C·Act

Co-designing Citizen Social Science for Collective Action

#1.2

DATA MANAGEMENT PLAN

● ● ● ●

Date – June 30, 2020

Dissemination level - Public

Responsible Partner – Open Knowledge Foundation (OKF)

Author(s) - Cédric Lombion (OKF)

Reviewers - Isabelle Bonhoure (UB), Josep Perelló (UB)

Document History

Version	Date	Contributor	Comments
1.0	2020-23-06	Cédric Lombion	Shared with reviewers
1.1	2020-29-06	Cédric Lombion, Isabelle Bonhoure	Shared with reviewers
1.2	2020-30-06	Cédric Lombion, Isabelle Bonhoure, Josep Perelló	Ready for publication

Table of Contents

List of figures

Figure 1: CoAct R&I life cycle for all R&I actions and Research Pilots

List of tables

List of abbreviations

CoAct	Co-Designing Citizen Social Science for Collective Action
CSS	Citizen Social Science
EC	European Commission
EU	European Union
ECSA	European Citizen Science Association
FARN	Fundación Ambiente y Recursos Naturales
FHP	Fachhochschule Potsdam
FSMC	Federació Salut Mental Catalunya
GA	Grant Agreement No. 873048 with the European Commission
GIG	Global Innovation Gathering e.V.
OKF	Open Knowledge International
UB	Universitat de Barcelona
UNIVIE	Universität Wien
UNSAM	Universidad Nacional de San Martín
WPs	Work Packages
ZSI	Zentrum für Soziale Innovation (Centre for Social Innovation)

1. Executive Summary

CoAct proposes a radically new participatory approach through "four wicked" social issues in which citizen groups act as co-researchers. The overall objective of CoAct is to develop and demonstrate the scientific relevance and social impact of Citizen Social Science, which is to date an underexplored area of citizen science.

As a Horizon 2020 project, CoAct is expected to produce a data management plan to validate the results of future scientific publications. The inclusion of non-research data in the data management plan is voluntary (https://www.openaire.eu/open-research-data-the-new-norm-in-h2020).

The CoAct projects includes a Data Management Plan as Deliverable D1.2. This deliverable is part of Task T1.4 of the Work Package WP1, focused on Project Management and Coordination. The main objective of this task is to make Research Data FAIR, meaning Findable, Accessible, Interoperable and Reusable, in accordance with the version 3 of the EC's Guideline on Fair Data Management in Horizon 2020 (https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).

The Data Management Plan is expected to be delivered in 3 phases: A draft version at Month 6, a revision at Month 18 and a final version at M36. The current document is the draft version. It aims to:

- Provide an overview of the research data that is expected to be collected, generated and processed across the lifecycle of the CoAct Work Package and Research Phases.
- Provide similar information related to non-research data.
- Detail our existing plans to make our data FAIR and embed personal data protection practices in our activities.
- Outline the current gaps and unknowns to be clarified by the Month 18 revision of the Data Management Plan.

For Horizon 2020 projects, a data management plan is required for the data needed to validate the results presented in a scientific publication, other data can be provided on a voluntary basis. Both kind of data are covered in this document, across 2 groups of WPs, which nonetheless overlap:

- WP 3,4,5,7 which mainly involve data used directly in research activities and whose FAIR data practices will be tailored to the specifics of their activities
- WP 2,6,8 which mainly involve personal data (contact information) collected as part of communication efforts

2. Introduction

2.1. The CoAct project and the Data Management Plan

This Data Management Plan (DMP) has been prepared by mostly following the version 3 of the 'Guidelines on Fair Data Management in Horizon 2020' (https://www.openaire.eu/open-research-data-the-new-norm-in-h2020), with the help of the DMPOnline tool (https://dmponline.dcc.ac.uk) recommend in the guideline document. As stated by the mentioned guidelines, the goal of this first version of the DMP is not to generate an extensive and definitive document but rather to set the basis of data management in the CoAct H2020 Project. This document will be kept therefore alive during the evolution of CoAct project. It will be periodically updated and completed during the whole duration of the project while a revised version will be submitted at M18 and the final version delivered at the end of the project (M36).

In this SWAFS project, the general goal is to deploy and demonstrate the scientific relevance and the social impact of Citizen Social Science (CSS). CSS combines equal collaboration between citizen groups (co-researchers) that are sharing a social concern and academic researchers. Such an approach enables to address pressing social issues from the bottom-up, embedded in their social contexts, with robust research methods.

CoAct is proposing a new understanding of CSS as participatory research co-designed and directly driven by citizen groups sharing a social concern, in which they become co-researchers in processes commonly dominated by academic researchers. CoAct aims to bring together and further develop methods to give citizen groups an equal 'seat at the table' through active participation in research, from the design to the interpretation of the results and their transformation into concrete actions.

This collaborative process involves the collection, generation and processing of data. The lifecycle of this data is addressed in this document, through the angle of FAIR data principles as well as personal data protection ones. Beyond supporting the EC's drive toward Open Science, this document will be actively used as a coordination tool for the CoAct Consortium in order to embed and align data management best practices across partners and stakeholders.

The Data Management Plan complements and reuses key elements of the Ethics Requirement Deliverable (D9.3) which describes the technical and organizational measures that will be

implemented to safeguard the rights and freedoms of the data subjects/research participants, including security measures.

2.2. Relevant Work Packages

The CoAct consortium will engage in data collection and processing as part of seven of its WPs.

- WP 2: CSS Foundations.
- WP 3: Research action #1: Mental Health, Barcelona.
- WP 4: Research action #2: Youth Unemployment, Vienna.
- WP 5: Research action #3: Environmental Justice, Buenos Aires.
- WP 6: Endeavoring new Citizen science spaces: Gender Equality.
- WP 7: Evaluation and Impact Assessment.
- WP 8: Communication, Dissemination and Exploitation.

WP 3,4,5,7 will mostly deal with research data (defined as data used specifically to validate research results) while WP 2,6,8 will mostly process non-research data (see section 3.3.4). They are consequently covered in separate sections.

In both groups, data collection and processing has been identified as necessary in order to deliver research and project outcomes. Beyond the data processing needed for communication purposes, the research activities of WP3,4,5,7 have data as a central component of their strategy: to map the complexity and diversity of the mental health community (WP3), to document the institutional environment of early school leavers (WP4) or to crowdsource testimonies about environmental issues (WP5). Indeed, the process of identifying, collecting, analyzing and presenting data is essential to CSS research projects.

2.3. Definitions

2.3.1. CoAct-specific terms

Co-Researchers: The Co-Researchers are citizens in a vulnerable situation, due to their lived experience in relation to the social concerns that motivates the collective R&I Actions (Mental Health Care, Youth Employment and Environmental Justice). They constitute the case studies of the CoAct's

CSS research process. Co-Researchers co-create, with academic researchers, the collective research tools (either digital or non-digital) through co-design mechanisms that allow to reach consensus and agreement among participants while including different perspectives and viewpoints. They participate in the research data collection (that may be upscaled to other collectives and individuals) or in some cases do most of the data collection themselves and, whenever possible, in the evaluation of the project process and results. In certain cases, and always with their explicit agreement, they may participate in the research tools launching and promotion and/or the presentation of the results. They analyse and interpret the research results, deliver them to the Knowledge Coalition, and/or autonomously trigger specific collective actions. They are co-owners of the research data and results and they may be listed as co-authors if they express their willingness to have their names disclosed for (scientific)publications.

Knowledge Coalition: The Knowledge Coalitions in CoAct are formed by representatives of Public Administrations, CSOs, educative organisations and Co-researchers—to name a few. Although all research projects involve people from different areas of the political and social realm, their involvement and collaboration vary according to the specific Research and Innovation Action planned. The participation of the individual parties of the Knowledge Coalition can take place in various forms: creating a structural framework for research, participating in the actual research process, informing on the corresponding problem or implementing and discussing possible solutions. Regardless of the role, each R&I Action creates a network between the different actors and promotes an exchange between them.

(Communities of) Citizen Scientists: Once the research is co-created with the Co-Researchers, participatory research will be eventually further conducted through digital platforms in order to collect massive robust scientific evidence to respond to the Co-Researchers concerns. The Citizen Scientists will be called for participation through communities' event, social media or local digital and non-digital press.

Research Subjects: Research subjects participate in the research project as interviewees, in group discussions, in participatory observation settings or similar. They do not participate in the design, conduct or analysis of the project but rather are in the position of informants and experts. This group of research participants will be especially important when it comes to the youth-led research projects of R&I Action #2 Youth Employment.

2.3.2. DMP terms

The types and formats of data generated will vary for each Work Package. In the Deliverable D9.3 (Ethics Requirement), the CoAct Consortium defines two types of data: personal data and special categories of personal data (referred to in this document as 'sensitive data').

Personal data refers to any information that relates to an individual who can be directly or indirectly identified. Names and email addresses are obviously personal data. Location information, ethnicity, gender, biometric data, religious beliefs, web cookies, and political opinions can also be personal data. Pseudonymous data can also fall under the definition if it's relatively easy to ID someone from it.

Special categories of personal data are subject to more stringent data-protection safeguards. They include 'personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation' (Article 9(1) GDPR).

The Data Management Plan will also include references to '**non-personal data**' which includes all data not covered by the above definitions such as aggregated data, anonymized data, open government data etc.

Due to the nature of the CoAct project, which involves research activities around topics such as mental health, youth employment or environmental justice, a lot of the research activities include moments where co-researchers share their personal perspective on the topic of the research. This leads us to define two types of situations wherein personal or sensitive data is generated and collected:

- Situations where the academic researcher collecting the data aims or has no choice but to collect or process these types of data: those situations are identified by the use of the modal verb '**will**' (e.g. "personal data will be collected for analysis purposes").
- Situations where the academic researcher collecting the data do not aim to collect these types of data: those situations are identified by the use of the modal verb '**may**' (e.g. "sensitive data may come up during the workshop").

2.4. The CoAct project lifecycle

The Research Action activities found in WP3, 4 and 5 can be categorized into five phases: Research & Innovation preparation, Research Co-design, Conducting research, Collective Data Analysis and Transformation of results into Action (see Figure 1). Those phases form together the CoAct Research and Innovation Cycle and are used throughout this document as a framework to breakdown the data lifecycle of the project.
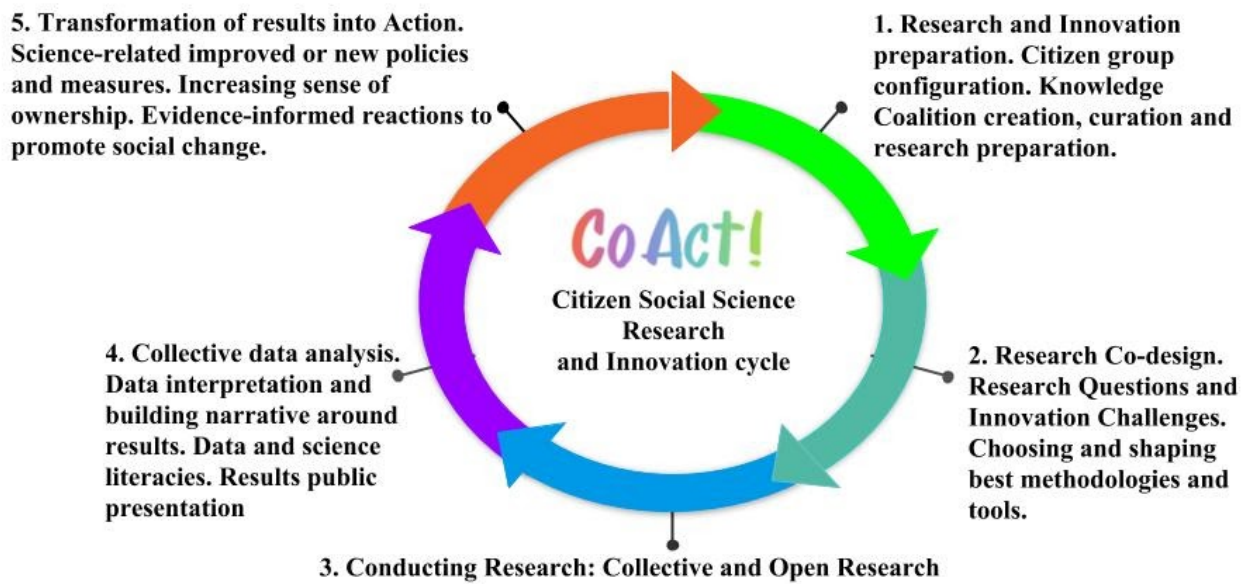

Figure 1: CoAct R&I life cycle for all R&I actions and Research Pilots.

3. Data Summary

3.1. State the purpose of the data collection/generation

Data collection, generation and processing serves 2 main purposes in the CoAct project:

1. Support the research activities and validate research results
2. Support the communication with project stakeholders

3.2. Explain the relation to the objectives of the project

The data collection, generation and processing within CoAct is linked to all of the project's objectives. Specifically, the data used to support research activities will be mainly relevant to objectives 1,2,3,4,6 while the data used to support the communication with project stakeholders will be mainly relevant to objectives 2,5,7,8 (see Table 1).

Objective	
O1	To generate new ground-breaking and open scientific outcomes by means of CSS
O2	To engage vulnerable citizens and local civil society groups in R&I initiatives and to place them at the centre of the R&I cycle
O3	To produce scientific evidence -informed reactions and thereby create new policies and to improve existing ones
O4	To build a common and validated transdisciplinary CSS methodological framework for a variety of end -users
O5	To promote Open Science and scientific research integrity in methods and data
O6	To create and validate a robust and inclusive R&I evaluation framework
O7	To increase scientific literacy, skills, competences and public awareness regarding science
O8	To disseminate CoAct results and build a global sustainable CSS community of Practice

Table 1: CoAct Objectives

3.3. Specify the types and formats of data generated/collected

The data processed by the CoAct consortium will be collected across a variety of formats and is expected to reach a few terabytes in size by the end of the project.

The tables below break down the different types and formats of data encountered across the project's lifecycle:

3.3.1. Types of data (summary)

Phases	Types of data	WP3	WP4	WP5	WP7
Research and Innovation Preparation	non-personal	will	will	will	
	personal	will	will	will	
	sensitive	may	may	may	
Research co-design	non-personal	will	will	will	will
	personal	will	may	may	will
	sensitive	may	may		will
Collective Research	non-personal	will	will	will	will
	personal	will	may	may	may
	sensitive		may	may	may
Collective Data Analysis	non-personal	will	will	will	will
	personal	will	will	may	will
	sensitive				will
Transformation of results into action	non-personal	will	will	will	will
	personal	will	may	may	may
	sensitive	may			

Table 2: Summary of Types of data collected in CoAct, by WPs and Research phases

3.3.2. Data formats (summary)

Phases	WP3	WP4	WP5	WP7
Research and Innovation Preparation	→ Audio recordings → Text transcriptions → Contact Information and Questionnaire answers spreadsheets	→ Audio recordings → Text transcriptions → Contact Information and Questionnaire answers spreadsheets	→ Audio recordings → Text transcriptions → Contact Information and Questionnaire answers spreadsheets	

Research co-design	→ Text notes → Contact Information spreadsheet	→ Audio recordings → Photos → Text transcriptions	→ Text notes → Contact Information spreadsheet → Photos	→ Text notes → Contact Information spreadsheet → Audio and video files → Text transcriptions → Contact Information and Questionnaire answers spreadsheets
Collective Research	→ Text transcriptions → Contact Information spreadsheet → Statistical analysis spreadsheets → SQL Database → Image files for data presentation	→ Audio recordings → Photos → Text transcriptions → Contact Information and Questionnaire answers spreadsheets	→ SQL Database → Photos	
Collective Data Analysis	→ Audio recordings → Contact information spreadsheet	→ Audio recordings → Text transcriptions and analysis outputs → Questionnaire answers spreadsheet → Image files for data presentation	→ Contact Information spreadsheet → Photos → Image files for data presentation	→ Non-personal data will be generated, analysed and preserved → Personal data will be processed for anonymisation → Sensitive data will be processed for anonymisation
Transformation of results into action	→ Audio recordings → Photos → Contact information spreadsheet → textual and tabular data	→ Audio or video recordings → Photos → Contact information spreadsheet → textual and tabular data	→ Audio or video recordings → Photos → textual and tabular data	→ textual and tabular data

Table 3: Summary of types data formats encountered across the CoAct WPs and Research Phases

3.3.3. Full Description of data types and formats

Phases	WP3	WP4	WP5	WP7
Research and Innovation Preparation	**Types of data** ● Non-personal data will be collected as part of the documentation process ● Personal data (contact information) will be collected to communicate with knowledge coalition members ● Sensitive data may come up during interviews **Data formats** ● Audio recordings ● Text transcriptions ● Contact Information and Questionnaire answers spreadsheets	**Types of data** ● Non-personal data will be collected as part of the documentation process ● Personal data (contact information) will be collected to communicate with knowledge coalition members ● Sensitive data may come up during interviews **Data formats** ● Audio recordings ● Text transcriptions ● Contact Information and Questionnaire answers spreadsheets	**Types of data** ● Non-personal data will be collected as part of the documentation process ● Personal data (contact information) will be collected to communicate with knowledge coalition members ● Sensitive data may come up during interviews **Data formats** ● Audio recordings ● Text transcriptions ● Contact Information and Questionnaire answers spreadsheets	
Research co-design	**Types of data** ● Non-personal data will be collected as part of the documentation process ● Personal data (contact information) will be collected to communicate with co-researchers and personal data	**Types of data** ● Non-personal data will be collected as part of the documentation process ● Personal data and sensitive data may come up and be recorded during the workshop discussions **Data formats**	**Types of data** ● Non-personal data will be collected for documentation and analysis purposes ● Personal data may be used for coordination purposes (registration) and personal data regulation compliance	**Types of data** ● Non-personal data will be collected for documentation and analysis purposes ● Personal data will be needed for evaluation instrument design and personal data regulation compliance

	- regulation compliance - Sensitive data may emerge from the design the process Data formats - Text notes - Contact Information spreadsheet	- Audio recordings - Photos - Text transcriptions	Data formats - Text notes - Contact Information spreadsheet - Photos	- Sensitive data will be needed for evaluation instrument design Data formats - Text notes - Contact Information spreadsheet - No paper data currently planned. - Text transcriptions - Contact Information and Questionnaire answers spreadsheets
Collective Research	Types of data - Non-personal data will be collected for statistical purposes and generated through automated analysis - Personal data (contact information) will be collected to communicate with co-researchers Data formats - Text transcriptions - Contact Information spreadsheet	Types of data - Non-personal data will be collected as part of the documentation process - Personal data and sensitive data may come up and be recorded during the research activities Data formats - Audio recordings - Photos - Text transcriptions - Contact Information and Questionnaire answers spreadsheets	Types of data - Non-personal data will be generated, analysed and preserved - Personal data may be used for coordination purposes (registration) - Sensitive data may be shared by citizen scientsists Data formats - SQL Database - Photos	Types of data - Non-personal data will be collected as part of the documentation process - Personal data and sensitive data may come up and be recorded during the research activities Data formats - Audio recordings - Photos - Text transcriptions - Tabular data - Statistical analysis spreadsheets - Contact Information and

	• Statistical analysis spreadsheets • SQL Database • Image files for data presentation			Questionnaire answers spreadsheets
Collective Data Analysis	**Types of data** • Non-personal data will be generated, analysed and preserved • Personal data and sensitive data will be anonymised then analysed **Data formats** • tabular data to shared aggregated and anonymized data • Images files for data presentation	**Types of data** • Non-personal data will be generated, analysed and preserved • Personal data and sensitive data will be anonymised then analysed **Data formats** • Audio recordings • Text transcriptions and analysis outputs • Questionnaire answers spreadsheet • Image files for data presentation	**Types of data** • Non-personal data will be generated, analysed and preserved • Personal data may be used for coordination purposes (registration) **Data formats** • Contact Information spreadsheet • Photos • Image files for data presentation	**Types of data** • Non-personal data will be generated, analysed and preserved • Personal data will be processed for anonymisation • Sensitive data will be processed for anonymisation **Data formats** • textual and tabular data
Transformation of results into action	**Types of data** • Non-personal data will be shared as part of the presentation of research results • Personal data will be collected as contact information about policy stakeholders	**Types of data** • Non-personal data will be shared as part of the presentation of research results • Personal data or sensitive data may be shared by co-researchers	**Types of data** • Non-personal data will be shared as part of the presentation of research results • Personal data or sensitive data may be shared if consent is given **Data formats**	**Types of data** • Non-personal data will be shared as part of the presentation of research results • Personal data or sensitive data may be shared by co-researchers

		during public events	Audio or video recordings Photos textual and tabular data	during public events
	• Sensitive data may be collected as part of the interview process Data formats • Audio recordings • Photos • Contact information spreadsheet • textual and tabular data	Data formats • Audio or video recordings • Photos • Contact information spreadsheet • textual and tabular data		Data formats • textual and tabular data • photos

Table 4: Full description of data types and formats encountered across the CoAct WPs and Research Phases

3.3.4. Data used to support communication with stakeholders

As part of WP2, WP6 and WP8 the only data expected to be collected is contact information used for communication purposes.

In WP2, the capacity and community building actions will encompass the collection of contact details of experts and researchers interested in CSS and from outside the Consortium. These contact details will serve to build a new community of practitioners and researchers. These contact details will be collected following the GDPR rules and an explicit agreement will be always be sought in case of eventual contact details which will always have a clear and specific purpose: such as the building of mailing lists for knowledge sharing.

In WP6, several Open Calls addressing research projects on gender equality will be prepared. The contact details of the applicants will be collected following the GDPR standards, safely stored and destroyed at the end of the project. In the Open Calls rules, special importance will be given to Ethics and Personal Data Protection. These items and how the applicants will work with them will constitute one of the eligibility criteria for the grants. Once the Open Calls are resolved, the grantees will be responsible to take the legal and ethical measures regarding personal data collection of research participants described in their application.

In WP8, contact details of persons interested by CoAct communication and dissemination outputs as well as by CoAct own events will be collected. This is for example the case of recipients of newsletters, participants of the PhD summer school or of the final CoAct conference. These contact details will be collected following GDPR standards, be safely stored during CoAct and destroyed at the end of the project.

3.4. Specify if existing data is being re-used (if any)

Most of the data collected or processed as part of the CoAct research project is sourced from humans: testimonies, opinions, interviews. The two exceptions include:

- Data sourced externally: as part of the data analysis phase of all Work Projects, openly available data may be used.
- Data collected in the "Main Research Activities" phase of WP5: the crowdsourced data will include environmental data.

3.5. Specify the origin of the data

As mentioned above, most of the data will be sourced from humans, with the exceptions of open data sourced externally and crowdsourced data in WP5.

3.6. State the expected size of the data (if known)

The data collected across project activities is expected to reach a few terabytes in size by the end of the project.

3.7. Outline the data utility: to whom will it be useful

In order to serve the project's objectives (see Table 1), the data will need to be useful to the main project stakeholders:

- The CoAct Consortium members themselves (O1, O4)
- The research community (O1, O4, O8)
- The co-researchers, citizen scientists and research subjects (O2, O7, O8)
- The policy and decision-makers (O3)

4. FAIR Data

4.1 Making data findable, including provisions for metadata

4.1.1. Outline the discoverability of data (metadata provision)

At this stage in the project the Consortium has not yet aligned its data documentation practices but plan to do so by the end of the first phase of the research projects (Research Coalition Building).

Discoverability of the data will be made possible through:

- The publication of data on open repositories such as Zenodo
- The use of unique identifiers for each dataset
- Consistent naming and versioning of files
- The creation of a data inventory to list all datasets
- The creation of metadata for each dataset

4.1.2. Outline the identifiability of data and refer to standard identification mechanisms

Datasets will be assigned a Digital Object Identifier (DOI) in order to facilitate discoverability and identification.

4.1.3. Outline naming conventions used

Naming conventions have not been enforced yet across Consortium members, but will be, in order to allow consistent naming and versioning conventions.

4.1.4. Outline the approach towards search keyword

Search Keywords have not been created yet but will be defined with a focus on discoverability.

4.1.5. Outline the approach for clear versioning

The project follows the semantic versioning standard (https://semver.org/).

4.1.6. Specify standards for metadata creation (if any)

We have not identified a need to follow a specific standard or ontology to structure the data as different work projects will generate different types of data. The topic is still being discussed among partners.

4.2 Making data openly accessible

4.2.1 Specify which data will be made openly available

We plan to publish all non-personal data (with its associated metadata) on open repositories. All relevant research data will be made available on the Zenodo open research platform.

4.2.2. Specify how the data will be made available

All relevant research data will be made available at least on the Zenodo open research platform.

4.2.3. Specify what methods or software tools are needed to access the data

Research data will be published in standard formats such a CSV, XLS, JSON, SQLITE, meaning that any standard software will be able to access it, including open source software.

4.2.4. Specify where the data and associated metadata, documentation and code are deposited

Zenodo will be the main hosting platform for open research data. Additional platforms may be considered later in the project.

4.2.5. Specify how access will be provided in case there are any restrictions

We do not currently anticipate having to deal with restricted access datasets. Nonetheless a public data inventory will be created with the publication status and an explanation if the data was kept private, allowing re-users to contact the data controller to ask about unpublished datasets.

4.3 Making data interoperable

4.3.1. Assess the interoperability of your data

All our datasets will be published using open source and/or standard formats. Additionally, the datasets which can exported in a tabular format (such as CSV) will be published using the Frictionless Data Standard developed by the Open Knowledge Foundation.

The frictionless data standard (https://specs.frictionlessdata.io/) is designed to maximise interoperability for data produced by research projects. These specifications are a standard vocabulary of patterns for describing datasets, files, and tables. As these specifications are in JSON, they are machine readable and interoperable and are also compatible with other field-specific standards.

4.3.2. Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability

The Consortium members are still defining the type of data that they expect to generate. From these discussions will emerge the standard vocabulary to be used across the research activities, which we will align as much as possible to commonly used social science vocabularies.

4.4 Increase data re-use

4.4.1. Specify how the data will be licensed to permit the widest reuse possible

To ensure compliance with GDPR regulation, all personal data will be anonymized. We aim to publish the resulting datasets under a Creative Commons Public Domain Dedication (CC0), in accordance with the OpenAire guidelines (https://www.openaire.eu/research-data-how-to-license/)

4.4.2. Specify when the data will be made available for re-use

All the publishable research data will be made available before the end of the project, by December 2022.

4.4.3. Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project

The CC0 License that we intend to use allow any kind of re-use of the data, without restriction.

4.4.4. Describe data quality assurance processes

Data quality will be ensured through a 2-step process

The first step takes place before the launch of the collective research activities (see figure 1). The Open Knowledge team, responsible for the maintenance of the DMP, will consult with each partner to help them review their planned data structures, types, ontologies etc.

The second step will take place before the publication of the dataset. Using the Frictionless Data's GoodTables pipeline, partners will be able to validate their data and ensure that their data exactly matches their expectations.

GoodTables (https://frictionlessdata.io/tooling/goodtables/) is a tool to validate tabular data. It can check the structure of a dataset (e.g. all rows have the same number of columns), and its contents (e.g. all dates are valid). Internally, it uses the Data Quality Spec for common tabular data errors, and it is completely customizable with an API for creating custom checks.

4.4.5. Specify the length of time for which the data will remain re-usable

The CC0 License that we intend to use for published data will allow any kind of re-use of the data, without restriction.

5. Allocation of Resources

5.1. Estimate the costs for making your data FAIR.

Most of the costs are already covered by the participating universities as they have already implemented FAIR data processes internally. Some costs, yet to be defined, may be accrued as part of the implementation of the Frictionless Data validation pipeline.

Due to the Covid-19 situation, several changes are planned in the deployment strategies of the CoAct WPs. Once a final decision is taken for each relevant Work Package, we will be able to evaluate the allocation of resources more precisely.

5.2. Clearly identify responsibilities for data management in your project

Each WP leader is responsible for their data management. Inside each WP, one person will be acting as the representative data controller:

- WP2: David Scheller (FHP)
- WP3: Josep Perelló (UB)
- WP4: Veronika Woehrer (UNIVIE)
- WP5: Valeria Arza (UNSAM)
- WP6: Cédric Lombion (OKF)
- WP7: Katja Mayer (ZSI)
- WP8: Kersti Wissenbach (GIG)

5.3. Describe costs and potential value of long-term preservation

No cost is presently foreseen as Zenodo is a zero cost long term service.

6. Data Security

We distinguish two types of storage:

- **Short term storage**, which refers to the intermediary devices used to collect data and transfer it to longer-term storage before deletion (e.g. audio recorder, camera, laptop)
- **Long term storage**, which refers to the servers where the data will be stored in until its deletion (e.g. University servers)

Further, storage devices are grouped into personal and professional devices:

- **Personal devices** are under the sole responsibility of the individual operating it and cannot be assumed to follow any kind of organizational policy
- **Professional devices** are operated by the researcher by managed by their affiliate organization and are consequently operated and secured following the guidelines of the relevant organization.

Lastly, several projects will include workshops where information will be recorded on paper. We will consequently talk about digital and paper data to distinguish these two formats.

Activities	WP3	WP4	WP5	WP7
Short term storage	Digital data ● Portable computers of the University of Barcelona (UB) Paper data ● Key-locked storage in the office of Prof Josep Perelló.	Digital data ● Professional computers of the University of Vienna (UNIVIE) ● Professional Audio recorders of UNIVIE ● Personal devices used by non-UNIVIE-affiliated co-researchers. A review of the devices used by the co-researchers will have to be done. Paper data ● No paper data currently planned.	Digital data ● Personal computers. May be subject to change after the Covid epidemic. ● Personal smartphones (audio recording). May be subject to change after the Covid epidemic. Paper data ● Not yet determined	Digital data ● Computers of the Zentrum für Soziale Innovation (Centre for Social Innovation, ZSI) Paper data ● No paper data currently planned
Long term storage	Digital data ● Servers of the University of Barcelona Paper data ● Key-locked storage in the office of Prof Josep Perelló.	Digital data ● Dedicated storage space on the servers of UNIVIE Paper data ● No paper data currently planned.	Digital data ● Personal Google Drive accounts. Will be reviewed. Paper data ● Not yet determined	Digital data ● Servers of the ZSI Paper data 1) No paper data currently planned
Storage of personal and sensitive data	Digital data ● The data will be immediately pseudonymized by the Data Controller. Inside UB, only the Principal Investigator (PI) (Josep Perelló) will have access	Digital data ● In the dedicated server space, two folders: one for anonymized data used for analysis, and one for non-anonymized data for GDPR	Digital data ● Not yet determined Paper data ● Not yet determined	Digital data ● The data will be immediately pseudonymized by the research team. Transcribed Audio and video files will be deleted after

	to the Personal Data and he will be the only person involved in the data pseudonymizatio n process Paper data ● No personal or sensitive data will be collected in paper format	compliance purposes. Paper data ● No paper data currently planned.		transcription and documentation. Paper data ● No paper data currently planned
Transfer of personal and sensitive data	Digital data ● Personal and sensitive data collected by the researchers themselves will be transferred directly from their professional devices to the UB servers using their institutional account on the GDPR compliant Microsoft OneDrive Cloud computing service. ● Personal data or sensitive data may be transferred through a Telegram platform-hosted chatbot. The design of the chatbot will have to consider the fact that conversation channels with chatbots are not	Digital data ● Personal and sensitive data collected by UNIVIE researchers themselves will be transferred directly from their professional devices to the UNIVIE servers. ● The process for transferring the data collected by other co-researchers will need to be clarified. Paper data ● No paper data currently planned.	Digital data ● Not yet determined Paper data ● Not yet determined	Digital data ● Personal and sensitive data collected by the researchers themselves will be transferred directly from their professional devices to the ZSI servers. Paper data ● No paper data currently planned

	encrypted on Telegram. <u>Paper data</u> ● No personal or sensitive data will be collected in paper format			

Table 5: Summary of Data management measures through the lens of data security

7. Ethical Aspects

Three Ethics Deliverables (D9.1, D9.2, D9.3) have been submitted at M3. Although the main information related to research data have been extensively detailed in the present DMP, some other Ethical aspect are closely connected:

2) The definition of the research participants of each R&I action has been carefully elaborated, together with the recruitment process of the different actors of the research. Consequently, the different research participants, that will generate the research data of the project are carefully categorized and recruited following a fair procedure (see D9.1 for more details).

3) CoAct policy follows the recommendations done in several H2020 official publications (Ethics and Data Protection, Ethics in Social Science and Humanities). In agreement with these publications, whenever the CoAct partners will collect personal data directly from research participants, they must seek their informed consent by means of a procedure that at least meets the minimum standards of the GDPR law. This will take the form of a written statement, which may be collected by electronic means or, very seldomly, it may take the form of an oral statement. Records documenting the informed consent procedure, including the information sheets and consent forms provided to research participants, including their consent to data processing, will be kept safely for all research activities.

Additionally, as CoAct is involving sensitive data-processing in the case of the Co-Researchers, a specific informed consent process covering the data-processing component of CoAct will be implemented. See D9.2 for more details and for consulting the Informed Consent templates.

4) Finally, the CoAct data framework and especially the personal data framework has been carefully described in D9.3, which was used as a starting point for the preparation of this deliverable. In this deliverable, the measures taken to protect the research participants, most of them in a vulnerable situation, are also described.

8. Other procedures for data management that you are using

Specific security policies determined by the research institutions involved (UB, UNIVIE, ZSI, UNSAM, FARN) apply to the professional devices used.