

Project Number: 687818  
 Project acronym: O4C  
 Project title: OPEN4CITIZENS - Empowering citizens to make meaningful use of open data  
 Contract type: H2020-ICT-2015 - RIA

Deliverable number:	<b>D4.6</b>
Deliverable title:	<b>Data Management Plan (Midterm)</b>
Work package:	WP4
Due date of deliverable:	M17 – April 2017
Actual submission date:	28/04/2017
Start date of project:	01/01/2016
Duration:	30 months
Reviewer(s):	Nicola Morelli (AAU), Tomas Edman (EXPERIO)
Author/editor:	Anne Sofie Juul Sørensen (DATAPROCES)
Contributing partners:	Anne Sofie Juul Sørensen (DATAPROCES), Janice S. Pedersen (ANTRO), Kjartan Jensen (DATAPROCES), Mehdi Ben Taarit (DATAPROCES), Nicola Morelli (AAU)

Dissemination Level of this Deliverable:	<b>PU</b>
------------------------------------------	-----------

<i>Public</i>	<i>PU</i>
<i>Confidential, only for members of the consortium (including the Commission Services)</i>	<i>CO</i>

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 687818. Further information is available at [www.open4citizens.eu](http://www.open4citizens.eu).

## Document history

Version no.	Date	Authors	Changed chapters
0.1	06/02/2017	Anne Sofie Juul Sørensen	Document structure aligned with version 3.0 of Guidelines on FAIR Data management in Horizon 2020 template
0.1	11/04/2017	Anne Sofie Juul Sørensen	Reusable contents from D4.5 transferred
0.1	19/04/2017	Janice S. Pedersen	Input throughout doc. re: research data
0.1	20/04/2017	Anne Sofie Juul Sørensen	Inputs throughout doc on Platform details
0.1	21/04/2017	Anne Sofie Juul Sørensen	Edited Executive Summary
0.1	24/04/2017	Nicola Morelli Tomas Edman	Internal review
0.1	26/04/2017	Anne Sofie Juul Sørensen Janice S. Pedersen	Internal corrections
0.2	27/04/2017	Anne Sofie Juul Sørensen Janice S. Pedersen	Final version

## Statement of Originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License, 2015. For details, see <http://creativecommons.org/licenses/by-sa/4.0/> (just one possible option)

## Contributors

Part. No.	Part. short name	Name of the Contributor	E-mail
	DP	Anne Sofie Juul Sørensen	as@datapoces.dk
	DP	Kjartan Jensen	k@datapoces.dk
	DP	Mehdi Ben Taarit	mbt@datapoces.dk
	ANTRO	Janice S. Pedersen	jp@antropologerne.com
	AAU	Nicola Morelli	nmor@create.aau.dk

**Table of Contents**

List of Abbreviations..... 7

Glossary ..... 7

1. Executive Summary ..... 8

2. Introduction..... 8

    2.1. Structure of the document..... 8

    2.2. Relation to other deliverables..... 9

3. Data Summary ..... 9

    3.1. What is the purpose of the data collection/generation and its relation to the objectives of the project? ..... 9

        3.1.1. Research data supports evaluation and dissemination activities ..... 10

        3.1.2. Data in the O4C platform is available for use in hackathons and reuse beyond these events 10

    3.2. What types and formats of data will the project generate/collect?..... 11

        3.2.1. Research data ..... 11

        3.2.2. Data in the O4C platform ..... 11

    3.3. Will you re-use any existing data and how?..... 11

    3.4. What is the origin of the data?..... 11

    3.5. What is the expected size of the data? ..... 12

    3.6. To whom might it be useful ('data utility')..... 12

4. FAIR Data ..... 13

    4.1. Making data findable including provisions for metadata ..... 14

        4.1.1. Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)? ..... 14

    4.2. Making data openly accessible..... 14

        4.2.1. Which data produced and/or used in the project will be made openly available as the default? 14

        4.2.2. How will the data be made accessible (e.g. by deposition in a repository)? ..... 14

        4.2.3. What methods or software tools are needed to access the data? ..... 15

    4.3. Making data interoperable..... 15

        4.3.1. Are the data produced in the project interoperable?..... 15

        4.3.2. What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?..... 15

    4.4. Increase data re-use (through clarifying licences) ..... 15

        4.4.1. How will the data be licensed to permit the widest re-use possible? ..... 15

        4.4.2. When will the data be made available for re-use? ..... 15

        4.4.3. Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why..... 15

        4.4.4. How long is it intended that the data remains re-usable?..... 16

        4.4.5. Are data quality assurance processes described?..... 16

5. Allocation of Resources ..... 16

    5.1. What are the costs for making data FAIR in your project? ..... 16

5.2.	How will these be covered? .....	16
5.3.	Who will be responsible for data management in your project? .....	17
5.4.	Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)? .....	17
6.	Data Security .....	17
6.1.	What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)? .....	17
6.2.	Is the data safely stored in certified repositories for long term preservation and curation? .....	18
7.	Ethical Aspects.....	18
7.1.	Are there any ethical or legal issues that can have an impact on data sharing? .....	18
7.2.	Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data? .....	19
8.	References.....	21

**List of Figures**

Figure 1: Draft consent form used by all five pilots in the first hackathon cycle of the O4C project....20

**List of Tables**

Table 1: Open4Citizens objectives, as stated in the project’s Description of Action (pg. 2).....9

## List of Abbreviations

---

The table below shows the abbreviations used in this document, presented in alphabetical order and their meaning.

Abbreviation	Description
DCAT	Data Catalogue Vocabulary
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable and Reusable (data)
MB	Megabyte
O€C	Open4Citizens
PDF	Portable Document Format
TB	Terabyte
CSV	Comma Separated Value

## Glossary

---

The table below provides a glossary of concepts used in this document that may require additional explanation. Terms are in alphabetical order.

Abbreviation	Description
O4C Platform	The technologies, methodologies and communities of the Open4Citizens project
VBN	Aalborg University's online research portal, <a href="http://vbn.aau.dk/en/">http://vbn.aau.dk/en/</a>

## 1. Executive Summary

---

This midterm document – deliverable D4.6 Data Management Plan (M17) – describes how the research data will be collected and generated throughout the project and how it will be handled during and after the Open4Citizens project. It describes the standards and methodologies used for data collection and elaborates on how the data will be shared and preserved. It also covers the data that is handled in the Open4Citizens Platform. This midterm version Data Management Plan is written by following the Guidelines on FAIR Data Management in Horizon 2020 v3.0, 26 July 2016 provided by the EU Commission. A third and final version of the DMP is contractually bound to appear by M30 and will include the final and updated plans for data management in the project.

## 2. Introduction

---

This deliverable is the updated and second version of the Data Management Plan (DMP). By answering the questions outlined in the updated version 3.0 Guidelines on Data Management in Horizon 2020 July 2016 it describes:

- the handling of research data during and after the end of the project
- what data will be collected, processed and/or generated
- which methodology and standards will be applied
- whether data will be shared/made open access and
- how data will be curated and preserved (including after the end of the project)  
(Guidelines on Data Management in Horizon 2020, version 3.0)

The first version of the DMP (M6) followed the ‘Guidelines on Data Management in Horizon 2020, version 2.0, July 2015 and focused mainly on the categories of data sets that would be generated and used during the two hackathon cycles. This second version has a shifted focus to include the research data that is generated throughout the project and answers to how the project will make the research data FAIR (Findable, Accessible, Interoperable and Reusable) (Guidelines on Data Management in Horizon 2020, version 3.0). Since the Open4Citizens (O4C) Platform also setup to handle data the document also covers the data that is related to the O4C Platform where applicable. A third and final version of the DMP is contractually bound to appear by M30 and will include the final and updated plans for data management in the project.

### 2.1. Structure of the document

The document is based on the template H2020 Programme Guidelines on FAIR Data Management in Horizon 2020 version 3.0, 26 July 2016 and is structured after the questions presented in the template. Section 3-6 focuses primarily on how the research data is managed throughout the project, while section 7-9 also includes questions that are relevant for the data in the O4C platform. Thus - where relevant - the sections will cover both topics.

## **2.2. Relation to other deliverables**

The datasets that were used in the first round of hackathons are presented in detail in deliverable D3.2 ‘Data Mapping and Integration’ (delivered in M15). The more specific details about the hackathons; their themes, associated challenges, pre-hack activities and key actors are described in detail in deliverable D3.4 ‘First Hackathon Report’ (M17) whereas the datasets that will be identified in the upcoming second round of hackathons will be included in D3.3 ‘Data Mapping and Integration’ (final version M26).

## **3. Data Summary**

---

### **3.1. What is the purpose of the data collection/generation and its relation to the objectives of the project?**

There are two main groups of data in this project:

1. the research data that is generated throughout the project and
2. the data that is uploaded to and handled in the Open4Citizens (O4C) platform.

The data that the platform will handle is in the form of datasets generated from Open Data and authorization data for user logins.

Data collected for and generated in the O4C platform ensures that there is sufficient and appropriate open data to support the activities related to the project’s objectives listed in the table below. In the first year, this especially relates to objective 2 - Exploring hackathons and objective 3 - Overcoming citizens’ cognitive gap regarding Open Data. In the second year of the project, the second cycle of hackathons in the five pilot locations, O4C platform data will additionally be used and generated in the development of OpenDataLabs (objectives 1, 3, 4, 5 and 6).

The research data generated and collected in the project during the first year of activity relates especially to formative evaluation of objectives 2 and 3. In the second year of the project research data collected will support both formative evaluation related to the development of OpenDataLabs (objectives 1, 3, 4, 5 and 6) as well as summative evaluation elements, regarding the project’s achievements over all.

<b>Number</b>	<b>Objective</b>
1	Creating OpenDataLabs where citizens can design new services, or improving the existing ones, in a collaborative environment, and by using open data.
2	Exploring hackathons as new forms of collaboration among citizens, technical experts and public institutions that enable citizens, interest groups and grassroots communities to understand and use the potential of open data.



3	Overcoming the cognitive gap citizens have with respect to Open Data by making that knowledge available in form of consultants in the OpenDataLabs where citizens will experience the practical value of Open Data in the conception, modification, adaptation and maintenance of urban services.
4	Combining two specific models of OpenDataLabs, specifically the solutions development lab and the incubator models.
5	Exploring and driving opportunities for further exploitation and implementation of the developed and tested solutions through social network.
6	Creating an international network of cities and organisations where the Open Data Labs model implemented by Open4Citizens can be replicated and transferred so generating an international movement based on network cooperation and learning.

**Table 1: Open4Citizens objectives, as stated in the project's Description of Action (pg. 2)**

### **3.1.1. Research data supports evaluation and dissemination activities**

Research data in the Open4Citizens project is primarily generated by the members of the five pilots to support formative and summative evaluation of the extent to which we are achieving the project's stated aims. Most of the data collected for evaluation purposes is focused on the hackathon pressure-cooker event, held over the course of a weekend in most pilots. Data collection is carried out consistently across the pilots using standardised data collection templates consisting of PowerPoint slides. Further data is generated in line with the evaluation framework described in deliverable D4.1 'Evaluation Framework' to support reflections on individual pilots' contributions and the overarching Open4Citizens project's contributions to achieving its objectives. In addition, some of this material is used for pilot-specific and project-level dissemination activities.

### **3.1.2. Data in the O4C platform is available for use in hackathons and reuse beyond these events**

The purpose of generating and uploading datasets to the O4C Platform is to make it possible for participants, curious citizens and other interested stakeholders to locate and find the data that has been produced from the activities in the emerging OpenDataLabs.

In general 'Open Data' means that the data is publically available and can be used, modified, and shared freely by anyone for any purpose (<http://opendefinition.org/>). These open data have a large potential to enhance several aspects of human life, including transport, health care, climate and even human behaviour, if they for instance are implemented in new applications. The O4C project intends to include citizens as a driver for innovation in the Open Data arena. However many citizens are not necessarily able to handle Open Data, or even to imagine what to do with them. To help closing this

gap between users and data, the datasets generated and used in the Open Data Labs is intended to inspire others to see the potential in open data.

### ***3.2. What types and formats of data will the project generate/collect?***

#### **3.2.1. Research data**

Research data generated throughout the project includes primarily qualitative material, as well as some quantitative information, e.g. about participants in the hackathons and partners in the OpenDataLabs. Curated and anonymised materials will be made publicly available. This includes the following:

- Data gathering and analysis templates (originally produced in Microsoft PowerPoint format)
- Templates and guidelines for evaluation data gathering and analysis (originally in Microsoft PowerPoint and Microsoft Word/PDF)
- Completed templates from each pilot containing evaluation material from hackathons (PowerPoint)
- Questionnaire responses (originally Microsoft Excel documents)
- Intermediate evaluation analysis outputs based on data gathered in hackathons (Excel and PDF format)

All hackathon participants sign consent forms, allowing the project to use audio, video, visual and textual material about them. Aliases are used unless project participants explicitly wish to be quoted or identified using their real names.

#### **3.2.2. Data in the O4C platform**

The datasets that have been chosen for the hackathon cycles and uploaded to the O4C Platform are in CSV format which allows them to be used with the various tools on the platform.

### ***3.3. Will you re-use any existing data and how?***

**Research data** will consist primarily of new qualitative information generated in the project and will therefore not re-use existing data. Some quantitative data will also be generated as part of the evaluation data collected, complementing qualitative data collected about users of the O4C platform.

The data that is uploaded to the **O4C platform** will reuse existing open data. In addition, datasets generated through O4C activities are intended to be available for re-use through the platform.

### ***3.4. What is the origin of the data?***

The research data is original material and will derive from evaluation and dissemination-related activities. The material to be shared will be produced by the Open4Citizens consortium members.

The datasets uploaded to the platform will derive from open data repositories and the internet in general from where the data is 'publically available and can be used, modified, and shared freely by anyone for any purpose'. Certain datasets may be made available by O4C stakeholders who had not

previously made this data available, i.e. generating new open data for use in project activities and beyond.

### ***3.5. What is the expected size of the data?***

The total amount of research data to be made available is not yet clear. For the first hackathon cycle, year one of the project, the total size of evaluation and dissemination materials varies between the five pilots, from between about 200MB to 3TB, depending on the amount of media (video and photos) produced.

The size is not yet known for the data in the platform.

### ***3.6. To whom might it be useful ('data utility')***

#### **Research Data**

We intend to make curated research data from the five pilot projects available, as well as cross-cutting material reflecting on the evaluation of the project as a whole. This data will be useful for researchers, practitioners and others wishing to duplicate or adapt the Open4Citizens model for empowering citizens to make appropriate use of open data for improved service delivery. The analysis material included in the research data will highlight the strengths and challenges of the approach, allowing others to learn from the experiences of the project. The availability of templates and guidelines used in the project will also allow for better adoption of the approach by others interested in making open data a common good, whether they belong to academia, the public sector, the private sector groups or civil society.

Materials that are specifically intended for use by ordinary citizens working in the O4C platform or as part of an OpenDataLab will be specifically highlighted in the platform and the labs to ensure ease of access and understanding. These materials include the toolkits produced as part of the project (Preliminary Hackathon Starter Kit and Citizen Data Toolkit), as well as any additional templates and guidelines to facilitate working with and understanding open data.

#### **Data in the O4C Platform**

The O4C platform is focused around helping its users gain understanding of Open Data as well as aiding the development of new services/improve existing services during the hackathon cycle. The data in the platform is aimed at being used as:

- Components in digital mobile or web applications – a dynamic product to access personally meaningful or context-aware data, such as a weather or route planner app.
- Elements in concepts - i.e. mock-ups of mobile or web applications.
- Data examples for the participants to gain a greater understanding of Open Data.
- Visualisation – a statistical representation of data, such as an infographics or a narrative told as a news article (data journalism). The main objective is to communicate what is otherwise “raw numbers in a table”.
- Digital service – a product-service system with various touch points ingrained with open data. For example, a service where citizens can report faulty street objects (broken lamppost, etc.) using a smartphone application, and the government is notified about these problems and can fix them.

## 4. FAIR Data

---

The O4C project is a grant recipient under Horizon 2020 and is therefore required to deposit peer-reviewed publications into an open access repository.

In addition, we wish to participate in the European Commission's Open Research Data (ORD) pilot to the extent that this is feasible and useful. This is in light of the fact that the bulk of research data produced in the project is qualitative information that cannot easily be used outside the project given the need for contextual knowledge for its interpretation. By providing open access to a curated selection of the research data we intend to give access to anonymised and consolidated evaluation materials, tools and guidelines produced as part of the project, and the underlying datasets used to address challenges in urban services during the Open4Citizens hackathon cycle. This is beneficial for the public, but also for researchers ([www.openaire.eu](http://www.openaire.eu), 2017a). Providing Open Access will however not remove the author's copyright ([www.openaire.eu](http://www.openaire.eu), 2017a).

### Open Research Data

It is yet to be decided by the project partners whether we will deposit in an institutional repository of the research institutions with which we are affiliated, a subject-based/thematic repository or if we will make use of a centralised repository, like Zenodo that is hosted by CERN, available to all and set up by the OpenAIRE project ([www.openaire.eu](http://www.openaire.eu), 2017b). We are likely to use the latter option to ensure that the data is easier to find than if it is solely deposited in an institutional repository. In addition, to increase accessibility and findability, we are likely to make the same research data available through the institutional repositories of the universities that are part of the Open4Citizens consortium: Aalborg University, Politecnico di Milano and Technische Universiteit Delft. This use also depends on the compatibility of the institutional platforms with the type and size of data to be uploaded. For example, the Aalborg University Research Portal, VBN (Aalborg University Research Portal, 2017), while compatible with OpenAire (Open AIRE, 2017c) has not previously been used for the type of research material which we are producing; video, photos and slides. Aalborg University is a signatory of the Berlin Declaration on Open Access in the Sciences and Humanities (Berlin Declaration, 2003), whose principles the Open4Citizens project subscribes to. Signatories to the declaration aspire to 'promote the Internet as a functional instrument for a global scientific knowledge base and human reflection and to specify measures which research policy makers, research institutions, funding agencies, libraries, archives and museums need to consider' (Berlin Declaration, 2003, pg. 1). Decisions regarding the core platform to be used and the extent to which partners' national institutional repositories will be used will be made in good time to make initial research data available in the second year of the project (2017) and all relevant research data to be available by the end of the project (project month 30, June 2018).

### Regarding Open Access Journals

Where it is not possible to publish final peer-reviewed publications in Open Access Journals, the project partners aim to make publications available in institutional open access repositories such as VBN at Aalborg University. This, however, will depend on the individual journals' policies regarding open access.

### **Open Access through the OpenDataLabs**

An intended the lasting and living legacy of the Open4Citizens project is the OpenDataLabs. As we are identifying the business and sustainability plans for the OpenDataLabs, we will be prioritising openness and co-creation within the labs as a standard approach across the five pilot locations. In the second hackathon cycle, in 2017, this will involve identifying which guidelines and standards for openness are applicable across all pilots and can be adopted as standard.

### **Regarding Data on the O4C platform**

The approach to data storage in the platform is also inspired by the FAIR principles to make it easier for the participants and other interested stakeholders to find, access and re-use the datasets and make them interoperable with other datasets. This will be elaborated in the sections below.

## ***4.1. Making data findable including provisions for metadata***

### **4.1.1. Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?**

For research data collected and generated in the project, a fit-for-purpose file naming convention will be developed in accordance with best practice for qualitative data, such as described by the UK Data Archive (2011). This will involve identifying the most important metadata related to the various research outputs. Key information includes content description, date of creation, version, and pilot location.

To make the datasets in the platform easily findable searchable tags have been added to the metadata. When uploading the data the creator of the dataset also have the option to create new tags that corresponds to the contents of the dataset making it easier for other users to find and re-use the data.

## ***4.2. Making data openly accessible***

### **4.2.1. Which data produced and/or used in the project will be made openly available as the default?**

Certain pilots use medical data as part of the open data in their hackathons and OpenDataLabs. This includes the Karlstad and Barcelona pilots. The consortium partners will be guided by the legal and ethical restrictions that their partners adhere to, i.e. the data owners, who make relevant open data available. Restriction regarding data availability will come into play if the partners of the two pilots deem it necessary.

### **4.2.2. How will the data be made accessible (e.g. by deposition in a repository)?**

The research data will be made accessible through an Open AIRE compatible repository which has yet to be decided upon.

#### **4.2.3. What methods or software tools are needed to access the data?**

At present Microsoft Office is used for producing research data. However, the research data will be made available in the most appropriate open source formats.

### ***4.3. Making data interoperable***

#### **4.3.1. Are the data produced in the project interoperable?**

During the first hackathon cycle, research data, consisting of templates, data gathered from hackathon participants and analysis of materials has been produced using Microsoft Office. This includes PowerPoint and Excel-formatted files in particular. The consortium will explore the best open source software to use instead of these formats, and will make relevant research data placed in repositories available in these formats instead.

#### **4.3.2. What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?**

Depending on the choice of repositories we are however likely to follow the standard of The Data Catalog Vocabulary (DCAT) (Data Catalog Vocabulary, 2014) as it defines a standard way to publish machine-readable metadata about a dataset where appropriate. We also intend to use common ontologies and vocabularies for data.

### ***4.4. Increase data re-use (through clarifying licences)***

#### **4.4.1. How will the data be licensed to permit the widest re-use possible?**

The Open4Citizens project aims to be as open as possible. We take the guidelines developed by Open Knowledge as our starting point. Specifically, we will explore the applicability of the Open Data Commons Open Database License (ODbL) for data created in the project. A project partner in Milan, OnData (<http://ondata.it/>), is currently developing openness guidelines that may be applicable data used and generated in the Open4Citizens project.

#### **4.4.2. When will the data be made available for re-use?**

The curated research data related to the first round of hackathons is sought to be made available by August 2017. By the end of the project in M30 all data that is not affected by embargo will be made available through the appropriate repositories.

#### **4.4.3. Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.**

The curated research data made available will be available for re-use. However it may not be feasible for researchers outside the project to interpret and use any anonymised qualitative research data.

Raw social science research data of this nature generally requires deep contextual knowledge to be appropriately analysed and interpreted.

The O4C Platform will form a virtual component of the five Open Data Labs. The aim is that it will be able to exist after the O4C project has been completed. This means that the Open Data that has been collected, generated and uploaded to the Platform during the project lifetime will be accessible both after each hackathon cycle and after the end of the funding period of the O4C project. The datasets that are uploaded to the O4C platform will be shared through the website [www.opendatalab.eu](http://www.opendatalab.eu) where everyone will have access to them.

#### **4.4.4. How long is it intended that the data remains re-usable?**

We will adhere to the repository standard of the chosen repository or repositories for the project.

#### **4.4.5. Are data quality assurance processes described?**

The data quality assurance processes are not described yet. These will be developed based on experiences finding, using and creating open data sets in the project, including the feasibility of their use in the O4C platform for creating data-led solutions to service challenges. At the end of the first hackathon cycle, as this document is being written, a pragmatic approach is being taken to the quality of the open data being used in the project: at this stage of the project quality standards are less important than making relevant data available and using it to the extent possible, thereby learning about what is needed in order to improve its quality. In further defining the O4C project's data management plan with respect to research data, we will seek inspiration in the plans of similar projects and identify good practice guidelines for social science research data. We will use this input to ensure that we make good quality research data available.

## **5. Allocation of Resources**

---

### ***5.1. What are the costs for making data FAIR in your project?***

Any minor costs for Open Access publishing will be covered by the overall project dissemination budget. This will be used at the discretion of the consortium and Aalborg University as the primary investigator. Wherever possible, free institutional repositories will be used for data and publications. Expected costs will be further identified in the final version of the Data Management Plan, based on project needs to the end of the project.

### ***5.2. How will these be covered?***

There is no significant budget allocation in the project for making the data FAIR. The project therefore expects to cover costs related to open access to research data as eligible costs as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions). For any additional costs, sponsorship outside of the project would need to be sought, for example by partners interested in developing OpenDataLabs or by securing additional research funding.

### **5.3. Who will be responsible for data management in your project?**

Aalborg University (AAU) and Antropologerne (ANTRO) are responsible for the overall collection and handling of research data while Dataproces is responsible for the data management of data in the O4C Platform. The pilot coordinators will be responsible for the collection and handling of research data in each pilot.

### **5.4. Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?**

This will be further developed as the business cases for the OpenDataLabs are defined in the project. The consortium expects the main questions relating to data preservation to be identified and answered as these business cases are developed.

All datasets that are uploaded to the O4C Platform will be stored on a server at Dataproces who will ensure preservation and backup throughout the project. The aim is that the Platform will continue to be available after the O4C project has been completed. This means that the Open Data that has been collected, generated and uploaded to the Platform during the project life-time will be accessible both after each hackathon cycle and after the end of the funding period of the O4C project. The data in the O4C platform will be available for as long as the internal server at Dataproces is up and running and costs covered by the business case by Dataproces.

## **6. Data Security**

---

### **6.1. What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?**

Research data is shared between project partners and stored in collaborative online working platforms during the project's lifetime. These are BaseCamp (<https://3.basecamp.com>), Google Drive (<https://drive.google.com>), and Dropbox (<https://www.dropbox.com>). Some intermediate and all final versions of evaluation data collected in the project and analysis outputs of this material are saved in a standardised filing system with dedicated naming conventions in the project's BaseCamp account.

Uncurated and unanalysed material created during the project is stored locally by the Open4Citizens partners according to their institutional data management and storage guidelines. This locally stored research data includes un-anonymised questionnaire data from hackathons, as well as consent forms signed by hackathon and other project participants allowing for the use of personally identifiable information about them. Consent forms will be kept beyond the end of the Open4Citizens project. Additional research data such as personal notes, unused photos and video clips etc. will be safely deleted and discarded after the end of the project. This research includes all data not made publicly available for the long term.



In the finished form of the platform the uploaded data is secure and recoverable with daily backups, where Dataproces can go back to any file from any day for all 365 days a year. Dataproces also fulfil all data management comments regarding the European Personal Data Protection Act.

## ***6.2. Is the data safely stored in certified repositories for long term preservation and curation?***

At the time of writing this deliverable, the Open4Citizens project partners are identifying the most appropriate repository or repositories in which to make research data available for the long term. An update on decisions made will be provided in the project's final Data Management Plan, due in M30.

## **7. Ethical Aspects**

---

### ***7.1. Are there any ethical or legal issues that can have an impact on data sharing?***

#### **User-generated data**

In the second hackathon cycle it is likely that user-generated data will be increasingly important for developing open-data driven solutions in Open4Citizens hackathons. Examples include crowd-sourced data such as that made available in platforms like Open Street Map ([www.openstreetmap.org](http://www.openstreetmap.org)), public social media data, as well data scraped from service and product review sites. The Open4Citizens consortium intends to adopt existing standards and approaches regarding the use of this kind of data. Our primary guide in defining an Open4Citizens guideline to using this type of data will be the licences and guidelines available through the Open Data Commons (<https://opendatacommons.org>) developed by Open Knowledge International (<https://okfn.org/>). In essence, the Open Data Commons Open Database License (ODbL) states that the databases to which it applies allow users to freely share, create and adapt from the database, as long as public use is attributed, shared alike and kept open. However, as these guidelines are voluntary, the consortium intends to engage with members of our advisory board to determine which current approaches to openness we should adopt. We intend, for example, to review and appropriately adopt guidelines produced by the Open Data Institute (<https://theodi.org/guides>) and Open Knowledge International, as leaders in the field, as well as any other relevant guidelines.

We foresee that balancing participants' rights to the intellectual property entailed in their hackathon outputs with the desire for openness may be challenging. We will therefore stay informed about developments in the area of rights, ethics and openness throughout the life of the project to ensure that we are transparently applying best practice procedures and guidelines. We expect that debates regarding use of user-generated data in data journalism will be one useful source of guidance.

#### **Data in the O4C platform**

When organising the events Open4Citizens will collect information from public repositories, which contain Open Data. Since Open Data consist of information databases that are public domain, the data can be freely used and redistributed by anyone. Open4Citizens is thus not subject to any regulations regarding proper data storage, including principles stipulated in The Data Protection Directive 95/46/ECP and the General Data Protection Regulation (EU) 2016/679.

When providing Open Data for the hackathons and related events, Open4Citizens, its employees, and any contributing partners of Open4Citizens or its employees, shall not be liable for any harm arising from the use of the collected datasets shared through the O4C Platform, including but not limited to, how participating parties handle and develop the Open Data available on the O4C Platform.

In regards to the Open Data from various data sources that are made available on the O4C Platform, Open4Citizens does not guarantee that this data has been published with the prior, necessary and informed approval that it requires.

## ***7.2. Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?***

The gathering and analysis of research data in the project is guided by standard ethics guidelines for the social sciences (e.g. as discussed in [http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/ethics-guide-ethnog-anthrop\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/ethics-guide-ethnog-anthrop_en.pdf)).

For research data collected in relation to Open4Citizens hackathons, as well as questionnaires and other personally identifiable information generated, informed consent is sought. All participants in hackathons are requested to provide their consent for all data produced to be used by the project. Figure 1 represents the template consent form used for the first hackathon cycle below. In addition all users of the platform will be asked for consent when creating a profile in the O4C platform.



**CONSENT FORM | OPENDATALAB [INSERT YOUR CITY] HACKATHON | [INSERT HACKATHON DATES]**

I hereby give my consent for all videos and photos of me, direct quotes, as well as any other material that I have made available to be used by OpenDataLab [insert your city] and the Open4Citizens project, provided that it has been anonymised. If I am happy for my name to be used in relation to this material, I have indicated this at the bottom of the form.

I also give permission for the Open4Citizens project to use these materials for dissemination and communication on project websites and social media channels (e.g. Facebook, Twitter, Instagram, LinkedIn and others) as long as my name is not used. The material may also be used in other dissemination information, such as workshop material, conference posters, and in both popular and academic articles.

The Open4Citizens project partners may use this anonymised material for project dissemination, reporting and communications with project funders, primarily the European Commission, as well as advisory and other boards. In addition, it may be used in reference material as an example of work carried out by the project partners\*.

This material may not be used by any third party in any other context, including for commercial purposes (e.g. advertisement and marketing).

The Open4Citizens project partners may use the material described in this document indefinitely. I can revoke this consent as it applies to future use of materials at any time by contacting the Open4Citizens project or OpenDataLab [insert your city]: See [www.open4citizens.eu](http://www.open4citizens.eu) or [insert preferred contact information – e.g. Name of main O4C pilot partner at your location plus e-mail address, website, physical office address etc.]

*Any additional comments about how the material can be used can be added here:*

---



---

I am happy for my name to be used in relation to material gathered (please tick the box)

Date: \_\_\_\_\_

Full name: \_\_\_\_\_

Signature: \_\_\_\_\_



\*O4C project partners are: Aalborg University; Antropologeme ApS; DataProcess ApS; Experio Lab, Västerås; Elna; Fundación Privada UCAIT; Internet i Innovació Digital a Catalunya; Politecnico di Milano; Technische Universiteit Delft; Telecom Italia SPA.

Figure 1: Draft consent form used by all five pilots in the first hackathon cycle of the O4C project

## 8. References

---

Aalborg University Research Portal, 2017. Retrieved April 27, 2017, from <http://vbn.aau.dk/en/>

Berlin Declaration 2003. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. (2003). Available at the Max Planck Society Open Access website: <https://openaccess.mpg.de/Berlin-Declaration>, Retrieved April 20, 2017

Data Catalog Vocabulary 2014, Marking up your dataset with DCAT | Guides. (n.d.). Retrieved April 27, 2017, from <https://theodi.org/guides/marking-up-your-dataset-with-dcat>

Group, O. K. The Open Definition. Retrieved April 27, 2017, from <http://opendefinition.org/>

Guidelines on FAIR Data Management in Horizon 2020. (n.d.). Retrieved April 27, 2017, from [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

ODbL, Open Data Commons Open Database License (ODbL). (2016, July 06). Retrieved April 27, 2017, from <https://opendatacommons.org/licenses/odbl/>

Open Aire 2017a, Open Aire FAQ. Retrieved April 27, 2017, from <https://www.openaire.eu/support/faq#article-id-234>

Open Aire 2017b, Principe, P. Open Access in Horizon 2020. Retrieved April 27, 2017, from <https://www.openaire.eu/open-access-in-horizon-2020>

Open Aire 2017c, OpenAIRE. Retrieved April 27, 2017, from <https://www.openaire.eu/>

UK Data Archive 2011, Managing and Sharing Data: Best practice for researchers. Retrieved April 24, 2017, from <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>