

Case configuration V

Project	Results of primary atomization in air-blast atomizers
Relevant aspects	-
Codes	Paris, Alya
WPs involved	6
Description	The present DMP is devoted to the numerical simulations using DNS for the study of the primary atomization process in airblast atomizers. Case configuration is that of fundamental experiments in the literature dealing with the atomization of a planar liquid film induced by an air stream.

Dataset name	***Description***	***Data Category***	***Repository location***	***F***	***A***	***I***	***R***	***References to other datasets/ software***
Case configuration	Description of the geometry and nominal conditions of the experiment to be simulated	Scientific data	EUDAT, BSC, UPV	0.75	0.875	0.67	0.575	No
Model set-up	Set-up of the simulations including models and their parameters...	Scientific data	EUDAT, BSC, UPV	0.75	0.875	0.67	0.575	No

Simulation results	Fields and variables obtained from the simulations including droplet size, velocity and statistical quantities describing the droplet population.	Scientific data	EUDAT, BSC, UPV	0.75	0.875	0.67	0.575	No

Dataset sheets

For case configurations I to IV

Name	Case configuration
Description	Description of the geometry and nominal boundary conditions of the experiment
Data Category	Scientific data
Licence	To be agreed with Topic Manager
Repository location	EUDAT, BSC, UPV, TUE, TUD, IVLR, KIT
Author	-
Naming Conventions	-
Versioning	-
Format	csv, vtk, txt
Size	Of the order of few megabytes
Storage	Physical support
Archive path	Folders structure
Associated metadata	-
Provenance	-
Backups needs	-
Access permissions	ESTiMatE partners
Legal/ethical restrictions	None
Reproducibility	Yes
Data transfer needs	-
Long term preservation	10 years
Metadata management	Website
Resources need	-
References to other datasets	-

Name	Experimental data
Description	Experiments measured data including images and other relevant information
Data Category	Scientific data
Licence	To be agreed with Topic Manager
Repository location	EUDAT, BSC, UPV, TUE, TUD, IVLR, KIT
Author	-
Naming Conventions	-
Versioning	-
Format	csv, vtk, txt
Size	Of the order of several gigabytes
Storage	Physical support
Archive path	Folders structure
Associated metadata	-
Provenance	-
Backups needs	-
Access permissions	ESTiMatE partners
Legal/ethical restrictions	None
Reproducibility	Yes
Data transfer needs	-
Long term preservation	10 years
Metadata management	Website
Resources need	-
References to other datasets	-

Name	Model set-up
Description	Set-up of the simulations including models and their parameters, etc.
Data Category	Scientific data
Licence	To be agreed with Topic Manager
Repository location	EUDAT, BSC, UPV, TUE, TUD, IVLR, KIT
Author	-
Naming Conventions	-
Versioning	-
Format	csv, vtk, txt
Size	Of the order of several gigabytes
Storage	Physical support
Archive path	Folders structure
Associated metadata	-
Provenance	-
Backups needs	-
Access permissions	ESTiMatE partners
Legal/ethical restrictions	None

Reproducibility	Yes
Data transfer needs	-
Long term preservation	10 years
Metadata management	Website
Resources need	-
References to other datasets	-

Name	Simulations results
Description	Fields and variables obtained from the simulations including velocity, temperature, etc. fields and soot mass, etc.
Data Category	Scientific data
Licence	To be agreed with Topic Manager
Repository location	EUDAT, BSC, UPV, TUE, TUD, IVLR, KIT
Author	-
Naming Conventions	-
Versioning	-
Format	csv, vtk, txt
Size	Of the order of several gigabytes
Storage	Physical support
Archive path	Folders structure
Associated metadata	-
Provenance	-
Backups needs	-
access permissions	ESTiMatE partners
Legal/ethical restrictions	None
Reproducibility	Yes
Data transfer needs	-
Long term preservation	10 years
Metadata management	Website
Resources need	-
References to other datasets	-

For case configuration V

Name	Case configuration
Description	Description of the geometry and nominal conditions of the experiment to be simulated
Data Category	Scientific data
Licence	To be agreed with Topic Manager
Repository location	EUDAT, BSC, UPV
Author	-

Naming Conventions	-
Versioning	-
Format	csv, vtk, txt
Size	Of the order of few megabytes
Storage	Physical support
Archive path	Folders structure
Associated metadata	-
Provenance	-
Backups needs	-
Access permissions	ESTiMatE partners
Legal/ethical restrictions	None
Reproducibility	Yes
Data transfer needs	-
Long term preservation	10 years
Metadata management	Website
Resources need	-
References to other datasets	-

Name	Model set-up
Description	Set-up of the simulations including models and their parameters, etc.
Data Category	Scientific data
Licence	To be agreed with Topic Manager
Repository location	EUDAT, BSC, UPV
Author	-
Naming Conventions	-
Versioning	-
Format	csv, vtk, txt
Size	Of the order of several gigabytes
Storage	Physical support
Archive path	Folders structure
Associated metadata	-
Provenance	-
Backups needs	-
Access permissions	ESTiMatE partners
Legal/ethical restrictions	None
Reproducibility	Yes
Data transfer needs	-
Long term preservation	10 years
Metadata management	Website
Resources need	-
References to other datasets	-

Name	Simulation results
Description	Fields and variables obtained from the simulations including droplet size, velocity and statistical quantities describing the droplet population.
Data Category	Scientific data
Licence	To be agreed with Topic Manager
Repository location	EUDAT, BSC, UPV
Author	-
Naming Conventions	-
Versioning	-
Format	csv, vtk, txt
Size	Of the order of several gigabytes
Storage	Physical support
Archive path	Folders structure
Associated metadata	-
Provenance	-
Backups needs	-
access permissions	ESTiMatE partners
Legal/ethical restrictions	None
Reproducibility	Yes
Data transfer needs	-
Long term preservation	10 years
Metadata management	Website
Resources need	-
References to other datasets	-

FAIR metrics

Value		
	F	FINDABLE
1	**F.1**	Unique and persistent identifiers (PDI)
0.5	**F.2**	Rich metadata
1	**F.3**	Metadata specify the PDI
0.5	**F.4**	Data registered in searchable resources
	0.75	

Value		
	A	ACCESSIBLE
0.5	**A.1**	Retrievable by the PDI with a standardized protocol
1	**A.1.1**	Protocol is open, free
1	**A.1.2**	Protocol allows authentication and authorization
1	**A.2**	Metadata accessible beyond the data availability
	0.875	

Value		
	I	INTEROPERABLE
1	**I.1**	Language is formal, accessible, shared and applicable
0.5	**I.2**	Vocabulary is FAIR
0.5	**I.3**	Metadata includes qualified references to other metadata
	0.67	

Value		
	R	REUSABLE
0.5	**R.1**	(Meta)data have plurality of accurate and relevant attributes
0.8	**R.1.1**	Released with a clear and accessible data usage licence
0.5	**R.1.2**	Provenance information
0.5	**R.1.3**	Domain-relevant community standards
	0.575	

Software sheet

Reference name of the program or workflow	Alya
Description	Alya is a CFD code of the PRACE Benchmark Suite for HPC applications that has been highly optimized and tested independently in most of the European supercomputer platforms.
Author	Guillaume Houzeaux, Mariano Vázquez, Oriol Lehmkuhl, Daniel Mira, Eduardo J. Pérez
Programming language	Fortran 90
Rules and best coding practices	https://gitlab.bsc.es
Access permissions and license	Open science software
Code size	> 1,000,000 lines
Repository type	GitLab
Repository structure	Branches
Provenance information	Containers, virtual environments
Backup and Archiving needs	None
Legal/ethical restrictions	None
Versioning control and rules/workflows managing	Internal version control system
Code transfer needs and security	None
Long term preservation needs	None
Documentation and inline comments rules	https://gitlab.bsc.es/alya/alya/wikis/home
Metadata management	None
Resources need	None

Reference name of the program or workflow	Paris-Simulator
Description	PARIS Simulator is a free code, or software, for the computational fluid dynamics (CFD) of multiphase flows, or computational multiphase fluid dynamics (CMFD) . Combines the Volume-Of-Fluid (VOF) and Front-Tracking methods to create simulations of interfacial fluid flow, such as droplets, bubbles or waves.
Author	Wojciech Aniszewski, Tomas Arrufat Jackson, Marco Crialesi Esposito, Sadegh Dabiri, Daniel Fuster, Yue "Stanley" Ling, Leon Malan, Sagar Pal, Ruben Scardovelli, Gretar Truggvason, Phil Yecko, Stephane Zaleski
Programming language	Fortran 90
Rules and best coding practices	None
Access permissions and license	Open free code
Code size	> 40000 lines
Repository type	Darcs Hub
Repository structure	Patches
Provenance information	Containers, virtual environments
Backup and Archiving needs	None
Legal/ethical restrictions	None
Versioning control and rules/workflows managing	Development version and stable version
Code transfer needs and security	None
Long term preservation needs	None
Documentation and inline comments rules	http://www.ida.upmc.fr/~zaleski/paris/index.html
Metadata management	None
Resources need	None

Formats definition

Code	Extension	Definition
Alya	csv	Comma-separated values
	vtk	Visualization Toolkit file format
	txt	Text file

PARIS	csv	Comma-separated values
	vtk	Visualization Toolkit file format

Data repositories information

Short Name	Extended name	Location	URL	Permissions
EUDAT			https://eudat.eu/catalogue	
BSC	Barcelona Supercomputing Center	Archive in MN4	https://www.bsc.es/marenostrum/marenostrum	
IVLR	Universität Stuttgart		https://www.ivlr.uni-stuttgart.de/	
KIT	Karlsruher Institut für Technologie		http://www.kit.edu/	
TUB	Technische Universität Berlin		https://www.tu-berlin.de	
TUD	Technische Universität Darmstadt		https://www.tu-darmstadt.de/	
TUE	Technische Universiteit Eindhoven		https://www.tue.nl/en/	
UPV	Universitat Politècnica de València	Archive at CMT	https://www.cmt.upv.es	