



Data Management Plan

Josef Seethaler and Olha Drozd

DELIVERABLE 1.2, V2.0 (Post-review revision)

MeDeMAP – Mapping Media for Future Democracies

Grant Agreement number: 101094984



**Funded by
the European Union**

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

Document information

Project information	
Grant Agreement no.	101094984
Funding scheme	Horizon Europe Research and Innovation Actions
Project title	Mapping Media for Future Democracies
Project acronym	MeDeMAP
Project starting date	01/03/2023
Document information	
Work package no.	1
Work package title	Coordination
Work package lead beneficiary	OEAW
Task(s)	1.1 – Legal, financial and administrative coordination
Deliverable no.	1.2
Deliverable title	Data management plan
Deliverable type	OTHER
Contractual date of deliverable	31/08/2023
Actual date of deliverable	31/08/2024 (V2.0)
Editor(s)	
Author(s)	Josef Seethaler (OEAW), Olha Drozd (OEAW)
Reviewer(s)	-
Version	2.0
Status	Final
Total number of pages (including cover)	16
Dissemination level	Public

Document history

Version	Date	Submitted by	Changes
V1.0	31/08/2023	OEA	Initial version
V2.0	31/08/2024	OEA	Addition of metadata schema, repository and executive summary; Post-review revision: extension of <i>Chapter 4.4 - Increase Data Re-Use</i> , addressing the potential misuse of shared data and research results

Table of Contents

1	Executive Summary	5
2	Introduction and Overview	5
3	Data Summary	6
4	FAIR Data	8
4.1	Making Data Findable, Including Provisions for Metadata	8
4.2	Making Data Accessible	9
4.3	Making Data Interoperable	11
4.4	Increase Data Re-Use	11
5	Allocation of Resources	13
6	Data Security	13
7	Ethics	14
8	Other Issues	15

1 Executive Summary

As coordinator of the Horizon Europe Project 101094984 “Mapping Media for Future Democracies (MeDeMAP)”, OEAW has created this Data Management Plan (DMP), which will be **regularly updated during the project lifetime**. Research data management is concerned with the process within the research lifecycle that includes the organisation, storage, preservation, security, quality assurance, allocation of persistent identifiers (PIDs) and rules and procedures for sharing of data. The MeDeMAP DMP addresses the requirements for research data management of Horizon Europe as described in article 17 and analysed in the Annotated GA, article 17, and promotes the application of open science and open data practices and compliance with the FAIR principles. It was developed using the «Horizon Europe» template (version 1, 5 May 2021) provided by the European Commission.

2 Introduction and Overview

MeDeMAP aims to clarify the extent to which certain media under which conditions perform which democratic functions for which audiences, thus setting out future-proof pathways to strengthen democracy through improving accountability, transparency and effectiveness of media production while also expanding active and inclusive citizenship. With the overall goal of MeDeMAP being to deliver comprehensive knowledge on the relationship between media and democracy, sharing of insights and data stands as important factor in achieving the project's targeted impact. Following the guideline that research data should be ‘as open as possible, as closed as necessary’ and in accordance with the Grant Agreement (GA), Art. 17, MeDeMAP will promote the application of open science and open data practices and the respect of the FAIR principles – Findability, Accessibility, Interoperability, and Re-use – across all project activities.

The projects funded under Horizon Europe calls should apply the FAIR principles through development and efficient implementation of Data Management Plans (DMPs). The MeDeMAP DMP is prepared within *Task 1.1 Legal, financial and administrative coordination* and subsequently submitted as *Deliverable 1.2*. The DMP is a living document that will guide all MeDeMAP consortium members in adhering to the conventions for management of data generated during the project in a way that data can be found, accessed, made interoperable and reusable. It will be regularly updated through the project's life cycle as decisions have been made and new data occur, in any case, two months after the end of the first reporting period. A final version of the DMP that describes how the data is managed and shared will be delivered at the end of the project.

The DMP provides the key elements of data management in the MeDeMAP project. It describes the data management life cycle encompassing data collection, processing, and generation within the MeDeMAP project. Additionally, the DMP covers issues related to new and re-used data generated through implementation of activities in the context of the various work packages of the project, including desk studies, document analysis, surveys, interviews, focus groups, and mapping exercises among others. Such activities are essential to fulfil the objectives of the project and its expected impact by communicating, disseminating and exploiting the outcomes in different formats, which frequently include various data sources.

In order to make research data findable, accessible, interoperable and reusable (FAIR), this DMP includes information on:

- the handling of research data during and after the end of the project,
- what data will be collected, processed and/or generated,
- which standards will be applied,
- whether data will be shared/made open access and
- how data will be curated and preserved.

Accordingly, the remainder of the document is structured as follows:

Chapter 2 describes characteristics of the data obtained throughout the project, Chapter 3 deals with the implementation of the FAIR principles, Chapter 4 addresses the allocation of resources, Chapter 5 is concerned with issues of data security, and Chapter 6 discusses ethical aspects.

3 Data Summary

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

Data are at the core of any research project. They contribute heavily to the project's development and provide input to deliverables, scientific publications, and dissemination activities. In terms of MeDeMAP's objectives, data serve a fivefold purpose:

- (1) Establishing an up-to-date overview and understanding of the state of democracy in the European Union, its challenges, strengths and threats, and the roles and functions of media and journalism that can be derived from it.
- (2) Mapping legal and (self-)regulatory frameworks within which traditional and new media organizations and journalists operate.
- (3) Mapping democratic qualities of media supply and media-enabled participation across all media channels, mandates & sources of financing.
- (4) Mapping people's media use, information needs and democratic attitudes in diverse societies and high-choice media environments.
- (5) Mapping future pathways for building a more resilient democratic media system and enhancing political participation and civic engagement.

In fulfilling its objectives, the project follows a two-stage approach. At the first stage, research is mainly concerned with secondary analyses of existing literature and data. The various approaches to find, collect and process unstructured and structured data include desk research, literature reviews, content-analytical and source-critical methods, data and text mining. Statistical analysis will be employed to extract information and insights from both survey-based microdata and expert- and statistics-based country-level data.

At the second stage, the project will collect data during the following operations:

- (1) interviews with representatives of national media authorities, media councils, media organizations and relevant European associations,
- (2) national surveys (CAWI) of members of media houses (management and newsrooms),
- (3) focus groups, ethnographic research and 'citizen parliaments' (guided by the concept of 'audience councils' representing the interest of readers, listeners, viewers and online media users).

Taking both stages together, research in this project will eventually result in the creation of a multilayered map that describes political information environments in the EU in such a way that guidelines can be derived for all relevant stakeholders to support developments that promote democracy and counteract phenomena that may jeopardize democracy.

*Will you re-use any existing data and what will you re-use it for?
What is its origin / provenance?*

A broad range of existing data will be re-used in work packages 2 to 6. They include data on methodological issues, the state of democracy; legal provisions and (self-)regulatory standards in the field of media and public communication; the composition of media supply and 'opportunity structures' to access political information; and audience structures and media use patterns. These are data of international, European and national origin, including other EU-funded projects and initiatives (e.g., Eurobarometer, Media Pluralism Monitor, V-Dem, Media Ownership Monitor, Council of Europe's Platform to Promote the Protection of Journalism, Mavise, Digital News Report).

What types and formats of data will the project generate or re-use?

Preferred data formats are:

- Textual data: .odt, .txt, .docx
- Tabular data with extensive metadata (variable labels, code labels, and defined missing values): .sav, .por
- Tabular data with minimal metadata (column headings, variable names): .csv, .xlsx
- Image data: .tif (uncompressed), .jpeg (if originally created in this format)
- Audio data: .flac (lossless), .mp3 (if originally created in this format)
- Documentation and reports: .pdf
- Presentations: .pptx

Recorded data (interviews, focus groups) will be transcribed and stored in a text format (.odt, .txt, .docx).

What is the expected size of the data that you intend to generate or re-use?

The projected data size within the project could potentially reach up to 100 GB.

To whom might your data be useful ('data utility'), outside your project?

Beyond the project, the data is expected to provide utility to the scientific community through the access to the scientific results described in publications and the intended establishment of elaborated standards in research on media systems. Additionally, stakeholders, including policymakers, regulators, self-regulatory bodies, media organizations, journalists, and civil society organizations are likely to benefit from the insights offered by the data.

4 FAIR Data

4.1 Making Data Findable, Including Provisions for Metadata

Will data be identified by a persistent identifier?

Persistent identifier (PID) is a long-lasting reference to a resource that provides the information required to reliably identify, verify and locate the resource. In a digital environment, PIDs have the form of URLs which will be used for the digital content of the MeDeMAP project.

Apart from digital resources, PIDs can also relate to researchers (e.g., ORCID) and institutions (e.g., ROR). Researchers are encouraged to register with ORCID and include their ORCID ID in any project output.

For MeDeMAP datasets and publications, the use of a Digital Object Identifiers (DOIs) is recommended. DOIs are typically automatically assigned by trustworthy repositories once data are deposited there.

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Metadata should answer basic questions regarding the data allowing others to understand, discover and share the data. This includes information about how data was produced, who was involved in the making, what the data is about and how it is preserved. Since there is no disciplinary standard for media and communication studies (see [Index of subjects – Metadata Standards Catalog \(bath.ac.uk\)](#)), the MeDeMAP consortium decided at its general assembly on September 8, 2023 to use [Dublin Core](#), which offers intuitive metadata generators, as the metadata schema of the MeDeMAP project. Dublin Core provides a set of vocabulary terms developed for describing digital objects. It comes in a simple (15 elements) and a larger set. The Simple Dublin Core Metadata Element Set (DCMES) includes 15 metadata elements (terms). Each element describes a property of a resource:

1. Title	6. Contributor	11. Source
2. Creator	7. Date	12. Language
3. Subject	8. Type	13. Relation
4. Description	9. Format	14. Coverage
5. Publisher	10. Persistent identifier	15. Licence (CC 0 or CC BY by default)

All elements are optional and repeatable.

Depending on the file format used, metadata will be integrated directly into a file (e.g., .odt, .pdf, .docx) or provided in an accompanying file. Tables are particularly suitable for this purpose, as they facilitate the entry of metadata with the help of specific input masks, a targeted search, and a structured backup of metadata. In addition, they allow (partial) automation of the metadata. Trusted repositories usually require filling in a metadata form about the data or files that should be published.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Keywords are particularly important for dissemination activities outside the project:

- Relevant keywords will be defined for every deliverable, dataset and publication and checked for consistency. Preferably, metadata records will make use of standard vocabularies (e.g., EuroSciVoc, a multilingual taxonomy that represents all the main fields of science and is managed by the Publications Office of the EU).
- For scientific publications, the keyword list of the respective publisher must be used.

Will metadata be offered in such a way that it can be harvested and indexed?

The metadata of each record will be indexed and made searchable directly in the repositories' search engines upon publication.

4.2 Making Data Accessible

Repository:

Will the data be deposited in a trusted repository? Have you explored appropriate arrangements with the identified repository where your data will be deposited? Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

In Horizon Europe, data should be deposited as soon as possible after its generation and, at the latest, by the end of the project. Data underpinning a scientific publication should be deposited at the time of publication, and in line with standard community practices. However, generated data should only be deposited if data protection rules have been complied with (see Chapters 5 and 6).

Horizon Europe also requires that, at the time of depositing research data in a trusted repository, access to information must be provided (via the repository) about any research output or any other tools or instruments needed to re-use or validate the research data.

The MeDeMAP project will adhere to the aforementioned requirements with research data being made open access by default and licensed under the latest version of CC BY or CC 0 (public domain). This will inform the selection of the repository or repositories for archiving the datasets, particularly the final datasets on which the map of European political information environments will be based. Accordingly, the MeDeMAP consortium decided at its general assembly on September 8, 2023 to use [Zenodo](#) for data archiving. Zenodo is a multi-disciplinary open repository maintained by CERN and supported by the EU. It is compliant with the data management requirements of Horizon Europe and accepts data in any file format, enables the assignment of metadata to datasets and automatically assigns a digital object identifier (DOI) to all Zenodo files.

Partners may also use their own repositories for the national data they collect, provided that these repositories meet the above-mentioned requirements. In Austria, for example, the Austrian Social Science Data Archive has been certified with the Core Trust Seal (CTS). Additionally, OEAW offers repositories such as ARCHE and EPUB.OEAW for long-term archiving and publication, as well as for reference and subsequent use according to the access rights defined by the researchers.

The project website will contain a list of the project datasets and publications, including those linked to repositories, along with the deliverables declared as public in the GA.

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Final data will be made openly available unless there is an incompatibility with personal data protection regulations (see Chapters 5 and 6). Considering a good balance between open access and protection of intellectual property, data underpinning scientific publications will be deposited only after the works have been published. If an embargo is applied to give time to publish, it will be as long as required by the publisher. Currently, no beneficiary has the intention to close their data. If the situation changes, the DMP will be updated accordingly.

Will the data be accessible through a free and standardized access protocol?

Final datasets will be accessible through a free and standardized access protocol.

How will the identity of the person accessing the data be ascertained?

The identification of the person accessing the data depends on the procedures specific to each repository.

Is there a need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?

No. Personal/sensitive data will not be published.

Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the GA? Will metadata contain information to enable the user to access the data?

All metadata will be made openly available and licenced under CC0. They will contain information to enable the user to access the data (see Chapter 3.1, question: *Will rich metadata be provided to allow discovery?*).

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

Data and metadata will be retained for the lifetime of the repository.

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g., in open source code)?

The tools required to read the data are common and free to use. Given the widespread availability of such software, documentation will not be necessary.

4.3 Making Data Interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones? In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Common data and metadata standards and formats are a key aspect for technological and semantic data operability. Standardization makes the data discoverable. Data and metadata in the repositories mentioned above (see Chapter 3.2) use a formal, accessible, broadly applicable language for knowledge representation, as well as vocabularies that follow FAIR principles. The project is currently in the process of determining the vocabularies, standards, formats, and methodologies that will ensure data interoperability. These choices will be made with consideration of community-endorsed best practices for interoperability.

The MeDeMAP project will not use uncommon or generate project-specific ontologies or vocabularies.

Will your data include qualified references to other data (e.g., other data from your project, or datasets from previous research)?

Research data will include qualified references to other data including those from the project.

4.4 Increase Data Re-Use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g., readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

Documentation with information required to validate the data analysis will be provided in commonly used formats such as Word and PDF and available together with the research data.

Describe all relevant data quality assurance processes.

Existing quantitative data sources will be subject to the quality assurance processes of the organization that originally gathered the data. However, the data will be reviewed prior to use regarding issues such as missing values which may have an impact on the study results. The respective WP leaders are responsible for the review (as stipulated in Deliverable D1.1 – Coordination Plan).

Quality of newly generated data will be assured through routine monitoring by the respective work package leader, and a mandatory screening and cleaning of the dataset after completion to check for outliers and errors. National datasets, which will consist of existing and generated data and will be fed into the map of European information environments, are checked again for integrity in a second step to ensure that the data meet the methodological requirements for inclusion in the final consolidated file.

A general quality assurance process has been set out in the Coordination Plan (Deliverable 1.1, Chapters 3.2 and 3.3).

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the GA?

Data produced in the project will be made freely available in the public domain and licensed using standard re-use licences, in line with the obligations set out in the GA.

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Final research data will be openly available for use by third parties. This, however, raises the question of how possible misuse of shared data and research results, e.g. by political parties and other actors, can be made more difficult, or at best prevented – because, as Pasquetto et al. (2024) have put it, possible “misuse should not be used as an excuse not to release critical data”.¹ Data-generating projects therefore face the challenge of taking the potential harm caused by deliberate or unintentional data misuse seriously. The MeDeMAP consortium will consider activities (such as discussions with the leaders of other Horizon Europe projects or consultations with experts such as fact-checkers and professional science communicators) to raise awareness of this issue. The measures recently proposed by Pasquetto et al. (2024) could be one starting point. They include:

- carefully documenting data provenance, formats, granularity, quality, etc., following the FAIR principles (as set out in this document) and disciplinary standards;
- adopting different forms of data descriptions in addition to metadata, for example, data papers or readme files providing contextual information about a dataset;
- applying self-reflexive research practices, especially when there is a potential for the data to be biased or harmful;
- displaying links on the dataset landing pages to papers based on the dataset, as examples of how the dataset has been “properly” used;
- encouraging data replication by reusers, while being explicit about how eventual errors found in the data or in the analyses should be communicated;
- applying prebunking and debunking interventions.

Nevertheless, one must be aware that “intentional misuse is hard to point at and even harder to prevent [...]. These interventions can, at least, set the reference standards for the intended meaning and scope of the data in case a controversy arises.”

¹ Pasquetto, I. V., Cullen, Z., Thomer, A., & Wofford, M. (2024). What is research data “misuse”? And how can it be prevented or mitigated? *Journal of the Association for Information Science and Technology*, 1–17. <https://doi.org/10.1002/asi.24944>

Will the provenance of the data be thoroughly documented using the appropriate standards?

The origin of the data will be thoroughly documented in all relevant publications and documents.

5 Allocation of Resources

What will the costs be for making data or other research outputs FAIR in your project (e.g., direct and indirect costs related to storage, archiving, re-use, security, etc.)? How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the GA conditions).

The costs related to storage, archiving, re-use, security, etc., are not anticipated in the MeDeMAP project, as the services and offerings of repositories are typically free of charge.

The costs for scientific open access publications with the research results of the project and for the project website will be covered by the available project budgets of the partners.

Who will be responsible for data management in your project?

The coordinator is responsible for the overall data management process. The lead partners are responsible for storage (see Chapter 3.2) and security (see Chapter 5) of the national data they collect.

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

TBD

6 Data Security

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

As long as datasets are not deposited in a trusted repository, each partner is responsible for storage and recoverability of their generated data (backup plans have to be established in accordance with the respective institution's practice). During the research process, access to data is only provided to project members. Standard security measures, such as authentication, are applied.

The final research datasets uploaded to open repositories would also be kept on the website of the project website (particularly for deliverables) and on the servers of the responsible partners' institutions. In the case of the datasets for the map of European information environments, they will be stored on a server of OEAW, which is responsible for setting up the map.

Processing of personal data, primarily collected through interviews, surveys, focus groups and events, is subject to EU and national laws on data protection. The collected data will be limited, in line with the 'data minimization principle', to only those necessary for research and communication with the participants. In general, unnecessary collection and use of personal data will be avoided. In all cases,

consent will be sought from the persons involved and information about the data processing will be provided.

If online surveys are conducted by providers such as SoSci Survey, they must offer a white-label solution that adheres to the strict requirements of data protection laws (the data are deleted from the provider's server after 90 days), and data processing agreements will be concluded.

All names of the persons participating in surveys, interviews, focus groups and events, such as citizens' parliaments, will be anonymized, and all related data will be published in aggregated form only. This will be assured through monitoring by the respective WP leader, and screening of the dataset after completion of the transcription. Recordings, protocols and datasets will be stored on national servers owned by the institutions of the partners, with daily encrypted backups ensured. Access to this kind of data will be restricted to selected staff members of the partners responsible for the relevant research or event organization. Furthermore, such data will be stored on password-protected computers. The files containing the keys for re-identification of the persons and the recordings of the interviews will be destroyed after transcription to safeguard the privacy of individuals concerned (see Chapter 6).

Sensitive documents containing even anonymised data (e.g., deliverables declared as SEN) will be kept at the internal page of the project website, where they will be clearly labelled as sensitive with a note explaining the meaning of this level of dissemination.

Each partner remains responsible for setting up internal processes to comply with the General Data Protection Regulation (GDPR). Personal data protection will be assessed with the support of the Data Protection Officer of the respective partner.

Will the data be safely stored in trusted repositories for long term preservation and curation?

Yes. All data will be preserved for the long term, except those that allow identifying respondents and/or their institutions. These data will be removed and, if possible, replaced by generic proxies.

7 Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

All partners comply with the highest ethical standards and EU, international and national laws on ethical principles. Detailed rules are outlined in Annex 5 of the GA.

Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?

In the context of qualitative interviews, focus groups and other participatory methods, personal data may become an issue. All collected data will be subject to appropriate safeguards and to freely given and fully informed consent of the persons concerned. Therefore, in conformity with data protection, the names of the participants will be replaced by pseudonyms such as person A, person B, person C etc., and other obviously identifying information will be removed. The participants' personal data

and/or at least the re-identification keys of the persons and the recordings of the interviews will be destroyed after transcription to safeguard the privacy of the individuals involved (see Chapter 5). This is relative anonymisation, as individuals present during the interview or focus group may have knowledge of the statements made. After a longer period of time, however, even those present can usually no longer recall the statements. Consequently, researchers will no longer be able to ascertain the identity of the person from the transcript should he/she wish to assert his/her rights months or years later. In the opinion of the data protection authorities and the courts, such relative anonymisation meets data protection law requirements. All participants in research activities and events will be informed accordingly.

Signed lists of participants in workshops and meetings will contain names and affiliation details. These lists serve as event verification records and, as such, they will be provided to the EC if necessary. They will not be kept for any other purpose, and there will be no further processing of collected personal data.

8 Other Issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

No procedures other than those described in this document will be used for data management.

MeDeMAP – Mapping Media for Future Democracies

Deliverable 2.1

Version 2.0

(August 2024)

