



COESO

connecting research and society

COLLABORATIVE ENGAGEMENT ON SOCIETAL ISSUES

WP1 - Project coordination and management

Data Management Plan

first version

16.07.2021



This deliverable is licenced under
a Creative Commons Attribution 4.0 International Licence.



COESO has received funding from the EU Horizon 2020 Research and Innovation Programme (2014-2020)
SwafS-27-2020 - Hands-on citizen science and frugal innovation, under Grant Agreement No.101006325

The content of this publication is the sole responsibility of the COESO consortium and can in no way be taken to reflect the views of the European Commission. The European Commission is not responsible for any use that may be made of the information it contains.

Deliverable 1.2

Data Management Plan

Grant Agreement number	: 101006325
Project acronym	: COESO
Project title	: Collaborative Engagement on Societal Issues
Funding Scheme	: H2020-EU.5. - SCIENCE WITH AND FOR SOCIETY
Topic	: SwafS-27-2020 - Hands-on citizen science and frugal innovation
Project's coordinator Organization	: EHESS-OpenEdition
E-mail address	: pierre.mounier@openedition.org , alessia.smaniotto@openedition.org
Website	: https://coeso.hypotheses.org
WP and tasks contributing	: Project Coordination and Management
WP leader	: EHESS-OpenEdition
Task leader	: EHESS-OpenEdition
Dissemination level	: Public
Due date	: 16 July 2021
Delivery date	: 16 July 2021
Author	: Alessia Smaniotto (EHESS-OpenEdition)
Contributors	: Tiziana Lombardo (Net7), Dominique Boullier (FNSP-CEE)
Reviewers	: Arnaud Gingold (CNRS-OpenEdition), Laure Desponds (EHESS, DPO)

Contents

Deliverable 1.2 at a glance	4
Data summary	5
COESO Project	5
Data definition	5
Data description	6
Project Data	6
Project Management Data	7
Research Data	7
Project Results	8
VERA Platform Data	8
FAIR Data	9
Data security	10
Project Management Data	10
Research Data	11
Results Data	11
Ethical aspects	11
Other issues	12
Supporting the partners in the implementation of a DMP at the pilots' level	12

I. Deliverable 1.2 at a glance

The COESO Data Management Plan (DMP) provides information relating to the managing and processing of the data that are produced and collected during the project's life, together with the measures for making the data Findable, Accessible, Interoperable and Reusable (FAIR). COESO's DMP further updates will provide information relating to the managing and processing of the data that will be produced and collected by the Virtual Ecosystem for Research Activation (VERA), among which the framework developed during the project life.

COESO's DMP is a living document in support of the project management and it provides core information in support of the Ethical requirements (POPD) and in support of the Plan for Exploitation and Dissemination of Results (PEDR).

During the six first months of the project, the main elements of the Data Management Plan have been collected, with the identification of the main datasets produced by the different project's tasks. The responsibilities for the managing and processing of each dataset has been identified, and the measures to comply with GDPR has been set up. More details for the FAIRification processes for the datasets produced during the project will be addressed further, and provided in the updated version of the DMP due at M12. In order to efficiently support the Exploitation Plan, the DMP will be constantly updated during all the project duration and an additional update of the DMP will be provided by M30.

The DMP includes information on:

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

This deliverable is based on the Horizon2020 generic template for DMPs.

II. Data summary

COESO Project

COESO aims to develop and sustain citizen science research in the social sciences and humanities. COESO’s mission is to enable strong growth of citizen science projects in the social sciences and humanities and to support participatory research through a service-first approach.

COESO is committed to the development of Citizen Science in the fields of Social Sciences and the Humanities (SSH). The project gathers 15 partners from 6 different European countries (France, Portugal, Italy, Spain, Belgium, and Germany), including academic research centres, public and private foundations, and small to medium-sized enterprises and associations.

COESO is a double level research project. It provides a framework for field research through ten collaborative experiences between researchers from different SSH disciplines (history, philosophy, political science, sociology and anthropology) and artists, journalists, associative members and local authorities. It then enables the study of these collaborations in order to develop a "collaboratory", a common platform allowing the different stakeholders to co-create new knowledge and solutions and disseminate them to society. The collaboratory will be enriched by a "cooperation analytics" system, allowing stakeholders to provide reflective feedback on these participatory experiences.



The project structure brings to the production of two bundles of data: (1) the project data on one side and (2) the VERA platform data on the other. A shared Mind Map of the data produced and collected during the project life has been drafted to visualise the dependencies and for allowing a collaborative cartography of the data produced.

The VERA platform data will be subject of the further updates of the DMP.

Data definition

A very broad definition of data can be found in a joint document by the United Nations Statistical Commission and the Economic Commission for Europe, produced for the Conference of European Statisticians Statistical Standards and Studies held in Geneva in 2000: “Data is the physical representation of information in a manner suitable for communication, interpretation, or processing by human beings or by automatic means.”¹

OECD members, in 2007, agreed on common recommendations concerning the access to

¹ Economic Commission for Europe of the United Nations (UNECE), "Terminology on Statistical Metadata", Conference of European Statisticians Statistical Standards and Studies, No. 53, Geneva, 2000).

research data produced from activities supported by public funding², and they defined “research data” as follow: “factual records (such as numerical scores, textual records, images, and sounds) resulting from research that is partially or fully funded by public funds, used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. This term does not cover laboratory notebooks, preliminary analyses, or drafts of scientific papers, plans for future research, peer reviews, personal communications with colleagues, or physical objects, (e.g., laboratory samples, strains of bacteria, or test animals).” The 2021 revision expands the scope of the Recommendations to cover “not only research data from public funding, but also other research-relevant digital objects, such as metadata and bespoke algorithms, workflows, models, and software (including code)”.

In the same recommendations OECD precises what is meant by “Research data management”: “the part of the research process that deals with organisation and handling of research data, including data management planning, structured storing, description, curation, preservation and provision of metadata and complementary algorithms, code, software, and workflows, and compliance with internal, national and international privacy legislation.”

This very first version of the COESO Data Management Plan drafts a general overview of the types and scopes of the “data” handled within the project. The further DMP updates will concentrate the attention on those of these data that will appear most valuable either for contributing to research as a common good or as a part of the knowledge economy. This may potentially exclude the project management data.

Data description

Project Data

COESO project data are separated into three bundles of data:

- the project management data, collected for the project coordination and management purposes
- the research data produced during the project life within the work packages
- the results data

² OECD, “Recommendation of the Council concerning Access to Research Data from Public Funding”, Adopted on: 14/12/2006, Amended on: 20/01/2021, available online: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347>



Project Management Data

For the purposes of the project coordination and management, a set of data will be produced such as documents, emails, chat conversations. These data will be generated and collected only for management purposes. No diffusion of such data is envisioned outside the consortium.

Research Data

COESO research activities will mainly collect qualitative data, gathering information that is not in numerical form but expressed in natural language, in a textual or visual form. Qualitative data collection methods within COESO project may include photography, audio recordings, video, unstructured interviews, semi-structured interviews, surveys, open-ended questionnaires, diary accounts, focus groups (group discussions), individual interviews and unstructured observations.

Furthermore, the research activity carried out for WP5 will collect information from the VERA Platform prototype in order to build Cooperation Analytics. The data collection will be executed through Natural Language Processing methodology and tools, and through the automatic recording of traces of uses of the various platform functionalities (logs). This will include: minimized IDs of each participant, which means that the personal data collected for identifying a person is reduced to relevant and limited information, only related to the purposes for which they are processed; the recovery of texts published on the platform, traces of activity monitored through the “cooperation analytics” themselves. The data collected through the Cooperation Analytics will be presented to the users in a pseudonymised form, in order to further refine the Cooperation Analytics and the VERA Platform.

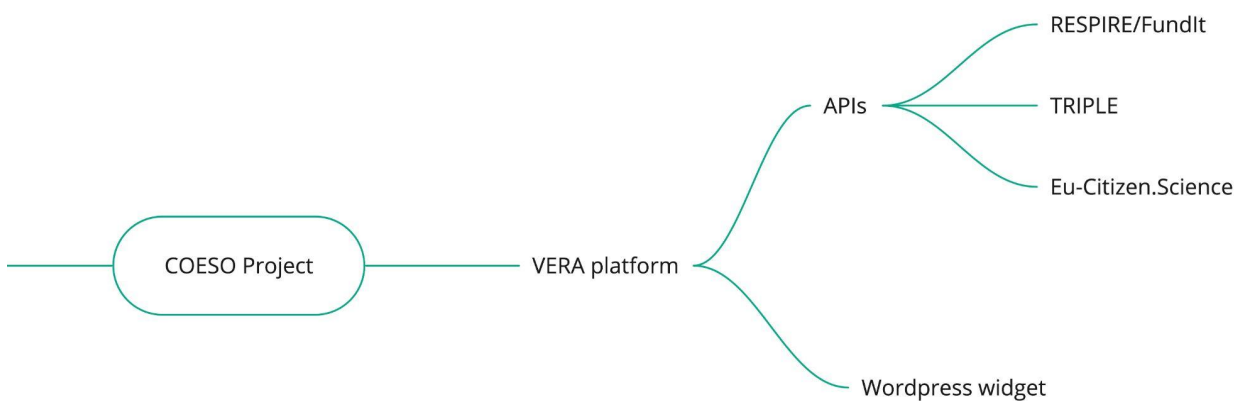
The purpose of the WP5 data collection is two-fold: (1) building and testing the Cooperation Analytics; (2) provide feedback to the VERA users (the pilots’ participants) about their own collective activity.

Project Results

The project results includes project deliverables, project and pilots’ website content, MemoRekall capsules related to Pilot 2 and transmedia products related to WP6.

VERA Platform Data

The VERA platform data will be described more in detail in the further updates of the DMP. A draft mind map has been produced to envision how the handling of the generated or collected data will flow.



III. FAIR Data

As one of its major goals, the project will ensure Open Access to its publications and make its data as open as possible. COESO project will act proactively to address FAIRification challenges in all its parts by facilitating discovery, using interoperable formats and metadata, securing hosting and storage, and supporting reuse through clear methodology and appropriate licensing.

The project reports will be safely stored and made openly accessible in appropriate repositories (e.g. zenodo.org). The deposit on Zenodo will enrich the records with Dublin Core metadata and a standard persistent identifier (DOI). All research outputs of COESO project will be released under Creative Commons Licenses CC-BY 4.0 international.

COESO project website and the single pilots' website are hosted on the Hypotheses.org platform, ensuring open access to the published content and worldwide outreach, relying on a solid public infrastructure.

The VERA platform hosting and data storage will be taken in charge by TGIR Huma-Num (HN) in France, which provides all the requirements for sustainability, high availability, secure storage, and long-term preservation³. VERA platform development will rely on the W3C standards for Open Web Platforms.

COESO project will actively work for the interoperability of VERA platforms with several platforms, by developing specific APIs to connect VERA with existing endeavours such as the federated databases of research funding RESPIRE and the european citizen science platform EU-Citizen.Science.

Moreover, the VERA development will use and produce open source tools and services. The addition of discovery interfaces for advanced discovery will follow the same rules and the code files will be deposited and updated on sustainable repositories. The discovery interfaces production will rely on VERA databases.

The development of the Cooperation analytics will provide data including a reference library on open source cooperation analytics and the configuration of the indicators will be thoroughly described. The data collected will be available in the most common formats (tab-delimited or spreadsheets).

More information about the research data FAIRification, including information about formats, metadata standards, size of the data and possibility of reuse will be included in the further updates of the DMP.

³ <https://www.huma-num.fr/services-et-outils/diffuser>

IV. Data security

Each partner involved in generating or collecting data for the project purpose has the responsibility for a secure data storage, data recovery and secure transfer of data, whether they are sensitive or not.

The partners have the responsibility of identifying or appointing their Data Protection Officer within their organization.

The project coordinator can support partners whether they need specific information about common procedures and general regulation.

COESO project will not implement activities or results raising security issues, and it does not include “EU-classified information” as background or results.

Project Management Data

The project coordinator is responsible for the secure storage of the project management data.

The project mailing lists are stored and archived on a dedicated server at EHESS headquarters. Backup of the servers are made on a regular basis. The mailing list archives are solely accessible by the mailing lists subscribers.

An appropriate shared workspace has been created by the Project Coordinator, through a GSuite Basic Google workspace. This workspace uses the Google Drive repository to facilitate the accessibility and exchange of the whole COESO documents.

Google workspaces rely on data centers located around the world ([locations list](#)) and, according to Google, “service data may be processed on servers located outside of the country where our users and customers are located because Service Data is typically processed by centralized or regionalized operations like billing, support, and security”. Google claims to comply with certain [legal frameworks](#) when transferring Service Data outside of the European Economic Area (EEA). In particular, they declare that “as of 16 July 2020, [they] no longer rely on the EU-U.S. Privacy Shield to transfer data that originated in the EEA or the UK to the US” and to be compliant with the European GDPR regulation ([more information about Google Workspaces and GDPR regulation](#)).

The access right to documents and folders can be regulated at the single item level.

OpenEdition Google Workspace is complemented by a backup process including all the data stored (emails, documents, calendar and contact). The backup relies on Backupify service. The common Drive backup includes all documents owned by [openedition.org](#) users and all the documents included in a directory owned by an [openedition.org](#) user.

The French Commission on Informatics and Liberty (CNIL-*Commission nationale de l'informatique*

et des libertés), that is the national data protection authority for France, recently⁴ discouraged the use of US collaborative work tools for research purposes, due to the invalidation of the privacy shield by the European Court of Justice in July 2020. The CNIL recognizes the need of a transitional period to allow switching complex information systems toward new European solutions and it declares being willing to help identify possible alternatives. OpenEdition is currently waiting for more information from the CNIL about the concrete support they can provide.

The chatting tool based on the Mattermost chatting service has been provided for all the consortium members. The tool is provided by the French infrastructure Huma-Num that ensures secure data storage.

Research Data

Each partner can rely on the common project workspace to store if no secured storage is provided by their organisation.

If partners need to store data potentially rising security issues for the people or organizations involved, and for which they cannot ensure a secure storage by themselves, an additional storage system can be provided through the Huma-Num (CNRS) services ShareDocs and Huma-Num Box ([information in French about the Huma-Num infrastructure services](#)) or through the EHES [Didomena](#) repository.

Whenever necessary, research data will be anonymized and the correspondence table will be securely stored in a separated file.

Results Data

The pilots' dedicated websites and the COESO project website are both hosted on the academic blogging platform Hypotheses.org. Hypotheses.org platform data are securely hosted and stored on the OpenEdition servers, located at the IN2P3, computing center of the CNRS (Centre National de la Recherche Scientifique), in Villeurbanne (France). The access to the servers is possible via SSH protocol with key authentication. Only OpenEdition system administrators have access to the servers. Hypotheses.org data are backed up both on disk and on tape.

COESO main deliverables will be uploaded on Zenodo repository. As stated on the Zenodo [General Policies v1.0](#), "All data files uploaded on Zenodo are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis." and "Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least."

⁴ See "[La CNIL appelle à des évolutions dans l'utilisation des outils collaboratifs étatsuniens pour l'enseignement supérieur et la recherche](#)"; CNIL website, published on May 27th 2021

V. Ethical aspects

Within the field of social sciences and the humanities, we will often work with data originating from human participants and thus we will often handle personal data, whether they are sensitive data or not. Sensitive personal data deserve special attention and COESO project implementation will ensure that data processing and management will respect the General Data Protection Regulation (GDPR) provisions.

An informed consent form will ensure the human participants voluntarily participate in this social sciences research project. When dealing with personal data, the informed consent for data sharing and long term preservation is included in questionnaires and explained to the participants.

More details on ethical issues and the examples of informed consent forms are included in the dedicated POPD - Ethical requirements deliverable no.3 deliverable (confidential).

The COESO websites are hosted on the Hypotheses.org platform, that it's part of the OpenEdition infrastructure. The [OpenEdition infrastructure privacy policy](#) is available on the OpenEdition website and it is directly accessible through a link provided in the footer of each Hypotheses.org website, including COESO project website and the single pilots' websites.

VI. Other issues

Supporting the partners in the implementation of a DMP at the pilots' level

COESO is a citizen science project involving partners coming from different sectors and professional areas, with a huge diversity in the organization sizes and governance structures.

Sensibilisation on data management issues and support in the implementation of a DMP plan at the pilot level will be provided for all the partners that are not already aware of all the data management obligations and opportunities.

The "Data Management Plan template for H2020 projects (Version v01.100419)", by Elias P. Koumoulos, Marco Sebastiani, Nikolaos Romanos, Maritini Kalogerini, & Costas Charitidis [2019, April 10, Zenodo: <http://doi.org/10.5281/zenodo.2635768>] provide a useful tool in such support and sensibilisation action.