This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776503
- CIRC4Life-H2020-IND-CE-2016-2017/CIRC-2017/TwoStage

A circular economy approach for lifecycles of products and services

Data Management Plan - #1

Deliverable 10.2.

PROJECT INFORMATION	
Type of Project	European Commission Horizon 2020
Topic	CIRC-01-2016-2017 Systemic, eco-innovative approaches for the circular economy: large-scale demonstration projects
Grant Agreement No.	776503
Project Duration	01/05/2018 – 30/04/2021 (36 months)
Project Coordinator	Nottingham Trent University (NTU)
Project Partners	Enviro Data (ENV), Jonathan Michael Smith (JS), Kosnic Lighting Limited (KOS), Centre of Research for Energy Resources and Consumption (CIR), European EPC Competence Center GmbH (EECC), The Institute for Ecology of Industrial Areas (IETU), SWEREA IVF AB (SWE), Make Mothers Matter (MMM), ONA PRODUCT (ONA), INDUMETAL Recycling (IND), GS1 Germany GMBH (GS1G), Laurea University of Applied Science (LAU), Center for European Policy Studies (CEPS), Institute of Communication and Computer Systems (ICCS), Recyclia (REC), S.A.T. Alia (ALIA)

DOCUMENT INFORMATION	
Title	Data Management Plan - #1
Version	V1.0
Release Date (dd.mm.yy)	<25.10.18>
Work Package	WP10
Dissemination Level	PU

DOCUMENT AUTHORS AND AUTHORISATION	
Document Responsible	Dr Wu, You (NTU)
Contributors	Wu, You (NTU), David White (NTU)
Reviewed by	Karin Wilson (SWE), Juan Costa (ONA)
Approved by	Professor Daizhong Su (NTU)

DOCUMENT HISTORY			
Version	Date (dd.mm.yy)	Description	Implemented by
V0.1	11.10.18	First draft	Wu, You
V0.2	12.10.18	Second draft, sentences are corrected, references are updated.	Wu, You
V0.3	16.10.18	Responded the reviewer (SWE) comments	Wu, You
V1.0	25.10.18	Final version	Wu, You

Summary

The Horizon 2020 FAIR Data Management Plan (DMP) template is designed to be applicable to any Horizon 2020 project that produces, collects or processes research data. CIRC4Life consortium has developed a single DMP for the project to cover its overall approach.

Deliverable 10.2 aims to develop the CIRC4Life DMP, in which the open data policy will be specified: detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved.

Table of Contents

Acronyms and abbreviations

Abbreviation	Description
EC	European Commission
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable and Re-usable
ICT	Information and Communications Technology
CEBM	Circular Economy Business Model
EPCIS	Electronic Product Code Information Services
RFID	Radio-Frequency Identification
EPC	(Electronic Product Code)
XML	Extensible Markup Language
CSV	Comma-separated values
FNC	File Naming Convention
DOI	Digital Object Identifier
IR	Institutional Repository
IP	Intellectual Property
PC	Project Coordinator
GDPR	General Data Protection Regulation

1 Introduction

CIRC4Life project comply the FAIR data management concept to develop this DMP. FAIR data management requires the project data should be 'FAIR', that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution.

Deliverable 10.2 is not intended as a strict technical implementation of the FAIR principles, it is inspired by FAIR as a general concept. The following documents have been referred in order to develop the CIRC4Life DMP.

- Guidelines on FAIR Data Management in Horizon 2020 (EC, no date)
- FAIR data principles (FORCE11, no date)
- FAIR principles (Wilkinson *et al.*, 2016)

The FAIR DMP template (EC, no date) is a set of questions that should be answered with a level of detail appropriate to the project. It is not required to provide detailed answers to all the questions in the first version of the DMP that needs to be submitted by month 6 of the project.

Rather, the DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Therefore, DMPs should have a clear version number and include a timetable for updates. As a minimum, the DMP should be updated in the context of the periodic evaluation/assessment of the project. If there are no other periodic reviews envisaged within the grant agreement, an update needs to be made in time for the final review at the latest.

2 Data Summary

The CIRC4Life project involves an ICT (Information and Communications Technology) solution using a variety of technologies (e.g. traceability technology), which are important to provide the means to implement and assess the effectiveness Circular Economy Business Model (CEBM). These technologies will drive the data generation and collection processes and help inform the testing against the proposed methodologies of using an eco-point approach, sustainable product design and production, real-time tracking and monitoring technology based on EPCIS (Electronic Product Code Information Services), information logistics sharing infrastructure across the supply chain and associated data security and privacy.

Given that the ICT solution will use a number of tracking technologies, potentially including barcode, RFID (Radio-Frequency Identification) and EPCIS. The project process tasks will include full documentation on barcode code types, RFID data models and EPCIS; and with regards the latter the EPC (Electronic Product Code) Core Business Vocabulary identifiers used. The documented descriptions will also contain an explanation on why particular identifiers or models were used for all the tracking technologies used, and how they related to the intended outcomes.

Example datasets with these descriptions will be supplied in an XML as well as Word DOCX format for the descriptions, and example tracking data datasets are in CSV and XLSX format that will be compatible with as many software applications as possible.

No comparable datasets have been found to exist, and so the importance of making these datasets FAIR (Findable, Accessible, Inoperable and Reusable) is recognised and planned for, but taking into account the demands of the 'living' nature of a data management plan as the project progresses, the final form of the datasets cannot yet be declared in detail.

Subsequent research will benefit from analysing the chosen identifiers used in attempting to improve the improved sustainability performance model. A clear description of the identifiers used would be critical to this type of analysis. Similarly to, the accurate documentation and example datasets of tracking technologies used.

With regards anticipating the expected size of data to be produced, the nature of the CIRC4Life project involving a number of collaborators makes an accurate guess difficult at this early stage, but it is proposed that the data generated will be in the order of Gigabytes (approx. 100), rather than Terabytes.

3 FAIR Data

3.1 Making Data Findable, Including Provisions for Metadata

Through the active phase of the project, data management will include simple organisational measures such as following a file naming convention (FNC) and will take the form of CIRC4Life_ DocumentName_ ResponsiblePartner _ YYYYMMDD_ Version.docx. Document filenames will be kept short to avoid unnecessarily long paths, always include the last person to edit the document, and a version indicator.

For example, CIRC4Life_ProjectManagementPlan_NTU_20181009_V0.1.docx

All project documentation will be stored in a Microsoft project SharePoint site (as part of an Office 365 for Business cloud service instance) that will enable full control over editing permissions of project participators. SharePoint as a collaboration and document management platform also offers good functionality for platform-wide metadata control of content.

For additional knowledge management, an excel document will be included and promoted that contains a list of keywords, and the accepted definitions to be used by all the participating project partners. It will also act as a controlled vocabulary to ensure consistent knowledge organisation that will aid subsequent retrieval and re-use. This document will be easily accessible for project collaborators using the project SharePoint site and included on the project website.

3.2 Making Data Openly Accessible

The final datasets to support published outcomes will be deposited in Arkivum, the Nottingham Trent University (NTU) instance of the Archive-as-a-Service data safeguarding and long-term lifecycle management solution. A Digital Object Identifier (DOI) will be minted using DataCite service, a global provider of DOIs for research data, and included in the description (with additional metadata) in the NTU Institutional Repository (IR) to facilitate persistent identification and discovery. The metadata included in the IR record will identify the file formats the datasets are available in, which will indicate to potential users the software application required to effectively use the datasets. Any software code will also be described in the IR and archived, but intellectual property (IP) implications may restrict access. The development of the ICT solution will be accompanied by software documentation that will provide instructions for reuse, and these will be archived and made available with the proviso again of potential IP implications.

NTU uses DataCite Metadata Schema to describe the data and is actively updated and promoted by interactively coordinating with community standards, like ORCID (https://schema.datacite.org/).

Access to datasets will be open and currently available as a request/mediated service via the Library Research Team at NTU, which will necessarily identify the individual requiring access to the datasets and provide statistical information on dataset usage. However, availability will always be determined by licence conditions, and that will allow for some granularity of access (including restrictions) to be potentially specified by different project collaborators. DataCite also has a reciprocal dataset registry service to aid discoverability and potential reuse of datasets that have been determined to be openly available.

The data will be preserved for a minimum of ten years in the Arkivum service, or in any other subsequent solution used. Ten years is specified to cover all institutional/publisher and funder policy requirements.

3.3 Making Data Interoperable

NTU uses DataCite metadata schema as DataCite looks to community practices that provide data citation guidance. The Project Coordinator (PC) will foresee that SharePoint project site will include appropriate mandatory DataCite metadata elements for project files. Such as (Creator, Title, Description, Access to the dataset, Data Collection Method, Data Processing/Preparation Activities), and are included as mandatory metadata classification fields for file inclusion on the SharePoint site.

Using standard vocabularies for all data types is not initially anticipated as the data definition is integral to the tracking technologies of barcodes, RFID and the EPCIS Standard already. It will be clearly stated what barcodes code type used, the data model of RFID and EPCIS Standard used. However, the PC will ensure that any datasets that are included as CSV will have a description definition for all the data elements included, and these are mapped to the industry standard description of the identifiers used for the different tracking technologies.

3.4 Increase Data Re-use (Through Clarifying Licences)

All openly available data will be offered on a CC-BY SA share-alike basis, so attribution will be required and that any repurpose, or re-use will be shared on the same basis. IP requirements may determine that some datasets are not open and will remain closed but archived in Arkivum. The description of the individual datasets in the NTU Institutional Repository will make the conditions of the licence clear.

Archival in Arkivum is for ten years, and will always be available, allowing for licence conditions, for external viewing during that period.

During the active phase of the project, data quality assurance process will be organised by the PC (responsible individual for data management) to discover inconsistencies and other anomalies in the data, as well as performing data cleansing activities (e.g. removing outliers, missing data interpolation) to improve the data quality. It is foreseen that this will involve sampling datasets initially, with a thorough assessment of particular grouped datasets if significant inconsistencies are discovered.

4 Allocation of Resources

The costs of using Arkivum Archiving-as-a-Service solution available at NTU at the end of the project is £0.50 per GB of archived data. As the projected final amount of data is expected to be in the order of 100 (estimation) Gigabytes, this will cost approximately £50 per year, or £500 for ten years.

NTU will support the archiving of funded research datasets to promote the open access and data agenda.

The responsibility of data management on the CIRC4Life project is the PC of the leading collaborating institution (NTU). Everyday workflow tasks will be delegated, but the PC will ensure that consistent data management is performed for the duration of the project and will conduct six monthly reviews on the use of controlled vocabulary, file naming and versioning conventions and that the organisational logic of the SharePoint site is adequate.

The PC will be responsible for selecting the datasets to be archived, for a period not less than ten years. Preservation will support publication outcomes, research deemed of long-term value, as well as project communication channels dissemination output and literature. Should personal details be included in the data preserved, then anonymity need to be maintained, but traceable should there be a need for source data verification.

Any data deemed to be not worth saving in the active and archival stage needs to be destroyed in accordance with NTU Information Systems data destruction policy.

5 Data Security

Access to project documentation and data is only available to those who have access to the SharePoint site, which is determined by the PC and implemented by NTU research staff participating in the project. Additional backup and data restore procedures will be agreed using this solution full functionality. A further backup will be performed daily to an on-premise storage solution hosted by NTU.

As stated NTU has a long-term storage solution (i.e. Arkivum), available to enable preservation and curation. It is an Archiving-as-a-Service solution that combines a hybrid local and cloud storage synchronised combination that guarantees data integrity and redundancy with full security. This ensures that funder, institutional or publisher retention compliance is satisfied, as is the authenticity of the original data for open data requirements or post research review if necessary.

6 Ethical Aspects

Any personal data gathering within the CIRC4Life project will conform to informed consent expectations that are expected with regards current Data Protection legislation, and the EU General Data Protection Regulation (GDPR) that started to implement on 25th May 2018.

It may be the case, that some Intellectual Property implications may influence communication dissemination such as through the project website, on what project deliverables could be communicated publicly.

7 Outlook and Conclusions

With regards, conducting research project NTU expects staff and students involved in research to adhere to the policies pertaining to 'Code of Practice for Research', 'Research Ethics Policy' and the 'Research Data Management Policy'. NTU has also sought to follow the requirements and recommendation of Horizon 2020 EU funding as described in this document. Lastly, as a UK HE organisation, NTU has provided support for research active members to be aware of national guidelines on open data, and to follow the principles wherever possible.

8 References

EC (no date) *Data management - H2020 Online Manual*. Available at: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm (Accessed: 10 October 2018).

FORCE11 (no date) *The FAIR Data Principles*. Available at: https://www.force11.org/group/fairgroup/fairprinciples (Accessed: 10 October 2018).

Wilkinson, M. D. *et al.* (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data*, 3, p. 160018. doi: 10.1038/sdata.2016.18.