



Horizon 2020
Programme

CICERONE

Coordination and Support Action (CSA)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 820707

Start date : 2018-11-01 Duration : 24 Months
<http://cicerone-h2020.eu>



Data Management Plan

Authors : Dr. Gabor SZENDRO (LGI)

CICERONE - Contract Number: 820707

Project officer: Eleni Magklara

Document title	Data Management Plan
Author(s)	Dr. Gabor SZENDRO
Number of pages	9
Document type	Deliverable
Work Package	WP6
Document number	D6.3
Issued by	LGI
Date of completion	2019-11-18 09:57:02
Dissemination level	Public

Summary

The main goal of the Data Management Plan is to describe how the research data will be handled during and after the end of the project, what kind of 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.

Approval

Date	By
2019-11-18 09:57:08	Dr. Gabor SZENDRO (LGI)

Table of Contents

1	Introduction	3
1.1	Purpose of the Data Management Plan	3
1.2	Data set properties.....	3
1.3	Data licensing.....	4
2	Description of the data	4
3	Summary of identified datasets	5
3.1	Format/license	5
3.2	Archiving/preservation.....	5
4	Data Protection Officer.....	5
5	Ethical aspects.....	5
6	Restrictions for re-use	6
6.1	Anonymization of data	6
6.2	Pseudonymisation of data	6
7	Personal data transfer and processing	7
8	References	7
9	Annex I.: Detailed information sheets for identified datasets	7





Summary

The main goal of the Data Management Plan is to describe how the research data will be handled during and after the end of the project, what kind of 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.

Any peer-reviewed research publications, abstracts and presentations for international conferences may also be shared through the website of the project at <http://CICERONE-h2020.eu>.

Abbreviations

Abbreviation	Definition
DMP	Data Management Plan
EC	European Commission
DRM	Digital Rights Management
PO	Programme Owner
RTO	Research and Technology Organisation
SME	Small and Medium Enterprise



1 Introduction

CICERONE brings together programme owners, research organizations and other stakeholders to create a platform for efficient Circular Economy programming. The priority setting and the organization of the future platform will be driven by Programme Owners (POs), involved either as project partners, or via a stakeholder network, and the work will be carried out in close cooperation with research & technology organisations (RTOs), which contribute with their expertise of the main scientific and technological challenges. Consultation mechanisms will also ensure that all stakeholders will be able to actively contribute (civil society, industry, innovative SMEs, startups, cities, investors, networks, etc.).

1.1 Purpose of the Data Management Plan

The purpose of this document is to lay out a plan for the management of data generated/collected in CICERONE. It covers the following:

- Identification of data to be collected/processed/generated
- Methodology and standards to be applied
- Data handling during and after the project
- Sharing, curating and preserving data

At the time of this writing, CICERONE partners have identified 4 data sets to be included in the DMP at this stage. All of these data sets identified at this stage will be made openly accessible to the public through repositories such as Zenodo, and they will be preserved after the end of the project. The DMP is a living document – if necessary, it will be updated throughout the project's lifetime.

1.2 Data set properties

Following the guidelines of the EC (EC, 2016), this document contains the following properties for each of the identified data sets:

1. Name
2. Short description
3. Standards to be applied, metadata
4. Data sharing
5. Curation/archiving/preservation

A short description of each of these properties is provided below.

1.2.1 Name and reference code

In order to imbue the names of datasets with easily identifiable meaning that conveys important information, the following naming convention shall apply:

CountryCode.DataOwner.Openness.Title

CountryCode: this string identifies the country to which the data pertains/where the data was collected using the ISO 3166 Alpha-2 coding system.

DataOwner: this string identifies the project partner in CICERONE that is associated with the dataset (data collector/custodian) using the official abbreviated partner names.

Openness: this string determines whether a given dataset is intended to be shared with the public as Open Data. It may take the following values:

1. Open: can be accessed, used and shared by anyone without limitations, accessible on the internet in a machine-readable format, free of restrictions on use in its licensing)
2. Shared: available to use, but not under an open data license. Restrictions on its use or reproduction may apply (limited to a given group of people or organisations, may not be reproduced without authorisation, etc.)
3. Closed: can only be accessed by its subject, owner or holder

Title: a short and descriptive string to identify the contents of the data

Using these strings, the name of a dataset would look like this:



FR.LGI.Open.CommuteHouseholdSurvey

A dataset with this name would describe a household survey on commuting preferences conducted in France and curated by LGI.

1.3 Data licensing

Without a license to set out the terms of use, data is not truly open. Data without a license may be publicly accessible, but users do not have the certainty that they can use and share the data, leaving them in a legal grey area. Data licensing standards are used to lay out the openness of data sets in concrete terms, and an open data license gives explicit permission to use the data both for commercial and non-commercial purpose. There are many types of licenses to choose from, and this document will not cover them in depth. The table below provides a summary of common data licenses that will be considered for use in the project (based on definitions from opendefinition.org):

Name	Domain	Attribution	Share-alike*	Notes
Creative Commons CCZero (CC0)	Content, data	N	N	All rights (including those of attribution) waived
Open Data Commons Public Domain Dedication and Licence (PDDL)	Data	N	N	All rights (including those of attribution) waived
Creative Commons Attribution 4.0 (CC-BY-4.0)	Content, data	Y	N	Credit must be given, a link to the license must be provided, changes made must be indicated. If these terms are not followed, license may be revoked
Open Data Commons Open Database License (ODbL)	Data	Y	Y	Credit must be given, share-alike must be assured, data may be redistributed using DRM as long as a DRM-free version is also released

*Share-alike is the requirement that any materials created using the given dataset must be redistributed under the same license

2 Description of the data

The following detailed information sheet will be produced for every dataset to be produced/collected/curated in the project:

Name of the dataset	A name to identify the data, see 1.2.1 for details.
Description of the dataset	<ul style="list-style-type: none"> • A brief, easy to understand description of what the dataset contains and what it will be used for in the project • A list of institutions to whom the data set could be useful outside the project • Whether the dataset has been/will be used for a scientific publication (if yes, brief details about the content and journal) • If the dataset is collected, a brief description of its origin and how it was collected will be provided • Openness of the dataset • Whether the dataset is anonymised or not
Format/license	The format in which the data will be available (e.g. .xls, .csv, .txt) will be provided. The license to be used will also be provided.



Archiving/preservation

Efforts and means to keep the data available after the end of the project will be described here, including where/how the data will be preserved, the duration of preservation, the associated costs and the plans of the consortium to cover these costs.

3 Summary of identified datasets

This DMP contains 4 datasets identified by the CICERONE partnership. The following tables provide information on the various aspects of these datasets. The sheets completed by partners are provided in Annex I.

3.1 Format/license

Name of the dataset	Format	License
FRLGIOpenPOsSurvey	.xls	ODbL
FIVTTKICOpenPOsSurvey	.xls	ODbL
ESC-KICOpenPOsSurvey	.xls	ODbL
World.Juelich.Open.International benchmark on CE(T1.2, D.1.3)	.xls	ODbL

3.2 Archiving/preservation

Name of the dataset	Sharing medium	Duration of preservation	Costs	How costs will be covered
FRLGIOpenPOsSurvey	Zenodo.org	Perpetual	N/A	N/A
FIVTTKICOpenPOsSurvey	Zenodo.org	Perpetual	N/A	N/A
ESC-KICOpenPOsSurvey	Zenodo.org	Perpetual	N/A	N/A
World.Juelich.Open.International benchmark on CE(T1.2, D.1.3)	Zenodo.org	Perpetual	N/A	N/A

4 Data Protection Officer

In accordance with applicable regulations, the host institution (Climate-KIC) is not required to appoint a Data Protection Officer. A detailed data protection policy for the project is kept on file.

5 Ethical aspects

This Data Management Plan (DMP) was drafted and updated taking into account the General Data Protection Rules (GDPR) for the collection, storage and re-use of the data, in line with the following general principles. :

Personal data shall be:

1. processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');
2. collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), however not be considered to be incompatible with the initial purposes ('purpose limitation');



3. adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');
4. 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');
5. 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; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');
6. processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').

6 Restrictions for re-use

Data generated through interviews and surveys will not be re-used directly due to privacy concerns. To allow re-use and avoid loss of research data, two different techniques could be used to disseminate its data while abiding by regulations on privacy.

6.1 Anonymization of data

"Anonymization" of data means processing it with the aim of irreversibly preventing the identification of the individual to whom it relates. Data can be considered anonymised when it does not allow identification of the individuals it is related to, and no individuals can be identified from the data by any further processing of that data or by processing it together with other information which is available or likely to be available.

There are different anonymization techniques. Here are the two most relevant:

- Generalisation : generalising data means removing its specificity. For example, in the case of a table containing household income levels, with 4 figures mentioned: €135,000, €60,367, €89,556, and €365,784. One way of generalising this numbers would be to write that the values are "more than €150,000, less than €80,000, between €90,000 and €120,000, and more than €300,000" respectively. Essentially it means taking exact figures, establishing a baseline category, and then obfuscating the data by assigning it to one of the categories in order to remove any sense of specificity from it.
- K-anonymity; A release of data is said to have the k-anonymity property if the information for each person contained in the release cannot be distinguished from the other individuals whose information also appear in the release. For instance, in a table composed of six attributes (Name, Age, Gender, State of Domicile, Religion and Disease), removing the name and the religion column while generalising the age is a way to effectively k-anonymise the data.

Other techniques, such as "masking" or "pseudonymisation", which are aimed solely at removing certain identifiers, may also play a role in reducing the risk of identification. In many cases, these techniques work best when used together.

6.2 Pseudonymisation of data

"Pseudonymisation" of data means replacing any identifying characteristics of data with a pseudonym, or, in other words, a value which does not allow the data subject to be directly identified. Although pseudonymisation has many uses, it should be distinguished from anonymization, as it only provides a limited protection for the identity of data subjects in many cases as it still allows identification using indirect means. Where a pseudonym is used, it is possible to identify the data subject by analysing the underlying or related data.



Task leaders will be responsible for the anonymization of data in CICERONE for all datasets where this is deemed necessary.

7 Personal data transfer and processing

In case personal data will be transferred from the EU to a non-EU country (NCKU is an international partner in the project) or international organisation, such transfers will be made in accordance with Chapter V of the General Data Protection Regulation 2016/679, and such transfers will comply with the laws of the country in which the data was collected.

In case of further processing of previously collected personal data, CICERONE will ensure that the beneficiary has legal grounds for the data processing and that the appropriate technical and organisational measures are in place to safeguard the rights of the data subjects.

8 References

EC, 2016: European Commission (2016), Guidelines on Data Management in Horizon 2020. [Available here](#) (accessed 5 November 2018)

9 Annex I.: Detailed information sheets for identified datasets

All detailed information sheets filled in by the partnership for the datasets identified are provided below.

Name of the dataset	FRLGIOpenPOsSurvey
Description of the dataset	<p>Collecting the inputs of POs about the interests and requirements of the platform, a value proposition:</p> <ul style="list-style-type: none"> - Data will be collected at workshops in a written format; - Then transferred to a database (.xls) and statistically evaluated using SPSS, if need data can be evaluated using concept maps - Data will be open, so it could be used for scientific research and decision-making process (e.g. higher education and research, POs, ...) - Data will be stored within Zenodo – once obtained it will be uploaded - Results will be disseminated via scientific journals (Sustainability (publication in 2020))
Format/license	<p>The data will be available in the following format(s):</p> <p>The license used for this dataset:</p> <p><input type="checkbox"/>CC0 <input type="checkbox"/>PDDL <input type="checkbox"/> CC-BY-4.0 <input checked="" type="checkbox"/>ODbL</p> <p><input type="checkbox"/>Other, please specify:</p>
Archiving/preservation	<p>The data will be made available through the following platform(s) and/or repositories: Zenodo</p> <p>The duration of the preservation will be: perpetual</p> <p>Foreseen costs of the preservation: Zenodo is free of charge</p> <p>Means to cover preservation costs:</p> <p>If the dataset contains personal data, will it be anonymised? No.</p>

Name of the dataset	FIVTTKICOpenPOsSurvey
---------------------	-----------------------



Description of the dataset	<p>Collecting the inputs of POs about the strategic role, also identifying current gaps, prepare recommendations with the POs, list of services to be included in the platform</p> <ul style="list-style-type: none"> - Data will be collected at workshops in a written format; - Then transferred to a database (.xls) and statistically evaluated using SPSS - if need data can be evaluated using concept maps - Data will be open, so it could be used for scientific research and decision-making process (e.g. higher education and research, POs, ...) - Data will be stored within Zenodo – once obtained it will be uploaded - Results will be disseminated at the Scientific conference (ERSCP 2019)
Format/license	<p>The data will be available in the following format(s):</p> <p>The license used for this dataset: <input type="checkbox"/> CC0 <input type="checkbox"/> PDDL <input type="checkbox"/> CC-BY-4.0 <input checked="" type="checkbox"/> ODbL</p> <p><input type="checkbox"/> Other, please specify:</p>
Archiving/preservation	<p>The data will be made available through the following platform(s) and/or repositories: Zenodo</p> <p>The duration of the preservation will be: perpetual</p> <p>Foreseen costs of the preservation: Zenodo is free of charge</p> <p>Means to cover preservation costs:</p> <p>If the dataset contains personal data, will it be anonymised? No.</p>

Name of the dataset	ESC-KICOpenPOsSurvey
Description of the dataset	<p>Collecting the inputs of POs about the data on capacity building (e.g. competencies required).</p> <ul style="list-style-type: none"> - Data will be collected at workshops in a written format; - Then transferred to a database (.xls) and statistically evaluated using SPSS, if need data can be evaluated using concept maps - Data will be open, so it could be used for scientific research and decision-making process (e.g. higher education and research, POs, ...) - Data will be stored within Zenodo – once obtained it will be uploaded - Results will be disseminated via scientific journals (Sustainability or Journal of Cleaner Production (publication in 2020))
Format/license	<p>The data will be available in the following format(s):</p> <p>The license used for this dataset: <input type="checkbox"/> CC0 <input type="checkbox"/> PDDL <input type="checkbox"/> CC-BY-4.0 <input checked="" type="checkbox"/> ODbL</p> <p><input type="checkbox"/> Other, please specify:</p>
Archiving/preservation	<p>The data will be made available through the following platform(s) and/or repositories: Zenodo</p>



	<p>The duration of the preservation will be: perpetual</p> <p>Foreseen costs of the preservation: Zenodo is free of charge</p> <p>Means to cover preservation costs:</p> <p>If the dataset contains personal data, will it be anonymised? No.</p>
--	---

Name of the dataset	World.Juelich.Open.International benchmark on CE(T1.2, D.1.3)
Description of the dataset	Survey on activities supporting Circular economy in extra-European countries. May be of interest for programm owners or other institution involved in supporting circular economy. The publication will occur in the frame of CICERONE as part of Deliverable 1.3. The data haben collected by the members of the project consortium involved in task 1.2 International benchmark.
Format/license	<p>The data will be available in the following format(s):</p> <p>The license used for this dataset: <input type="checkbox"/> CC0 <input type="checkbox"/> PDDL <input type="checkbox"/> CC-BY-4.0 <input checked="" type="checkbox"/> ODbL</p> <p><input type="checkbox"/> Other, please specify:</p>
Archiving/preservation	<p>The data will be made available through the following platform(s) and/or repositories: Zenodo</p> <p>The duration of the preservation will be: unlimited</p> <p>Foreseen costs of the preservation: zero</p> <p>Means to cover preservation costs: no need</p> <p>If the dataset contains personal data, will it be anonymised? yes</p>