



Co-designing Citizen Social Science for Collective Action

## #1.4

### FINAL DATA MANAGEMENT PLAN



This document is shared under Creative Commons Attribution 4.0 International License (**CC BY 4.0**).

Cite as: Lombion, Cédric (2021). CoActD1.3: Data Management Plan. Zenodo. <http://doi.org/10.5281/zenodo.7650029>

*DISCLAIMER: The present Project Deliverable has been submitted to the European Commission for review. The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.*



The CoAct project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No. 873048

**Date** – December 31, 2022

**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	2022-12-21	Cédric Lombion	Shared with reviewers
1.1	2022-12-23	Cédric Lombion, Valeria Arza, Teresa Wintersteller, Josep Perello, Isabelle Bonhoure, Guillermina Actis, Veronika Wöhrer	Revisions
1.2	2022-12-29	Cédric Lombion	Validation of suggested edits



## Table of Contents

### 1. Executive summary

### 2. Introduction

2.1. The CoAct project and the Data Management Plan

2.2. Relevant Works Packages

2.3. Definitions

2.3.1. CoAct-specific terms

2.3.2 DMP terms

2.4. The CoAct Project lifecycle

### 3. Data Summary

3.1. State the purpose of the data collection/generation

3.2. Explain the relation to the objectives of the project

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

3.3.1. Types of data (summary)

3.3.2. Full Description of data types and formats

3.3.3. Data used to support communication with stakeholders

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

3.5. Specify the origin of the data

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

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

### 4. FAIR data

4.1 Making data findable, including provisions for metadata

4.1.1. Outline the discoverability of data (metadata provision)



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

4.1.3. Outline naming conventions used

4.1.4. Outline the approach towards search keyword

4.1.5. Outline the approach for clear versioning

4.1.6. Specify standards for metadata creation (if any)

#### 4.2 Making data openly accessible

4.2.1 Specify which data will be made openly available

4.2.2. Specify how the data will be made available

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

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

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

#### 4.3 Making data interoperable

4.3.1. Assess the interoperability of your data

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

#### 4.4 Increase data re-use

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

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

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

4.4.4. Describe data quality assurance processes

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

## 5. Allocation of Resources

5.1. Estimate the costs for making your data FAIR.

5.2. Clearly identify responsibilities for data management in your project

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



## **6. Data Security**

## **7. Ethical Aspects**

## **8. Other procedures for data management that you are using**



## List of figures

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

## List of tables

- Table 1. CoAct objectives
- Table 2. Summary of Types of data collected in CoAct , by WPs and Research phases.
- Table 3. Summary of types of data formats encountered across the CoAct WPs and Research Phases
- Table 4. Extended description of data types and formats encountered across the CoAct WPs and Research Phases
- Table 5. Summary of Data management measures through the lens of data security

## 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 was 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 project 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](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 M36 revision. It aims to:

- Document the changes in the data workflow across WP3, 4, 5 and 7 which occurred since M18 and the previous version of the Data Management Plan (<http://doi.org/10.5281/zenodo.6078330>)
- Produce the final version of the Data Management Plan

From a data lifecycle point of view, CoAct’s WPs can be split into two groups:

- 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

The main change factor between M18 and M36 has been the completion of the last phases of the research cycle by the Research and Innovation teams, which led to the collection of more data for certain teams. Additionally, the WP6 team (OKF and FHP) oversaw research projects delivered by the Gender Equality Open Call grantees, and FHP conducted some research on the mentoring process.

Overall, these changes are minor from a data lifecycle point of view, and this final revision of the Data Management Plan mostly confirms what had been stated before.

The main changes to the document are summarized below:



Updated section	Update summary
2	Text updated to remove the sections that were only relevant when the DMP was not complete
3.3.1	Table updated to include the final status of the data review
3.3.2	3.3.2 and 3.3.3 have been merged to only include the table with the full description of data types and formats
3.3.3	Text updated to reflect completed activities
3.5	Added detailed data sources
3.6	Updated content with known data size
4	All sections have been updated with the most recent information
5	Text updated to reflect completed activities





## 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>) recommended in the guideline document.

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) sharing a social concern and academic researchers. Such an approach makes it possible 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 Employment, 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 challenges of early school leavers (WP4), to crowdsource testimonies about environmental issues (WP5) or to evaluate the outputs and outcomes of the research activities (WP7). 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. Their names may be disclosed any other sort of publication (communication and dissemination) only if they express their willingness.

**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' events, social media or local digital and non-digital press. They may have their name disclosed in any sort of publication (communication and dissemination) only if they express their willingness.

**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.

### 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. Whenever personal or sensitive data has been collected, regardless of the amount, a 'yes' was used in the summary table. 'no' was used otherwise.

## 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.



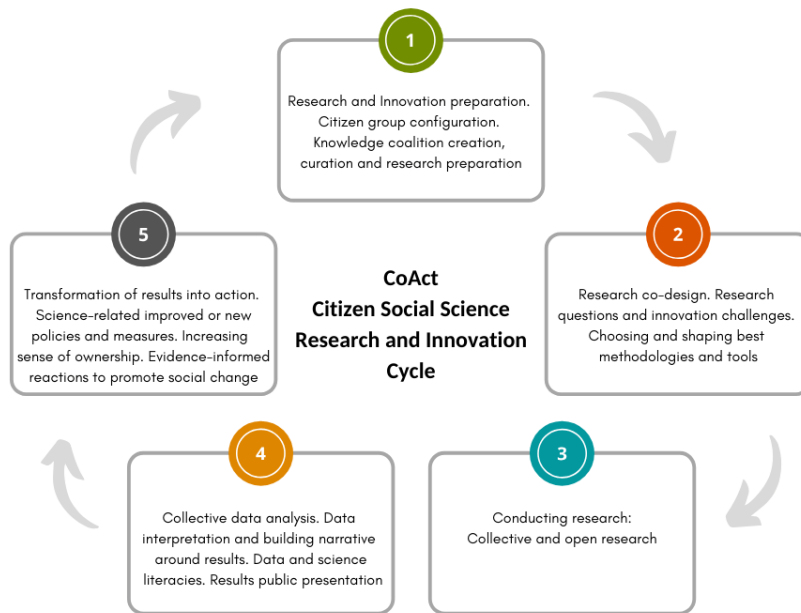


Figure 1: CoAct R&I life cycle for all R&I actions and Research Pilots.



### 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	yes	yes	yes	
	Personal	yes	yes	yes	
	Sensitive	no	no	yes	
Research co-design	non-personal	yes	yes	yes	yes
	Personal	yes	yes	yes	yes
	Sensitive	yes	yes	yes	yes
Collective Research	non-personal	yes	yes	yes	yes
	Personal	yes	yes	yes	yes
	Sensitive	no	yes	no	yes
Collective Data Analysis	non-personal	yes	yes	yes	yes
	Personal	yes	yes	yes	no
	Sensitive	yes	yes	no	no
Transformation of results into action	non-personal	yes	yes	yes	yes
	Personal	yes	yes	yes	yes
	Sensitive	yes	yes	yes	yes

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

### 3.3.2. Data types and formats

Phases	WP3	WP4	WP5	WP7
Research and Innovation Preparation	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>• Non personal data (meetings analysis)</li> <li>• Personal data: Knowledge coalition members</li> </ul>	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>• Non-personal data will be collected as part of the documentation process</li> <li>• Personal data will be collected</li> </ul>	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>• Non-personal data will be collected as part of the documentation process</li> <li>• Personal data: contact</li> </ul>	



	<p>contact info and personal opinions</p> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>Contact info spreadsheets</li> <li>Video recordings</li> <li>Quantitative analysis of meetings stored on CSV files</li> </ul> <p>Text transcripts anonymised</p>	<p>during interviews and to communicate with knowledge coalition members</p> <ul style="list-style-type: none"> <li><u>Data formats</u></li> <li>Audio recordings</li> <li>Text transcriptions</li> <li>Contact Information</li> </ul>	<p>information from knowledge coalition members</p> <ul style="list-style-type: none"> <li>Sensitive data: recordings and text transcripts from group meetings include co-researchers personal or other people’s experiences of environmental injustice.</li> </ul> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>Audio recordings</li> <li>Video recordings</li> <li>Text transcriptions</li> <li>Contact Information spreadsheet</li> </ul>	
<p>Research co-design</p>	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>Non personal data (meetings analysis)</li> <li>Contact info spreadsheets</li> <li>Sensitive data: text or recordings where co-researchers share their personal experiences or diagnosis.</li> </ul> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>Video recordings</li> </ul>	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>Non-personal data will be collected as part of the documentation process</li> <li>Personal data and sensitive data were recorded during interactions with co-researchers: workshop discussions</li> </ul> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>Audio recordings</li> <li>Text transcriptions</li> <li>Memos</li> </ul>	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>Non-personal data will be collected as part of the documentation process</li> <li>Contact information</li> <li>Recordings from workshops group meetings</li> <li>Sensitive data: text and recordings where co-researchers shared personal or other people’s experiences of</li> </ul>	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>Non-personal data provided by other WP teams</li> <li>Interview data with specific co-evaluation questions</li> <li>Audio transcripts from group discussions</li> </ul> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>Digital text data</li> <li>Audio files and transcripts</li> </ul>





	<ul style="list-style-type: none"> <li>• Text transcripts anonymized</li> <li>• Co-Researchers Micro stories</li> </ul>	<ul style="list-style-type: none"> <li>• Pictures with no faces</li> <li>• Contact Information</li> </ul>	<p>environmental injustice.</p> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>• Audio recordings</li> <li>• Video recordings</li> <li>• Text transcriptions</li> <li>• Contact Information spreadsheet</li> <li>• Testimonials</li> </ul>	
Collective Research	<p><u>Types of data</u></p> <p><u>Personal data:</u></p> <p><u>Telegram Identification number</u></p> <p><u>Non personal data:</u></p> <p><u>data collected through the chatbot.</u></p> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>• Database of anonymized chatbot results</li> <li>• Separated data base with Telegram Identification numbers</li> </ul>	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>• Non-personal data will be collected as part of the documentation process and as part of the use of the gamification app for co-researchers</li> <li>• Personal data and sensitive data collected as part of the interaction co-researchers (workshop discussions, roundtables, posters))</li> <li>• Personal data collected during interviews with external experts</li> </ul> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>• Audio recordings</li> <li>• Text transcriptions</li> <li>• Memos</li> </ul>	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>• Non-personal data will be collected as part of the documentation process</li> <li>• Contact information</li> <li>• Recordings from group meetings</li> </ul> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>• Audio recordings</li> <li>• Video recordings</li> <li>• Text transcriptions</li> <li>• Contact Information spreadsheet</li> </ul>	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>• Non-personal data provided by other WP teams</li> <li>• Interview data with specific co-evaluation questions</li> <li>• Audio transcripts from group discussions</li> <li>• Data collected through online whiteboard platforms like Miro and Padlet</li> </ul> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>• Digital text data</li> <li>• Audio files and transcripts</li> </ul>



		<ul style="list-style-type: none"> <li>• Pictures with no faces</li> </ul>		
Collective Data Analysis/Interpretation	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>• Non-personal data will be generated, analysed and preserved</li> <li>• Personal data and sensitive data (personal opinions of Co-Researchers) will be anonymised</li> </ul> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>• Video and audio recordings</li> <li>• Quantitative analysis of meetings stored on CSV files</li> <li>• Text transcripts anonymized</li> <li>• Pictures with no faces</li> <li>• Digitised manual meeting notes</li> </ul>	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>• Non-personal data will be generated, analysed and preserved</li> <li>• Anonymized data was used for analysis</li> <li>• Personal and sensitive data was collected during the workshops</li> </ul> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>• Audio recordings</li> <li>• Text transcriptions and analysis outputs</li> <li>• Posters from collaborative analysis</li> <li>• Graphic Recordings</li> <li>• Videos</li> <li>• Pictures with no faces</li> </ul>	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>• Non-personal data will be collected as part of the documentation process</li> <li>• Contact information</li> <li>• Recordings from group meetings</li> </ul> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>• Audio recordings</li> <li>• Video recordings</li> <li>• Text transcriptions</li> <li>• Contact Information spreadsheet</li> </ul>	<p><u>Types of data</u></p> <p>* Anonymised and summarised data shared from other WP teams</p> <p><u>Data formats</u></p> <p>Digital text data</p>
Transformation of results into action	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>• Personal opinions of co-researchers (personal and sensitive data)</li> <li>• Non-personal data: actions proposals</li> </ul> <p><u>Data formats</u></p>	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>• Non-personal data and anonymised data was shared as part of the presentation of research results</li> <li>• Personal data or sensitive data was shared by</li> </ul>	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>• Non-personal data will be collected as part of the documentation process</li> <li>• Contact information</li> <li>• Recordings from group meetings</li> </ul>	<p><u>Types of data</u></p> <ul style="list-style-type: none"> <li>• Non-personal data provided by other WP teams in the forms of brochures and text materials</li> <li>• Interview data with specific co-evaluation questions</li> </ul>



	<ul style="list-style-type: none"> <li>• Pictures</li> <li>• Video and audio recordings</li> <li>• Text transcripts anonymized</li> <li>• Quotes</li> </ul>	<p>young co-researchers during workshops with decision-makers</p> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>• Audio recordings</li> <li>• Audio transcripts</li> <li>• Pictures with no faces</li> <li>•</li> </ul>	<p>and personal interviews</p> <ul style="list-style-type: none"> <li>• Sensitive data_ recordings and text transcripts of interviews and policy workshops.</li> </ul> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>• Audio recordings</li> <li>• Video recordings</li> <li>• Text transcriptions</li> <li>• Contact Information spreadsheet</li> </ul>	<ul style="list-style-type: none"> <li>• Audio transcripts from group discussions</li> <li>• Data collected through online whiteboard platforms like Miro and Padlet</li> </ul> <p><u>Data formats</u></p> <ul style="list-style-type: none"> <li>• Digital text data</li> <li>• Digitized materials</li> </ul>
--	---	---	--	---

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

Each activity where sensitive data could be recorded included an informed consent procedure, as detailed in the report on the Informed consent procedures and requirements (<https://zenodo.org/record/6078642>)

### 3.3.3. Data used to support communication with stakeholders

All the WP ended up collecting contact information to communicate with their stakeholders. For WP2, WP6 and WP8, this represents the only type of personal data collected.

In WP2, the capacity and community building actions encompassed 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. They were collected following the GDPR rules and an explicit agreement was always sought.

As part of WP6 several Open Calls were organised, leading to the collection of contact information from the applicants. This data was deleted after the completion of the Open Calls. The final Open Call



grantees did collect more data as part of their work, but this data was not accessible to the CoAct consortium members. Nonetheless, the WP6 team reviewed the Ethics and Personal Data Protection processes of each grantee's project to ensure that they aligned with CoAct policies.

As part of WP8, contact details of persons interested in CoAct communication and dissemination outputs as well as by CoAct own events were 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 were collected following GDPR standards. They are safely stored for the duration of the CoAct project and will be destroyed at the end of the project.

For WP3, WP4, WP5, sensitive data, in the form of testimonies, has a chance to be included as part of the communication to stakeholders, but only after express approval from the individual whose testimony was recorded.

### 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 was used in WP3 and WP5
- Data collected in the "Main Research Activities" phase of WP5: the crowdsourced data includes environmental data.

### 3.5. Specify the origin of the data

For WP3, the external data sources used are listed below:

- Survey by the Catalan Generalitat: Enquesta salut Catalunya (2021); Report/Survey Questions/Data Metadata: [https://salutweb.gencat.cat/ca/el\\_departament/estadistiques\\_sanitaries/enquestes/esca/resultats\\_enquesta\\_salut\\_catalunya/](https://salutweb.gencat.cat/ca/el_departament/estadistiques_sanitaries/enquestes/esca/resultats_enquesta_salut_catalunya/)
- Survey by Barcelona Health Agency (L'Agència de Salut Pública de Barcelona (ASPB)): Encuesta salud Barcelona (2020-2022): <https://www.aspb.cat/documents/enquesta-salut-barcelona/>



- Survey done by Obertament: " L'ESTIGMA I LA DISCRIMINACIÓ EN SALUT MENTAL a Catalunya 2016". (Stigma and discrimination in mental health in Catalonia, 2016) <https://obertament.org/images/Presentaci%C3%B3%20Acte/Estigma%20a%20Catalunya%202016.pdf>
- List of municipalities (with number of inhabitants and area) of Catalunya, Statistics Institute of Catalunya. <https://www.idescat.cat/pub/?id=aec&n=250>
- Population of Catalunya per municipality, age bracket, and sex, Dades obertes Catalunya (Open Data Portal of the Generalitat of Catalunya), <https://analisi.transparenciacatalunya.cat/en/Demografia/Poblaci-de-Catalunya-per-municipi-rang-d-edat-i-se/b4rr-d25b>
- Map of Spain with political division: <http://centrodedescargas.cnig.es/CentroDescargas/index.jsp> , BCN500, Descripción: Base Cartográfica Nacional a escala 1:500.000
- Gini coefficients of 63 Zooniverse projects according to the number of classification tasks accomplished per participant, Table 3 in Spiers, H., Swanson, A., Fortson, L., Simmons, B. D., Trouille, L., Blickhan, S. and Lintott, C. (2019). 'Everyone counts? Design considerations in online citizen science'. JCOM 18 (01), A04. <https://doi.org/10.22323/2.18010204> .

For WP5, public environmental data was used as follows:

- Hydrographic sub-basins within the Matanza-Riachuelo basin. Source: ACUMAR. Available at <https://mapas.acumar.gob.ar/datos/categoria/1>
- Water courses of the Matanza-Riachuelo basin. Source: ACUMAR. Available at <https://mapas.acumar.gob.ar/datos/categoria/1>
- Hydrographic sub-basins of Buenos Aires Province. Source: Provincial Authority of Water (ADA, from its acronym in Spanish). Available at <https://gis.ada.gba.gov.ar/gis/>
- Protected areas in the Matanza-Riachuelo basin. Source: ACUMAR. Available at <https://mapas.acumar.gob.ar/datos/categoria/1>
- Protected areas in Buenos Aires Province. Source: Integrated System of Environmental Information (SINIA, from its acronym in Spanish). Available at <https://ciam.ambiente.gob.ar/repositorio.php?tid=6#>
- Popular neighbourhoods registered in the country: Source: National Registry of Popular Neighbourhoods (RENABAP for its acronym in Spanish). Available at <https://datos.gob.ar/dataset/desarrollo-social-registro-nacional-barrios-populares>



- Shantytowns and “informal” settlements of the Matanza-Riachuelo basin. Source ACUMAR. Available at <https://mapas.acumar.gob.ar/datos/categoria/2>
- New buildings for social houses associated with the resettlement process in Matanza-Riachuelo basin. Source ACUMAR. Available at <https://mapas.acumar.gob.ar/datos/categoria/4>
- Work tables for citizen participation in the resettlement process in Matanza-Riachuelo basin. Source ACUMAR. Available at <https://mapas.acumar.gob.ar/datos/categoria/2>

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

The final size of the research data is not expected to go over 100 GB by the end of the project, with a current estimate across WPs at 40 GB mainly due to audio and video recordings obtained as part of WP3 (20 GB), WP4 (10 GB) and WP5 (50 GB).

### 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)



The only WPs expected to publish quantitative data will be WP3 and WP5. The WP4 team shared qualitative data from their collaborative analysis workshops.

Discoverability of the data will be made possible through:

- The publication of data on open repositories
- Consistent naming and versioning of files
- 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

All published documents follow the pattern of:

- Deliverable number
- Deliverable title.

#### 4.1.4. Outline the approach towards search keyword

The most used keywords are: citizen science, citizen social science, participation, participatory methods, collaboration, CoAct, co-creation. Each WP team also uses keywords specific to the nature of their published documents and data.

#### 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.



## 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. Only two WPs will have data to publish:

- WP3: anonymised chatbot database
- WP4: anonymised material from collaborative data analysis
- WP5: crowdsourced data

### 4.2.2. Specify how the data will be made available

All relevant research data will be made available on or linked to from the Zenodo open research platform.

WP3: the anonymous chatbot database will be published in Zenodo as well as a preprint of the data descriptor, that is planned to be submitted for publication in a Data Descriptor Open Access journal, (e.g., Nature Scientific Data).

WP4: The anonymised materials will be made available on Zenodo

WP5: The anonymised crowdsourced data will be published on Zenodo.

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

Research data will be published in standard formats such as CSV or Python language meaning that any standard software will be able to access it, including open-source software. Specifically, WP3 uses the Python-based Pickle database format.

### 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 for all WPs with the exception of WP3. GitHub will be used to deposit the source code of the platforms developed as part of WP3 and WP5.

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

In the case of WP4, anonymised transcripts cannot be published due to the risk of de-anonymisation. A contact information will be provided to allow other researchers to directly request the transcripts if needed.

WP3 and WP5 will publish the data by the end of the project but will make it accessible after an embargo of a few months.

### 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.

#### 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 WP3 team used several standards when designing its database:

- Timestamp standard ISO 8601, time difference in seconds (float);
- Spanish postal codes: 5-digit string, according to Sociedad Estatal de Correos y Telégrafos 1984;
- Country codes: ISO 3166 international standard: alpha-2 code

The WP5 team took inspiration from several standards to create its own data taxonomy

Both teams otherwise used standard vocabulary used in database management and data processing.

### 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. All CoAct documents are published under a Creative Commons Public Domain Dedication (CC0), in accordance with the OpenAire guidelines (<https://www.openaire.eu/research-data-how-to-license/>), unless otherwise stated. The exceptions are listed below:

#### WP3

- The database will be published under an ODBL licence
- The code source will be published under a GPL-3 licence
- The research content will be published under a Creative Commons SA-BY-NC licence

#### WP4

- The pictures and design used in the German report for policymakers (<https://zenodo.org/record/7390893#.Y6QtmOKZOE>) are not reusable under an open-source licence
- Graphic recordings and graphic recording videos are not reusable under an open-source licence

#### WP5

- The database will be published under an ODBL licence
- The code source will be published under a GPL-3 licence
- The videos and research content will be published under a Creative Commons BY-SA 4.0 licence

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

All the publishable research data produced during the project will be published under embargo before the end of the project, in March 2023. The embargo will last a few months.

More data will be published after the end of the project as the platforms built by the WP3 and WP5 teams will continue to operate.



#### 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 data will be freely reusable once the embargo is lifted.

#### 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 takes place during the data analysis, when each team checks that the data used for the analysis is correct.

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

There is no time restriction on the reuse of the data.

## 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.

### 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.

## 5. 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.

Included in the table below are the WPs which dealt with personal and sensitive data.

Activities	WP3	WP4	WP5	WP7
<b>Short term storage</b>	<p>Digital data</p> <ul style="list-style-type: none"> <li>• Portable computers of the University of Barcelona (UB)</li> </ul>	<p>Digital data</p> <ul style="list-style-type: none"> <li>• Professional computers of the University of Vienna (UNIVIE)</li> </ul>	<p>Digital data</p> <ul style="list-style-type: none"> <li>• Some data can be stored on personal computers and phones</li> </ul>	<p>Digital data</p> <ul style="list-style-type: none"> <li>• Computers of the Zentrum für Soziale Innovation</li> </ul>



	<p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>Key-locked storage in the office of Prof Josep Perelló.</li> </ul>	<ul style="list-style-type: none"> <li>Professional Audio recorders of UNIVIE</li> <li>ActionBound platform servers. The data is reviewed systematically to identify and remove sensitive information</li> </ul> <p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>Key-locked storage in the project office</li> </ul>	<ul style="list-style-type: none"> <li>Also served in the University's OneDrive</li> <li>Digitized notes</li> </ul> <p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>No</li> </ul>	<p>(Centre for Social Innovation, ZSI)</p> <p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>None</li> </ul>
<b>Long term storage</b>	<p><u>Digital data</u></p> <ul style="list-style-type: none"> <li>Servers of the University of Barcelona</li> </ul> <p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>Key-locked storage in the office of Prof Josep Perelló.</li> </ul>	<p><u>Digital data</u></p> <ul style="list-style-type: none"> <li>Dedicated storage space on the servers of UNIVIE</li> </ul> <p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>Key-locked storage in the office of Prof. Veronika Wöhrer</li> </ul>	<p><u>Digital data</u></p> <ul style="list-style-type: none"> <li>University's OneDrive storage</li> <li>FARN storage servers</li> </ul> <p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>No</li> </ul>	<p><u>Digital data</u></p> <ul style="list-style-type: none"> <li>Servers of the ZSI</li> </ul> <p><u>Paper data</u></p> <p>1) None</p>
<b>Storage of personal and sensitive data</b>	<p><u>Digital data</u></p> <p>De-anonymisable data is stored on university servers and will be deleted within one year</p> <p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>No paper data</li> </ul>	<p><u>Digital data</u></p> <ul style="list-style-type: none"> <li>In the dedicated server space, two folders: one for anonymized data used for analysis, and one for non-anonymized data for GDPR compliance purposes.</li> </ul> <p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>Some data was recorded on</li> </ul>	<p><u>Digital data</u></p> <ul style="list-style-type: none"> <li>Same as above</li> </ul> <p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>None</li> </ul>	<p><u>Digital data</u></p> <ul style="list-style-type: none"> <li>Audio and video files are deleted by the end of the project, and only transcripts are kept.</li> <li>Co-evaluation survey results stored on the ZSI secure storage</li> </ul> <p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>No paper data</li> </ul>



		paper and stored in the key-locked storage in the office of Prof. Veronika Wöhrer		
<b>Transfer of personal and sensitive data</b>	<p><u>Digital data</u></p> <ul style="list-style-type: none"> <li>Personal and sensitive data collected by the researchers themselves are transferred directly from their professional devices to the UB servers using their institutional account on the GDPR compliant Microsoft OneDrive Cloud computing service.</li> <li>Personal data or sensitive data is transferred through a Telegram platform-hosted chatbot. The database only stored the anonymised data.</li> </ul> <p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>No paper data</li> </ul>	<p><u>Digital data</u></p> <ul style="list-style-type: none"> <li>Personal and sensitive data collected by UNIVIE researchers themselves are transferred directly from their professional devices to the UNIVIE servers.</li> <li>Data collected by co-researchers is automatically uploaded to the ActionBound servers. Then the UNIVIE team downloads the data to transfer it to its own servers.</li> </ul> <p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>Paper data was immediately digitized and stored on UNIVIE servers.</li> </ul>	<p><u>Digital data</u></p> <ul style="list-style-type: none"> <li>Anonymized transfer to ZSI</li> </ul> <p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>No</li> </ul>	<p><u>Digital data</u></p> <ul style="list-style-type: none"> <li>No transfer of personal and sensitive data outside of CoAct consortium</li> <li>Data was communicated only with people that were present at group discussions or co-evaluation workshops where the data was relevant</li> <li>The controller of the data was always ZSI, and the data didn't change hands</li> </ul> <p><u>Paper data</u></p> <ul style="list-style-type: none"> <li>No paper data</li> </ul>

Table 4: Summary of Data management measures through the lens of data security

## 6. 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.

## 7. 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.

