Reflow | Co-Creating Circular Resource Flows in Cities

'constRuctive mEtabolic processes For material flOWs in urban and peri-urban environment across Europe'

Deliverable 7.2

DATA MANAGEMENT PLAN

Due date of deliverable: 30/11/2019
Actual submission date: 29/11/2019
Start date of project: 01/06/2019 Duration (36 Months)
Dissemination Level: Public ✓

DELIVERABLE

Work Package	WP7 Dissemination, Exploitation & Sustainability
Deliverable	D7.2 Data Management Plan
Document Name	Data Management Plan
Due Date	M6: 30 November 2019
Submission Date	M6: 29 November 2019
Dissemination Level	[x] P – Public [ ] CO – Confidential
Deliverable Lead	Institute for Advanced Architecture of Catalonia (IAAC)
Author(s)	Milena Juarez, Guillem Camprodon, Pablo Muñoz
Point of Contact	Milena Juarez, milena@fablabbcn.org; Guillem Camprodon, guillem@iaac.net; Pablo Muñoz, pablo@fablabbcn.org
Reviewers	Elena Japundzic (DYNE), Zartashia Ahmed (CBS), Yury Glikman (FOKUS)
Status	[ ] Plan [ ] Draft [ ] Working [ ] Final [X] Approved
Abstract (for public dissemination only)	The Data Management Plan (DMP) of the REFLOW project outlines how the project data will be handled during the lifetime of the project and after the project is completed. It gives an overview of which data will be collected, processed and/or generated, which methodology and standards will be applied, whether and how this data is shared and/or made open access, and finally how it is curated and preserved. This initial version of the DMP provides a set of standard guidelines to ensure that all the REFLOW partners manage their data following the FAIR Data principles while being compliant with the EU GDPR. The DMP is a living document and will be updated, whenever significant changes arise, to support the data management lifecycle for all data generated in the project.
Keywords	Data management, database, research, innovation
Statement of Originality	This deliverable contains original unpublished work except where clearly indicated otherwise.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.

Revision History

Revision	Date	Author(s)	Organisation	Description
D7.2 v0.1	01 November 2019	Milena Juarez; Guillem Camprodon; Pablo Muñoz	IAAC	*First draft*
D7.2 v0.2	15 November 2019	Elena Japundzic; Zartashia Ahmed; Yury Glikman	DYNE, CBS, FOKUS	*Change by partners*
D7.2 v1.0	29 November 2019	Milena Juarez; Guillem Camprodon; Pablo Muñoz	IAAC	*Final version by the consortium to be submitted to the EC*
D7.2 v0.1	01 November 2019	Milena Juarez; Guillem Camprodon; Pablo Muñoz	IAAC	*First draft*

Abbreviations

AI	Adobe Illustrator Artwork
API	Application Programming Interface
CE	Circular Economy
CVS	Comma Separated Values
DMP	Data Management Plan
DOC	Document File
DPO	Data Protection Officer
EPS	Encapsulated PostScript
EU	European Union
GA	Grant Agreement
GDPR	General Data Protection Regulation
HTML	Hypertext Markup Language
IPR	Intellectual Property Rights
IT	Information Technology
JSON	JavaScript Object Notation
ODF	Open Document Format
ORD	Open Research Data
OS	Operating System
PCO	Project Coordinator
PDF	Portable Document Format
PPT	PowerPoint Presentation
REST	Representational State Transfer
RTF	Rich Text Format
SGM	Standard Generalized Markup
SOAP	Simple Object Access Protocol.
TXT	Plain Text
WP	Work Package
XML	Extensible Markup Language

Table of Content

1. Introduction – Overview of REFLOW

REFLOW is an EU H2020 funded project, from 2019 to 2022, that seeks to understand and transform urban material flows and to co-create and test circular and regenerative solutions at business, governance and citizen levels. It has received funding from the European Union Innovation and Research Program, under the Grant Agreement No. 820937.

The vision of REFLOW is to develop circular and regenerative cities through the re-localisation of production and the re-configuration of material flows at different scales. More specifically, it will use Fab Labs and makerspaces as catalysers of a systemic change in urban and peri-urban environments, which enable, visualize and regulate "four freedoms": free movement of materials, people, (technological) knowledge and commons, in order to reduce materials consumption, maximize multifunctional use of (public) spaces and envisage regenerative practices. Concretely, REFLOW aims at providing realistic best practices aligning market and government needs in order to create favourable conditions for the public and private sector to adopt circular principles. In order to provide critical examples of ways in which cities can adopt a CE model and reach the 2030 Sustainable Development Goals, REFLOW will create new CE business models (Distributed Design Market model, On-Demand System, Corporate Hacking and Corporate Pyramid) within 6 pilot cities (Amsterdam, Berlin, Cluj-Napoca, Milan, Paris, Vejle) and assess their social, environmental and economic impact.

The project will make use of blockchain technologies in order to incentivise the circular practices in local ecosystems and data visualisation tools to enable continuous monitoring and optimisation of "urban metabolic" processes and rapid interventions management. Networks of sensors, urban computing and geo-localisation will capture data ensuring accuracy, integrity and interoperability of relevant data infrastructures, while data visualisation and standard templates will be available for effective communication, public consultation, and exchange of experiences.

2. Scope of the Data Management Plan

The REFLOW project plans to collect data from the pilot cities (Amsterdam, Berlin, Cluj-Napoca, Milan, Paris and Vejle) and from Europe in general. The data will be processed and treated only by project partners. This Data Management Plan (DMP) details the data the project will generate and manage, following what methodology and standards, whether and how the data will be exploited or made accessible for verification and re-use, and how it will be curated and preserved with respect to the FAIR principles and the EU GDPR. The signed Consortium Agreement defined the ownership of prior and post key knowledge (IPR, data etc.) of all involved parties (including partners and cities).

The Data Management Plan is a living document that will be regularly updated to support the data management lifecycle for all data that is collected, processed or generated by the project. The DMP will be updated whenever significant changes arise, such as the inclusion of a new category of data.

Data management is an important aspect of European projects as it ensures long-term preservation and accessibility of data during the project and after the project has ended. The DMP also outlines how the data is handled during the lifetime of the project and after it has been completed. In addition, the DMP seeks to reduce the risk of data loss or other threats that could render the data illegible or unusable.

The REFLOW consortium will conform to the DMP template and will make every effort to make this content discoverable, accessible, intelligible, and usable by all interested and legally permitted (based on this DMP and other legal documents) stakeholders (especially by other EU related projects).

2.1 Related Policies

The present DMP has been developed according to the following EU regulations regarding Research Data and Data Protection:
1. The Open Research Data Pilot ("ORD pilot")
2. The General Data Protection Regulation

2.1.1 Open Research Data Pilot (ORD pilot)

The European Commission is running a flexible pilot under Horizon 2020 called the **Open Research Data Pilot** (ORD pilot). The Open Research Data Pilot aims to make the research data generated by Horizon 2020 projects accessible with as few restrictions as possible, while at the same time protecting sensitive data from inappropriate access. According to this pilot, participating projects must submit a first version of the DMP (as a deliverable) within the first 6 months of the project. The DMP is a living document and will be updated whenever significant changes arise.

With the objective to provide open access for the public deliverables and to boost the visibility of the REFLOW outcomes as an European project, the Grant Agreement (GA) presents an entire Article for the dissemination of results and open access. Open access refers to the practice of providing online access to scientific information (e.g., peer-reviewed scientific research articles and research data) that is free of costs or other barriers to the end-users and reusable.

In relation to the above-mentioned, the following articles are important to consider and restate in the following deliverable.

Article 29.1 is related to the consortium's obligation to disseminate results. This Article states that, unless it goes against the beneficiaries' interests, each beneficiary must – as soon as possible – 'disseminate' its results by disclosing them to the public by appropriate means (GA, pp. 48).

Lastly, Article 29.3 focuses on the digital research data generated in the project and specifies the beneficiaries' obligations. It encompasses the commitment to deposit the results in a research data repository, making sure that third parties can access, mine, exploit, reproduce and disseminate the content (free of charge for any user). Article 29.3 also includes that the beneficiary must provide information — via the repository — about tools and instruments necessary for validating the results (and — where possible — provide the tools and instruments themselves) to ensure that the project results can be further exploited in the future (GA, pp. 49).

However, the Commission also recognises that there might be good reasons to keep some or even all research data generated in a project closed. It follows the principle "as open as possible, as closed as necessary".

2.1.2 General Data Protection Regulation (GDPR)

The **General Data Protection Regulation (GDPR)** [(Regulation (EU) 2016/679)](#) is a regulation through which the European Parliament, the European Council and the European Commission intend to strengthen and unify data protection for all individuals within the EU. It also addresses the export of personal data outside the EU and EEA territories. The purpose of the GDPR is to provide a set of standardised data protection laws across all member countries. This should make it easier for EU citizens to understand how their data is being used, and for them to raise any complaint regarding the treatment of their data, even if the citizen is not in the country where the data is located. GDPR therefore works on two levels, (1) the ethical use of personal data, and (2) keeping that personal data secured.

GDPR is therefore important because it improves the protection of European data subjects' rights and their personal data, including its ethical uses. REFLOW project will comply with GDPR in every action in order to apply appropriate technical and organisational measures to implement the data protection principles.

In order to comply with the GDPR, personal data including gender, age, social level, education, ethnic origin, family composition, or any other data that could potentially identify the research subject will be pseudonymised in REFLOW. It is the responsibility of the researcher to make sure that the personal data is pseudonymised as quickly as possible after its collection, and that it remains like that (see also D9.2).

Personal data will be collected in different WPs during REFLOW project, including observational, experimental and derived data, as defined in section 3.2 of the DMP. All partners shall manage their data following the FAIR data principles while being compliant with the EU GDPR. As stated in D9.2, "the Data Protection Officer (DPO) will be the contact point for the REFLOW consortium executing the tasks envisaged in Article 39 GDPR with regard to project implementation. He will help the REFLOW consortium ensure that all data collection and processing are carried out in such a way that it complies with relevant European Directives". Along the Project Coordinator (PCO), the DPO will ensure the proper management and processing of all data in the project, complying with the regulations set out by the EU.

Personal data will never be disclosed. Access will be performed via a cloud operated storage system (SharePoint), to which only the research team and the coordination team (CBS) will have access. If, due to the nature of data, any other platform (section 3.5 details several additional platforms that may be used) is used to store data, this must be communicated to the DPO.

Personal data collected directly from a research subject (e.g. an interview) must be pre-emptively accepted by the participants themselves through the procedures of informed consent, as detailed in D9.1. All offline data that includes personal data will be digitalised and uploaded to the main storage system (SharePoint) as soon as possible by the researcher responsible of its collection. Once this is done, offline personal data must be destroyed.
In REFLOW, personal data will not be kept for longer than 5 years after the project ends (unless otherwise required by relevant national legislation). More details on personal data storage and use for future research can be found in D9.2.

3. Data Summary

This section gives an overview of the scope of the project (purpose and objectives) in order to clarify the relation between it and the data generated, collected and/or processed in the project.

3.1. Purpose of the data (collection and generation)

REFLOW aims to support the development of regenerative and circular cities, and adopt CE practices, through the re-localisation of production and the reconfiguration of material flows. To do so, each work package has a critical role to play, and will collect with the purpose of providing guidance for public and private sector organisations on how to create favourable conditions for the adoption of CE practices.

In WP3 for instance, a clear understanding of the city material and information flows from the pilot cities has to be mapped in order to co-create a circular strategy that starts from local production ecosystems and scales up through the articulation of networks of stakeholders and cities.

In other WPs, experimental and observational data will be generated from the pilot cases and included in the project's dashboard tool to make it accessible to the rest of the project's partners. This data, along with other derived data from cases studies (e.g. WP4), literature review (e.g. WP1), etc. will be structured into information and uploaded to REFLOW's main storage system. Additionally, other storage platforms may be used, depending on the nature or purpose of the data. This information includes models and ICT infrastructure developed to test potential CE strategies for re-localisation of production and re-configuration of material flows at different scales. Compilation and analysis of this information will result in knowledge. This will be disseminated in the shape of toolkits, reports or scientific articles, accessible to public and private sector organisations. Third parties will be able to access, mine, exploit, reproduce and disseminate data and information generated or derived throughout the project (Article 29.3), according to REFLOW's ethics, privacy and data protection requirements as well as its knowledge dissemination strategy.
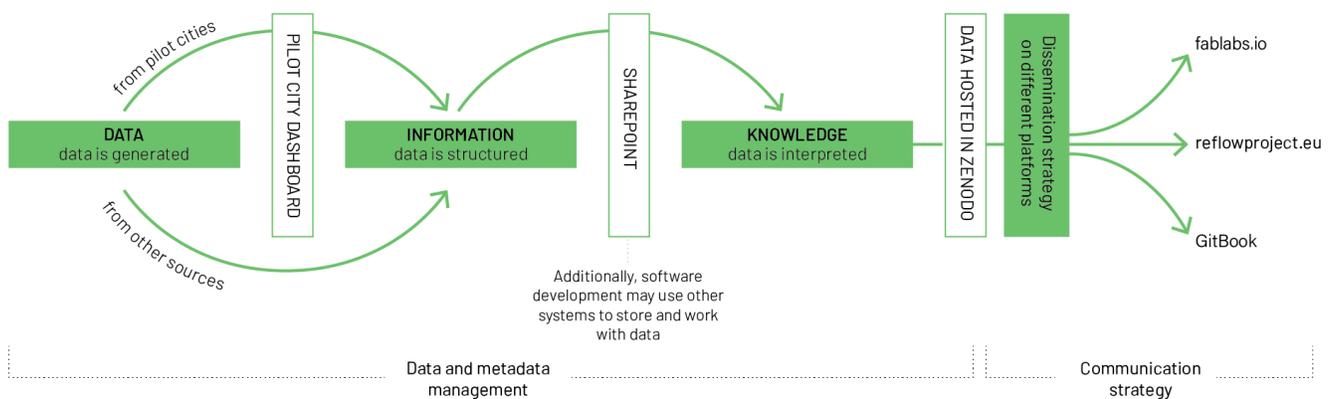

Figure 1: Data flow in REFLOW project

Figure 1 shows the data flow in REFLOW, considering the main steps it follows until public release. Experimental data generated by the pilot cases will be initially stored in REFLOW's Pilot City dashboard tool and, when structured, uploaded to SharePoint, along with its metadata information. Data from other types of sources will also be stored in SharePoint, however other platforms, such as GitHub, may be needed when the nature of data requires a different treatment (e.g. a source code from an OS developed for REFLOW). Once data is structured and analysed, it will be made publicly available on the Zenodo platform, as the research data repository recommended by the EC. Public data will also be disseminated through different online services, including REFLOW website (reflowproject.eu), GitBook and fablabs.io.

3.2. Types of data

Given the project's aim, different types of data coming from multiple sources will be collected, processed and/or generated in the project. This data will mainly – but not exclusively – consist of:

- **Project management data;** used for internal management of the project and its partners. It includes minutes, mailing lists or protocols among other types of datasets.
- **Observational data;** captured in real time by researchers or other stakeholders in the consortium. It includes social, economic and technical aspects of each pilot scenario. Observational data includes survey results, images, data collected on public events organised by the project, interviews or statistics generated by the project, among other types of datasets.
- **Experimental data:** from labs and equipment, can often be reproduced. It includes datasets with stakeholders' inputs about production and collection, sorting, recycling and end-of-life qualities and properties of the materials for each pilot city. Experimental data includes environmental and technical information obtained under experiments, monitoring, sensors, etc. It will be gathered through the open data dashboard, an internal project tool used for monitoring and exchanging information.
- **Simulations and technical data**: It is obtained from models. It can usually be reproduced if the input data is known. Simulations and technical data include material flow or economic models, among other types of data
- **Derived data**: including compiled data from other sources or case studies. It includes data derived from REFLOW's online platforms, databases generated after mining data or statistical analysis, toolkits or reports that use REFLOW's generated data, among other types of data.
- **ICT infrastructure data:** including the information related to technology (IT) infrastructure developed for REFLOW. It includes technical data related to the open data dashboard, the development of open source software and an operating system. It will involve the development of technologies and systems through the use of publicly shared design and technological information. ICT infrastructure data includes source code, APIs and graphical interfaces, among other types of data.

3.3. Origin of data

REFLOW is building on previous experience and knowledge from partners and connects with existing networks. Through experimentation and co-creation within pilot cities, REFLOW will investigate a broad range of topics from different sources in order to boost a new metabolism of cities through the reconfiguration of materials flows, and by creating CE

business models and assessing their social, environmental and economic impact. As specified in section 3.2, data will be collected from many sources, such as literature reviews, interviews, surveys, activities with pilots and stakeholders, online sources, workshops, meetings and events, field measurements, previous research projects, open source platform, legal documents and policies. Due to the nature of each material flow, pilot data will vary depending on each pilot case. The specific origin of data in each pilot case is not known yet, hence other sources may be used throughout the project.

3.4. Data formats

The DMP describes the data that will be generated by REFLOW project and how it will be published. This document is nevertheless written when most tasks are yet to be developed. Therefore, new data formats may appear during the project development. These will be included in the REFLOW dataset tool (see annex 1), and if significant changes arise, the DMP will also be updated.

The DMP follows the EC Open Research Data Pilot (ORDP). Therefore, it must be possible for third parties to access the data generated (in its various formats) during the project's development and in the future. Formats that allow the use of open source software are prioritised, as it ensures that re-usability of the data in the future. Moreover, the specific software needed to view or edit the datasets is to be specified in their metadata information. If multiple software are available for one data format, open source is preferred and should be mentioned.

Reports, surveys, interviews, databases, data collected in public events or statistics may use spreadsheets, text, and presentation formats as they are easily accessible by third parties and researchers who want to further use or reproduce the data. Nevertheless, raw data may not be possible to store under these formats. In these cases, or when data requires to be saved in other specific formats, datasets will be stored under open software formats preferably. Most data types, including project management, observational, experimental, simulations and derived data, will use mainly – but not exclusively – the following formats to store data:

- **Spreadsheets:** Many authorities store information in spreadsheets, for example Microsoft Excel. This data can often be used immediately, given the informations' descriptions in the rows and columns. However, in some cases there can be macros and formulas in spreadsheets, which may be somewhat more cumbersome to handle. It is therefore advisable to document such calculations within or alongside the spreadsheet, since it is generally more accessible for users to read.
- **Comma Separated Values (CSV):** CSV files can be a very useful format because it is compact and thus suitable to transfer large sets of data with the same structure. However, the format is so limited that data can often be useless without documentation since it can be almost impossible to guess the significance of the different columns. It is therefore particularly important for the comma separated values formats that documentation or meta-information of the individual fields is provided and is sufficient and accurate.
- **Text Documents:** Formats like DOCX, RTF, ODF, or PDF are appropriate for certain kinds of documents, e.g., scientific journals, deliverables, reports, etc. Templates may be used whenever possible, so that displayed data can be re-used.

- **Presentation (.ppt):** PPT is a file extension for a presentation file format used by Microsoft PowerPoint. The popular presentation software is commonly used for project slide shows.
- **Plain Text (TXT):** Plain text documents (.txt) are chosen because they are very easy to read and process via plain text parsers. They generally exclude structural metadata.
- **Image or photo formats (.jpg, .png):** Image file formats are composed of digital data in one of these formats that can be rasterized for use on a computer display or printer. An image file format may store data in uncompressed, compressed, or vector formats.
- **Audio formats (.wav, .mp3, .ogg):** An audio file format is a file format for storing digital audio data on a computer system. The audio coding format can be uncompressed, or compressed to reduce the file size.

Derived data used to disseminate knowledge will be accessed through online platforms. This will include mainly - but not only - the following formats to store data:

- **Web Services:** For data that changes frequently, and where each scope is of limited size, it is very relevant to expose data through web services. There are several ways to create a web service, but some of the most used is SOAP and REST.
- **Video formats (avi):** A video file format is a type of file format for storing digital video data on a computer system. A video file normally consists of a container, containing video data in a video coding format, alongside audio data in an audio coding format. The container can also contain synchronization information, subtitles, and metadata such as title.
- **Scalable Vector Graphics (.svg):** It is an Extensible Markup Language (XML)-based vector image format for two-dimensional graphics with support for interactivity and animation. It can be used and viewed with several software, including open source one. For this reason, it will be preferably used over .ai formats to store diagrams and graphic data.
- **Adobe illustrator (.ai):** Adobe Illustrator Artwork (AI) is a proprietary file format developed by Adobe Systems for representing single-page vector-based drawings in either the EPS or PDF formats.

Specific datasets are linked to ICT infrastructure data type, including OS, platforms and ICT tools developed by REFLOW, e.g. dashboard tool to store pilots' data. Even though these formats may be used in other data types as well, such as derived or experimental data, they are listed here as they will mainly be used for ICT infrastructure data. Due to their simplicity, JSON files will be preferred, but other formats will be used as well, depending on the data requirements. ICT infrastructure data type will use mainly – but not exclusively – the following formats to store data:

- **JavaScript Object Notation (JSON):** JSON is a simple file format that is very easy for any programming language to read. Its simplicity means that it is generally easier for computers to process than other formats, such as XML. If possible, JSON files will be used to store ICT infrastructure data.
- **Extensible Markup Language (XML):** XML is a widely used format for data exchange. It offers good opportunities to preserve the data's structure and how files are built while also allowing developers to write parts of the documentation in with the data without interfering with how they are read.

- **Standard Generalized Markup (SGM)**: SGM files are saved with the .sgm file extension and written in the SGM programming language. The SGM language is a document generation language that is used to share information digitally using custom tags and DTD file references.
- **Hypertext Markup Language (HTML):** Nowadays much data is available in HTML format on various sites. This may well be sufficient if the data is very stable and limited in scope. In some cases, it could be preferable to have data in a form easier to download and manipulate, but as it is cheap and easy to refer to a page on a website, it might be a good starting point in the display of data.

The file formats listed above are classified according to the data type which most likely will use them. Nevertheless, they may be used interchangeably in other data types or new formats will be used in the project, according to the partners' requirements. These will be specified in REFLOW dataset tool and, if significant changes arise, the DMP will be updated.

3.5. Data storage and sharing

REFLOW collaborates with 28 partners, hence a large amount of data is being - and will be - generated. To manage such quantities of data, and allow the partners (and other researchers) to share them with each other, suitable storage media shall be used to comply with European privacy legislation.

All possible datasets must be stored in the main storage system (REFLOW SharePoint). This platform will be the main system for storage of active research and project management data. Due to some datasets nature (e.g. a dataset that contains sensitive personal data, a source code of an OS being developed for the project), these may use additional offline or online storage systems. This will be communicated to the DPO (contact information can be found in D9.2) and the coordination team, as well as indicated in REFLOW's dataset tool, once the dataset is uploaded.

Zenodo will be used as the main platform for long-term preservation of results and research data. Additional platforms, such as fablab.io, GitBook or REFLOW's website will be used to publish and disseminate data, along with publications in scientific journals, according to the communication and dissemination strategy (D7.1).

Below we will explain the different storage systems used in REFLOW.

3.5.1. Storage systems

The storage systems will include mainly - but not only - the following platforms:

REFLOW SharePoint
All possible datasets will be uploaded and maintained on a cloud-operating server, REFLOW's SharePoint storage. SharePoint will act as the project repository. It is a web-based collaborative platform that integrates with Microsoft Office. It is limited to a core set of collaboration, file hosting, documents and content management scenarios, and is updated by all partners on a frequent basis, recording also the metadata of each uploaded dataset in REFLOW dataset tool. The primary account holder is held by the coordination team. SharePoint is furthermore GDPR compliant.

Access to REFLOW's SharePoint is limited to consortium members and the coordination team (CBS). A few folders (e.g. informed consent forms) are however restricted to the research teams and the coordination team only. CBS will be responsible for making such restricted folders.

The consortium will store all project data on the project's SharePoint created by CBS, including:
- Grant Agreement and Consortium Agreement;
- Project reports and deliverables (public and non-public);
- Personal and sensitive user data;
- Surveys and interviews (including digitised versions of physical forms);
- All dissemination-related material;
- Projects templates;
- Consent forms signed by participants;
- Notes and minutes from meetings;
- Pictures taken during conferences, workshops and seminars;
- REFLOW dataset tool: An updated live document of record keeping of the data collected.

As specified in D9.2, REFLOW data will be kept for enough time to allow for final publications (maximum of 5 years) unless otherwise required by relevant national legislation. Furthermore, information about the storage limitation for each specific personal data set shall be included in all data protection policies and notices.

Additional storage platforms in which REFLOW data will be stored may be:

Zenodo Data repository
In order to reach the research and professional community, additionally to the main storage system, all public data will be made available on the Zenodo digital repository during the development of REFLOW project and at the end of it. Zenodo was launched at the CERN Data Centre in May 2013 with a grant from the European Commission with a special commitment to sharing, citing and preserving data and code. As a digital repository, Zenodo registers DOIs for all submissions through DataCite to make them easily and uniquely citable. The platform is based on the Invenio open-source software, Zenodo profits from and contributes to the foundation of code used to provide Open Data services to CERN and other initiatives around the world. A specific Zenodo community for REFLOW has been created by the Coordination team and will store all public data related to the project.

On premises - Partners internal facilities
Some of the existing and ongoing project datasets are stored at the premises of the owning consortium partners, including software, tools and models. It is the responsibility of the partner to ensure that such software complies with the GDPR principles and respect the requirements presented in the Ethics Deliverables (D9.1, D9.2 and D9.3). To ensure that the software complies with GDPR, the partners are encouraged to send an email to the DPO (contact information can be found in D9.2) and the coordination team.

Where possible, the partners are encouraged to share the data from their internal facilities to SharePoint as this is the project's main repository.

3.5.2. Sharing systems

The sharing systems will include mainly - but not only - the following platforms:

REFLOW Website
The consortium will disseminate all applicable data on REFLOW's website (reflowproject.eu), which is regularly updated by the WP7 Leader (IAAC), Project Coordinating Team (CBS) and all the WP Leaders who have administrator rights. The website, setup and maintained by IAAC, will be a front-end platform to disseminate knowledge and public data developed in REFLOW. In order to make it possible for third parties to access, mine, exploit, reproduce and disseminate data, the REFLOW website will always link from Zenodo Data repository.

GitHub
The REFLOW open source software involves the development of technologies and systems through the use of publicly shared design and technological information. The open data dashboard will be co-created and maintained in collaboration with communities of users enabling citizens, grassroot organisations, associated with small and medium-sized enterprises (SMEs) and municipalities to develop customised technical solutions to emerging CE challenges and opportunities.

The sustainability of the adopted technology will be guaranteed by the inclusion of the open source code in GitHub, the largest online development platform, where it will be shared and further improved on by other developers.

Gitbook
REFLOW will develop an online asset using the Gitbook digital platform for its Handbook. This useful guide will be a practical resource to assist cities in adopting circular economy strategies. It will compromise a collection of tools and instructional resources to implement CE practices in design, innovation and digital fabrication processes. Data in the Gitbook will be shared and treated in accordance with this DMP and other regulations and legal documents that are relevant for the data of the citizens (subjects) in question.

Fablabs.io
Fablabs.io is the online social network of the international Fab Lab community. It is the current official list of Fab Labs that share the same principles, tools, and philosophy around the future of technology and its role in society. Fablabs.io is an exchange platform for people, labs, projects, machines, events and groups that operate around the Fab Lab Network, which collaboration and communication tools in order to align interests and to expand the global scale of this community. Data in the Fablabs.io will be shared and treated in accordance with this DMP and other regulations and legal documents that are relevant for the data of the citizens (subjects) in question.

Before any service contract with other external platforms used within the project, acceptance must be obtained by the PCO who is in charge of the overall communication of the project.

3.6. Data Utility

The data collected and generated in REFLOW will be useful for three main groups: citizens, who wish to participate in co-creating their city to make it environmentally, socially and economically vibrant; businesses, which include organisations, manufacturers and startups of all sizes that want to partake in future-proofing their industries by supporting social, economic and environmental transformation; and governments, which are seeking to adopt holistic, innovative policies that represent today and tomorrow's urban citizen in each of the Social-Technological-Economic-Environmental-Political (STEEP) aspects.

REFLOW will describe systemic solutions, co-design, test and implement them in the pilot cities. Each solution will embed CE principles to produce the regeneration of urban loops in terms of optimisation of flows and less consumption of resources. Following open source principles for each solution, documentation and data will be provided so that interested private and public organisations will be able to easily replicate and adopt them.

Throughout those processes, the DPO at CBS will make sure that the collected data respects the rules of this DMP and other regulations and legal documents that are relevant for the data of the citizens (subjects) in question.

4. FAIR data

Horizon 2020 beneficiaries should make their research data findable, accessible, interoperable and reusable (FAIR), to ensure it is soundly managed. Good research data management is not a goal in itself, but rather the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse.

REFLOW research data should be "F.A.I.R.". These principles precede implementation choices and do not necessarily suggest any specific technology, standard or implementation-solution.

Below, the FAIR Data Principles are further described (Wilkinson et al. 2016).

1. To be Findable:
a. Metadata are assigned a globally unique and eternally persistent identifier.

b. Data are described with rich metadata.

c. Data (and metadata) are registered or indexed in a searchable resource.

d. Metadata specify the data identifier.

2. To be Accessible:
a. Data (and metadata) are retrievable by their identifier using a standardised communications protocol.

b. The protocol is open, free, and universally implementable.

c. The protocol allows for an authentication and authorization procedure, where necessary.

d. Metadata are accessible, even when the data are no longer available.

3. To be Interoperable:

a. Data (or metadata) use a formal, accessible, shared, and broadly applicable language for knowledge representation.

b. Data (or metadata) use vocabularies that follow FAIR principles.

c. Data (or metadata) include qualified references to other (meta)data.

4. To be Re-usable:

a. Data (or metadata) have a plurality of accurate and relevant attributes.

b. Data (or metadata) are released with a clear and accessible data usage license.

c. Data (or metadata) are associated with their provenance.

d. Data (or metadata) meet domain-relevant community standards.

It should be noted that, participating in the ORD Pilot does not necessarily imply opening up all REFLOW research data. Rather it follows the principle "as open as possible, as closed as necessary". In fact, as per the GA Article 36.1, during implementation of the action and for 4 years after the period set out in Article 3, the parties must keep any data, documents or other material confidential that is identified as confidential at the time it is disclosed.

Below we will explain how we will apply the FAIR principles in practice.

4.1 Making data findable, including provisions for metadata:

The key to making data findable is equipping datasets with metadata that meets basic standards and adheres to uniform, consistent schema. Provisions and actions that shall be taken to ensure the discoverability of REFLOW data include:

- Accompanying datasets with properly structured and accurate metadata, according to FAIR data principles specified above.
- Providing proper documentation identifying their content and potential uses
- Making data identifiable by using standard identification mechanisms (e.g. DOI, where applicable)
- Publishing papers and reports with references to the data

All data and metadata involved in REFLOW project will comply with Zenodo Data repository FAIR data principles. See https://about.zenodo.org/principles/

This will be done in respect to Article 29.2 in the GA, and by assigning a DOI to each dataset in the project. Moreover, as specified in D8.2, file naming convention and versioning will be used for all files generated during the project.

REFLOW dataset tracking database

Datasets must be documented and organised in order to be accessible. A dataset tool was developed to collect and trace all relevant datasets and shared with the REFLOW consortium. This tool will allow the collection, identification and classification of the data generated during the lifespan of the project and will be applicable to all cases and data types.

Each REFLOW member will be responsible for regularly adding and classifying the data generated throughout the project's duration to the dataset tool, sharing the following information:

- Dataset name
- Trace Code
- Type
- Data source
- Format
- Storage platform
- Currently accessibility
- Archiving access policy
- WP related
- Data controller responsible partner
- Data controller person
- Data controller person email
- Data collection status
- Dataset description
- GDPR condition (for personal data)

The database will automatically provide a **dataset unique trace code** and a registry date and it will be updated regularly.

The repository will be stored on SharePoint and it will be only accessible by the consortium members. The REFLOW dataset tool will be used to identify project datasets and partners responsible for each dataset, and is aimed to guarantee the traceability of all the project data.

The dataset tool configuration is attached in Annex 1.

4.2 Making data openly accessible:

REFLOW will identify and consolidate existing open datasets in cities and integrate them in an open data platform, which will be securely stored using blockchain technologies, respecting the transparency and open source principles of the true sharing economy. The core of the REFLOW platform and its modules will each be accompanied by an open specification that others can use to integrate into their code by using a royalty free license.

In respect to making data openly accessible, and as described in the Grant Agreement, all project software deliverables will be licensed under open source principles, unless otherwise specifically mentioned in the consortium agreement or decided by the general assembly. The preferred open source licenses will be GPLv3.0 and Apache 2.0 while existing code bases will preserve their own existing open source licenses (GA, pp. 41).

Public documents will be stored on Zenodo platform as a backup and external links, to access the data, will be embedded on the REFLOW website. The documents will be available in protected PDF files.

As stated earlier, this document will be updated along the project with which data will be made accessible and which will not, along with the reasons for opting out.

4.3 Making data interoperable

The data standards related to the project activities might vary depending on partners' internal tools and methodologies. As part of the project's open data access and sustainability strategy, once the project finishes, the data that can be published as open data will be selected and published. To ensure that project data is made available using the same standards towards consistency and usability both during and after the project ends, all datasets should provide at least the following metadata (see section 4.3 Making data findable, including provisions for metadata):

- Digital Object Identifier (DOI)
- Publication date
- Title
- Authors
- Description
- Source
- Keywords

4.4 Increase data re-use (through clarifying licenses):

Further exploitation and uptake of REFLOW will be a full, holistic, complete open source environment with all required open source widgets, application examples, interfaces, documents and other items. These deliverables will be based on internet standard developer environments and services (such as GitHub) for easy exploitation.

All non-software deliverables of the project will be licensed as Creative Commons where this is legally possible, with the exception of academic publications that will have to comply with the licensing terms and rights of the academic publication venue. Where applicable, all media deliverables will also be placed on Wikimedia Commons for easy exploitation.

According to the Art. 28.1 in the Grant Agreement, each beneficiary must — up to four years after the period set out in Article 3 — take measures aiming to ensure 'exploitation' of its results (either directly or indirectly, in particular through transfer or licensing) by:
(a) using them in further research activities (outside the action);
(b) developing, creating or marketing a product or process;
(c) creating and providing a service, or
(d) using them in standardisation activities.

Data archiving and sustainability will be guaranteed by the Zenodo digital repository. As a European Commission supported initiative, and technically supported by CERN, we trust this as the best way to ensure access to the generated data continues long after the project ends.

5. Allocation of resources

The Data Management Plan is a living document and it will be updated regularly by the lead beneficiary and reviewed by WP-leads and the Coordination Team with the purpose of supporting the data management lifecycle for all data that will be collected, processed and/or generated by the project. Before updating the DMP, the Coordination team shall be notified and the final update shall be approved by them.

The responsibilities of the lead beneficiary is to update the DMP as a deliverable when significant changes arises, and – together with the DPO – assist the consortium partners with specific questions in relation to the DMP.

The Coordination team is responsible for ensuring proper management and processing of all data in the project, complying with the EU data protection regulations. Moreover, the Coordination team is in charge of uploading the deliverables to the Participant Portal and to place a copy on SharePoint. The public deliverables will also be uploaded by the Coordination team to the Zenodo Platform, including the datasets related to the documents, at the same time they are made available on the Participant Portal. Other types of public data also must be uploaded to Zenodo by the coordination team.

The responsibilities of the project members include:

● Regular update and classification of the data collected and generated in the REFLOW dataset tool (available on SharePoint);

● Implementing and respecting the Data Management Plan;

● As stated in D9.2, before the start of any task that requires processing of any kind of personal data, an analysis of the types of personal data necessary will be made with the participation of the respective partners, who will also request the opinions of the ethical manager and the PCO.

6. Data security

All personal data collected and processed during the project will be stored on SharePoint as it conforms with the EU GDPR. The Regulation (EU) 2016/679 (GDPR) defines 'personal data' as "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". The purposes of processing personal data in REFLOW include, but are not limited to, project management and coordination, project outreach, communication and dissemination activities, cases studies, research activities, administration and compliance with the Grant Agreement and statutory obligations of the project partners. Personal data collected by the REFLOW partners will be processed in a way that safeguard the privacy and rights of individuals following the applicable national laws and regulations in compliance with the EU GDPR.

Project participants through the different pilots and activities will provide some personal information to allow their identification by the consortium. All recordings, transcripts and notes from workshops and interviews will be pseudoanonymised. If personal data is collected in a physical form (e.g. on paper), the researcher is expected to store it

in the cloud-operating server in a digitalised version as soon as possible. The physical copies shall be destroyed securely. Moreover, as specified in the D9.2, all identifiable data will be dissociated from the rest of the collected data in a separate database during data collection, in order to respect the regulations on personal data. This also accounts for public released, where data will contain personal information, such as those collected in WP1 (social impact), WP3 (material flow analysis), WP4 (cases studies) and WP5 (pilots) and handled by the IT team (WP2).

As specified in the D9.2, each researcher is responsible for making sure that personal data is pseudonymised as quickly as possible after its collection, and that it remains safe.

7. Ethical aspects

Research and innovation activities conducted by REFLOW's consortium team play an essential role in achieving the project's objectives. Therefore, a specific work package is dedicated to establishing a consolidated approach for REFLOW research ethics. By complying with the ethics deliverables (D9.1, D9.2 and D9.3), and the ethics requirements as set out in the DoA, REFLOW project aims to demonstrate that the consortium members are fully committed to the advancement of ethical research. The D9.1 - H Requirement No. 1 presents the procedures and criteria that will be used to identify and recruit research participants. The deliverable 9.2 - POPD Requirement No. 2 sets out the procedure of the data protection policies implemented, and measures taken to ensure compliance throughout the project's lifecycle. Finally, examples of templates for the informed consent/assent forms as well as information sheets that will be used to identify and recruit research participants are provided in the Deliverable 9.3 - POPD - H - Requirement No. 3.

As specified in the ethics requirements deliverables, the project partners conducting research activities will obtain formal ethical approval from the DPO (Data Protection Officer) and the Ethical Manager, following their internal procedures as long as their practice is in compliance with applicable international, EU and national law and follow the standards and best practices of ethical research.

All WP leaders are expected to supervise and check the work and procedures by its team in accordance with the ethics deliverables, and report non-compliance, if there is any, to the Project Coordinator and the Ethical Manager/Data Protection Officer (depending on the nature of non-compliance).

By following the ethics requirements, REFLOW project ensures that all research activities undertaken by the project partners, which involve human subjects and personal data, are undertaken in a way that safeguards the dignity, rights, safety, and privacy of those involved.

8. Plan Maintenance

The Data Management Plan is intended to be a living document in which information can be made available through updates as the implementation of the REFLOW Project progresses and when significant changes occur. The DMP will be

updated over the course of the project whenever significant changes arise by the lead beneficiary of the deliverable (IAAC), reviewed by WP-leads and approved by the PCO.

9. References

Wilkinson, M.D. et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3, p.160018.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

10. Annexes

Annex 1 – REFLOW Project dataset tool configuration

The 'REFLOW Project dataset tool' is used as a log file in which all metadata is described and associated to a specific uploaded dataset. It is a .xlsx file, stored on SharePoint. This tool generates a dataset code, which must be included in the dataset file's name, so it can be traceable within the storage system. The following information is organised in the columns of the main sheet. Further instructions and examples are included in the "Read-me" tab of the file.

Figure 1. REFLOW Project dataset tool on SharePoint

Trace Code and Trace code label

The Trace Code works as an identifier to trace data in the main storage system (SharePoint). In order to be able to filter datasets and locate them easily, **Trace Codes** are sorted according to the following classification:

Trace Code	Trace Code Label	Description
Personal Data	PD	Personal data are all data of an identifiable natural person. This involves information that is either directly about someone or can be traced back to a person

Single Value	SV	Single value refers to one value of any data type
Maps	MA	Visual representation data of an entire area or a part of an area, typically represented on a flat surface
Time Series	TS	Times series refers to a sequence of well-defined data points measured at consistent time intervals over a period. The common use of time series and lifetime models makes material flow analysis (MFA) a suitable tool for assessing long-term trends in material use, including production, consumption and distribution
Source Code	SC	Source code refers to a digital coding system for data. Data as code allows computers to treat instructions in a programming language as data handled by a running program
Multiple Values	MV	Multiple values refer to any data that contains more than one value, but it is not personal data, maps, time series or source code

Table1. Classification of dataset trace codes

This a preliminary classification. The DMP is an ORDP, therefore if new categories are proposed or needed by the project partners according to the datasets used or developed by the project, the DPO will evaluate whether to enlarge this list.

Dataset code

This code will be used to easily locate the dataset in REFLOW's main storage platform (SharePoint) and, at the same time, access dataset metadata in the Project dataset tool. Therefore, this code must be included in the dataset file's name. It is composed by DS + Trace Code Label + Correlative Number. The number must be correlative to the last uploaded dataset within the same Trace Code, e.g. "DS_MV_02"

Dataset name

If applicable, dataset name must indicate the pilot case it is associated with, e.g. "Amsterdam_Textile waste material flow data". This information must be also in the dataset file name, next to the dataset code, e.g. "DS_MV_02_Amsterdam_Textile waste material flow data.xlsx".

Dataset type

It refers to the type of data, according to 3.2 of the DMP.

Dataset Type
Project management data
Observational data
Experimental data

Operations data
IT infrastructure data

Table 2. Classification of dataset type

Data Source

If data has been derived or reused, the original source(s) of the dataset must be specified. Ideally, data source is identified with a DOI or URL to the original source. If not available, details about it must be provided, including (if possible) date of original data. Some examples (if DOI or URL are not available) are: "Online meeting (12/11/2019)", "survey on SMEs in Berlin metropolitan region (02/10/2019)", "Wilkinson, M.D. et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3, p.160018", "Cluj-Napoca Pilot material flow data on dashboard (05/07/2019)", "Regulation (EU) 2016/679", etc.

Data Format

Extension(s) of the files in the dataset. If the dataset is packaged, the original extensions of the dataset files inside the packaged must be specified, avoiding using ".zip", ".rar" or any other compression extensions.

Start and end of data collection

Knowing the date of data collection gives an idea on how updated is the information obtained from it and if a new collection of data is needed during the project. If data has been mined or collected for a period of time this must be specified.

Software

The specific software required to use, view or access the stored data must be specified. If multiple are available, at least one must be mentioned, preferably open source software.

Additional Storage Platforms

All datasets must be uploaded in REFLOW's main storage platform (REFLOW SharePoint). Nevertheless, due to their nature and specific needs (e.g. a source code may be developed on Github, or a poster which is an outcome from a workshop may be kept also on the partner's facilities) there may be additional platforms where datasets are stored (see DMP 3.4 Data storage and sharing). If dataset is only stored in SharePoint, this field is left blank.

Additional Storage Platform
On premises - Partner internal facilities
Zenodo
Other platforms

Table 3. Classification of alternative Project data storage platforms

Current Accessibility

This defines dataset accessibility within the consortium when uploaded. The project partner who uploads it, along with DPO, must ensure the dataset access is granted accordingly to project members and partners.

Current Accessibility
Public
All the consortium
Partner / Owner
Project coordinator and WP leader

Table 4. Classification of dataset current accessibility

Specific Accessibility

A description of the specific dataset accessibility must be included, according to the classification below. This classification is developed according to Project's ethics, privacy and data protection requirements as well as its knowledge dissemination strategy, as explained in chapters "4. FAIR data" and "7. Ethical aspects" of the DMP.

Specific Accessibility
Publicly Accessible Anytime
Publicly Accessible After Pending Publication
Publicly Accessible After Anonymization and Aggregation
Publicly Accessible Once the Project Finishes
Under discussion
Confidential - Non public disclosure (Personal Data)
Confidential, only for members of the consortium (including the Commission Services)

Table 3. Classification of dataset specific accessibility

Work Package and Task

In order to understand where the dataset lies within the project's structure, the related work package(s) and task(s) must be identified. If the dataset belongs to or will be used in different tasks, these must be included. Neither the work package nor the task will be used to classify the dataset. Data set code and name fulfil this purpose.

Person who uploads the dataset, responsible researcher and partner

The name and email of the person who uploaded the information, as well as the responsible researcher and project partner of the task that the dataset was developed for must be specified in the dataset tool.

Status

The Status indicates whether the dataset may be updated throughout the course of the project or it will remain.

Status	Description
Completed	The final version of the dataset and it will not change during the project
Revisable	A preliminary version of the dataset. It is developed by one or more project partners and it is pending revision by other project or external partner
Ongoing	The dataset is still being developed and will change during the project

Table 5. Classification of dataset status

Dataset description

A short description of the dataset shall include: what its purpose is, a short description of the methodology used to generate or derive the dataset, what files it is composed of and how it can be used.

GDPR compliance notes

Personal data must indicate whether it complies with GDPR Regulation.