



# COALESCE

Co-creating the EU Competence Centre for  
Science Communication

## Deliverable 1.2

# Data Management Plan

Revision: v1.0



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

# DELIVERABLE DETAILS

**Due date:** September 30th, 2023

**Actual submission date:** September 30th, 2023

**Project start date:** April 1st, 2023

**Duration:** 48 months

**Work Package concerned:** WP1

**Concerned work package leader:** EUR

**Concerned work package co-leader:** SfC

**Dissemination level:**

**PU:** Public (must be available on the website)

**CO:** Confidential, only for members of the consortium (including the Commission Services)

**CI:** Classified, as referred to in Commission Decision 2001/844/EC

**Authors:** Jason Pridmore (EUR), Simone Driessen (EUR), Charlotte Bruns (EUR), Emmy Hermans (EUR), Joanna Mania (EUR), Lucía Recio (Science for Change), Joana Magalhães (Science for Change), Rosa Arias (Science for Change)

**Revision history:**

revision	date	Author/ Contributor	Description
v0.1	25/07/2023	Joanna Mania, Charlotte Bruns, Emmy Hermans, Simone Driessen (EUR)	First Draft
v0.2	30/08/2023	Jason Pridmore, Gai Umali (EUR)	Internal review
v1.0	27/09/2023	Lucía Recio, Joana Magalhães, Rosa Arias (Science for Change)	Final review

## STATEMENT OF ORIGINALITY

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

## DISCLAIMER

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

## Acronyms and definitions

Acronym	Meaning
DMP	Data Management Plan
EC	European Commission
EUR	Erasmus University Rotterdam
FAIR	Findable, Accessible, Interoperable, Re-usable
GDPR	General Data Protection Regulation
WP(s)	Work Package(s)
RDM	Research Data Management
Centre	European Competence Centre for Science Communication
DCMI	Dublin Core Metadata Initiative
CoP	Community of Practice
SeP	Stakeholder Engagement Panel

## SUMMARY

This document outlines the data management plan (DMP) prepared for the COALESCE project. It informs about the intentions of the consortium partners concerning data collection, its documentation, sharing and short- and long-term storage. It represents the most up-to-date knowledge and awareness of research data management, as well as data to be collected in the scope of the COALESCE project. The strategy aims to enable Findable, Accessible, Interoperable, Re-usable (FAIR) data use, which will be actively evaluated throughout the project. The DMP will evolve during the lifespan of the project, as a living document and will be regularly updated.

### About the COALESCE Project

COALESCE will consolidate, further develop, and mainstream generated knowledge and connections on science communication to build up the European Competence Centre for Science Communication. To achieve measurable and sustainable long-term impact of the Centre, project objectives will pivot on co-creation and co-design building on cooperative relationships with multiple stakeholders including scientific and journalist networks, as well as university alliances. COALESCE operates in relation to international, national and regional hubs and builds on an interdisciplinary approach. It will demonstrate the means for rapid mobilisation of scicomm in times of crisis while fighting misinformation and engendering trust in science. Supporting all of this will be an accessible library of critical resources, toolkits, handbooks and training opportunities to R&I actors across the ERA.

The COALESCE project ensures the effectiveness of and best practices for science communication through a concerted and focused effort made by an experienced, knowledgeable, and well-connected consortium. We build on and add to already existing forms of excellence in science communication, public engagement, and co-creation practices developed from the joint efforts of the coordinators of all previous SwafS-19 calls and contributions from other key partners. COALESCE continues and strengthens connections between academic researchers and practitioners representing the state of the art of scicomm in Europe and defines and trains on best practices within Europe and beyond.

The consortium represents several key features to make it well positioned in relation to the call as it: 1. Integrates universities and higher education providers with SMEs; 2. Connects participatory engagement methodologies and journalism practices with new channels for scicomm; 3. Participates with agendas of national scientific bodies; and 4. Is supported by existing scicomm reference and university alliance networks.

## Consortium partners

Partner organisation name	Acronym	Country
Erasmus University Rotterdam	EUR	NL
Science for Change, SL	SfC	ES
Venice International University	VIU	IT
The Provost, Fellows, Foundation Scholars & the Other Members of Board, of the College of the Holy & Undivided Trinity of Queen Elizabeth Near Dublin	TCD	IE
Lappeenranta-Lahden Teknillinen Yliopisto Lut	LUT	FIN
Formicablu SRL	FB	IT
Universitat de Valencia	UVEG	ES
Stichting VU	VU	NL
Stickydot SRL	SD	BE
Fundacion Espanola Para La Ciencia y La Tecnologia, F.S.P., FECYT	FECYT	ES
S.I.S.S.A. Medialab SRL	SML	IT
Tallinn University	TLU	EE
University of the West of England, Bristol	UWE	UK

# TABLE OF CONTENTS

## Contents

1. Purpose of the Data Management Plan .....	7
2. Summary: Data and Research Outputs .....	7
Table 1: <i>Data Summary</i> .....	7
Table 2: <i>Research Output Summary</i> .....	8
<b>3. FAIR Data</b> .....	9
<b>3.1 Making data findable, including provisions for metadata</b> .....	9
3.1.1 Persistent Identifier: DOI.....	9
3.1.2 Metadata schema: Dublin Core Metadata Initiative (DCMI) .....	9
<b>Table 3: List of DCMI metadata</b> .....	9
<b>3.2 Making data accessible</b> .....	10
3.2.1 Repository: Zenodo.....	10
3.2.2 Accessibility of datasets and research output .....	10
3.2.3 Accessibility of metadata.....	11
<b>3.3 Making data interoperable</b> .....	12
<b>3.4 Increase data reuse</b> .....	12
3.4.1 Licence.....	12
3.4.2 Project-level documentation.....	12
3.4.3 Naming and versioning .....	12
3.4.4 Data quality assurance .....	13
Table 4: <i>Data quality assurance</i> .....	13
<b>4. Allocation of resources</b> .....	13
<b>4.1 FAIR data and infrastructure</b> .....	13
<b>4.2 Expertise</b> .....	14
<b>5. Data security</b> .....	14
<b>5.1 Protection of personal data through anonymisation/pseudonymisation</b> .....	14
<b>5.2 Controlled data access</b> .....	14
<b>5.3 Secure storage infrastructure</b> .....	14
Table 5: <i>Specifications of data storage solutions</i> .....	14
<b>6. Ethics</b> .....	15
<b>6.1 GDPR compliance</b> .....	15
<b>7. Other issues</b> .....	16
<b>8. Closing remarks</b> .....	16

## 1. Purpose of the Data Management Plan

This DMP aims to provide a strategy regarding data within the scope of the COALESCE project. It specifies what datasets and research outputs will be generated or re-used, how these will be stored, accessed, archived and made available for reuse according to FAIR principles.

This document draws heavily on the template [HORIZON EUROPE Data Management Plan](#). The first section summarises the relevant data within the project, and it is followed by FAIR data provisions and allocation of resources. In addition, the document covers aspects of data security. The management of the ethical and privacy aspects are outlined in Deliverable 1.3 “Report Ethics Management”.

It should be noted that, for this first version, the document concerns mostly predicted or expected datasets, since empirical data collection has not fully started yet.

## 2. Summary: Data and Research Outputs

A general overview of the data collected in COALESCE is presented in Table 1: *Data Summary*. It covers data information such as types, formats, the purpose of collection, expected size, and origin for each of the Work Packages (WP). This information was gathered with the support of the consortium in a shared document. The overview is non-exhaustive and will be updated throughout the project.

Table 1: *Data Summary*

WP	Type	Formats	Size	Origin	Data Sharing
WP 1	<b>Project management and coordination</b>				
	Contact details consortium members	.xlsx	<1 GB	G	NA
	Contact details external stakeholders	.xlsx	<1 GB	G	NA
WP 2	<b>Pro-active Knowledge Consolidation and Co-development</b>				
	EU-funded project documentation, policy documents, secondary datasets	.pdf, .docx	<10 GB	R	NA
	Interviews, transcripts	.mp3, .mp4, .docx	<10 GB	G	R
	Workshops (recordings, notes, minutes, images, contact details)	.jpeg, .docx, .mp3	<10 GB	G	R
WP 3	<b>COALESCE, the European Competence Centre for Science Communication</b>				
	Open source code	tba	<10 GB	G	O
	Secondary data from previously funded EU-projects	.pdf, .docx	<10 GB	R	O
	Contact details external stakeholders	.xlsx	<1 GB	G	NA
WP 4	<b>COALESCE Community Engagement and Capacity Building</b>				
	Database Communities of Practice	.xlsx	<1 GB	G	NA

	Case Clinics / Workshops (recordings, notes, minutes, images, contact details)	.jpeg, .docx, .mp3	<10 GB	G	R
	Interviews, transcripts	.mp3, .mp4, .docx	<10 GB	G	R
WP 5	<b>COALESCE Impact pathways and sustainability</b>				
	Database of involved University Alliances	.xlsx	<1 GB	G	NA
	Database of key actors in Horizon EU Missions, Clusters, and wider ERA activities. Database of international networks of researchers.	.xlsx	<1 GB	G	NA
	Database National and regional hubs and their Communities of Practice	.xlsx	<1 GB	G	NA
WP6	<b>Mainstreaming the European Competence Centre for Science Communication</b>				
	Database of communication officers of partners' institutions such as N&R Hub	.xlsx	<1 GB	G	NA
	List of participants to the COALESCE events	.xlsx	<1 GB	G	NA

*Origin:* R – reuse, G – generate, *Data Access:* O – Open, R – Restricted, TBU – to be updated, NA – not applicable, A/A – as above

In addition to the above-mentioned data, project partners will generate research outputs in the different tasks (Table 2) that will be managed and made available for re-use, in line with the FAIR principles.

Table 2: *Research Output Summary*

Related Tasks	Research Output	Format	Data Access	Origin
2.2	Evaluation Process Design	.pdf	TBU	G
2.3, 4.5,	Workshops and training material	.pdf	O	G
2.4, 5.5, 5.6	Roadmap, Action and Business Plan	.pdf	TBU	G
2.6	Library of Resources and Handbooks	.pdf, .html	O	G
3.3	Matchmaking Tool	TBU	TBU	G
3.5	Self-evaluation Tool	TBU	TBU	G
4.2, 4.3, 5.1, 5.4	Policy Briefs	.pdf	O	G



### 3. FAIR Data

#### 3.1 Making data findable, including provisions for metadata

##### 3.1.1 Persistent Identifier: DOI

Every new record (such as datasets, research materials and other research outputs) previously settled to be published and/or disseminated as an Open format will be assigned a DOI in Zenodo. Metadata will include the DOI of its concerned dataset, material or output.

##### 3.1.2 Metadata schema: Dublin Core Metadata Initiative (DCMI)

Zenodo follows the DCMI metadata standard. On the day of DMP submission, next to indexing published records in its own search engine, Zenodo sends and indexes metadata in the DataCite, links data to OpenAIRE publications and to grants in the EC Participant Portal.

At the later stage of the project, a set of keywords will be selected for consistency and to optimise the discoverability and re-usage of the (meta)data.

**Table 3:** List of DCMI metadata

Name of metadata standard	Description
1. Title (M)	Title or name of the dataset, research output, presentation, etc.
2. Creator (M)	Main researchers responsible for producing data.
3. Subject (MA)	Keywords that will help to tag, find and index the data, for example “science communication”, “coalesce”.
4. Description (MA)	Clear and concise description of the dataset, research output, etc. It can cover its objective, contents, type of data, method of collection, etc.
5. Publisher (MA)	An institution who submitted the dataset or research output to the repository. Others can be listed as contributors.
6. Contributor (R)	An institution or a researcher who made contributions to collection, preparation, processing of the dataset.
7. Date (M)	Date on which the resource was made available (format: YYYY-MM-YY).
8. Type (M)/(R)	Enter what type of resource it is, for example: dataset, workflow, software, poster, image, publication, etc.
9. Format (R)	Specify a format of uploaded resource, for example: .pdf, .csv, .txt

10. Identifier (M)	An unambiguous reference to the resource within a given context, such as DOI
11. Source (R)	Indicate the original source of the data.
12. Language (R)	English is the default language.
13. Relation (R)	Specify related publication, datasets, presentations by providing its persistent identifier (for example, a DOI), but also information about grant agreement.
14. Coverage (O)	Spatial topic and spatial applicability may be a named place or a location specified by its geographic coordinates.
15. Rights (R)	Specify 1) access rights that apply to the dataset or other research output, for example: open access, restricted access, 2) licence of re-use

M – Mandatory, MA – Mandatory when applicable, R – Recommended, O – Optional

## 3.2 Making data accessible

### 3.2.1 Repository: Zenodo

Data and research outputs are deposited in Zenodo under the COALESCE Community. All new uploads should follow the data-sharing manual. They will be revised and accepted by the project coordinator.

The consortium does not foresee the necessity for any special arrangement with Zenodo to deposit their data.

The record identifiers and the collection name of metadata can be harvested using OAI- PMH protocol.

### 3.2.2 Accessibility of datasets and research output

COALESCE consortium partners will make datasets and other research output accessible to third parties for reuse. However, the openness of data is dependent on the type of data.

All data gathered within the COALESCE project and planned to be published in a data repository will be anonymised.

Non-personal data which has been gathered and used in research reports, project deliverables and articles will be made open access (OA) for reuse by the consortium partners. However, original data containing personal information will receive pseudonymisation and/or anonymisation. This data will be published under restricted access, and metadata will include



information about how and under what conditions external parties can be granted access to this data.

Further, in the COALESCE project contact information of SciComm stakeholders will be collected to create a database for the European Competence Centre for Science Communication. Those interested in becoming part of the database are asked to subscribe and provide their personal information voluntarily in accordance with the GDPR regulations. The exact process of how this subscription will be developed will be updated in a later version of the DMP. As an initial cohort for our database, and through previous coordinators, all previous SwafS-19 sister projects (TRESKA, NEWSERA, RETHINK, PARCOS, CONCISE, ENJOI, QUEST and GLOBALSCAPE) participants (up to 5000 people) will be informed and invited to enrol in COALESCE, in order to foster continuity of the communities of practice (CoPs) early created. This database will not be made available for reuse because it includes personal data such as names, email addresses, phone numbers and job titles, such data is therefore not suitable for anonymisation to protect participants' privacy. Access to the database is restricted, for internal use only, and for the duration of the project. One database will be created for the External Stakeholder Panel (ESP), to which the research teams of EUR and SfC as well as specific assigned partners will have access. Another database for the CoPs will be accessible to the partners involved in WP1, WP4 and WP6.

If any privacy risk arises during the project, consortium partners can consult EUR Privacy Officer to help reduce or remove them.

The need of establishing a data access committee will be discussed at a later stage of the project. Nevertheless, data with restricted access will be evaluated on a case-by-case basis if permission is requested.

Produced data that does not consist of personal data, such as research reports and policy briefs, is usable for third parties and accessible through the repository of the European Competence Centre for Science Communication after the COALESCE project has been completed (see table 3).

### 3.2.3 Accessibility of metadata

Metadata will be publicly accessible and licensed under the public domain (Creative Commons - CC0), meaning that no authorization will be necessary to retrieve it. If there will be datasets published under restricted access, metadata will contain information on how to access the dataset.

Data and metadata will be retained for the lifetime of the host laboratory CERN, which currently has an experimental program defined for the next 20 years at least. Metadata is stored in high-availability database servers at CERN, which is separate from the data itself.

Information about software (and their versions) needed to open, process or analyse data will be included in the dataset documentation through readme files.

### 3.3 Making data interoperable

Data formats listed in Table 1 will be revised and converted to non-proprietary formats before publication. The relation between resources can be made through the “Relation” element in the metadata. While uploading data to the repository, consistent keywords will be used.

Descriptions of the datasets and metadata are in English to increase their discoverability and accessibility.

In terms of interoperable metadata, Zenodo uses JSON Schema as an internal representation of metadata and offers export to other popular formats such as Dublin Core or MARCXML.

### 3.4 Increase data reuse

#### 3.4.1 Licence

In principle, the research data, materials, and research outputs generated within the COALESCE project can be of interest to researchers in the field of SSH research and policymakers but not exclusively. The COALESCE data can be relevant for disciplines and research groups working around the topic of science communication and competency centres.

Research data will be labelled with Creative Commons CC BY 4.0 licence to maximise its reusability of all data deposited in Zenodo. This licence places the copyright in the public domain, so that others may freely build upon, enhance, and reuse the works for any purposes without restriction under copyright or database law.

#### 3.4.2 Project-level documentation

To validate data analysis and facilitate data reuse, shared and accessible data is provided with README files with information on the investigators, methodology and access to files to explain the content, folder and/or file. A README file template, named as README\_template.txt, is added to the SURF Research Drive Folder and Google Drive for Education under WP 1 - Project management and coordination.

#### 3.4.3 Naming and versioning

For the naming and versioning of documents (if not otherwise specified) the following guidelines apply:

- [date]\_[name-of-the-file]\_[version].[format]

The file name consists of a date, name of the file, version number and format. Date follows *yyyymmdd* format. Names of the files are short and descriptive, and written out in small letters. Version consists of *v* and a number, no space in-between. Different types of chunks are separated by underscore (*\_*), and part of the same chunk by hyphen (*-*).

- *yyyymmdd\_name-of-the-file\_vnr.F*

Example: For a document with the title: use cases definition (V3). Date of creation: July 30, 2023

- File name: *20230730\_use-cases-definition\_v3.docx*

### 3.4.4 Data quality assurance

The consortium partners have the ambition to produce reliable data that yield reproducible research results. To promote quality assurance throughout the project, following steps can be taken:

Table 4: Data quality assurance

Research phase	Measures
Data collection and data analysis (secondary data such as policy briefs and other literature, interviews, co-creation sessions, workshops)	<ul style="list-style-type: none"> <li>- Check authenticity of secondary material and systematically note its sources</li> <li>- Using standardised methods and protocols for capturing observations (precise recording of audio and visual data)</li> <li>- Provide good quality of collected data (especially sound, video, images)</li> <li>- Data is curated systematically so that data not allowed by consent or not relevant to research question is excluded from analysis</li> </ul>
Digitalization and data entry	<ul style="list-style-type: none"> <li>- Use of templates for documentations</li> <li>- Logical folder structure to organise data in files</li> </ul>
Data checking	<ul style="list-style-type: none"> <li>- Double-checking coding of observations or responses</li> <li>- Checking data completeness</li> <li>- Correcting errors made during transcription</li> <li>- Cross-check notes</li> </ul>
Data authenticity	<ul style="list-style-type: none"> <li>- Keep and maintain original data files</li> <li>- Keep information of sources of data</li> <li>- Keep history of data analysis files</li> </ul>

## 4. Allocation of resources

### 4.1 FAIR data and infrastructure

All partners of COALESCE are responsible for the implementation of FAIR practices when collecting and storing research data. The data steward can be approached if assistance is needed. Specifically, it is important that enough time is considered in the research process to ensure the quality of the data, to prepare transcripts, to pseudonymise them when needed, to document the data, to format files into sustainable formats, and to digitise analog material. No extra costs are foreseen to occur to cover these activities. The costs for data storage can be found in table 5: Technical specifications of data storage solution.

## 4.2 Expertise

The implementation of the best Research Data Management (RDM) practices and FAIR principles is supported by professional expertise of data stewards, privacy officers, and legal advisors available through the project coordinator.

## 5. Data security

### 5.1 Protection of personal data through anonymisation/pseudonymisation

When conducting research involving personal data, principles of anonymisation and pseudonymisation will be followed in order to protect the privacy of research participants. Researchers are also aware of the responsibility to pseudonymise the data before depositing it to Zenodo.

### 5.2 Controlled data access

As indicated in the table below, access to the different storage tools is controlled by members of the Coordination and Co-Coordination teams. As part of the project management strategy, an overview of the access of the single consortium members is maintained and regularly updated in a file accessible only to the Coordinator and Co-Coordinator.

### 5.3 Secure storage infrastructure

The specifics of the chosen data storage solutions for the project can be found in the table below.

**Table 5: Specifications of data storage solutions**

Storage solution	<u>Google Drive for Education</u>	<u>EUR SURF Research Drive</u>	<u>Zenodo</u>	<u>EUR Yoda Vault</u>
<b>Stored data</b>	Documentation of the project, administrative data	Research data, administrative data with personal and sensible information	Research data, metadata, project outputs	Project documentation, research data, meta data, research outputs
<b>Access Management</b>	Coordinator and Co-Coordinator	Coordinator	Coordinator (accepts submissions)	Coordinator
<b>Access Rights</b>	All consortium members, single folders accessible by selection of consortium	Selection of Consortium Members	Public	To be updated

<b>Preservation period</b>	Short-term: project duration. Will be transferred to EUR Yoda Vault when project ends	Short-term: project duration. Will be transferred to EUR Yoda Vault when project ends	Long-term: min. 10 years after completion of the project, max. lifetime of repository	Long-term: 10 years after completion of the project*
<b>Backup and security measures</b>	<a href="#">Security specifications.</a> Version history and <a href="#">trashbin (30 days)</a>	<a href="#">Security specifications.</a> <a href="#">Trashbin (30 days)</a> and versioning functionality	<a href="#">Security specifications.</a> (Meta)data are <a href="#">backed up nightly and replicated</a> into multiple copies in the online system	<a href="#">Security specifications.</a> <a href="#">Incremental backup, full monthly backup</a> retained for 30 days
<b>Costs</b>	Free of charge, provided by EUR	Free up to 500GB, provided by EUR	Free of charge	Free up to 500GB, provided by EUR

\*The archive duration will be 10 years after either the assignment of a persistent identifier or the publication of a related work following the project completion, whichever is later. The decision about deletion or destruction will be carried out only after considering all legal and ethical perspectives, as well as confidentiality and security aspects. The considerations also include the interests of funders, stakeholders, consortium partners and employees. The rationale for deletion or destruction of research data should be explained in the updated metadata file.

## 6. Ethics

All ethical requirements and procedures of the COALESCE project can be found in deliverable 1.3: Report on Ethics Management.

### 6.1 GDPR compliance

The Consortium is dedicated to upholding the data privacy rights of its research subjects, following the EU Charter of Fundamental Rights, GDPR, e-Privacy Directive and other relevant data privacy laws. Privacy by design and by default (Article 25 GDPR) is integral in the research methodology.

Guided by the research question, only essential and pertinent personal data will be processed. This practice of data minimisation ensures that research data is adequate, relevant and necessary within the parameters of the research project. The same data will be used solely to serve the research project's purpose, in harmony with the principle of purpose minimization.

Furthermore, appropriate technical and organisational measures will be adopted throughout the research project. For example, research data will be stored in a secure facility which was assessed to be safe, ensuring the protection of sensitive information. Security incident responses are in place to detect, contain, handle, and report data breaches. Processes for notifying data breaches to the Dutch Data Protection Authority (Autoriteitpersoonsgegevens) and the research subjects are also set in place.

Additionally, research participants' personal data will undergo either pseudonymization or anonymization, depending on project applicability and relevance. Anonymization ensures irreversible subject identification prevention, while pseudonymization reduces data linkage with subjects. On this note, personal data remains raw for a limited period during data collection phase until processed by means of pseudonymization/anonymization, ensuring subject de-identification. During this temporary phase, security measures like encryption in transit/situ, access logs, and multi-factor authentication, restricted access based on assessment of necessity, among others are implemented. Regular maintenance and review of these measures ensure ongoing data safety.

If deemed necessary, a Data Protection Impact Assessment (DPIA) will be carried out in accordance with Article 35(1) GDPR. The DPIA evaluates potential high risks to the rights and freedoms of research subjects arising from the processing. It documents identified risks to subjects' rights and freedoms, along with measures to mitigate identified harms and prevent their occurrence.

The parties have signed Joint Controller Agreement (JCA) pursuant to Article 26 GDPR

Further specifics concerning the GDPR compliance of the COALESCE project are defined in deliverable 1.3: Report on Ethics Management.

## **7. Other issues**

In the initial data management strategy and information document no other issues are to be reported upon its submission in month 6. As this document is a living document, this section might be updated in later versions following developments in the project.

## **8. Closing remarks**

This document has outlined the initial data management strategy and information for the COALESCE project. At month 6, the date of submission for the first version of this deliverable, most of the datasets and outputs related to the project are still to be generated. Therefore, as a living document, this Data Management Plan will be updated periodically according to developments in the project as well as reporting periods.