HORIZON –CL4-2022-TWIN-TRANSITION-01-06
HORIZON Innovation Actions
Grant Agreement No.: 101092295

Circular and Dynamic Manufacturing Supply Chain Orchestration and OptimiSation

D1.2 Data Management Plan			
Report Identifier:	D1.2		
Work-package:	WP1	**Task:**	T1.4
Responsible Partner:	European Dynamics Luxembourg SA (ED)	**Version Number:**	1.0
Due Date	29/02/2024	**Document Date:**	08/03/2024
Distribution Security:	PU	**Deliverable Type:**	DMP
Keywords:	Data management, FAIR data, IPR management, Datasets		
Project website: https://circuloos.eu/			

Document History

Version	Content & Changes	Issue Date
0.1	Document created	23/02/2024
0.2	Document sent for review	26/02/2024
0.3	Document reviewed	29/02/2024
0.4	Document reviewed	29/02/2024
0.5	Reviews are combined	01/03/2024
0.6	Sent for Quality Assurance	06/03/2024
1.0	Quality Assurance and Submission	08/03/2024

Quality Control

	Name	Organisation	Date
Editor	Vivi Plagaki	ED	23/02/2024
Peer review 1	Daniela Corredor	ALA	29/02/2024
Peer review 2	Alex Carmona	CAN	29/02/2024
Authorised by (Technical Coordinator)	Alkiviadis Giannakoulias	ED	07/03/2024
Authorised by (Quality Manager)	Themis Kolyvas	ED	08/03/2024
Submitted by (Project Coordinator)	Anastasia Garbi	ED	08/03/2024

Legal Disclaimer

CIRCULOOS is an EU project funded by the Horizon Europe (HORIZON) research and innovation programme under grant agreement No. 101092295. The information and views set out in this deliverable are those of the author(s) and do not necessarily reflect the official opinion of the European Union. The information in this document is provided "as is", and no guarantee or warranty is given that the information is fit for any specific purpose. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. The CIRCULOOS Consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law.

Table of Contents

List of Figures

List of Tables

Abbreviations

Acronym	Description
AI	Artificial Intelligence
CE	Circular Economy
CSV	Comma Separated Values
D	Deliverable
DMP	Data Management Plan
DoA	Description of Actions
DOIs	Digital Object Identifiers
EC	European Commission
EPC	European Patent Convention
ERP	Enterprise Resource Planning
EU	European Union
EXDs	Experiments for Demonstration
FAIR	Findable, accessible, interoperable and re-usable data
GA	Grand Agreement
GDPR	General Data Protection Regulation
GIS	Geographic Information System
GRETA	Green Targets
IoT	Internet of Things
IPR	Intellectual Property Rights
JPEG	Joint Photographic Expert Group
JSON	JavaScript Object Notation
KPI	Key Performance Indicators
LCA	Life Cycle Assessment
LCI	Life Cycle Inventory
M	Month
MSMEs	Manufacturing SMEs
OA	Open Access
ORDP	Open Research Data Pilot
PEDR	Plan for Exploitation and Dissemination of Results
PNG	Portable Network Graphics
RAMP	Robotic and Automation MarketPlace
REST	Representational state transfer
RDMS	Relational Database Management Systems
TBD	To Be Determined
SCDT	3D Digital Twin of Supply chain / production / products
SCODT	AI and Data-driven Supply Chain Optimization
SMEs	Small and medium manufacturing enterprises
S-LCA	Social Life Cycle Assessment
SSFTP	Sustainable Sourced Fresh Timber Platform
TXT	Text only
URI	Uniform Resource Identifier

URL	Uniform Resource Locator
WP	Work Package

Executive Summary

The Data Management Plan (DMP) is a document that outlines the procedures and methodologies of data treatment during the project and how data will be used and shared after the project ends. The DMP has to describe the observed data that are collected and processed during the lifespan of the project, while providing the overview of available research data, access procedures, data management and terms of use. The first version of the CIRCULOOS DMP that is presented in this document, reflects the current state of the discussions, plans and ambitions of partners. It includes the preliminary scenario of data set definitions and will be updated and implemented with new datasets and results in the following months. Partners have been asked to identify and characterize the data they are going to use and have been involved in a brainstorming session aimed at discussing the data utility. Moreover, CIRCULOOS strives to enhance access and reusability of research data, with a particular focus on achieving a balance between openness and safeguarding scientific information. Given the involvement of workers in the technical activities foreseen in the project, technical and organisational measures that safeguard the rights and freedoms of the data subjects/research participants will be developed and informed consent procedures in regard to data processing will be designed, in conformity to the GDPR requirements and guidelines. The DPM is a live document that will be kept update all along the project duration.

1 Introduction

1.1 Project Introduction

The overall vision of CIRCULOOS is to deliver the tools to enable MSMEs become full members of the Circular Manufacturing value chain. These tools orchestrate and continuously optimise the supply-chain end-to-end and integrate planning and execution monitoring to enable transparent and on-time communication. Combining these with direct calculation of the product sustainability and circularity profile, for both internal and external partners, this environment will enable them to configure and execute disruptive circular manufacturing processes for sustainable production that covers the entire life cycle of products; either by recovering the value of product that ended-up as waste or from recycled and remanufactured products.

To achieve this objective the project aims to deploy:

- Circular end-to-end supply chain orchestration for collaborative workflows which incorporates planning and execution metrics and integrates advanced and multimodal visualisation and analytics. The visualisation is delivered by comprehensive Digital Twins of the supply chains formulated, the factory processes and product design phases.
- Supply Chain Optimisation that monitors the global (across the supply chain) and local (within the factory) processes and execution, inputs and outputs and configuration parameters, to enable data-driven AI decision making, this way supporting continuous optimisation of targeted and measured performance and sustainability parameters.
- Dynamic Sustainability Assessment functionalities that investigate alternative supply-chain scenarios (varying in terms of materials used, processing technologies, suppliers involved and/or activated circular economy practices) in place of the existing schemes, quickly measuring their performance in terms of environmental sustainability and circular economy profile.
- Supply Chain Data Spaces for seamless, multi-level data flow across the supply chain partners, supporting the reuse of materials in novel products, the extension of the life-cycle of finished products (remanufacturing), and data-driven decisions for collaboration of parties offering matching services in the most dynamic and efficient way.
- Cybersecure and trustworthy data sharing across the supply chain by employing a distributed, trusted and efficient Identity and Access management system, that together with the associated trust framework will coordinate the identities of all IoT objects and ensure trustworthy data sharing among its members, aligned with the trust framework that is being implemented in EBSI.
- CM specific tools for the automatic recognition of recyclable parts by modern Machine Vision tools and Advanced Robotics, to enable optimised flows in the selection process.
- Novel circular business processes will be demonstrated supporting reusing, reducing, and recycling material in production and consumption systems. The new collaborative production models will provide quantifiable results on the sustainability increase across the supply chain, in terms of efficient use of raw materials, of by-products, of waste and energy and of emissions reduction. CIRCULOOS leverages the above with the RAMP integrated innovation IOT platform and the European network around it to deliver a CM ecosystem and platform for Manufacturing SMEs.
- Skills upskilling and reskilling will be provided in RAMP and through online courses, webinars, and best practice guides and success stories based on the pilots and Experiments for Demonstration (EXDs).

1.2 Deliverable Purpose

Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the

data publication process. The amount of data generated by research projects continuously increases, however re-using the data for further research purposes and thus maximising the benefit deriving from the research investments still represents a hard challenge. Data needs to be properly collected, annotated, filed in such a way that they will be available in the long-term and can be re-used for downstream investigations, either alone, or in combination with newly generated data, in line with the European Commission and H2020 guidelines. The Data Management Plan fills this purpose as it describes and characterises the observed data that are collected and processed during the life-time of a project and how they are managed after the project completion. This document presents the status quo of the CIRCULOOS Data Management Plan and the expectations of the partners with first assumptions of data set definitions and draft results that will be updated and improved continuously during CIRCULOOS lifespan.

1.3 Data Protection Legislative Framework

The CIRCULOOS consortium is fully aware of the ethical implications of the proposed research and respects the ethical rules and standards of Horizon Europe, and those reflected in the Charter of Fundamental Rights of the European Union. Where necessary, the CIRCULOOS consortium confirms its abidance to national and international laws including Regulation (EU) 2016/6792 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, the Directive on Privacy and Electronic Communications (2002/58/EC)[1], Directive on Protection of Privacy in the Telecommunication Sector (97/66/EC)[2], The Universal Declaration of Human Rights[3] and the Convention 108 for the Protection of Individuals with Regard to Automatic Processing of Personal Data[4]. Article 19 "Ethical principles" of Regulation No. 1291/2013/EC of the European Parliament[5] and of the Council which states the fundamental principles of the HORIZON Ethics in research.

1.4 The Data Management Plan (DMP)

The DMP is a document that outlines the procedures and methodologies of data treatment during the project and how data will be used and shared after the project ends. It deals with the generation and discovery of data, their collection, and evaluation by quality assurance, classification, organization and documentation. Dissemination and sharing policies are also addressed within the DMP. The process of creation of the DMP respects some golden rules that ensure the successful management of research data arising from the project. W. K. Michener[6] in his article "Ten Simple Rules for Creating a Good Data Management Plan" well outlines the building process of the DMP, which starts with the identification of the data to be collected (1), followed by the definition on how the data will be organized (2), documented (3), how data quality will be assured (4), the strategy adopted to preserve and store the data collected (5),

[1] Directive - 2002/58 - EN - eprivacy directive - EUR-Lex (europa.eu)
[2] Directive - 97/66 - EN - EUR-Lex (europa.eu)
[3] https://www.un.org/en/about-us/universal-declaration-of-human-rights
[4] Convention 108 and Protocols - Data Protection (coe.int)
[5] Regulation (EU) No 1291/2013 of the European Parliament and of the Council of 11 December 2013 establishing Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020) and repealing Decision No 1982/2006/ECText with EEA relevance (europa.eu)
[6] Michener, William K. "Ten simple rules for creating a good data management plan."PLoS computational biology11.10 (2015): e1004525.

the data policies adopted (6), ending with the rules for the dissemination of the data and the appointment of the roles and responsibilities of data management (7).

Even if the CIRCULOOS DMP is still at an early stage and much information are still not available, the plan presents a first characterization of the data partners expect to collect. Namely:

- **the types of data**: text, spreadsheets, software, algorithms, pictures, videos, audio files;
- **the volume of data:** data management activities might be influenced by the number of the data to be handled;
- **sources:** stakeholders or research centres that wish to re-use the data generated by the project are interested in the source of the data, if they are proprietary, derived from other research studies, are subject to restrictions or can be freely used;
- **data and file formats:** non-proprietary formats are preferred as they ensure the accessibility to the data for the long term; examples could be Comma Separated Values [CSV].

As already anticipated, the DMP is a living document that must be revised during the project lifecycle. Therefore, contents and information on data will be periodically (about every 8-12 months) updated with the missing data in order to achieve, by the end of the project, a comprehensive roadmap on the use of data generated and used within the project.

The definition of the way the data will be organized should be addressed as well and it is strongly influenced by the volume of the data generated and used within the project. Larger data volumes and usage constraints may require the use of relational database management systems (RDBMS) for linked data tables like ORACLE or MySQL, or a Geographic Information System (GIS) for geospatial data layers like ArcGIS, GRASS, or QGIS, while a small number of data could be effectively managed with commercial or open-source spreadsheet programs like Microsoft Excel and OpenOffice Calc. The data organization will be defined once we will have a clearer scenario of the types of data and the volume of data collected within the project.

The documentation of the data by means of proper metadata is fundamental to ensure that data will be discoverable, usable and properly cited by those who will look for them and use them for further research purposes. Indeed, metadata means "data about data". Metadata is defined as the data providing information about one or more aspects of the data; it is used to summarize basic information about data which can make tracking and working with specific data easier. In this project, structural and descriptive metadata will be deployed to define formats, types, versions and relationships between digital data and to identify the data through some of their characteristics, like data of collection, volume, name, authors/owners and keywords.

Quality control and quality assurance of research data will be defined at a later stage, when the characterisation of research data will be more complete, and it will be in line with the quality assurance procedures defined in D1.1 Project Management Handbook.

Data storage and preservation strategy will be an important topic that the DMP will address, by answering the following questions:

- How long will the data be accessible?
- How will data be stored and protected over the duration of the project?
- How will data be preserved and made available for future use?

The answers to those questions are expected to vary from one partner to another due to several factors. The internal data policy of each partner will influence the possibility to make some of the research data collected public or not. Moreover, if the partners will make the data available to be consulted by other researchers after the end of the project, the storage may depend on the type of data generated, indeed some of them may need to be kept for a short time as they are extremely repeatable, while if they are strongly variable from one experiment to another might need to be stored for a very long time. The data that will receive the authorization by the owners to be published and available for others to use will be stored within online collaborative platforms accessible by the project consortium and will also be uploaded on online repositories like Zenodo (http://zenodo.org/).

The DMP will also include explicit policy statements about how data will be managed and shared. Such policies include the licensing or sharing arrangements that pertain to the use of pre-existing materials; the plans for licensing, sharing, and embargoing (i.e., limiting use by others for a period of time) data, code, and other materials; the legal and ethical restrictions on access and use of human subject and other sensitive data. The policies will be in line with what has been stated in the Consortium Agreement with respect to the Intellectual Property Rights (IPR) and the joint ownership of the results and the data beyond them, and to what the GDPR and the Ethics Requirements deliverables report with regards to the Personal Data treatment and the treatment of data deriving from the involvement of human beings in the project experiments.

The IPR and non-disclosure statements will be very important for the dissemination actions taken by the project partners. The DMP will be compliant with the statements of the Consortium Agreement (CA) and the Ethics Deliverables in defining when, how and which data will be made available. Dissemination actions will range from posting the data on social media and the project website, to publishing the data in online repositories and in scientific journals. Specific information regarding dissemination actions can be located in D6.1, the "Initial Plan for the Exploitation and Dissemination of Results," which was submitted in M3.

1.5 Data Management Strategy

A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon Europe project. As part of making research data findable, accessible, interoperable, and reusable, a DMP should include information about the handling of research data during and after the end of the project:

- What (kind of) data will be collected, processed and/or generated and to whom might they be useful later on?
- Which methodology and standards will be applied?
- What metadata will be required to enable data to be found and understood, ideally according to the standards of a scientific discipline?
- Whether data will be shared/made open access.
- How will data be preserved (including after the end of the project)?
- How to archive and preserve the open datasets of the project?

More specifically, for Horizon Europe projects a FAIR DMP template[7] has been designed to be applicable to any project that produces, collects, or processes research data. The FAIR data principles towards promptly disseminating the data outcomes of a research project[8] can be seen below in the table below.

Table 1: FAIR Data Principles

FAIR Data Principles	
Data should be Findable	F1. (Meta)data are assigned a globally unique and persistent identifier. F2. Data are described with rich metadata (defined by R1 below). F3. Metadata clearly and explicitly include the identifier of the data they describe. F4. (Meta)data are registered or indexed in a searchable resource
Data should be Accessible	A1. (Meta)data are retrievable by their identifier using a standardised communication protocol. A1.1 The protocol is open, free, and universally implementable. A1.2 The protocol allows for an authentication and authorisation procedure, where necessary. A2. Metadata are accessible, even when the data are no longer available.
Data should be Interoperable	I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (Meta)data use vocabularies and definitions that follow FAIR principles. I3. (Meta)data include qualified references to other (meta)data.
Data should be Reusable	R1. (Meta)data are richly described with a plurality of accurate and relevant attributes. R1.1. (Meta)data are released with a clear and accessible data usage license. R1.2. (Meta)data are associated with detailed provenance. R1.3. (Meta)data meet domain-relevant community standards.

1.6 Open Research Data Pilot

[7] Guidelines on FAIR Data Management in Horizon 2020, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[8] https://www.go-fair.org/fair-principles/

Since 2017, all the thematic areas of the H2020 and HORIZON programmes have been included in the work programme Open Research Data Pilot (ORD Pilot) which is a flexible pilot on open access to research data and scientific publication in H2020 and HORIZON EUROPE. The pilot considers the need to balance the possibility to make the data open with the protection of scientific information, commercialization and IPR, privacy concerns, and security, as well as questions of data management and preservation. The reason for the creation of the ORD Pilot is that the free access to scientific publications and data should improve the quality of the results on which further research will hopefully be built, encourage collaboration and avoid duplication of efforts, speed up the entrance of the results into the market and should boost and underline the benefits of public investments in research funded under the HORIZON programme also among citizens and society. From a high-level perspective, open access (OA) consists in providing on-line access to scientific information free of charge to the users to promote the reusability of the data. Within the context of the R&D actions there are two main categories of data that OA addresses: the scientific papers and the research data collected from the experiments conducted in the laboratories. Research data refer in particular to facts or numbers, collected to be examined and considered as a basis for reasoning and discussion on project results, examples 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. Indeed, the ORD Pilot run by the European Commission applies to the datasets at the basis of scientific publication and to the peer reviewed papers released within the context of the H2020 and HORIZON projects. According to the Budapest declaration (2002) and Berlin Declaration (2003), within the context of ORD Pilot, open access means not only the basic rights to download, save and print a document, but also to copy, distribute, search, link, crawl and mine data. The DMP will be useful to state which type of data will be made available and which restriction may apply to some data.

Open Access[9] to scientific publications means free online access for any user. The two main routes to Open Access are::

- **Self-archiving / 'green' open access** – the author, or a representative, archives (deposits) the published article or the final peer-reviewed manuscript in an online repository before, at the same time as, or after publication. Some publishers request that open access be granted only after an embargo period has elapsed.
- **Open access publishing / 'gold' open access** - an article is immediately published in open access mode. In this model, the payment of publication costs is shifted away from subscribing readers. The most common business model is based on one-off payments by authors. These costs, often referred to as Article Processing Charges (APCs) are usually borne by the researcher's university or research institute or the agency funding the research. In other cases, the costs of open access publishing are covered by subsidies or other funding models.

To ensure visibility and openness of CIRCULOOS resources, some platforms have been considered to be used, where general public, researchers and other investigators can discover and download information, data and documents on the project's results. They could be Zenodo and the EC's OpenAIRE platforms. Furthermore, the EFFRA Innovation Portal is also taken into consideration since being provided by the European Factories of the Future Research Association (EFFRA), is a unique resource database combining

[9] Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-datamanagement/data-management_en.htm

various projects' databases' information about the latest projects' outputs, together with reports and demo materials.

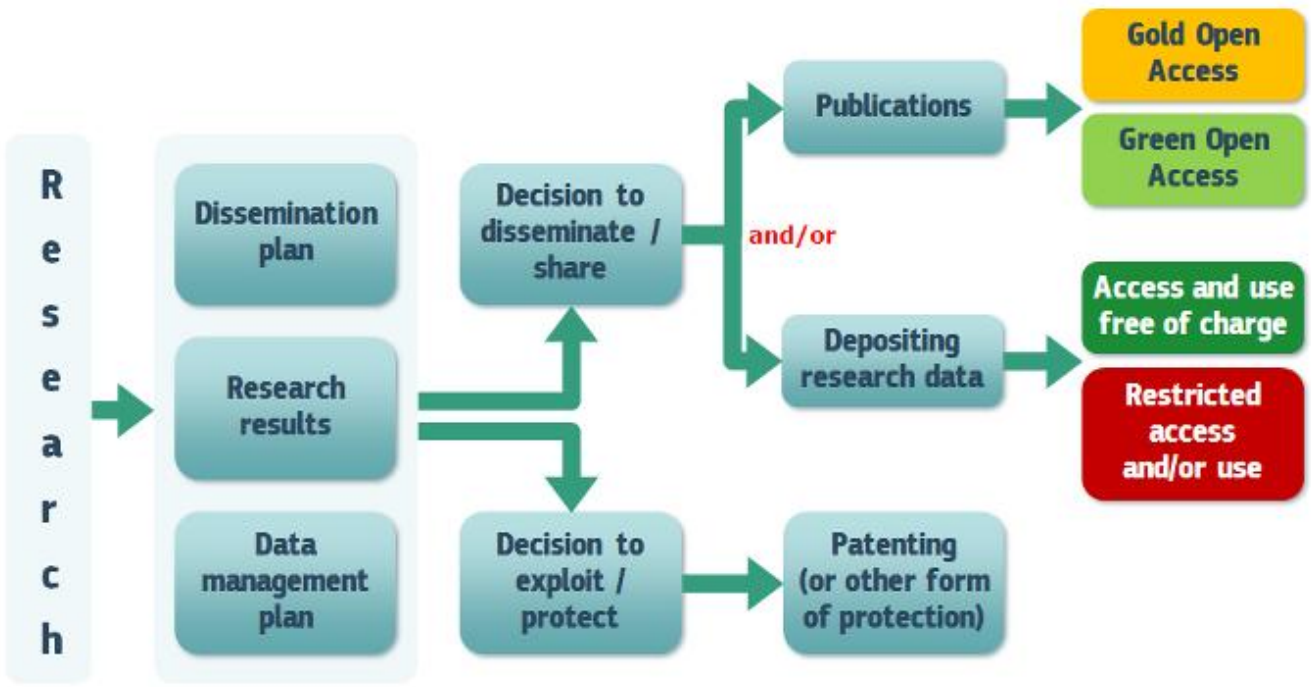

Figure 1: Open Access strategy for publications and research data

1.7 Objectives of the Data Management Plan

CIRCULOOS will deliver the tools to enable Manufacturing SMEs (MSMEs) become full members of the Circular Manufacturing value chain. The project will ensure that these tools orchestrate and continuously optimise the supply-chain end-to-end and integrate planning and execution monitoring to enable transparent and on-time communications. The project will ensure that the kits are widely distributed to a wide audience of MSMEs and midcaps in Europe.

The adoption of these tools will be facilitated with the use of systems like ERP, PLM, CAD, data sharing platforms used with suppliers, external database containing LCI and LCIA information, as well as IoT sensors and wearable devices, robots, and other factory data sources.

Considering the whole amount of data that CIRCULOOS is expected to generate, the purpose of the DMP is to define the proper management of project research data and of data subjects in compliance with the EC recommendations and national and international regulations and guidelines on the use of data.

The Plan is intended as a roadmap illustrating how data arisen from project research lines will be treated throughout the project lifetime and beyond, once it will be finished. CIRCULOOS DMP will provide a vehicle for conveying information to and setting expectations for the project team during the different stages of the project. The plan will be a living document that is periodically reviewed according to the new data gathered, the needs and any changes in protocols (e.g. metadata, QA/QC, storage) and policies.

1.8 CIRCULOOS Data Management Plan

CIRCULOOS project is the result of a successful proposal (number 101092295) submitted on 10 Oct 2023 to the European Commission. The consortium has signed the Grant Agreement and thus jointly undertaken to execute the project.

The mandate of the project is to execute the work described in accordance with the signed contract (Grant Agreement) and its annexes, within budgetary and schedule constraints.

This document is related to what is described in the Description of Actions (DoA) (CIRCULOOS, 2023) as deliverable D1.2 Data Management Plan and is related to T1.4 Data Management Plan.

The CIRCULOOS ecosystem is organized around a platform offering added value services to MSMEs who are willing to share information coming from their production systems and human resources. Open management and distribution of data is thus relevant to the generation of services while, on the other hand, data security and data ownership have to be carefully addressed. This aspect of the platform will be reflected in the data management strategy both during project duration, as well as in the following platform exploitation activities. In particular, it is of paramount importance to ensure an appropriate quality management of data collection and sharing in-line with European privacy and data protection regulations when pilot-related datasets or data deriving from human beings are dealt with. Furthermore, attention will be paid to research data management starting from the beginning of the project so to make sure that results can be findable, accessible, interoperable and re-usable (FAIR). This includes careful screening of relevant standards (and appropriate selections that ensure the widest possible compatibility and outreach), but also developing models for data sharing, while protecting privacy of users and providing security of IP and business of the involved companies.

A first analysis allows the identification of the following categories of data that will be relevant for the CIRCULOOS project:

- **Individual data**: users' data have to be shared with the platform in order to build the customized service packages and do matchmaking between different entities (users or companies); experiments that include the use of cobots may require some human monitoring activity to provide realtime support and enhance human-robot interaction. The profile may contain, on the one hand, static info such as psychophysiological features or skills and preferences; on the other hand, dynamic and physiological data collected real-time during the execution of activities.
- **Production system data**: in order to offer real-time support for an improved quality control, a more agile reconfigurability, or for a symbiotic interaction between operators and the production system, also data from the production system have to be shared with the platform.
- **Supply Chain Data:** multi-level data flow across the supply chain partners, supporting the reuse of materials in novel products, the extension of the life-cycle of finished products (remanufacturing), and data-driven decisions for collaboration of parties offering matching services in the most dynamic and efficient way.
- **Assessment data:** the dashboard of KPIs collected from the concluded project are elaborated to both understand if the single solution has been successfully implemented and derive statistics and trends useful to improve the design of future solutions and, hence, the provided service. Registry data: data providing information about the registered companies.

The DMP of the CIRCULOOS project has been prepared following the template provided by the European Commission of the "Guidelines on Data Management in HORIZON".[10] In the following chapters, there will be a preliminary version of the data sets that will be generated within the project, even if only a few partners were already able to provide information on the data they expect to generate. Also, general information about data (types, format, re-use, origins, size, etc); data strategies (FAIR data); allocation of resources and data security are provided.

[10] Horizon-Europe-Data-Management-Plan-Template.pdf (enspire.science)

2 Data Set identification

The CIRCULOOS DMP describes the observed data that are collected and processed during the lifetime of the project while providing the overview of available research data, access policies, data management and terms of use. The DMP reflects the current state of the discussions, plans and ambitions of the partners. It includes the preliminary scenario of data set definitions and expected results and will be updated and implemented with new datasets and results during the lifespan of CIRCULOOS. Moreover, CIRCULOOS project takes part to the Open Research Data Pilot (ORDP) in HORIZON that aims to improve access to and re-use of research data with a special focus on the need to balance openness and protection of scientific information. This version of the DMP contains a preliminary dataset where information listed below reflects the conception and design of the different work packages by the individual partners at the beginning of the project.

The data register will deliver information according to information detailed in Annex 1 (Part A) of the Grant Agreement Document (GA):

- **Data set reference and name:** identifier for the data set to be produced.
- **Data set description:** descriptions of the data that will be generated or collected, its origin or source (in case it is collected), nature, scale, 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.
- **Partners' activities and responsibilities:** partner owner of the device, in charge of the data collection, data analysis and/or data storage, and WPs and tasks it is involved.
- **Standards and metadata:** reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created. Format and estimated volume of data.
- **Data exploitation and share:** description of how data will be shared, including access procedures and policy, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of 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.) and if this information will be confidential (only for members of the Consortium and the Commission Services) or public. In case a dataset cannot be shared, the reasons for this should be mentioned (e.g., ethical, rules of personal data, intellectual property, commercial, privacy-related, security related).
- **Archiving and preservation (including storage and backup)**: description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered. Such data can be anonymised for statistical or other dissemination purposes and shared with open access, which could be further analysed and provide the possibility to extract information and knowledge from them. Each dataset can be accompanied by several metadata (e.g. type, gender, age, etc.) which can support various kinds of historical data analysis.

2.1 Data set identification by partner organisations

All partners need to identify the data that will be produced in the different project activities they are involved in and must provide an overview on the nature and details for each dataset.

To do so partners have been asked to fill a table with data they expect to treat in their activities according to the WPs and tasks they are involved in. The table template to be used is shown in Table 2.

Table 2: Data Identification template

Data Identification	
Data set description Type of data: qualitative or quantitative? Order of magnitude	Describe the existing or intended data, indicating their origin, nature and order of magnitude. Motivate the creation of new data sets and their added value.
Provenance of data: sources	Describe whether the data come from interviews, surveys or are extracted from disciplinary archives, databases and / or other projects, devices, machines...
Nature and formats of data	Describe nature and format of data: a) text documents (DOC, ODF, PDF, TXT, etc.); b) images (JPG, GIF, SVG, PNG, TIFF); c) video / film (MPEG, AVI, WMV, MP4); d) audio recordings (MP3, WAV, AIFF, OGG, etc); e) structured data (HTML, JSON, TEX, XML, RDF); f) tables (CSV, ODS, TSV, XLS, SAS, Stata, SPSS portable); g) source codes (C, CSS, JavaScript, Java, etc.); h) configuration data (INI, CONF, etc.) i) database (Microsoft Access, MySQL, Oracle, etc.)
New data set value	Motivating the creation of the new dataset, defining its added value for the scientific community or other recipients
Audio-visual material	In case of video, indicate the duration (the data is useful for planning the costs design and archiving)
Partners Activities & Responsibilities	
Partner owner of the device producing the data	Partner of the project owner of the device/software producing the data
Partner in charge of the data collection (if different)	

Partner in charge of the data analysis (if different)	
Partner in charge of the data storage (if different)	
WPs and tasks	
Standards and Metadata	
Metadata standards and data documentation	Describe the type of metadata with reference to standards and documentation
Methodology for data collection/generation	Describe the methodologies of data collection and production during the research process. 1. Who and how it collects data; 2. Who and how it structures and stores them; 3. Who and how he processes them; 4. Who and how he distributes them. Refer to regulations or practices in force in the scientific community of reference
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	How can data be exploited? For what purpose? Will data be exploited in their raw form coming from the dataset or will data analysis be exploited?
Data ownership	Who is the owner of the data? Is another organization contributing to the data development? Are you re-using some pre-existing data?
Suitability for sharing	Public/confidential/limited access
Data utility	How will this data shared/made accessible for verification and re-use?
Open research data pilot	Can data be uploaded in an open research data pilot? When?
Embargo periods (if any)	
Archiving & preservation (including storage and backup)	
Managing, storing and curating data	Please describe the modality of: ● storage; ● backup; ● transmission;

	● data processing in the short and medium term, with references to practices, standards and regulations where applicable.
Data Storage	Please indicate: ● where data will be stored ● if the conservation concerns the whole collected data or only part of them, ● for how long data will be stored

2.1.1 Partners' data identification

The following tables report partners' feedback on data that will be generated during the project's activities. The data analysis follows the structure presented in Table 2

Table 3: ED Data identification

Data Identification	
Data set description Type of data: qualitative or quantitative? Order of magnitude	Registration data for companies that participate in RAMP. Some personal data will be collected from representative personnel for verification purposes (contact points). The data is collected during the registration process and the users are required to consent to the storage and processing of their data. The data consists of textual information in the magnitude of <1Mb. A minimal set of user's behaviour inside the marketplace may be collected to support matchmaking and the creation of new business opportunities between the registered users.
Provenance of data: sources	Provided by the humans (during the registration process).
Nature and formats of data	Raw text data, formatted in JSON files.
New data set value	1. Understand the users' needs and improve/ adjust the functionality of RAMP 2. Matchmake end-users to solution providers 3. Inform end-users about the 'offering' 4. Validate the offering of solution providers in the Marketplace
Audio-visual material	N/A
Partners Activities & Responsibilities	
Partner owner of the device producing the data	ED (RAMP)
Partner in charge of the data collection (if different)	
Partner in charge of the data	

analysis (if different)	
Partner in charge of the data storage (if different)	
WPs and tasks	WP3
Standards and Metadata	
Metadata standards and data documentation	N/A
Methodology for data collection/generation	Form is presented to user during the registration process. Participants of Open Calls will be required to register to RAMP
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	See New Data Set value above
Data ownership	The users own the data. They can retrieve / remove the data by sending a request to RAMP administrators
Suitability for sharing	limited access to other RAMP residents (see Questionnaire below
Data utility	N/A
Open research data pilot	No
Embargo periods (if any)	
Archiving & preservation (including storage and backup)	
Managing, storing and curating data	Data is securely stored and accessible in ED's ISO27001:2013-certified infrastructure. End-to-end encryption for data in transit. Data in transit is protected against active (e.g., replays, traffic injection) and passive attacks (e.g., eavesdropping), thus ensuring data integrity
Data Storage	Data is securely stored and accessible in ED's ISO27001:2013-certified infrastructure. Data is deleted once the users request to leave RAMP. Data coming from the Open Calls are expected to stay at least for the duration of the individual Open Call projects.

Table 4: Dutch Pilot 3A - 1st Platform - Plennid - Houthub (Sustainable Sourced Fresh Timber Platform (SSFTP)

Data Identification	
Data set description Type of data: qualitative or quantitative? Order of magnitude	No existing data will be used. For pilot A, **Local timber hub** - (Sustainable Sourced Fresh Timber Platform (SSFTP) will create demand and supply data. With this platform we are coordinating a market for locally sourced fresh wood. Important parameters are: 1. Locally sourced supply - Reduction of transportcosts - Characteristics of supply (quantities, location, quality, ect) - Argis data from municipalities 2. demand coordination - Market insights - Functional characteristics of demand (application, quantities, durability ect) - Market development (creating awareness) - Interviews, marketresearch, dataanalysis and deskstudies
Provenance of data: sources	Our data is collected with databases like ArcGis and Airtable. Also interviews and surveys are necessary for market insight and demanded parameters. Desk-research and physical research will give important characteristics of supply.
Nature and formats of data	Although the precise format of the data is unknown at this time, the following formats should be anticipated: a) text documents (DOC, PDF, TXT, etc.); b) images (JPG, PNG, TIFF); d) audio recordings (MP3); e) structured data (unknown) f) tables (CSV) Arcgis data: ● Geodatabase rasters. ● TIFF (. tif) ● Esri Grid. ● CRF raster (. CRF) ● ERDAS IMAGINE (. img)
New data set value	Match making *"Supplydemand".* We use the dataset to get knowledge about the characteristic/parameters that the end-users want to see in the platform. Such as: - Buzzwords for searching specific application of source - insights in harverst moment (pre-ordering based on Gis) - Insights in marketprice (price setting)

Audio-visual material	N/A
Partners Activities & Responsibilities	
Partner owner of the device producing the data	(unknown at the moment)
Partner in charge of the data collection (if different)	
Partner in charge of the data analysis (if different)	
Partner in charge of the data storage (if different)	
WPs and tasks	WP3
Standards and Metadata	
Metadata standards and data documentation	Unknown
Methodology for data collection/generation	Plennid is collecting, structuring, processing, distributing data about fresh locally timber streams. We need guidance and from Circuloos partners.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	See New Data Set value above
Data ownership	Plennid
Suitability for sharing	Limited access, but data will always be available on platform (No extract of dataset possible)
Data utility	Unknown but we strive towards Pre-ordering possibilities
Open research data pilot	Developing an algorithm which can provide an optimal match making in the locally fresh timber market. Data will be open-source through the platform to provide research information (No extract of full dataset possible)
Embargo periods (if any)	N/A
Archiving & preservation (including storage and backup)	
Managing, storing and curating data	Safe and user friendly storage needed

Data Storage	Not clear yet

Table 5: Leather recycling pilot – leather waste platform of supply chain (Khoani Ltd. – B&A Ltd. – ITI Hungary Ltd.)

Data Identification	
Data set description Type of data: qualitative or quantitative? Order of magnitude	Data of produced and potentially available leather leftovers and forecast about it. The data is collected during the manufacturing and leftover selection processes. The data consists the characteristics of usable leftovers, and the characteristics of raw materials that the potential re-users needs.
Provenance of data: sources	Provided by the members of supply chain.
Nature and formats of data	b) images (JPG, GIF, SVG, PNG, TIFF); i) database (Microsoft Access, MySQL, Oracle, etc.)
New data set value	1. Understand the users' needs 2. Matchmake end-users to leftover producers 3. Provide information on remaining material stocks 4. Inform end-users about the 'offering' 5. Inform leftover producers about future needs
Audio-visual material	N/A
Partners Activities & Responsibilities	
Partner owner of the device producing the data	KHOANI
Partner in charge of the data collection (if different)	N/A
Partner in charge of the data analysis (if different)	N/A
Partner in charge of the data storage (if different)	Unknown
WPs and tasks	WP3
Standards and Metadata	
Metadata standards and data documentation	Unknown
Methodology for data collection/generation	KHOANI is collecting, structuring, processing, distributing data about own leftovers and forecast about it. B&A producing information about own raw material needs. ITI making the platform, that communicating between the actors.

Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	See New Data Set value above
Data ownership	KHOANI
Suitability for sharing	Limited access for platform users, based on their registration, but data will always be available on platform
Data utility	It is unknown, but we are trying to expand the user base (more leftover re-users)
Open research data pilot	No
Embargo periods (if any)	N/A
Archiving & preservation (including storage and backup)	
Managing, storing and curating data	Data is securely stored and accessible in the KHOANI's webpage. End-to-end encryption for data in transit. Data in transit is protected against active (e.g., replays, traffic injection) and passive attacks (e.g., eavesdropping).
Data Storage	Data is securely stored and accessible. Data is deleted once the if the residual leather is used or if the demand for raw materials no longer exists.

Table 6: Plastic material recycle/remanufacture (Thermolympic S.L. – Contenedores Lolo S.L – Canonical Robots S.L)

Data Identification	
Data set description Type of data: qualitative or quantitative? Order of magnitude	The data originates from an array of plastic byproducts produced by Thermolympic. These plastics are transported to Contenedores Lolo, where they undergo classification through an automated sorting system. The information derived from this sorting process is then systematically cataloged, serving as a valuable resource for subsequent utilization by other market participants. This process ensures a meticulous collection of data, aligning with the requirements for reuse and recycling within the industry.

	Order of magnitude for data produced during a year: 10GB
Provenance of data: sources	This source of the dataset is the system of automatic classification using hyperspectral images.
Nature and formats of data	e) structured data (HTML, JSON, TEX, XML, RDF);
New data set value	The goal is to refine the "Supply demand" platform's match-making functionality, specifically for the RAMP ecosystem. This dataset is pivotal for: 1. Enhancing User Experience: provide insights into end-users' preferences, such as market pricing. 2. Streamlining Matchmaking: more accurately connect end-users with appropriate solution providers. 3. Educating and Validating: inform users about offerings and to validate the solutions provided in the Marketplace.
Audio-visual material	N/A
Partners Activities & Responsibilities	
Partner owner of the device producing the data	CANONICAL ROBOTS S.L. (CAN)
Partner in charge of the data collection (if different)	N/A
Partner in charge of the data analysis (if different)	N/A
Partner in charge of the data storage (if different)	Unknown
WPs and tasks	WP3
Standards and Metadata	
Metadata standards and data documentation	Unknown

Methodology for data collection/generation	CANONICAL is gathering, organizing, analyzing, and disseminating information on plastic byproducts from Thermolympic, which undergo automatic classification at the Contenedores Lolo facilities.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	To improve the "Supply demand" platform's efficiency in the RAMP ecosystem, a key dataset is utilized to refine match-making by understanding user preferences, like market pricing. This approach facilitates accurate pairing of users with solutions, enhances user education about offerings, and ensures solution credibility, boosting platform effectiveness and trust.
Data ownership	Thermolympics S.L.
Suitability for sharing	Access to the platform is restricted for users based on their registration status, yet the data will remain constantly accessible on the platform.
Data utility	The data offers producers insights into user behavior and preferences for targeted enhancements and offerings, and enables informed decision-making and personalization for registered users.
Open research data pilot	No
Embargo periods (if any)	N/A
Archiving & preservation (including storage and backup)	
Managing, storing and curating data	Data is securely stored and accessible in the KHOANI's webpage. End-to-end encryption for data in transit. Data in transit is protected against active (e.g., replays, traffic injection) and passive attacks (e.g., eavesdropping).
Data Storage	Data during transmission is safeguarded Through encryption and secure transmission protocols; from both active attacks, such as replays and traffic injection, and passive attacks, like eavesdropping.

Table 7: Data identification - Open calls

Data Identification F6S

Data set description Type of data: qualitative or quantitative? Order of magnitude	Proof that MSMEs are legal entities and abide by the instructions of the Guide for Applicants that will be created in the project , for example not declared bankruptcy. Description of the technical solutions, level of maturity and potential for market use. Generic information will be collected in textual or numeric format, while the data regarding value propositions will be collected in a multiple choice format. Personal data, including the names and contact details, of the organisation's coordinator and team members will also be processed.
Provenance of data: sources	Applicants fill-out the templates during the application process. F6S will collect data from academia, research institutes, innovative SMEs and hi-tech startups through an online form.
Nature and formats of data	Text data collected via the online form PDF/ Word file corresponding to a filled-in template. The datasets to be collected during the Open Calls in order to facilitate a good analysis of proposals include (non-exhaustive list): country; organisation name; address; number of team members; funds requested; application domain; solution type; name of coordinator and name of the team members; abstract; brief description; company years of experience. The information will be accessible through an online Dashboard application and it will be downloadable in CSV and XLS formats.
New data set value	Evaluate the eligibility of applicants for the Open Calls. The type of data collected will include specific indicators to evaluate the potential of the open call's proposals. Such indicators include measurements of the innovation potential and maturity of proposals, the team and the organisation proposers, the technology used and technology experience, the market orientation, the financial aspects, and the benefits expected.
Audio-visual material	Links to online resources
Partners Activities & Responsibilities	
Partner owner of the device producing the data	F6S
Partner in charge of the data collection (if different)	F6S
Partner in charge of the data analysis (if different)	F6S.
Partner in charge of the data	F6S

storage (if different)	
WPs and tasks	WP6
	Standards and Metadata
Metadata standards and data documentation	N/A
Methodology for data collection/generation	Online Form and application collection platform hosted by F6S All the datasets will be represented in a mapping of submitted proposals. The selected proposals are a subgroup of the submitted proposals and are identified with the field. It is expected to generate a deal-flow of [total number of expected applicants] applicants through [total number of open call to implement] Open Calls for the project which will contribute to create [number of data sets] main datasets: (1) third parties that start an application but don't submit a proposal; (2) submitted proposals; (3) deleted/withdrawn applications; (4) evaluated proposals; (5) winning start-ups; and (6) follow up metrics. The information will be captured through online F6S forms and will be recorded and reported. The information will be accessible through an online Dashboard application and it will be downloadable in CSV and XLS formats. Only authorised users will be allowed to access the data sets via authentication.
	Data exploitation & sharing
Data exploitation (purpose/use of the data analysis)	Explicit consent for data preservation and sharing will be obtained from applicants through an online form. The data will be exploited for three main purposes: (1) evaluation of proposals; (2) impact assessment; (3) research/statistics. The individual registrants' data will only be accessible for evaluation purposes carried out by accredited and authorised evaluators. Each evaluator will be granted limited access to a restricted number of registrants from the dataset. The access will be online for a limited period of time, using a secure mode via authentication. The datasets will be anonymized for impact assessment and research purposes. The anonymized datasets will be exploited through the creation of maps and charts that will be updated at the end of the selection process of each Open Call. The maps and charts generated will be publicly shown as part of the dissemination activities of the project. The full dataset of anonymized data will also be available for third parties that

	request access to the information for research purposes. A 'Data Management Plan' will be developed as indicated in WP6 detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use.
Data ownership	CIRCULOOS Consortium; records are kept for reasons of accountability to the EU.
Suitability for sharing	Confidential; access only to EU representatives.
Data utility	N/A
Open research data pilot	No
Embargo periods (if any)	The data is stored for accountability purposes to the EU.
Archiving & preservation (including storage and backup)	
Managing, storing and curating data	F6S
Data Storage	F6S

Table 8: MWCB Data Identification

Data Identification MWCB	
Data set description Type of data: qualitative or quantitative? Order of magnitude	MWCB plan to collect data from the following sources: newsletters, video – interviews, website contact forms, events – Networking (online and in-person), social media (LinkedIn,) and blogs. MWCB plans to collect names and e-mails from interested third parties in different formats (.csv, .txt, .xlsx, .pdf, .doc).
Provenance of data: sources	Provided by the humans (during the registration process)..
Nature and formats of data	Textual (.txt, .docx, .pdf), Tabular (.csv, .xlsx), Video (.mp4, .ogv)
New data set value	The dataset is used to inform and engage identified target groups about the project and its outcomes according to the impact maximisation plan. Furthermore, it serves to attract interested parties, encouraging their participation in open calls and other planned activities such as the trainings.
Audio-visual material	yes
Partners Activities & Responsibilities	
Partner owner of the device producing the data	MWCB

Partner in charge of the data collection (if different)	
Partner in charge of the data analysis (if different)	
Partner in charge of the data storage (if different)	
WPs and tasks	WP6
Standards and Metadata	
Metadata standards and data documentation	N/A
Methodology for data collection/generation	Data from interviews, trainings, website contact form, events, social media, and blog information will be stored on MWCB servers. The collection of personal data will be gathered according to MWCB privacy policy (https://mobileworldcapital.com/privacy-policy/) and will follow guidelines agreed with the consortium.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	See New Data Set value above
Data ownership	MWCB
Suitability for sharing	The data gathered by MWCB from the activities of Impact Maximization (WP6) will not be shared with any other third parties outside MWCB to comply with current regulations.
Data utility	MWCB also fulfil their commitment to storing them and therefore take all the necessary measures to prevent any loss, unauthorized processing or access or alteration thereto, as established in the applicable data protection regulations.
Open research data pilot	N/A
Embargo periods (if any)	N/A
Archiving & preservation (including storage and backup)	
Managing, storing and curating data	The data will be stored on MWCB servers with access to members of the team only. This information will not be shared with any other partner or third party. The data will be stored in MWCB servers with access to members of the [project's name] team only. MWCB undertakes to use and process users' personal data while respecting their confidentiality and to avail of them in accordance with the purpose for which they were initially collected

Data Storage	The collection of personal data will be gathered according to MWCB privacy policy (https://mobileworldcapital.com/privacy-policy/).

Table 9: SUPSI Data identification

Data Identification SUPSI	
Data set description Type of data: qualitative or quantitative? Order of magnitude	Qualitative and quantitative data will be sourced from SUPSI, with a rigorous scientific approach employed for data acquisition. Quantitative data will be derived through the calculations performed in GRETA, a customized application developed by SUPSI for comprehensive sustainability evaluations. These data involve sustainability indicators (social, environmental, economic, and circular economy metrics) set with specific units. The magnitude of these indicators may exhibit variance, necessitating potential normalization for comparative analysis. Qualitative data will be harnessed from the utilization of quantitative data in the formulation of sustainability reports, derived from both pilot studies and open calls.
Provenance of data: sources	Both quantitative and qualitative data will be generated from GRETA. The inputs required to produce GRETA results can be sourced from interviews, surveys, disciplinary archives, or databases. However, the data products are systematically derived from calculations and analyses conducted within the GRETA software environment.
Nature and formats of data	Nature and format of data: a) text documents (PDF) for reports; b) structured data (JSON) as exchanged messages obtained exploiting the REST API exposed by GRETA.
New data set value	Compiling a list of sustainability indicators involving social, environmental, economic, and CE provides a significant added value for the scientific community and other stakeholders by contributing to the expansion of existing databases focused on sustainability.
Audio-visual material	None.

Partners Activities & Responsibilities	
Partner owner of the device producing the data	SUPSI
Partner in charge of the data collection (if different)	SUPSI for LCI data and process characterization and ED as RAMP manager
Partner in charge of the data analysis (if different)	SUPSI
Partner in charge of the data storage (if different)	ED as RAMP manager
WPs and tasks	WP3 - Sustainability and LCA Assessment tools WP5 - DEM - Demonstrator, pilot, prototype
Standards and Metadata	
Metadata standards and data documentation	DOI, author's name list, keywords
Methodology for data collection/generation	1. Data Collection: This process involves researchers, analysts, and experts. These individuals use GRETA to calculate relevant sustainability data pertaining to different sectors, such as wood, plastic, and leather. In particular, the data collection carried out by sustainability experts will be fundamental in order to model processes and products from a sustainability point of view, creating the customization spaces on which the users will be able to generate their product alternatives. 2. Structuring and Storage: Once collected, the data are structured and stored within the GRETA software. 3. Processing: Data processing occurs within the GRETA software environment. The data from pilots undergo thorough analysis to provide the sustainability indicators. 4. Distribution: The processed data, along with the qualitative findings in the form of reports, are disseminated through RAMP. Through RAMP, stakeholders, policymakers, and the scientific community can access and use the data generated by GRETA for further research, decision-making, and policy formulation.
Data exploitation & sharing	

Data exploitation (purpose/use of the data analysis)	The data provided can be exploited in different ways: in their report form they can be used to understand conceptual insights. Alternatively, the quantitative data can be applied in further studies by the scientific community, and comparative assessment's data can inform decision-making processes. Reports can also be accessed by various sectors to improve sustainability practices, expanding the impact beyond the scientific community. Overall, data analysis and interpretation, whether in raw form or through reports, serve diverse purposes, from understanding concepts to guiding practical improvements across sectors.
Data ownership	SUPSI
Suitability for sharing	Limited access for specific platform users
Data utility	Publications, Deliverables, RAMP, Sustainability certification
Open research data pilot	Yes
Embargo periods (if any)	
Archiving & preservation (including storage and backup)	
Managing, storing, and curating data	Please describe the modality of: ● Storage: GRETA leverages on different DBs with different natures according to the nature of data. ● Backup: GRETA cluster is running on SUPSI virtual machine which is daily backup; ● Transmission: all data exchanged with third-party applications is in JSON format and is exchanged over a secure channel (HTTPS); ● data processing in the short and medium term: short term for the basic functionalities such as assessment and comparison (assessment usually takes less than a second). Some advisory functionalities (still under development) might require a long time. with references to practices, standards and regulations where applicable.
Data Storage	Please indicate: · where data will be stored: GRETA environment and RAMP platform · if the conservation concerns the whole collected data or only part of them: the whole data set in GRETA, part of them in RAMP platform · for how long data will be stored: unlimited time (or at least for the period of the project).

Table 10: INN Data identification

Data Identification INN

Data set description Type of data: qualitative or quantitative? Order of magnitude	INN collected qualitative data from the pilot partners during the completion of T2.1 where the pilot partners provided details about their production processes, goals and vision of the project, along with potential cooperation and production optimisation opportunities.
Provenance of data: sources	The data had been collected and edited on miro boards, excel files and in D2.1. All collected data is from online interviews with the pilot and task partners.
Nature and formats of data	Nature and format of data: a) text documents: D2.1 deliverable b) images (JPG): provided by the pilot partners c) video: provided by pilot partners, stored on ProofHub d) tables (XLS): filled by INN and pilot partners, describing use cases
New data set value	The collected data is used to analyse the pilots in more depth and provide an overview to the technical team. New data can be collected regarding follow-up specifications.
Audio-visual material	N/A
Partners Activities & Responsibilities	
Partner owner of the device producing the data	N/A
Partner in charge of the data collection (if different)	INN
Partner in charge of the data analysis (if different)	INN
Partner in charge of the data storage (if different)	INN, ED as owner of ProofHub project space
WPs and tasks	WP2
Standards and Metadata	
Metadata standards and data documentation	N/A
Methodology for data collection/generation	Methodologies of data collection and production during the research process: 1. INN and SUPSI collected qualitative data during interviews with pilot partners 2. INN stores the collected data in ProofHub 3. INN and SUPSI analysed and documented the information 4. The collected data will not be shared with any third parties INN handles personal data according to its privacy policy: https://innomine.com/privacy-policy/ and according to the agreed guidelines with the consortium.

Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	The collected data is used to analyse the pilots in more depth and provide an overview to the technical team. The data is not used in its raw form, the interview answers have been structured and categorised, to provide a well-defined input for T2.2 and T2.3.
Data ownership	INN, ED as owner of ProofHub cloud project space
Suitability for sharing	Confidential, the collected data is for internal use only
Data utility	INN fulfil their commitment to storing the data and therefore take all the necessary measures to prevent any loss, unauthorized processing or access or alteration thereto, as established in the applicable data protection regulations. All documents, meeting memos, txt and xls files are stored in ProofHub.
Open research data pilot	N/A
Embargo periods (if any)	N/A
Archiving & preservation (including storage and backup)	
Managing, storing and curating data	All documents, meeting memos, txt and xls files are stored in ProofHub. This information will not be shared with any other third party.
Data Storage	Data storage is according to ProofHub's and INN's privacy policy.

Table 11: Thermolympic (PILOT1 Plastic recycling)

Data Identification – PILOT1 Plastic recycling

Data set description Type of data: qualitative or quantitative? Order of magnitude	Machine configuration, homologation, production and logistic data for comparative purposes with the newly created recipes. For Thermolympic and his customers, the opening of multiple configuration possibilities for using recycled material within normal productions while guarantees the dimensional, mechanical and technical characteristics. Once the quality and stability of the process is guaranteed with full traceability: ● Raw material origin certificate from the supplier. ● Production order serial releases with parameters, recipe mixture, quality controls, training matrix and shift data. ● SPC controls This will trace through the production chain creating a declaration of conformance that met all the regulations from customer specific requirements. With that objective met, the main focus will be moved to logistics with Lolo and local suppliers where additional data will be used and shared for the supply chain analysis like: ● Customer specific requirements with the criteria to be met for each new material included into the system. ● Data form the existing costs for including new materials on existing lines (homologation, functional test, energy costs, production cycle impacts…) ● Logistic insights and costs (Km/€) with comparative data for economic viability and supplier relay. ● Production planification and capabilities on all supply chain for a JIT or batched production.
Provenance of data: sources	The data will come basically from production lines, so most of them will relate their origins from IOT devices, local Apps and MES/ERP systems on the company.
Nature and formats of data	Until the relationship within the consortium platform where this can change it will be: -Text documents (DOCX, ODF, PDF, TXT) -Tables (XLS) -Images (JPG, GIF, SVG, PNG, TIFF); -Mail messages (.msg). -Structured data on DataBases (SQL Server and Oracle DB) -Structured data on the apps (HTML, JSON,XML);
New data set value	-Cover the gaps between the current production systems for enable the use of recycled materials in a heavily restricted environment.
Audio-visual material	-Training material for awareness and new methodologies.
Partners Activities & Responsibilities	
Partner owner of the device producing the data	TBD: -Thermolympic will be the owner for production and endcustomer data. -Contenedores Lolo will be the owner for recycling data produced from the material separation, and the origin/traceability certificates.

Partner in charge of the data collection (if different)	-
Partner in charge of the data analysis (if different)	-
Partner in charge of the data storage (if different)	-
WPs and tasks	-
Standards and Metadata	
Metadata standards and data documentation	-
Methodology for data collection/generation	-Thermolympic production, quality, data is automatically collected through IOT or HMI devices, structured into tables and stored into databases and distributed through service apps. -Contenedores Lolo data is manually collected, stored into their facilities (paper and digital data) and distributed through their internal channels.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	Data value explanation.
Data ownership	Customer data should be anonymized for project purposes. All the data will be owned by their creator (or authorized entities) unless agreements are reached.
Suitability for sharing	Limited access. To be defined which data and how can be shared.
Data utility	TBD.
Open research data pilot	Can data be uploaded in an open research data pilot? When? It can be upload into open research's, but anonymized according a data protection act.
Embargo periods (if any)	N/A
Archiving & preservation (including storage and backup)	
Managing, storing and curating data	**Storage:** Digital mostly, into defined servers or storage services. **Backup:** Local backups for servers, critical data with backup on cloud systems like (ACENS) **Transmission:** TBD, secure protocols to be checked.
Data Storage	The data will be stored in local servers for digital data and internal storages for paper ones. The conservation plan is for three years on direct query for digital data and then will be archived into xml for preservation and recovery purposes (planned maintenance for the next 20 years). Paper data will be stored for 6 years.

Table 12: FIW Data identification

Data Identification – FIWARE	
Data set description Type of data: qualitative or quantitative? Order of magnitude	● N/A
Provenance of data: sources	● N/A
Nature and formats of data	FIWARE actions are cross-cutting to the rest of participants. Heterogeneous interfaces, data sources, and datasets will evolve so that seamless interoperability is achieved through the use of the Open Standard NGSI-LD API and harmonized open data model templates (Smart Data Models)
New data set value	● N/A
Audio-visual material	● N/A
Partners Activities & Responsibilities	
Partner owner of the device producing the data	- ED/RAMP, WP3 Module Owners, WP4 Data Sharing Framework developers, WP5 (hosted Pilots + FSTP experiments)
Partner in charge of the data collection (if different)	-
Partner in charge of the data analysis (if different)	-
Partner in charge of the data storage (if different)	-
WPs and tasks	- WP5→ Execution of Demonstrators
Standards and Metadata	
Metadata standards and data documentation	- NGSI-LD - Smart Data Models
Methodology for data collection/generation	● N/A
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	N/A
Data ownership	N/A
Suitability for sharing	N/A
Data utility	N/A
Open research data pilot	N/A
Embargo periods (if any)	N/A
Archiving & preservation (including storage and backup)	
Managing, storing and curating data	N/A
Data Storage	N/A

Table 13: ALA Data identification

Data Identification ALA	
Data set description Type of data: qualitative or quantitative? Order of magnitude	Fetching data from companies that want to deploy a node on Alastria's blockchain networks or access nodes via APIs. Some personal data (first and last name, company name, email, IP address) will be collected from representative personnel for verification purposes (contact points). The data is collected during the registration process and the users are required to consent to the storage and processing of their data. The data consists of textual information in the magnitude of <1Mb.
Provenance of data: sources	Provided by the humans (during the registration process).
Nature and formats of data	Text data collected in forms, stored in Alastria's online repository and the GitHub repository.
New data set value	Access to the blockchain networks that support the deployment of the project's pilots.
Audio-visual material	N/A
Partners Activities & Responsibilities	
Partner owner of the device producing the data	ALA
Partner in charge of the data collection (if different)	
Partner in charge of the data analysis (if different)	
Partner in charge of the data storage (if different)	
WPs and tasks	WP4, WP5
Standards and Metadata	
Metadata standards and data documentation	N/A
Methodology for data collection/generation	Form is presented to the user during the registration process.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	See New Data Set value above
Data ownership	The users own the data. They can retrieve / remove the data by sending a request to ALA administrators.

Suitability for sharing	N/A. Data is not accessible except by the owner.
Data utility	N/A. Data is not accessible except by the owner.
Open research data pilot	No
Embargo periods (if any)	N/A
Archiving & preservation (including storage and backup)	
Managing, storing and curating data	Alastria has technical, organizational and security policies and measures implemented to comply with the GDPR regulations. It also has a legal working group on privacy, which supports the association with any questions that may arise.
Data Storage	Data is securely stored and accessible in Alastria's online repository and the GithHub repository. Data is expected to be stored at least for the duration of the pilots. Data is deleted once the users request it.

Table 14: CUT Data identification

Data Identification	
Data set description Type of data: qualitative or quantitative? Order of magnitude	Models of intra-factory processes Visualizations of intra-factory processes Models of supply chains Visualizations of supply chains Optimized sequences of intra-factory processes Optimized supply chain arrangements Predictions of intra-factory evolutions Predictions of supply chain evolutions Predictions of combined intra-factory and supply chain evolutions Order of magnitude: Depending on the size of the problem
Provenance of data: sources	Provided by the humans Provided by factory logs Provided by models
Nature and formats of data	Raw text data, formatted in JSON files. CSV data Images (png, jpg, etc)
New data set value	1. Create, adjust and optimize models of intra-factory processes

	and supply chains 2. Visualize operations 3. Predict future outcomes 4. Use of data in scientific publications and to further improve the SCDT and SCOPT modules 5. Optimize intra-factory and supply chain efficiency
Audio-visual material	Images of raw materials, intermediate products, finished products, scrap materials, virtual environment of SCDT
Partners Activities & Responsibilities	
Partner owner of the device producing the data	CUT (SCDT, SCOPT)
Partner in charge of the data collection (if different)	
Partner in charge of the data analysis (if different)	
Partner in charge of the data storage (if different)	
WPs and tasks	WP3
Standards and Metadata	
Metadata standards and data documentation	N/A
Methodology for data collection/generation	Logs that are available on RAMP. Factory ERP/WMS/SCDT connections on RAMP Forms filled by users
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	See New Data Set value above
Data ownership	The users own the data. They can retrieve / remove the data by sending a request to RAMP administrators
Suitability for sharing	limited access to other RAMP residents (see Questionnaire below
Data utility	Improve SCDT and SCOPT, including through RAMP services
Open research data pilot	No
Embargo periods (if any)	Depends on the case
Archiving & preservation (including storage and backup)	

Managing, storing and curating data	Data is securely stored and accessible in ED's ISO27001:2013-certified infrastructure. End-to-end encryption for data in transit. Data in transit is protected against active (e.g., replays, traffic injection) and passive attacks (e.g., eavesdropping), thus ensuring data integrity
Data Storage	Data is securely stored and accessible in ED's ISO27001:2013-certified infrastructure. Data is deleted once the users request to leave RAMP. Data coming from the Open Calls are expected to stay at least for the duration of the individual Open Call projects.

3 Data Summary

3.1 Reasons for collecting and generating data within CIRCULOOS activities

The main goal behind collecting users' data is to better understand which features/parts of the CIRCULOOS platform need to be further improved. All collected information and data will be used during the CIRCULOOS project lifecycle and exclusively for the purpose of testing the platform, increasing knowledge, adapting to possible users' needs. The current version of the DMP is a preliminary version of the document, that will be updated and augmented with new datasets and results during the lifespan of CIRCULOOS. For completing the DMP deliverable, partners have provided input following the HORIZON Data Management Plan template [11] as well as the Data Set Template, which will be the basis to fill in and expand in the following months as consortium partners will be requested to investigate and provide more detailed information on their produced data and whether these are discoverable, accessible, assessable, and intelligible, useable (beyond the original purpose) and interoperable. A tenet of CIRCULOOS is that artificial intelligence (AI) has the potential to increase the profitability of small and medium manufacturing enterprises (SMEs) by improving product quality (e.g., through anomaly detection and correction), by optimising production line configuration (e.g., rapid reconfiguration to achieve lot-size-one production) and by levelling up staff productivity (e.g., through stress relief suggestions and personalised training). To enact such positive changes, the CIRCULOOS platform needs to acquire data from a variety of sources (e.g., sensors deployed on the shop floor, human resources databases, etc.) and have to translate them into information flow for the AI components that can process it. Apart from the general objective of collecting data for the improvement of the CIRCULOOS platform, other small and task-specific purposes were identified by the consortium as follows.

Data collected in the first months of the projects about end-users will help technology providers develop the CIRCULOOS solutions on the basis of their needs and to structure the future demonstration and validation activities.

From an exploitation/commercialisation point of view, collected data will be used to understand the users' needs and improve/ adjust the functionality of RAMP; matchmake end-users to solution providers, inform end-users about the 'offering', validate the offering of solution providers in the Marketplace (ED).

In the following version of the DMP, the Table 15 will be updated with a list of research data that partners are expected to generate in the project.

[11] h2020-tpl-oa-data-mgt-plan-annotated_en.pdf (europa.eu)

Table 15: Research Data

Research Data	Related Partners
KPIs from pilots	**Dutch Pilot 3A&B:** **Interviews:** 3A:5 interviews to get an overview of the market needs **Surveys:** 3A:30 surveys to get quantitative data about parameters and market needs **Desk research:** 3A&B: researches about characteristics of the supply **Field research:** 3 experiment about data collecting process -3A Onboarding Herso on Jutplaats with FSC tracking -3A Rijksmuseum : Boymans, exhibition with W.A.A.S. -3A Kava: Dutch Government office building -3A Processing impact -3B Rotterdamskrom (Modular housing parts) **Leather recycling pilot from Hungary:** **Desk research:** a desk research about the characteristics of leather (results will be incorporated into the leather database) **Pilot 1: Plastic material recycle Desk Research**: A study on the properties of plastics, findings to be integrated into the plastic reutilization database.
RAMP MSMEs registration data	ED
Production and sustainability data	Partner MSMEs SUPSI (LCA, LCC, S-LCA, CE) Pilots: HER, CON, THER, FICT, PLE, KHO, B&A, ITIH
3D Digital twins data of factories and products designs	CUT and all MSMEs 3D models, process models and predictive models and data
CIRCULOOS Stakeholders	F6S, ED, FIW
Processes and mission models	ED
Online courses and material	MWCB

3.2 Data types and formats collected and generated in CIRCULOOS

A non-exhaustive list of types and formats of data was provided to partners (see Table 2) in order to define the most common data types and format that will be generated within the project.

According to the consortium the main formats are:

- structured data (HTML, JSON, TEX, XML, RDF);
- tables (xlsx files) containing textual data (strings) and quantitative and qualitative information, including:
    - ① timestamp, barcode of the battery pack, charge, residue, forces and data that describes the location of the battery in the pack (face, cell, point);
    - ① data concerning the companies (manufacturing SMEs, automation solution and consultancy services providers (from the RAMP);
    - ① personal data about companies' profiles involved in the WP1 surveys, RAMP and CIRCULOOS Open calls;
- images (jpg or png) for training (dataset) and from pilots;
- raw text data (Json format) for module validation;
- PDF/ Word file corresponding to a filled-in template;
- database (Microsoft Access, MySQL, Oracle, etc).

Information regarding data types and formats are still provisional and will be more precise in the next version of the DMP.

Regarding the CIRCULOOS platform, it will be interfaced with different kinds of Internet of Things (IoT) devices and robots as well as factory information systems such as demand & supply management systems and human resources databases. Each IoT device or robot typically produces raw data in a proprietary format which may vary over time even within the same device and a similar degree of data format volatility can be expected of information systems. Thus, to acquire data from those environments, a plethora of diverse data formats will need to be understood by the platform, at least at its boundary where information is exchanged with external systems. Moreover, new formats may have to be accommodated as shop floors are connected to the platform.

3.3 Data characterisation: re-use of existing data, data origins and size

Some partners will reuse existing data. Regarding the demonstration activities within WP5, generally some already existing images and information coming from pilots manufacturing floors and machines, including images of defects, products and production lines, could be reused.

In the Dutch pilot 3A some GIS data from the municipality of Rotterdam may be used. Such datasets may be used to reveal future supply of virgin material. This provides the possibility to align demand.

Leather recycling pilot does not use secondary data.

Plastic recycling pilot does not use secondary data.

SUPSI plans to reuse existing data and methodologies developed by their own team for sustainability assessments (LCA, LCC, s-LCA and CE). This includes algorithms, codes and archetypes previously developed for this purpose. Additionally, we use commercial databases (such as Ecoinvent) for comprehensive data coverage.
PILOT1: Plastic recycling plans to use historical data retrieved from THER actual production lines and processes. Digitalized data are available from 2015 until now, with additional quality and processes data being stored since 2020.

CUT plans to re-use existing historical data from pilots and experiments for the needs of SCDT and SCOPT modules.

Some partners will use data of existing 'residents' that is companies already registered in RAMP. This data is provided by companies which register to the Marketplace, which is a mandatory step for the participation to the Open Calls. On general, the size of these data (textual description) is expected to be lower than 1 MB.

3.4 Types and formats of data that CIRCULOOS will generate/collect

To fulfil the purpose of the data collection/generation, the CIRCULOOS project will collect and generate the following types and formats of data:

Table 16: Types and formats of data

Data/Data Source	Data type	Data format	Data origin
Surveys, workshops/living labs data, validation cycles data	Electronic document	• Word document (.doc,.docx) • Excel document (.xls/.xlsx) • Pdf document	WP2, WP3, WP4, WP6
	Hardcopy	Paper/website surveys/Ramp surveys	
Deliverables	Electronic document	• Word document (.doc,.docx) • Excel document (.xls/.xlsx) • Pdf document	All WPs
	Hardcopy	Paper/Website	
Website public reports	Electronic document	• Word document (.doc/.docx) • Pdf document • Excel document (.xls/.xlsx) • .csv files • .txt	All WPs
Video files	Electronic document	.mov, .mpeg, .avi, .mp4, etc.	All WPs
Audio files	Electronic document	.mp3, .wav, etc.	All WPs
Images	Electronic document	.jpg, .png, .gif, etc.	All WPs
Software	Source Code	Source Code	WP3, WP4
Signed documents (eg. Consent forms, information sheets, attendance lists, Consortium Agreement, etc.)	Electronic document	• Word document (.doc,.docx) • Excel document (.xls/.xlsx) • Pdf document	WP1, WP2, WP5, WP6, WP7

	Hardcopy	paper	
Presentations	Electronic document	Powerpoint document	All WPs
	Hardcopy	paper	
Network and system related data	Electronic document		WP3, WP4, WP5, WP6

3.5 Information types in the CIRCULOOS project

Following the "Guidance Guidelines for the classification of research results" of the European Commission, the deliverables have three types of classification: PUBLIC, SENSITIVE, and RESTREINT UE/EU RESTRICTED.

3.5.1 PUBLIC

In Table 17, CIRULOOS deliverables are classified as PUBLIC.

Table 17: CIRULOOS Public Deliverables

Number	Deliverable Title
D1.1	Project Management Handbook
D1.2	Data Management Plan M6
D1.3	Data Management Plan M30
D1.4	Data Management Plan M42
D2.2	CMRA Specification M9
D2.3	CMRA Specification M18
D3.1	Supply Chain Orchestrator M12
D3.2	Supply Chain Orchestrator M24
D3.3	3D Digital Twin of supply chain/production/ products M12
D3.4	3D Digital Twin of supply chain/production/ products M24
D3.5	Sustainability and LCA Assessment tools M12
D3.6	Sustainability and LCA Assessment tools M24
D3.7	AI and Data-driven supply chain Optimisation M12
D3.8	AI and Data-driven supply chain Optimisation M24
D3.9	CV-based system for composition detection M12
D3.10	CV-based system for composition detection M24

D4.1	Data sharing layer M1
D4.2	Cyber secure data sharing layer M21
D4.4	First version of Integrated Platform
D4.5	Final version of Integrated Platform
D5.1	Early versions of 3 Pilot Demonstrators
D5.2	Final version of 3 Pilot Demonstrators
D5.4	Demonstration of EXDs
D5.5	EXDs scaled up demonstrations
D5.6	Lessons Learnt, feedback and recommendations handbook
D6.2	Periodical dissemination retrospect and follow-up action plan M12
D6.3	Periodical dissemination retrospect and follow-up action plan M30
D6.4	Periodical dissemination retrospect and follow-up action plan M42
D6.5	Open Calls report M12
D6.6	Open Calls report M24
D6.7	Open Calls report M42
D6.11	Report on training activities and platform M18
D6.12	Report on training activities and platform M42

SENSITIVE

In Table 18 , CIRCULOOS deliverables are classified as SENSITIVE.

Table 18: CIRCULOOS Sensitive Deliverables

Number	Deliverable Title
D2.1	Ambitious Scenarios for the pilots
D2.4	Technical components specification
D4.3	Integration plan and infrastructure
D5.3	Summary of EXDs and individual execution plans
D6.1	Initial Plan for the Exploitation and Dissemination of Results
D6.8	Market Analysis and Standardisation Plan M12
D6.9	Business plan (tools and platform, EXDs, and partners) M30

D6.10	Business plan (tools and platform, EXDs, and partners) M42
D7.1	OEI - Requirement No. 1

4 FAIR Data

The principles of FAIR (Findable Accessible Interoperable Reusable) data have been established by a set of different stakeholders like academia, industry, funding agencies, and scholarly publishers. According to Wilkinson et al. (2016)4, the FAIR Data Principles are a set of guiding principles in order to make data Findable, Accessible, Interoperable and Reusable (Figure below).

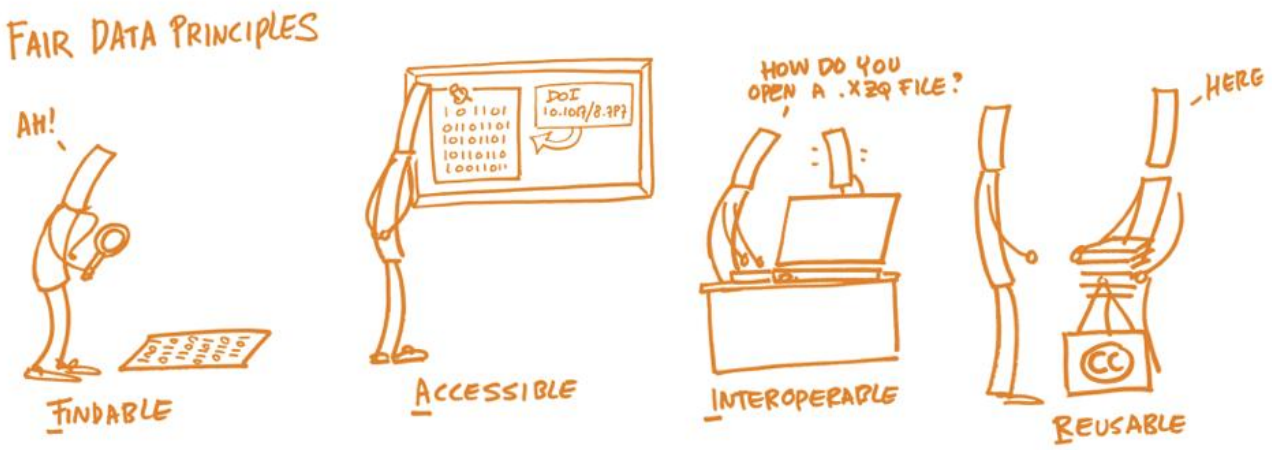

Figure 2: FAIR Data Principle

The FAIR Data Principles clearly provide a concise and measurable set of parameters that should be respected to ensure the availability and reusability of data for further research purposes by third parties that are not part of the project. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. The mentioned principles do not necessarily propose an explicit technology, standard or implementation solution, forego implementation choices and promote the maximum usage of data.

CIRCULOOS project complies with the principal priorities as defined below:

- The publications will be made available through the project's website, and the data produced will be discoverable with metadata and identifiable and locatable by the Digital Object Identifiers (DOIs). The metadata will include: title, data types/formats and software, data collection method and dates, geographic coverage, language, data processing details, funding details, ethics clearance details, a project abstract, keywords, and licensing.
- Data needed for the collaborative tools are expected to include a unique ID and a timestamp allowing for proper indexing and handling when stored. No specific standards or metadata have been identified for the time being for the datasets.
- The data from the surveys, workshops/living labs, and the validation cycles will not be published as primary data (data that is collected directly from the data source) due to privacy and security concerns.

More detailed information on how to make data findable, accessible, interoperable and re-usable will result from discussions among partners and will be included in the 2nd version of the DMP which is expected by M30.

4.1 Making data Findable, including provisions for metadata

Most of the data produced within the project are discoverable with metadata and identifiable and locatable by means of a standard identification mechanism. To make their identification easier, partners will decide upon a common naming procedure, taking into account internal project conventions. The conventions that will be followed will be further discussed in the upcoming months: some partners already rely on their internal company system for data identification and their model could be adopted by the whole consortium as the common project methodology.

Data regarding companies will be available in RAMP, other data derived from these companies since sensitive will be securely stored and accessible in ED's ISO27001:2013-certified infrastructure (ED).

Dutch Pilot 3A&B we need some help here, as we have no knowledge about this terminology.

Data regarding residual leather will be available in KHOANI's webpage, other data derived from the user companies since sensitive will be securely stored and accessible in a cloud storage.

Concerning the SUPSI's team, the publication of reports and deliverables on open-access platforms will ensure the broad accessibility. Additionally, metadata will be used to describe data following the platform requirements and standards such as authorship, keywords, and DOIs if published. RAMP will also be used for increasing visibility and accessibility.

PILOT1 Plastic recycling: Data is available and classified in our systems according to different parameters (job order/machine, customer, timestamps, workers…), to be defined with the consortium how to access it. Shareable data will be published through different platforms like our webpage, Linkedin, and through partners.

CUT will also publish on open-access platforms, following the platform requirements. RAMP will also be used for increasing visibility and accessibility.

Regarding the CIRCULOOS platform, data are stored and manipulated in accordance with the NGSI standard. NGSI is a protocol developed by OMA to manage Context Information. It provides operations like managing the Context Information about Context Entities, for example the lifetime and quality of information and accessing (query, subscribe/notify) to the available Context Information about Context Entities. NGSI extends well-known Web standards, such as Representational state transfer (REST) and Linked Data7, to develop an ontology and interoperability framework for IoT. In particular, entities and relationships which constitute the system data (the so-called IoT "context" in NGSI parlance) become Web resources, each identified by a unique Uniform Resource Identifier (URI) and retrievable through an HTTP call by constructing a suitable Uniform Resource Locator (URL) from the resource's URI. Those Web resources are made available through a Web service, the FIWARE Context Broker, which provides a standard mechanism for clients to discover what resources are available in the IoT context that it manages.

Regarding pilots, it still needs to be defined but mainly information related to the source, the type of piece and the defect could be made available. It will be agreed with other partners in the next months

4.1.1 Metadata

Metadata provides additional information that helps data consumers better understand the meaning of data, its structure, and to clarify other issues, such as rights and license terms, the organization that

generated the data, data quality, data access methods and the update schedule of datasets. Generally speaking, metadata will have a twofold nature: descriptive, therefore giving information on the data discovery and identification (titles, author, keywords) and administrative outlining when and how data was created, file type and other technical information, and who can access it (dataset name, version, description, format, License, keywords).

CIRCULOOS platform adopts the NGSI data standards. NGSI makes provisions to model plain metadata as well as complex relationships between metadata items.

4.1.2 Naming conventions

To (i) enhance data searchability and discoverability, and (ii) provide clues to the content, status, and versioning of the files, each set of data produced (dataset, deliverables, etc…) will be named in a uniform way and will include a table with a version control.

The recommendations to name the documents of the project are as follows:

- Choose easily readable identifier names (short and meaningful);
- Do not use acronyms that are not widely accepted;
- Do not use abbreviations or contractions;
- Avoid language-specific or non-alphanumeric characters;
- Add a two-digit numeric suffix to identify new versions of one document.
- Dates should be included back to front and include the four-digit years: YYYYMMDD.

For deliverables: Project's name - Dx.y - [Name of the deliverable as described in the DoA] being x - work package assigned to the deliverable y - the number of deliverables within the work package i.e.: D.1.2 - Data management plan M6.

For datasets: Project's name - WP [Work Package number] P [Pilot number; pilot activity number] - T [Task number; description of the activity] e.g., WP1 Task 1.4 Data Management Plan.

Easy-to-use search keywords will be used in CIRCULOOS to optimise the reuse of data by interested stakeholders. The metadata standards employed by CIRCULOOS provide opportunities for tagging the data collected/generated and its content with keywords.

In general, the keywords will comprise terms related to the topics addressed, such as energy efficiency, energy renovations, smart contracting, innovative business models, fair energy transition, capacity building in the energy sector, green currency, smart renovations, energy efficiency policies, as well as keywords specific to the project, such as CIRCULOOS, Horizon Europe, etc.

The keywords will accurately reflect the content of the datasets and avoid words used only once or twice within them.

4.2 Making data openly Accessible

According to Art. 17 of the Grant Agreement, the beneficiaries must

- ensure open access to peer-reviewed scientific publications relating to their results. They must ensure that:

- ① at the latest at the time of publication, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications.
- ① immediate open access is provided to the deposited publication via the repository, under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a licence with equivalent rights; for monographs and other long-text formats, the licence may exclude commercial uses and derivative works (e.g. CC BY-NC, CC BY-ND) and
- ① information is given via the repository about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication.

- manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles and by taking all of the following actions

All publications will be made available through ArXiv (https://arxiv.org/) repository or OpenAIRE (https://www.openaire.eu/).

The dataset obtained under WP3, WP4, WP5 and WP6 might contain classified information related to the existing mechanisms/processes/infrastructure, hence these datasets could be classified.

4.3 Making data Interoperable

According to the consortium, most of the data produced in the project will be interoperable, and data exchange and re-use between researchers, institutions, organisations, countries, etc. will be favoured. However, partners still have to define which metadata vocabularies, standards and methodologies will be used to this end. Generally, interoperable data will use a formal, standard, accessible, shared, and broadly applicable language for knowledge representation, vocabularies that follow FAIR principles and standard data-storage methods. On the contrary, non-interoperable data will use vocabularies, standards and methodologies that will primarily answer to internal efficiency and safety requirements.

4.4 Increasing data Re-use

As already shown before some data generated within the project will be prepared to be reused by third parties including academia and industry stakeholders. In the event that some project data will be licensed to permit their re-use by the external stakeholders, Creative Commons or GNU licenses will be used. As previously mentioned, the data that will be made available for individuals that are external to the project will be included in the project deliverables marked as public. Such data will be made public, through the project's website and open access repositories for scientific publications such as Zenodo and other Institutional Repositories connected with OpenAIRE, as soon as they will be approved by the project officer on the Funding and Tenders Platform of the EC.

With regards to the access rights to third parties, according to Art. 9.8 of the Consortium Agreement, it shall, as far as Needed for the Exploitation of the Party's own Results, comprise the right to grant in the normal course of the relevant trade to end-user customers buying/using the product/services, a sublicense to the extent as necessary for the normal use of the relevant product or service to use the Object Code alone or as part of or in connection with or integrated into products and services of the Party having the Access Rights

5 Allocation of resource

The estimation of the costs to be incurred when making the data FAIR according to the principles of the EC and the proper allocation of the resources to do that are two topics addressed by the DMP. Moreover, the plan identifies the responsibilities for data management in project describing the costs and potential value of long-term preservation. Even if this is an early stage of the project and not all the points could find an exhaustive description, this chapter of the DMP will address the costs for making data FAIR in CIRCULOOS, how these costs will be covered, who will be the responsible for the data management and who will decide and how what data will be kept and for how long.

5.1 Estimation of cost

The costs related to open access to research data are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions). Costs are eligible for reimbursement during the duration of the project under the conditions defined in the Horizon Europe Grant Agreement.

5.2 Responsibilities for data management

The partner who will be responsible for data management is ED as detailed in D1.1 Project Management Handbook. ED will be in charge to periodically update the DMP with further details and information provided by all project partners that should provide their contributions.

6 Data security

The data collected/generated throughout the whole period of the CIRCULOOS project will be held in data repositories in the servers of organisations. The servers will be kept in locked rooms with authorized access and will adhere to appropriate security standards as well as to state-of-the art security mechanisms. The data repositories will be accessed by authorized personnel both at physical and network levels. The databases will have a dedicated identification and authentication mechanism, name and password, for each authorized user and will adhere to different access privileges for the authorized users of the organisation in any case.

Access to the data repository will be achieved through the appropriate protocols. The transfer of the data (e.g. files, deliverables, raw data etc.) will be achieved using encryption methods (e.g. file transfers, encrypted disks, encryption keys, a physically protected and secure PC responsible for this process) when necessary.

As for the "physical" data storage, the documents will be kept in an office's secure environment, in computers with authentication and authorization mechanisms. If there are documents with restricted access, they will remain in a locked cabinet at the organisation's premises.

Backups of the databases will be stored encrypted and on the organisations' premises. Data backups of devices will happen every week and be stored in devices that follow the same security standards and procedures as the main server.

7 Legal framework and guidelines

Applicable legislation on the protection of personal data:

- Regulation (EU) 2016/6792 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons about the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).
- Regulation (EU) 2018/17253 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons about the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/20014 and Decision No 1247/2002/EC5.
- National data protection laws.

Applicable legislation on the protection of EU classified information:

- Commission Decision (EU, Euratom) 2015/4446 of 13 March 2015 on the security rules for protecting EU classified information.
- Directive (EU) 2016/11487 of the European Parliament and of the Council of 6 July 2016 concerning measures for high common level of security of network and information systems across the Union.

Applicable legislation on the exchange and use of data within the EEA.

- Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act)

Applicable legislation related to the use of AI

- As part of its digital strategy, the EU wants to regulate artificial intelligence (AI) to ensure better conditions for the development and use of this innovative technology. The AI Act, soon to be the first comprehensive AI law globally, regulates the development and use of AI in the EU to balance innovation with safety and ethical considerations. While not yet formally law, it's expected to be adopted by the EU in 2024.[12]

7.1 Personal Data Management

7.1.1 Important GDPR provisions

According to Article 4 par.1 GDPR, 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

According to Recital 26 GDPR, the principles of data protection should not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. Since GDPR does not apply to anonymous information, it is very important to distinguish between anonymised and pseudonymised data, as for the latter GDPR remains applicable.

12 EU AI Act: first regulation on artificial intelligence | Topics | European Parliament (europa.eu)

According to Article 5 par.1-2 GDPR, the principles relating to the processing of personal data are:

- lawfulness, fairness and transparency;
- purpose limitation;
- data minimisation;
- accuracy;
- storage limitation;
- integrity and confidentiality;
- accountability.

According to Article 6 par.1 GDPR, processing shall be lawful only if and to the extent that at least one of the following applies: (a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes; (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; (c) processing is necessary for compliance with a legal obligation to which the controller is subject; (d) processing is necessary in order to protect the vital interests of the data subject or of another natural person; (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child. Point (f) of the first subparagraph shall not apply to processing carried out by public authorities in the performance of their tasks.

Article 13 GDPR stipulates the information that must be provided by the controller to the data subjects when the personal data is collected by them.

Article 14 GDPR stipulates the information that must be provided by the controller to the data subjects when the personal data is not obtained by them. Paragraph 4 (b) provides for that paragraphs 1 to 3 shall not apply when the provision of such information proves impossible or would involve a disproportionate effort, in particular for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, subject to the conditions and safeguards referred to in Article 89 par.1 or in so far as the obligation referred to in paragraph 1 of this Article is likely to render impossible or seriously impair the achievement of the objectives of that processing. In such cases the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available.

Article 25 GDPR stipulates the “Data protection by design and by default”, including procedures for pseudonymisation and data minimisation.

Article 30 GDPR stipulates a “Record of processing activities”, according to which an accurate description of the protection activities shall be kept by the data controller and the data processor and, upon request, made available to supervisory authorities.

Article 32 GDPR stipulates the “Security of processing”. This article aims to ensure that the data are kept and processed in a secure manner in order to avoid their unlawful or accidental destruction, loss, alteration, unauthorised disclosure or access. Possible measures to enforce data security include pseudonymisation and encryption as well as regular testing of the implemented technical and organisational measures.

According to Article 35 par.1 GDPR, where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks.

According to Article 89 GDPR, processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with GDPR, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner.

8 Intellectual Property Rights (IPR) Management

All partners in the Consortium have agreed on explicit rules that need to be concerned as regards to the Intellectual Property (IP) ownership. In that way, access rights have been given to any Background and Foreground for the execution and protection of intellectual Property Rights (IPR) and confidential information before the project starts. All details have been addressed within the Consortium Agreement between all project partners.

8.1 Definitions

Intellectual Property Rights or IPR(s) means patents, patent applications and other statutory rights in inventions; copyrights (including without limitation copyrights in Software); registered design rights, applications for registered design rights, unregistered design rights and other statutory rights in designs; and other similar or equivalent forms of statutory protection, wherever in the world arising or available, but excluding rights in Classified Information and/or trade secrets.

Background means any data, know-how or information — whatever its form or nature (tangible or intangible), including any rights such as intellectual property rights — that:

- is held by the beneficiaries before they acceded to the Consortium Agreement, and
- is needed to implement the action or exploit the results.

Foreground or Result(s) means any tangible or intangible output of the Project, such as data, knowledge and information whatever their form or nature, whether or not they can be protected, which are generated in the Project as well as any rights attached to them, including intellectual property rights.

8.1.1 Applicable legislation

- European Patent Convention, 16th edition9 (published June 2016) which contains the Convention on the Grant of European Patents (EPC) as in force since 13 December 2007, the EPC Implementing Regulations as in force since 1 May 2016 but also including an amendment that entered into force on 1 November 2016, the rules of procedure of the EPO boards of appeal and Enlarged Board of Appeal, included for the first time in an edition of the EPC, the protocols forming integral parts of the EPC (Protocol on the Interpretation of Article 69 EPC, Protocol on Centralisation, Protocol on Recognition, Protocol on Privileges and Immunities, Protocol on the Staff Complement), an extract from the EPC Revision Act of 29 November 2000, the Administrative Council's decision of 28 June 2001 on the transitional provisions under Article 7 of the Revision Act, and the Rules relating to Fees.
- Regulation (EU) 2017/100110 of the European Parliament and of the Council of 14 June 2017 on the European Union trademark.
- Council Regulation (EC) No 6/200211 of 12 December 2001 on Community designs.
- Directive (EU) 2019/79012 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Directive on Copyright in the Digital Single Market).
- Directive 2001/29/EC13 of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society.

- Directive 96/9/EC14 of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.
- Directive 2009/24/EC15 of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs.
- Directive 2004/48/EC16 of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights.
- National laws on patent, design, trademark and copyright protection.

8.2 IPR Management in the CIRCULOOS Project

The CIRCULOOS Consortium Agreement expressly stipulates the rules related to the management of IP rights and distinguishes, on the one hand, the IP rights that are held by the partners prior to their accession in the Consortium Agreement and are needed for the Project (Background) and, on the other hand, the IP rights that are held by the partners during the lifetime of the Project (Results).

Section 8 Results and Section 9 Access Rights of the CIRCULOOS Consortium Agreement include all relevant clauses that have been agreed between the partners and refer to the IPR management. In Attachment 1 of the CIRCULOOS Consortium Agreement, the Parties have identified and agreed on the Background for the Project and have also, where relevant, informed each other that access to specific Background is subject to legal restrictions or limits. Therefore, we aim to work methodically to classify CIRCULOOS IPRs and define:

- the treatment of existing IPRs (background),
- the management of joint ownership. Partners will keep record of their contributions which are protected by IP Law and potentially Trade Secrets. This will permit the Consortium to discern the share of each owner in relation to the results of a joint effort. CIRCULOOS Partners aim to reach a point where exploitation of results will become possible,
- the protection and management of the results of CIRCULOOS (foreground),
- the exploitation and dissemination of the results of CIRCULOOS (foreground),
- the protection of know how created during CIRCULOOS.

Furthermore, section 4 and section 5 of the CIRCULOOS Consortium Agreement provide for the responsibilities of the partners and their liability towards each other (including for the management of IP rights).

All partners may settle any disputes in accordance with the clause 11.8 of the CIRCULOOS Consortium Agreement.

To effectively achieve the objectives regarding the overall management of IPRs, a cumulative IPR Control form will be circulated by the coordinator.

9 Ethical Aspects

Work package 7 (Deliverables 7.1) of the CIRCULOOS project deals with the ethics requirements with which CIRCULOOS's objectives, methods, processes, tasks and results must comply. These ethics requirements mainly relate to the processing of personal data and, with that, with data protection. Therefore, in this deliverable, requirements that deal specifically with the procedures with respect to data protection will not be covered.

In order to ensure that all ethical aspects are considered and that the CIRCULOOS project is compliant with all legal requirements and ethical issues, a general strategy has been designed based on the Ethics Requirements that were defined by deliverables from WP7. This strategy involves an ad hoc monitoring process of the project development by applying the privacy-by-design approach through a methodological design based on a "Socio-legal Approach." This is a risk-based approach to privacy and data protection issues in line with the new General Regulation for Data Protection (GDPR).

Ethics Requirements that were drawn by the EC, during the ethics check process before the signature of the GA, have been considered by the consortium; to address all these requirements the CIRCULOOS beneficiary Trilateral Research Limited, with its ethical and legal expertise will provide general guidelines to the project and the consortium on all aspects, covering data protection, privacy issues, research participants' safety, etc.

10 Conclusion

The initial DMP for CIRCULOOS is presented and delivered in this document describing how acquired data and knowledge will be shared and/or made open as well as how data will be maintained and preserved during and after the timeline of the CIRCULOOS project.

This deliverable defines any kind of information including scientific publications, white papers, Open-Source code, open datasets, anonymous interview results, or mock-up datasets used for gathering customer feedback that may be used or generated from the project. The collected datasets in the current version of the report are research data, related to the project's work packages and are managed according to their level of availability (public or Consortium). The CIRCULOOS Data Management also follows the Guidelines on FAIR Data Management in Horizon 2020, i.e., data must be findable, accessible, interoperable, and reusable.

All partners will be responsible to periodically update information on their research and subject data.

The current report will be a living document throughout the project. The DMP will be updated whenever significant changes arise, such as (but not limited to) new data, new innovations, patent filings, changes in the consortium members and others. Formal updates to the DMP will be submitted at specific project milestones, currently scheduled for M30 and M42.

CIRCULOOS

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101092295. The herewith information reflects only the author's view. The European Commission is not responsible for any use that may be made of the information herewith included.