

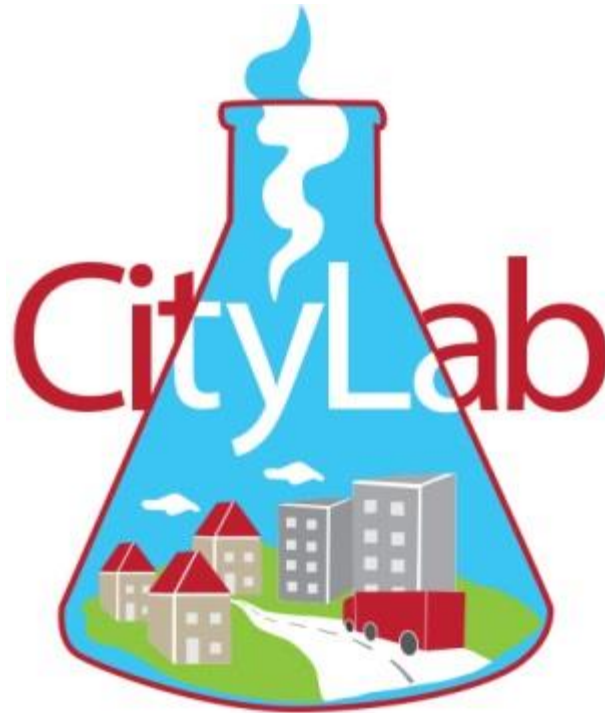
EUROPEAN COMMISSION

INNOVATION and NETWORKS EXECUTIVE AGENCY

HORIZON 2020 PROGRAMME for RESEARCH and INNOVATION

Reducing impacts and costs of freight and service trips in urban areas (Topic: MG-5.2-2014)

Grant agreement no: 635898



Deliverable 1.2

Data management plan



Document Control Sheet

Project no.:	635898	Acronym	CITYLAB
Project Title	City Logistics in Living Laboratories		
Work Package	WP1	Title:	Project management and coordination
Deliverable no.:	D1.2	Title:	Data management plan
Version	1	Revision	0
Issue Date	31 Aug 2015		
Dissemination Level	Public		
Future references	CITYLAB Deliverable D1.2 (Data management plan).		

Author(s)	Jardar Andersen and Olav Eidhammer (TOI)
Co-author(s)	Julian Allen (UoW) and Fraser McLeod (SOTON)
WP Leader	TOI
Internal Reviewer	WP leaders

Project Manager	Walter Mauritsch (INEA)
------------------------	-------------------------

CITYLAB consortium by Living Lab			
Living lab	Municipal partner(s)	Industry partner(s)	Research partner(s)
Brussels	Brussels Mobility	Procter & Gamble Services	Vrije Universiteit Brussel
London	Transport for London	TNT Gnewt Cargo	University of Westminster
Oslo	Oslo kommune	Steen & Strøm	TOI
Paris	Mairie de Paris		IFSTTAR DLR
Randstat	Gemeente Rotterdam	PostNL	TNO
Rome	Roma Capitale	Poste Italiane MeWare SRL	Università degli studi Roma Tre
Southampton	Southampton City Council	Meachers Global Logistics	University of Southampton
Networking and outreach partner			
POLIS			

Executive summary	4
1 Introduction	5
2 Data collection	6
2.1 Data in CITYLAB	6
2.2 Metadata	6
2.3 Important data management issues	7
3 Ethics and legal compliance	8
4 Storage, preservation and data sharing	9
5 Process and responsibilities	10
5.1 Process overview and responsibilities	10
5.2 Registration of data sets	11
5.2.1 Data set reference and name	11
5.2.2 Data set description	11
5.2.3 Standards and metadata	11
5.2.4 Data sharing	11
5.2.5 Archiving and preservation (include storage and backup)	12
5.2.6 Name of person responsible for data set	12
References	13
Appendix A: Data registration template	14
Appendix B: Example template for obtaining permission to use data	15

Executive summary

The objective of the CITYLAB project is to develop knowledge and solutions that result in roll-out, up-scaling and further implementation of cost effective strategies, measures and tools for emission free city logistics. In a set of living laboratories, promising logistics concepts will be tested and evaluated, and the fundament for further roll-out of the solutions will be developed. In Horizon 2020, the emphasis on data management and open data has been increased compared to earlier framework programmes.

In order to properly assess the urban freight transport environment and to understand the effects of the measures being implemented, CITYLAB deals with several types of data:

- Living lab data: Data and knowledge concerning the living lab cities will be collected and analysed in WP 2 and WP 3. These include open statistical urban freight data reflecting traffic and freight flows, location of facilities, environmental status, and data stemming from interviews with stakeholders.
- Data in models: Data will be collected to perform a robust evaluation and impact assessment.
- Implementation data: For each implementation, data will be collected in WP 4 to allow for before/after comparisons. These data relate to effects and impacts of implementations, as well as the processes themselves.
- Behavioural data: The behavioural modelling and analysis of willingness to pay requires surveys where the priorities of different actors are mapped. These data are at a more general level and neither contain personal nor commercially sensitive data.
- Transferability data: Data on critical indicators will be collected to check a possible transferability of the concept to another city.

Specific data sets within each of these groups will be further specified during the course of the project.

In this document CITYLAB establishes a first version of a data management plan (DMP) to make sure that the project data are managed in an appropriate manner. The DMP describes the data management life cycle for data sets that are collected, processed or generated by the project and defines a registration system for data sets that arise during the project, covering:

- A general description of data sets, including type of data, methods used to obtain them and file formats
- Plans for preserving and sharing data
- Storage and backup responsibilities

The basic principle is that data should be accessible to the public, and a dedicated area of the CITYLAB web site will be used for sharing publicly accessible data. Exceptions from access can be made when legitimate academic or commercial interests exist.

In cases where personal data are collected, plans for anonymisation must be defined before data collection takes place and informed consent has to be obtained from respondents of interviews or surveys.

DMPs should not be considered as fixed documents, as they naturally evolve during the lifespan of a project.

1 Introduction

The objective of the CITYLAB project is to develop knowledge and solutions that result in roll-out, up-scaling and further implementation of cost effective strategies, measures and tools for emission free city logistics. In a set of living laboratories, promising logistics concepts will be tested and evaluated, and the fundament for further roll-out of the solutions will be developed.

In Horizon 2020, the emphasis on data management and open data has been increased compared to earlier framework programmes. Some projects participate in the *Pilot on Open Research Data in Horizon 2020*, and these projects are obliged to develop a data management plan (DMP). CITYLAB is not amongst these projects, but nevertheless develops a DMP to make sure that the project data are managed in an appropriate manner.

Amongst the reasons for having a data management plan (DMP) are (Jones, 2011):

- It will be easier to find and understand the data we have in our possession, and we avoid reworking and re-collection of data
- Data sharing increases collaboration and advances research
- Increased visibility of the available data may increase the impact of the research project
- Data underlying publications are systematically maintained, allowing results to be validated

The DMP describes the data management life cycle for data sets that are collected, processed or generated by the project. DMPs should not be considered as fixed documents, as they naturally evolve during the lifespan of a project (European Commission, 2013a).

The establishment of a data management plan (DMP) for CITYLAB underlines an appreciation of the project's responsibility to manage relevant data in an appropriate manner. All CITYLAB partners have to collect, store and manage data in line with local laws and to treat data in line with the guidelines of this document.

Several principles have to be used while dealing with research data, amongst these are:

- Data protection and privacy has to be respected, and appropriate solutions for data storage and handling must be established
- Open access to data should be the main principle for projects funded by public money
- Data should be discoverable, accessible and interoperable to specific quality standards
- Integrity of the research depends on the quality of data and that data are not manipulated, and data should be assessable and intelligible.

In this document we set out a few principles for data management in CITYLAB, the structure is inspired by *DMP online* of the Digital Curation Centre¹, also recommended by the Consortium of European Social Science Data Archives (CESSDA).

The rest of this deliverable is organised as follows. Chapter 2 deals with data collection, data sets that are dealt with, and metadata. Chapters 3 and 4 deal with ethical issues and procedures for management and storing of data, respectively. Finally, Chapter 5 defines the additional data management process and responsibilities.

¹ <https://dmponline.dcc.ac.uk/>

2 Data collection

2.1 Data in CITYLAB

The European Commission (2013b) define research data as *“information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form.”*

In order to properly assess the urban freight transport environment and to understand the effects of the measures being implemented, CITYLAB deals with several types of data:

- Living lab data²: Data and knowledge concerning the living lab cities will be collected and analysed in WP 2 and WP 3. These include open statistical urban freight data reflecting traffic and freight flows, location of facilities, environmental status, and data stemming from interviews with stakeholders.
- Data in models: Data will be collected to perform a robust evaluation and impact assessment.
- Implementation data: For each implementation, data will be collected in WP 4 to allow for before/after comparisons. These data relate to effects and impacts of implementations, as well as the processes themselves.
- Behavioural data: The behavioural modelling and analysis of willingness to pay requires surveys where the priorities of different actors are mapped. These data are at a more general level and neither contain personal nor commercially sensitive data.
- Transferability data: Data on critical indicators will be collected to check a possible transferability of the concept to another city.

Specific data sets within each of these groups will be further specified during the course of the project. A registration procedure is defined for data sets in CITYLAB, see Section 5.2. To ensure that data sets are registered, the regular reporting from each living lab will contain information on data sets that are captured.

CITYLAB uses a harmonised approach for all living labs, which ensures standardisation of data collected from the different locations and implementations. This ensures interoperability of data and facilitates cross-simulation of data for improved understanding. Next, CITYLAB builds on previous projects and adapts parts of the evaluation frameworks of the FP7 projects STRAIGHTSOL and SMARTFUSION. By using similar indicator formats as previous projects, we allow for cross-comparison also with other initiatives. CITYLAB will follow established practice and international standards for data collection and preservation.

2.2 Metadata

Metadata can be defined as “structured or semi-structured information which enables the creation, management and use of records [i.e. data] through time and within and across domains” (Day, 2005). Metadata facilitates exchange of data by making them more detectable, and makes it easier to organise, reproduce and reuse data.

Metadata will be defined for data sets that are collected as part of the project.

² As much of this data will be provided by city authorities or companies, it is likely that the ownership of this data will reside with them rather than the CITYLAB project.

2.3 Important data management issues

The European Commission (2013a) defines a set data issues that should be addressed for data sets that are dealt with, these are summarised in Table 1.

Table 1. Key data requirements and DMP questions. *Source: European Commission (2013a).*

Data requirements	DMP question
Discoverable	Are the data and associated software produced and/or used in the project discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier)?
Accessible	Are the data and associated software produced and/or used in the project accessible and in what modalities, scope, licenses (e.g. licencing framework for research and education, embargo periods, commercial exploitation, etc.)?
Assessable and intelligible	Are the data and associated software produced and/or used in the project assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review (e.g. are the minimal datasets handled together with scientific papers for the purpose of peer review, are data provided in a way that judgments can be made about their reliability and the competence of those who created them)?
Usable beyond the original purpose for which it was collected	Are the data and associated software produced and/or used in the project usable by third parties even a long time after the collection of the data (e.g. is the data safely stored in certified repositories for long-term preservation and curation; is it stored together with the minimum software, metadata and documentation to make it useful; is the data useful for the wider public needs and usable for the likely purposes of non-specialists)?
Interoperable to specific quality standards	Are the data and associated software produced and/or used in the project interoperable allowing data exchange between researchers, institutions, organisations, countries, etc. (e.g. adhering to standards for data annotation, data exchange, compliant with available software applications, and allowing combination with different datasets from different origins)?

3 Ethics and legal compliance

In cases where personal data are collected, plans for *anonymisation* must be defined before data collection takes place.

For data that are collected or considered reused from existing sources, the necessary rights to use the data have to be obtained. If data are planned to be shared publicly, we have to make sure that we have the right to do so. *Informed consent* is crucial, where respondents of interviews or surveys are made aware of the plans for use of the data and the rights they have to withdraw, etc. Before data are collected, plans for future use have to be discussed, so that participants in surveys and interviews may be informed on these plans and agree to it. Appendix B contains a simple example template for obtaining the right to use data.

Data are owned by the party that generates them, principles for intellectual property rights are defined in the CITYLAB Consortium Agreement. Proprietary data gathered by a consortium member remains in the care of that consortium member, and will not be distributed to any other consortium member or any party outside of the consortium. Processing and use of data will follow Directive 95/46/EC (the data protection directive) and the “General Data Protection Regulations law”. In addition, each CITYLAB partner is obliged to collect and manage data in line with national legislation.

Integrity of the research depends on the quality of data and that data are not manipulated, it is required that all CITYLAB partners refrain from such manipulation.

4 Storage, preservation and data sharing

All non-public data will be stored in secure environments at the locations of consortium partners with access privileges restricted to the relevant project partners. Non-public data will not be stored through Dropbox, Google Docs or other third party cloud-based services.

CITYLAB is committed to distribute results and publications via Open Access publishing and has allocated dedicated resources for this. Consortium partners will seek to publish results in open access journals to widen the target audience of the project's results. Consortium partners will publish results in scientific journals that can assure such open access without restriction.

The basic principle is that data should be accessible to the public, and a dedicated area of the CITYLAB web site will be used for sharing publicly accessible data. Exceptions from access can be made when legitimate academic or commercial interests exist, and such issues will be handled by the Management Committee. One such example is financial implementation data where protection of information revealing, for instance, industry partners' general cost structure or competitive conditions may be needed. Possible methods by which proprietary data could be made publicly available include referring to relative changes rather than absolute values, aggregation and anonymization.

In CITYLAB's WP 2 it is planned to develop an observatory for urban logistics, and this will be one mechanism for sharing data. The observatory will be connected to the web site hosted by University of Southampton.

For many previous European projects, it has been difficult to reuse the findings because the web sites have closed down after the projects' end dates. The CITYLAB web site will be planned in such a way that before the project ends, a post-project phase version will be established to facilitate access to project data.

5 Process and responsibilities

This chapter describes the process for ongoing management of data in CITYLAB.

5.1 Process overview and responsibilities

Each CITYLAB partner has to respect the policies set out in this data management plan. Data sets have to be created, managed and stored appropriately and in line with national legislation. University of Southampton has a particular responsibility to ensure that data shared through the CITYLAB web site are easily available, but also that backups are performed and that proprietary data are secured.

Monitoring and registration of data sets is the responsibility of the partner that generates the data. In Section 5.2 the template for registration of data sets is described; the full template is available in Appendix A. When a partner is ready to register a new data set, they should send the requested information to the Project Coordinator who will update the template in CITYLAB's Sharepoint site. This can be done at any time, but it will also be possible to inform about new data sets as part of the regular living lab reporting.

The partner that generates the data is also responsible for obtaining the necessary rights to use and share the data. Appendix B contains a simple example template for obtaining the right to use data.

Quality control of the data is the responsibility of the relevant WP leader, supported by the Project Coordinator.

If data sets are updated, the party that possesses the data has the responsibility to manage the different versions and to make sure that the latest version is available in the case of publically available data. When data sets are registered, a person with responsibility for the data set has to be named. This can be changed later, for instance if the physical location of the data are changed.

Table 2 summarises the main data management responsibilities.

Table 2. Overview of data management responsibilities.

Activity	Responsible
Registration of data sets	Partner generating the data set
Ensure that rights to use (and if applicable share) the data are obtained	Partner generating/introducing the data set
Keep overview of data sets at Sharepoint	Project Coordinator
Quality control	Relevant WP leader
Version control for files	Person defined to have data set responsibility (see Section 5.2.6).
Backing up data	Organisation possessing the data. For data shared through the web site University of Southampton is responsible
Security and protection of data	Organisation possessing the data. For data shared through the web site University of Southampton is responsible

In the case of conflicts or issues that need discussion or voting, the Management Committee will be consulted.

5.2 Registration of data sets

Following the information in Chapter 2 and specific advice on data management of the European Commission (2013a), a template for registration of data sets has been established. The template can be found in Appendix A. Below we explain each of the elements that has to be described in the template.

The registration of data should not be a complicated or complex task, and we have therefore made a short version of the template emphasising what we believe is most crucial.

Completed templates should be sent to the Project Coordinator who will keep the information on the CITYLAB data sets up to date.

5.2.1 Data set reference and name

An identifier has to be included (data sets are numbered consecutively) as well as an appropriate name of the dataset. A data set can be defined as (Wikipedia) “*a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question*” or alternatively as “*data in a collection of closely related tables, corresponding to a particular experiment or event*” (also Wikipedia). Depending on the nature of the data or information covered, both alternatives can be applicable in CITYLAB.

5.2.2 Data set description

A proper description of the data should be included. The description should cover what the data represent, its source (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse should be described if applicable. If data from other sources are reused, this should be clearly specified.

The data have to be properly described in terms of:

- Type of data (for example experimental, observational, raw or derived)
- Methods used to obtain them (for example manual collections, models, simulations)
- File format (for example text files, images, audio, etc.) and whether non-standard software is needed for further processing of data

5.2.3 Standards and metadata

Metadata are “*data that provides information about other data*”³ describe the contents of data files and the context in which they have been established. Several metadata standards exist (see https://en.wikipedia.org/wiki/Metadata_standards). Proper metadata facilitates use of the data by others, makes it easier to combine information from different sources, and ensures transparency.

5.2.4 Data sharing

Describe plans for sharing data. Describe how data will be shared (including access procedures and embargo periods), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use. Please also define whether access will be widely open or restricted to specific groups. Identify the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).

³ Merriam Webster dictionary

If the dataset cannot be shared, the reasons for this should be elaborated (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related or security-related).

5.2.5 Archiving and preservation (include storage and backup)

Describe procedures that will be put in place for long-term preservation of the data. Indicate for how long the data should be preserved, and where the data will be stored. If applicable, plans for destruction of data should be described. This information should be available for each data set, but procedures for backup will most likely be similar for multiple data sets stored in the same location,

5.2.6 Name of person responsible for data set

For each data set a specific responsible person (and belonging institution) has to be defined. This person will be responsible for version control, answer questions related to the data set, and for ensuring data security and backup of the data. Responsibility for security and back-up can be transferred to other persons/organisations if appropriate, for instance if a data set is shared through the web site.

References

Day, M. (2005). DCC – Digital Curation Manual. Instalment on Metadata. Available from <http://www.dcc.ac.uk/resource/curation-manual/chapters/metadata>.

European Commission (2013a). Guidelines on Data Management in Horizon 2020. Version 1.0, 11 December 2013.

European Commission (2013b). Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Version 1.0, 11 December 2013.

Jones, S. (2011). 'How to Develop a Data Management and Sharing Plan'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides>.

Appendix A: Data registration template

Will be made available in the CITYLAB Sharepoint folder “/data”.

Data set reference:	
Name:	
Description ¹ :	
Standards and metadata ² :	
Data sharing ³ :	
Archiving and preservation ⁴ :	
Name of person responsible for data set (version control, etc):	

1: Description of the data, origin of data (if collected); nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.

2: Describe the contents of data files and the context in which they have been established.

3: Describe how data will be shared (including access procedures) and identify repository where data will be stored.

4: Describe the procedures that will be put in place for long-term preservation of the data (duration of preservation, approximated end volume, where data will be stored).

Further details and explanations can be found in Section 5.2 of Deliverable 1.2.

Appendix B: Example template for obtaining permission to use data

CITYLAB (City Logistics in Living Laboratories) is a 3-year research project funded by the European Commission through the Horizon 2020 program. The core of CITYLAB is seven living laboratories where cities work as contexts for innovation and implementation processes for public and private measures contributing to increased efficiency and sustainable urban logistics. The outputs from the living labs will include best practice guidance on innovative approaches and how to replicate them. Further information is available from www.citylab-project.eu.

The project depends on extensive information and data collections that will contribute to increased understanding of urban freight transport. We are grateful for your willingness to support our work.

Name of company and contact person:	
Description of data or information shared with CITYLAB:	
Conditions for sharing or particular restrictions applied to further use of the data:	
Date:	
Signature:	