



TIME4CS

SUPPORTING SUSTAINABLE INSTITUTIONAL CHANGES TO
PROMOTE CITIZEN SCIENCE IN SCIENCE AND TECHNOLOGY

D7.4 First version of the Data Management Plan



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006201

Work Package	Work Package 7
Task	Task 7.3
Due Date	30/04/2021
Submission Date	29/04/2021
Author(s)	Matteo Di Rosa, Claudia Iasillo and Karolina Jurkiewicz (APRE)
Contributor(s)	Muki Haklay (UCL), Eugenia Vilarchao (ESF), Gitte Kragh (AU), Teresa Schaefer (ZSI), Elena Maffia (UniSR), Andreea Petrea (CHX)
Reviewer(s)	Muki Haklay (UCL), Eugenia Vilarchao (ESF), Andreea Petrea (CHX)
Version	1.0
Dissemination Level	Public
DOI	10.5281/zenodo.4726899

Document Revision History

Version	Date	Description of change	List of contributor(s)
0.1	15/04/2021	First draft of the deliverable	Matteo Di Rosa, Claudia Iasillo and Karolina Jurkiewicz (APRE)
0.2	26/04/2021	Minor changes and suggestions on the text	Muki Haklay (UCL), Eugenia Vilarchao (ESF), Gitte Kragh (AU), Teresa Schaefer (ZSI), Elena Maffia (UniSR), Andreea Petrea (CHX)
1.0	29/04/2021	Final version	Matteo Di Rosa, Claudia Iasillo (APRE)



Disclaimer: 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 herein.



Table of Contents

EXECUTIVE SUMMARY	5
1. OVERVIEW OF THE PROJECT	6
2. THE DMP IN THE OVERALL TIME4CS APPROACH	7
2.1 OBJECTIVES OF THE DATA MANAGEMENT PLAN.....	7
2.2 THE OVERALL STRUCTURE OF THE DATA MANAGEMENT IN TIME4CS.....	7
2.3 DATA MANAGEMENT DURING TIME4CS LIFETIME	8
2.4 LONG-TERM DATA MANAGEMENT	8
3. DATA COLLECTION	9
3.1 DATA INFORMATION.....	9
3.2 ORGANIZATION OF DATA COLLECTED.....	10
3.3 DATA COLLECTED FOR EACH TIME4CS WORK PACKAGE	11
WORK PACKAGE 1	11
WORK PACKAGE 2	13
WORK PACKAGE 3	15
WORK PACKAGE 4	17
WORK PACKAGE 5	19
WORK PACKAGE 6	21
WORK PACKAGE 7	23
4. DATA STORAGE AND BACKUP.....	25
5. DATA DOCUMENTATION	26
6. DATA ACCESS.....	28
7. DATA SHARING AND REUSE.....	29
8. DATA PRESERVATION AND ARCHIVING	30
9. PRIVACY OF PARTICIPANTS	31
LIST OF REFERENCES	34



Executive Summary

This deliverable describes the preliminary version of the Data Management Plan (DMP) for the TIME4CS project. This document provides an initial summary of the main elements to take into account in the definition of the TIME4CS data management policy to be used by project partners throughout the project activities. The DMP will provide indications about the management of all the data generated during the project activities, including how data will be collected, managed, stored and made available during the project and how they will be shared upon TIME4CS completion.

The DMP is a living document that will be updated during the project lifetime according to the project's progresses and rising needs. The final version of TIME4CS DMP will be provided at the end of the project (M36) and not only will describe the procedures applied, but it will provide more details about the long-term management and preservation of project data.



1. Overview of the project

TIME4CS – Supporting sustainable Institutional Changes to promote Citizen Science in Science and Technology – is an H2020 project funded under the Science with and for Society Work Programme in the call SwafS-23-2020 “Grounding RRI in society with a focus on citizen science”.

The project aims at **supporting and facilitating the implementation of sustainable Institutional Changes in Research Performing Organizations (RPOs)** to promote Public Engagement and Citizen Science in science and technology. Institutional Changes can be defined as a type of change triggered in an organization and characterized by being *irreversible, comprehensive, inclusive and contextualized*. TIME4CS does not aim at exploring Institutional Changes as a general phenomenon but the project focuses on the specific Institutional Changes needed to embed Citizen Science in RPOs. Therefore, TIME4CS has identified 4 **Intervention Areas (IAs)** that alone or combined can stimulate the Institutional Changes necessary to promote Citizen Science in Research and Innovation processes: 1) *Research*, 2) *Education and Awareness*, 3) *Support resources and Infrastructure* and 4) *Policy and Assessment*. TIME4CS analyzes these areas to consolidate the knowledge about the institutional adoption, establishment and maintenance of Citizen Science capacity and to establish a model for Citizen Science expansion through Institutional Changes.

For each Intervention Area, TIME4CS has identified an organization that has already undergone some Institutional Changes to promote Citizen Science showing therefore a comprehensive knowledge and expertise. The knowledge of these partners, called **Front-Runners**, will contribute to the definition of a set of practical actions aimed at paving the way to Institutional Changes, defined **Grounding Actions**. Front-Runners will interact with and mentor four organizations willing to face the challenge of introducing Citizen Science more and more in their structures. These organizations, defined **Implementers**, within TIME4CS lifetime will develop tailored roadmaps including a specific set of Grounding Actions to carry out, benefitting of the constant support of Front-Runners. The interaction between Front-Runners and Implementers will take place during the whole TIME4CS lifetime through the development of a **mutual learning and knowledge exchange** framework designed to support the implementation and the evaluation of the implementation of grounding actions leading to institutional changes with the ultimate goal to encourage public engagement in science and technology.

2. The DMP in the overall TIME4CS approach

2.1 Objectives of the Data Management Plan

Accessibility of the data during the project lifetime and beyond, and long-term data preservation are the ultimate goals of a successful data management, which is therefore an essential aspect of each research project, and it must be thoroughly planned since the early stages of the project. This deliverable aims at describing the overall data management process within TIME4CS: how data will be collected, managed, stored and made available during the project, and how they will be shared upon TIME4CS completion. It is worth to point out that a proper data management plan can mitigate the risk of data loss or other threats that could render the data illegible or unusable (e.g. the obsolescence of software).

Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process¹. For this reason, all data collected and generated during the TIME4CS project will be managed following the FAIR principles: Findable, Accessible, Interoperable, and Reusable.

The management of the data produced during the project will aim to ensure open access, as stipulated in Article 29.3 of the Grant Agreement, taking also into consideration the “as open as possible, as closed as necessary” principle, to ensure the adoption of suitable measures to preserve project results that can be further commercially exploited in the future (e.g. through patenting). Thus, the project will seek a balance between openness and protection of information, commercialization and Intellectual Property Rights (IPR), privacy concerns, security etc. In case project partners need to keep specific parts of their research data closed, the choice will be explicitly justified in the DMP.

2.2 The overall structure of the Data Management in TIME4CS

This deliverable firstly describes the Data Collection within TIME4CS, presented in Section 3. This section describes the methods of data collection for each work package, as indicated by the work package leaders. For each work package, the following items are checked and described, as far as applicable: type of data, aim of the data, formats and estimated size. The Data Collection section is essential to give a brief description of the data, including any existing data or third-party sources that will be used, providing information about its content, type and coverage.

The Data Storage and Backup section (see Section 4) will then describe where and how the data generated by the project activities will be stored. Information about the IT infrastructure used will be also provided.

The Data documentation section (see Section 5) will detail how data will be documented to ensure that new members of the TIME4CS team and/or any possible (secondary) users are able to understand and reuse the

¹ Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

data. Additional information will be provided about the metadata used to provide indications about the identification and access to the data.

The Data access section (see Section 6) will mainly describe the level of data confidentiality and the responsibility of the individual organizations, within TIME4CS, involved in the data collection process while the Data Sharing and reuse section (see Section 7) will describe how the data will be shared and where they will be deposited.

Preserving all the material produced by the project will be a challenge. The data preservation and archiving section of the DMP (see Section 8) will therefore explain the general rules adopted by the project in order to preserve the main part of data generated.

Finally, Section 9 will detail the procedures to ensure the privacy of Participants to TIME4CS activities.

2.3 Data management during TIME4CS lifetime

The DMP will be updated at least once a year. APRE will be responsible for coordinating the data management with the active support of Work Package Leaders and Task Leaders. The content of the data management plan will be reanalyzed and enriched in each update of the deliverable according to the project implementation progresses.

2.4 Long-term data management

To be useful, findable and reusable over the long-term, data will be stored on ZENODO² that will automatically assign a persistent identifier, namely a DOI, to datasets to allow easy citation and discoverability. By archiving on ZENODO, the long-term preservation will be guaranteed, avoiding the risk of data loss. More information about the long-term management of data internal to TIME4CS consortium will be provided in the next updates of the Data Management Plan.

² <https://zenodo.org/>

3. Data Collection

3.1 Data information

During the TIME4CS project, various types of data will be collected. This section describes the methods of data collection for each work package as indicated by the work package leaders. For each work package, the following items are checked and described, as far as applicable.

Type of Data

In research projects, the following type of data may be collected and documented:

- **Observational:** data captured through observation of an activity and/or a behavior in real-time (and, therefore, usually irreplaceable). The collection usually may require human observation, surveys and instruments to record the information.
- **Experimental:** data obtained in controlled situation through the active intervention of researchers, who design an activity (called experiment) to collect data. They can be reproduced if needed, but it may be expensive.
- **Simulation:** data generated reproducing a real-world system, usually used to determine what would happen in certain conditions.
- **Derived or Compiled:** data derived upon transformation of existing data sets to create new data. They can be reproduced if lost, but it is very time-consuming and it could be expensive.
- **Reference or Canonical:** it refers to data used to categorize and/or classify other data. They could be static or changing over time.
- **Event-related:** data collected during events. As they are real-time events, they are irreplaceable. They can include personal data of participants, and GDPR will be applied.

It is worth to point out that these categories can overlap, and data can belong to more than one category depending on the specificities of the data itself (e.g. event-related data could also be considered observational data upon certain conditions).

Aim of the data and the process of data collection

For each Work Package a short description of the aim of the data produced linked to the objectives of the Work Package is provided, together with a short description of the process of the data collection that will be applied.

Format of the data and estimated size

Research data come in various formats: text, numeric, multimedia, models, software languages, discipline and instrument specific, etc. For each WP, it will be described the format of the data produced, with the ultimate goal to facilitate the accessibility and interoperability of the data, giving preferences to open and

standard formats. In this deliverable, a first estimation of the size of the data collected in each WP will also be provided. This information will be updated in the next versions of TIME4CS DMP according to the actual size of data collected.

3.2 Organization of data collected

The data collected within TIME4CS will be organized taking into consideration the following aspects, as far as applicable.

Version Control

To overcome the challenge of managing and tracking research materials during the course of research, especially in collaborative projects, data will be organized to keep track of different versions of the datasets, through the application of a version control system. A version control system (or revision control system) is a system that tracks incremental versions (or revisions) of files and, in some cases, directories over time. Improving the ability to consistently track and retrieve each version of a file can lead to more efficient collaboration and increased accuracy of research results. Through the version control system, the risk of losing information after modifications is minimized.

Versioning is important for long term-research data management where metadata and/or files are updated over time. For scientific projects wherein different parties are involved, it is advised to utilize a data storage system with version control function to prevent losing of data.

Naming conventions used

With the overall aim to make the data accessible (and ultimately foster reproducible research), TIME4CS datasets will be carefully named by choosing file names that are informative and useful for both humans and machines. Name should be meaningful so to make easier for others to understand what the file contains and how it should be used. Following, there are some suggestions about the naming conventions within TIME4CS:

Choose machine readable names

Use deliberate delimits. Common approach is using “_” and “-” to delimit units of metadata in the file names. A general rule could be to use “-” to separate words you want to glob together and “_” to separate different information within a file name. Don’t use spaces, punctuation, capital letters or special characters (Using \$, @, %, #, &, *, (,), !, etc. may have meanings in programming languages).

Choose human readable names

Choose names that explain the content. The more meaningful the name, the more useful it is for human users. The more metadata you store in the name, the less you need to explain elsewhere. Choose short names.

3.3 Data collected for each TIME4CS Work Package

Work Package 1

The information was collected by the Work Package leader (Muki Haklay/ UCL) through the use of data collection process that is based on templates and data collection forms. The results are summarized below.

Pre-Existing Data

Existing reports, web-pages, and openly available information in scientific publication or news reports and press releases

Source of Pre-Existing data

Open material

Conditions for use of pre-existing data

Each material that will be collected will be checked for permission to use and will be credited.

Type of data collected

- Observational data
- Experimental data
- Simulation data
- Derived or compiled data
- Reference or canonical data
- Event related data

Aim of this data

The overall objective of WP1 is to map and extract knowledge from successful institutional adoption of Citizen Science in order to codify best practices and lessons learnt as basis for the roadmaps developed in WP2. In particular, this process will contribute: i) to expand the knowledge related to Intervention Areas and Grounding Actions, ii) to build the knowledge base of the overall project, iii) to support the capacity building activities, iv) to support the mutual learning programme by providing content (WP3).

Process of data collection

The data will be collected with a specific data collection methodology that will be developed for the project. Where relevant, data about the project at the institution will follow the PPSR-core framework³. All the case

³ <https://core.citizenscience.org/>

studies will follow the same template, and an annotated empty template will be created as both a reference and documentation of the decisions that were made by the project team and to provide an explanation to the information that will be shared. The truth table will be provided with an explanation to the decisions and ranking that are provided for each case study. The template will include a version number, and with every adjustment, the version number will be updated. The case studies will be peer reviewed by another member of the team and the truth table will be created in a workshop and will be quality assured by members of the consortium.

Data will be collected as much detailed as possible on the Front Runner actions thought interviews, presentation and personal communications.

Format of data collected

- ✓ .XLM
- ✓ .DOC / .DOCX
- ✓ .XLS / .XLSX
- ✓ .PPT / .PPTX
- ✓ .PDF
- ✓ Photo Format (e.g., .JPEG, .PNG)
- ✓ Video Format (e.g., .MKV, .MP4, .AVI)

Estimated size of data

1.5 GB

Work Package 2

The information was collected by the Work Package leader (Eugenia Vilarchao/ ESF) through the use of questionnaires. The results are summarized below.

Pre-Existing Data

Data collected in WP1 and WP2

Source of Pre-Existing data

N/A

Conditions for use of pre-existing data

Grant Agreement

Type of data collected

- Observational data
- Experimental data
- Simulation data
- Derived or compiled data
- Reference or canonical data
- Event related data

Aim of this data

WP2 is primarily aimed at developing a roadmap framework supporting the achievement of sustainable institutional changes that assure the responsible involvement of society in science and innovation activities. In this view, it will first focus on the development of a roadmap framework that serves as a basis for structuring specific, need-based institutional roadmaps shape-cut for individual organizations. Then, the roadmap framework will be tailored for the 4 consortium Implementers to assure the attainment of sustainable institutional changes via a series of well-defined, timely and achievable short-term (project lifespan) and long-term (sustainable) goals.

Process of data collection

In order to elaborate individual roadmap framework leading to institutional change data about the specific institutional needs will be collected. It will be shaped according to the mapping process and state-of-the-art analysis on the policies and best practices carried out in WP1 (T1.1, T1.2).

Other data will be collected during the implementation of Grounding Actions. Each Implementer as documented and planned in own roadmap will implement at least 4 Grounding Actions. At the time of

drafting the first version of the DMP, it is not possible to design what kind of data will be collected and how. More details will be provided in the next DMP update.

Format of data collected

- ✓ .DOC / .DOCX
- ✓ .PPT / .PPTX
- ✓ .PDF
- ✓ .XLS / .XLSX
- ✓ Photo Format (e.g., .JPEG, .PNG)
- ✓ Video Format (e.g., .MKV, .MP4, .AVI)

Estimated size of data

300MB

Work Package 3

The information was collected by the Work Package leader (Claudia lasillo / APRE) through the use of questionnaires. The results are summarized below.

Pre-Existing Data

N/A

Source of Pre-Existing data

N/A

Conditions for use of pre-existing data

N/A

Type of data collected

- Observational data
- Experimental data
- Simulation data
- Derived or compiled data
- Reference or canonical data
- Event related data

Aim of this data

WP3 aims to support the implementation of Grounding Actions leading to Institutional Changes and to ensure a continuous knowledge transfer and mutual learning. It will provide the basis for the development of the actions described in the tailored roadmaps of Implementers and will support the knowledge transfer and capacity building between Front-Runners and Implementers through face-to-face interactions and online activities.

Process of data collection

The data collected will be related to the mutual learning events organized during the project activities. For each event, a list of participants will be compiled and reference documents (i.e. presentations and pdf) and video recording will be collected, stored and shared with the consortium. Additional observations data will be collected during all WP activities to shape the final report capitalizing the elements emerging from each mutual learning practice.



Format of data collected

- ✓ .DOC / .DOCX
- ✓ .XLS / .XLSX
- ✓ .PPT / .PPTX
- ✓ .PDF
- ✓ Photo Format (e.g., .JPEG, .PNG)
- ✓ Video Format (e.g., .MKV, .MP4, .AVI)

Estimated size of data

5GB



Work Package 4

The information was collected by the Work Package leader (Kristian Hvidtfelt Nielsen / AU) through the use of questionnaires. The results are summarized below.

Pre-Existing Data

N/A

Source of Pre-Existing data

N/A

Conditions for use of pre-existing data

N/A

Type of data collected

- Observational data
- Experimental data
- Simulation data
- Derived or compiled data
- Reference or canonical data
- Event related data

Aim of this data

WP4 aims at building capacity within institutions for the use of RRI, especially through CS and PES. This WP builds on the work done in WP1, 2 and 3 to enable the institutional and cultural changes needed for CS to be encouraged and supported. WP4 extends this work to the researchers and other staff members directly involved in conducting and supporting CS projects by developing the required capacity to design, execute, support and evaluate CS projects. The capacity building will include raising awareness and knowledge of all aspects of conducting responsible CS projects, as well as integrating knowledge from WP1, WP2 and WP3 on institutional change, roadmap and stakeholders to identify participants' role in these structures, initiatives and networks. The participants of the workshops and webinars will be primed for acting as CS ambassadors at their respective institutions (i.e. one of the GAs).

Process of data collection

Several types of data will be collected during the project activities related to WP4. First of all, data will be collected about existing citizen science training courses to enable extraction and analysis of a list of best practices in terms of content, structure, learning goals, and evaluation. This will be the basis for development

of the training course (D4.1), workshops (D4.2), and webinars (D4.3), and will support the project objective of creating a dynamic and inclusive community through engagement of relevant stakeholders through mobilization through training activities. The data collected will be about existing training courses on citizen science. This will include:

- Title of training course
- Institution/organization offering the course
- Length of training course, and dates offered
- Target audience for the course
- URL to access the training course / information about the training course
- Financing of training course, including course fee, if any
- Details of what is covered in the course
- Structure of training course
- Learning goals of training course
- Evaluation of training course

The data will be mainly text and stored in a spreadsheet for better overview. An analysis of the data will be performed and shared in an appropriate format.

Additionally, during the training course implementation other data will be collected and generated, such as:

- List of participants
- Slides of the presentations
- Video Recording of the interactions (recording of ZOOM or video-recording of site visit presentations)

Format of data collected

- ✓ .DOC / .DOCX
- ✓ .XLS / .XLSX
- ✓ .PPT / .PPTX
- ✓ .PDF
- ✓ Photo Format
- ✓ Video Format

Estimated size of data

5GB

Work Package 5

The information was collected by the Work Package leader (Teresa Schäfer / ZSI) through the use of questionnaires. The results are summarized below.

Pre-Existing Data

N/A

Source of Pre-Existing data

N/A

Conditions for use of pre-existing data

N/A

Type of data collected

- Observational data
- Experimental data
- Simulation data
- Derived or compiled data
- Reference or canonical data
- Event related data

Aim of this data

This WP has three main objectives. First, it aims to jointly assess the progress of implemented institutional changes in the participating organizations. Evaluation will also support TIME4CS understanding on mutual learning effects between Front-Runners and Implementers of the project. Second, this WP will develop a set of indicators to assess institutional change to promote citizens engagement in science along the four intervention areas of research, education & awareness, support infrastructure, and policy & assessment. This indicator set will investigate output, intermediate and long-term outcomes. It will be based on the experiences made from the project's evaluation activities and include latest insights from related research and innovation. Third, based on the developed indicator set and the experiences made with the implementation of data monitoring and collection instruments in the project, it aims to provide RPOs and RFOs with a monitoring toolkit to self-assess and monitor their progress towards organizational change.

Process of data collection



Data will be collected to capture the state-of-the art with regard to citizens' engagement in RPOs from implementers inside and potentially also outside of our consortium and capture changes in implementing institutions due to the roadmaps, Grounding Actions and trainings developed in WP1-WP4.

The data collected will be quantitative and qualitative data from the implementer organizations and individual participants in these organizations, who were engaged in the project activities (e.g. training participants).

Data collection instruments will be defined in more detail during the first 12 months of the project together with the consortium partners, and will be e.g. questionnaires, interviews, focus groups, workshops etc.

- Questionnaires will be set up in the software "LimeSurvey" which is hosted on a ZSI server and GDPR conform; data can be extracted from LimeSurvey as xls.- or csv.files,
- Interviews will be organized face-to-face or online and, if not consented by the interviewee, audio-recorded (e.g. MPEG-4 audio file); audio transcripts will be done in Microsoft Word or Open Office,
- Outcomes from focus groups and workshops will be documented in protocols, done in Microsoft Word or Open Office.

Participants to be enrolled will be exhaustively informed, so that they are able to autonomously decide whether they consent to participate or not. The purposes of the research, the procedures as well as the handling of their data (protection, storage) will be explained to them either face-to-face or in written format. For online interviews these explanations will be part of the initial briefing of interviewees, for face-to-face interventions informed consent shall be agreed and signed by both, the study participants as well as the respective research partner.

The data exploitation will be in line with the respective national data protection acts. The data gathered through questionnaires, interviews, focus groups, workshops and other possible data gathering methods during this research will be anonymised and therefore the data cannot be traced back to the individual. Data will be stored in anonymous form so the identities of the participants will only be known by the research partners involved. Finally, citizen science data reports based on interviews, focus groups and other data gathering methods will be based on aggregated information and comprise anonymous quotations respectively.

Format of data collected

- ✓ .DOC / .DOCX
- ✓ .XLS / .XLSX
- ✓ .PPT / .PPTX
- ✓ .CSV
- ✓ .PDF
- ✓ Photo Format (e.g., .JPEG, .PNG)
- ✓ Video Format (e.g., .MKV, .MP4, .AVI)

Estimated size of data

5GB

Work Package 6

The information was collected by the Work Package leader (Michael Browne/CHX) through the use of questionnaires. The results are summarized below.

Pre-Existing Data

N/A

Source of Pre-Existing data

N/A

Conditions for use of pre-existing data

N/A

Type of data collected

- Observational data
- Experimental data
- Simulation data
- Derived or compiled data
- Reference or canonical data
- Event related data

Aim of this data

WP6 aims at establishing the project's identity and visibility in Europe. It will develop the necessary support to the promotion of project outcomes and will sustain the expected achievements of the project. It structures all dissemination, communication and exploitation processes, both within the project and most importantly, towards the TIME4CS target groups: Citizen Science community, scientists, policy makers, the general public and other stakeholders. It also deals with the need to provide appropriate mechanisms to amplify TIME4CS' impact throughout Europe. Engagement, networking, promotional activities, and partnership building are key components of this WP. Dissemination is a key element for the proposal as the constant engagement of relevant stakeholders will draw attention to the project activities.

The main goals can be defined as follows: i) To communicate, create awareness and promote TIME4CS among the identified target groups and to be present at conferences, lectures and events open to the general public; ii) To create a bespoke Citizen Science Helix (Virtual Research Cluster) to be hosted on the CHX-IE Platform. This Helix will act as a 'home base' for engagement and dissemination activities stemming from the TIME4CS Project and post project impacts; iii) To disseminate the outputs of research activities from researchers, industry/SMEs, the Helix, media, publications (such a citizen science focused magazines, websites, journals,

TV/Web series). And to welcome input from citizen scientists by creating an organic workflow in which the two ends meet; iv) To create an anticipatory commercialization and engagement strategy by the end of the project to achieve the end goal of advancing the delivery of the project's impact.

Process of data collection

The main part of data collected in this WP is related to statistical data about the communication and dissemination activities.

Statistical data about communication and dissemination activities will be generated by the main communication channels, i.e. website, social media account and Helix (the virtual ecosystem/community developed for the project purpose).

Other data will be collected during the stakeholders and outreach events such as: list of participants and presentations, videos and photos (their collection will be subject the informed consent procedure described in detail in deliverable 8.2).

Format of data collected

- ✓ .DOC / .DOCX
- ✓ .XLS / .XLSX
- ✓ .PPT / .PPTX
- ✓ .CSV
- ✓ .PDF
- ✓ Photo Format (e.g., .JPEG, .PNG)
- ✓ Video Format (e.g., .MKV, .MP4, .AVI)

Estimated size of data

5GB

Work Package 7

The information was collected by the Work Package leader (Karolina Jurkiewicz / APRE) through the use of questionnaires. The results are summarized below.

Pre-Existing Data

Grant Agreement, Consortium Agreement

Source of Pre-Existing data

Project proposal preparation

Conditions for use of pre-existing data

Data is available for internal use within TIME4CS; no specific conditions applicable.

Type of data collected

- Observational data
- Experimental data
- Simulation data
- Derived or compiled data
- Reference or canonical data
- Event related data

Aim of this data

This WP aims at ensuring: i) effective consortium coordination and project management according to the description of action; including the relevant administrative and financial procedures, communication within the consortium and with EC as well project meetings organization; ii) appropriate data management and personal data protection in line with EU rules.

Process of data collection

The data collected is related to financial aspects of the projects (i.e. costs declarations of the consortium partners) and the internal consortium meetings, and it comprises meeting minutes, presentations and recordings. Additional data could be generated through internal survey to collect feedback regarding specific issues related to project management.

Format of data collected

- ✓ .DOC / .DOCX
- ✓ .XLS / .XLSX



- ✓ .PPT / .PPTX
- ✓ .CSV
- ✓ .PDF
- ✓ Photo Format (e.g., .JPEG, .PNG)
- ✓ Video Format (e.g., .MKV, .MP4, .AVI)

Estimated size of data

2GB



4. Data storage and backup

TIME4CS involves the collaboration between 11 beneficiaries and a moderate amount of data will be generated in various activities. To manage such quantities of data, with the aim to follow the FAIR principles of Findability, Accessibility, Interoperability, and Reusability and to stimulate the collaboration among partners, a few data storage tools have been selected. The selection was made having in mind the purposes and the end-users of the data stored, divided in tools appropriate for internal use and daily project activities, and tools for guaranteeing the long-term storage of TIME4CS data.

List of storage tools **NOT** appropriate for long-term storing of data generated by TIME4CS project but used to carry out daily project activities

- **OneDrive:** is a file hosting service and synchronization service operated by Microsoft as part of its web version of Office. It is a private space, connected to Teams Groups created for the project activities (as detailed in TIME4CS D7.1), used by all beneficiaries to share working files and store temporary master copy of raw data that should be available for use by other beneficiaries for data processing.
- **Local drives, company cloud storage and external portable storage devices:** these are storage facilities that do not fall under surveillance of this data management plan. These are very common and convenient for individual short-term storage of data.

List of storage tools appropriate for long-term storing of data and to guarantee the findability, accessibility, interoperability, and reusability of data generated by TIME4CS

- **Zenodo:** A TIME4CS Zenodo community has been created. TIME4CS datasets will be stored there. Data files are versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis. All data files are stored along with a MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant. Each beneficiary will be in charge to upload the data produced by their project activities in the community and a moderator will validate it. Each data-set will have a readme file.

Link: <https://zenodo.org/communities/time4cs/?page=1&size=20>

5. Data documentation

All data are documented to help interested users to clearly understand and reuse them. For this reason, descriptive and substantive (i.e. how the data should be read or interpreted) metadata will be elaborated and described in a readme.txt file complementing each dataset (see Table 1).

Creator*	Main researchers involved in producing the data.
Title*	Name or title by which the dataset is known.
Contributor	Institution where the data was created or collected. A person or organization responsible for making contributions to the dataset.
Publisher	A holder of the data (including archives appropriate) or institution which submitted the work. Any others may be listed as contributors.
Publication year*	The year when the data was or will be made publicly available.
Date created*	Date the resource itself was put together; this could be a date range or a single date.
Description*	Concise description of the contents of the dataset. Describe the research objective, type of research, method of data collection and type of data.
Subject	Subject, keyword, or key phrase describing the resource.
Temporal coverage	Indicate the dates to which the data refer. Enter the year or beginning and ending dates.
Spatial coverage	Describe the geographic area to which the data refer (e.g. municipality, town/city, region, country). The geographic coordinates of the area may be included, if desired.
Identifier	Zenodo automatically assigns a DOI to a dataset once the entire deposit procedure has been completed. In some cases, a dataset may be known by one or more other (persistent) identifiers.
Language*	The primary language of the resource.
Link to publication	Include the web addresses or DOIs for any publication, important internal reports or other datasets that are related to your dataset.

Table 1 - Metadata included in the readme file of TIME4CS data

File naming & Identifiers

In TIME4CS project, an identifier is used as a reference number or name for a data object. The identifier is a key part of our documentation and metadata. Table 2 shows the codes that can be used for making identifiers.

Description	Deliverables	Meetings	Conferences/Events
First Letters	TIME4CS	TIME4CS	TIME4CS
Underscore	–	–	–
Next letters	Deliverable number (Dx.y) [x=WP number, y=deliverable number]	Type of document (i.e. Agenda, Minutes, Presentation). In case of presentations, mention WP number as well.	Event title
Underscore	–	–	–
Next letters	Short explanatory title for the document	Location and date of the meeting separated by underscore	Location and date of the meeting separated by underscore
Underscore		–	–
Next letters		Short name of organisation and initials of presenter	Short name of organisation and initials of presenter
Underscore	–	–	–
Last letters	"v" and number of revision of this specific report [v0.1=draft version, v1.0=final version]	"v" and number of revision of this specific report [v0.1=draft version, v1.0=final version]	"v" and number of revision of this specific report [v0.1=draft version, v1.0=final version]

Table 2 - Examples of files identification

6. Data access

The main part of raw and processed data generated by the project will be accessible to all consortium partners. It will be stored in TIME4CS Teams Group. Only members of consortium, after validation from the group administrator (i.e. APRE), can access this space. The public data will be accessible by Zenodo.

The data containing personal data and the data gathered through questionnaires, interviews, focus groups, workshops and through other data gathering methods used during the project will be anonymised, and therefore the data cannot be traced back to the individual. Data will be stored only in anonymous forms so the identities of the participants will only be known by the research partners involved. The data will be used only within the project and will not be made accessible for any third party unless authorised by the individual informant via informed consent. Raw data like interview protocols and audio files will be shared within the consortium partners and with external open repositories only upon agreement of the participants (agreement via informed consent) and in anonymised form. The stored data will not contain the names or addresses of participants and will be edited for full anonymity before being processed (e.g. in project reports). In case of sharing the data in an open repository, the following formats would be shared: audio files, transcripts, aggregated files, interview guidelines. Finally, citizen science data reports based on interviews, focus group and other data gathering methods will be based on aggregated information and comprise anonymous quotations respectively.

Confidential data and data collected for internal purposes will be stored in the secure facilities of the organization responsible for collecting the data and will be retained for two years after the end of project. If requested and agreed by the participants, data can be shared with other consortium members through the online repository.

In case of data that cannot be accessible to public, it will be stored in Zenodo (to preserve it) but only the metadata will be public (In CC 0).

Details concerning the ownership, transfer and dissemination of projects results are defined in section 8 of the TIME4CS Consortium Agreement and shall be followed accordingly. The relevant rules of the Grant Agreement, in specific Article 26(2), Article 29(1) and Article 30, are also relevant and apply accordingly.



7. Data sharing and reuse

TIME4CS beneficiaries will be the main users of data produced in the project. In TIME4CS, all data may be classified into open, sensitive and closed. Access depends on the classification of the data.

- Open data must be acknowledged by citing the data set. Permission from the researcher is not required.
- Sensitive (or confidential/restricted data) may be made available by the researcher after any identifying information has been removed. Then, they can be used and cited. Permission from the researcher could be sought.
- Closed data are not available for sharing.

In any case, the main part of TIME4CS data will have the CC-BY license.



8. Data preservation and archiving

All relevant obtained data will be preserved and archived in ZENODO. All data and items on ZENODO will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.



9. Privacy of participants

Since TIME4CS activities involve the processing of personal data, Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (GDPR) and other relevant EU-level⁴ and relevant national legal acts shall be complied with and respected by the consortium as a whole, as well as by each project Beneficiary which will deal with personal data processing in the context of project implementation. To confirm compliance with the applicable EU and national legislation, each TIME4CS beneficiary signed a 'Declaration on Horizon 2020 Ethical Standards and Data protection' (all signed declarations are presented in the TIME4CS deliverable D8.1: POPD - Requirement No. 1).

According to Art. 4(1) of GDPR, **personal data** 'means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'. Examples of personal data protection typically cover: names and surname, home and email address, an identification card number, location data and IP address or a cookie ID. **Processing of personal data** is defined in Art. 4(2) of the Regulation and means 'any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction'.

In its Art. 5, the GDPR provides for the **principles of personal data processing**, stating that personal data must be:

- 'processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');
- collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes (...) ('purpose limitation');
- adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimization');

⁴ Regulation (EC) N° 45/2001 of the European Parliament and the Council of 18 December 2000 in the protection of individuals regarding the processing of personal data by the Community institutions and bodies and on the free movement of such data, Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons regarding the processing of personal data by competent authorities for the purpose of prevention, investigation, detection or prosecution of criminal offences, Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce in the Internal Market, Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector.

- accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy');
- kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed (...) ('storage limitation');
- processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures ('integrity and confidentiality').

In order to fulfill to the GDPR requirements, there is a set of indications that each TIME4CS Beneficiary dealing with data processing will respect during the project implementation:

1) Consent

The consent should be given by an individual (data subject) in a free, specific, informed and unambiguous manner, by way of a request presented in clear and plain language. Moreover, the consent should be given by an affirmative act, such as checking a box online or signing a form (this methodology is explained in TIME4CS Deliverable 8.2).

2) Providing transparent information

Individuals must be clearly informed on who is processing their personal data as well as what data will be processed, why and how. In addition to the above information, individuals should be also informed about who will receive the data, how long the data will be stored, the individual's data protection rights (i.e. right to access, correction, erasure, restriction, objection, portability, etc.), whether there is a statutory or contractual obligation to provide the data and how consent can be withdrawn.

3) Ensuring the right to access and right to data portability

Individuals must be ensured with the right to request access to their personal data, free of charge and in an accessible format. In addition, when the processing is based on consent or a contract, the individual can ask for their personal data to be returned or transmitted to another company (right to portability).

4) Ensuring the right to erasure (right to be forgotten)

An individual has the right to request the data controller to erase their personal data, such as when the data is no longer needed to fulfil the processing purpose. The data controller is not obliged to comply with such request if: the processing is necessary to respect one's freedom of expression and information, they must keep the personal data to comply with a legal obligation; there are other reasons of public interest to keep the personal data, such as public health or scientific and historical research purposes; they need to keep the personal data to establish a legal claim.

5) Ensuring the right to correct and right to object

If an individual thinks that their personal data is incorrect, incomplete or inaccurate, they have the right to have it rectified or completed without undue delay

6) Obligation to appoint a Data Protection Officer

An entity must appoint a DPO when: it regularly or systematically monitors individuals or process special categories of data; this processing is a core business activity; and it does it on a large scale. In the case of TIME4CS Beneficiaries, all relevant declarations are provided in Appendix 1 to this deliverable and the list of DPOs with their contact details is provided in Appendix 2.

7) Obligation of data protection by design and by default

In order to minimize privacy risks and increase trust, a data controller⁵ must take all necessary steps (e.g. pseudonymization) to implement the data protection principles and protect the rights of individuals (data protection by design). In accordance with the obligation of data protection by default, the entity processing personal data should ensure that the most privacy friendly setting is the default setting.

8) Obligation to provide proper notification in the case of a data breach

In case a data breach occurs, and the breach poses a risk to individual rights and freedoms, the entity processing personal data should notify the relevant Data Protection Authority within 72 hours after becoming aware of the breach. If the data breach poses a high risk to those affected, the entity may also be required to inform all individuals affected by the data breach.

Details on specific measures that will be taken by each TIME4CS Beneficiary to safeguard the rights and freedoms of TIME4CS data subjects/research participants and the security measures that will be implemented to prevent unauthorized access to their personal data are provided in TIME4CS deliverable D8.1: POPD - Requirement No. 1 (in Appendix 1 in the '**Declaration on Horizon 2020 Ethical Standards and Data protection**' of each Beneficiary, in the part related to the privacy policy of the Beneficiary).

⁵ According to Art. 4(7) GDPR, 'controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law.

List of References

- Data Management Plan - REPAiR REsourceManagement in Peri-urban AREas: Going Beyond Urban Metabolism
- Data Management Plan – ZeroBrine Industrial Desalination – Resource Recovery – Circular Economy
- Data Management Plan - StatusCities: Migrant legal STATUS diversity and diversity dynamics in European CITIES
- “Defining research data” (pp. 5-6), in Edinburgh University Data Library Research Data Management Handbook http://www.docs.is.ed.ac.uk/docs/data-library/EUJL_RDM_Handbook.pdf
- <https://help.osf.io/hc/en-us/articles/360019739014-Licensing>
- <https://libguides.library.cqu.edu.au/c.php?g=761037&p=5456900>

