Efficient Carbon, Nitrogen and Phosphorus cycling in the European Agri-food System and related up- and down-stream processes to mitigate emissions

Start date of project: 2018-09-01

Duration: 48 months

D6.4. DATA MANAGEMENT PLAN

Deliverable details	
Deliverable number	6.4.
Revision number	E-0219
Author(s)	IRTA
Due date	0219
Delivered date	
Reviewed by	Victor Riau
Dissemination level	PU
Contact person EC	Giuseppe La Ciura

Contributing partners	
1.	AREC
2.	KWB

Project Number: *773649*
Project Acronym: *CIRCULAR AGRONOMICS*
Deliverable name   *D6.4. Data Management Plan*

TABLE OF CONTENTS

Project Number: *773649*
Project Acronym: *CIRCULAR AGRONOMICS*
Deliverable name   *D6.4. Data Management Plan*

1. INTRODUCTION

This Data Management Plan (DMP) will guide the CIRCULAR AGRONOMICS' Consortium in managing the data quality and protection issues that will arise along the project life. The DMP focuses on encouraging good data management as an essential element of best research practices. This deliverable describes the data management life cycle for all datasets to be collected, processed and/or generated by the research project, always in accordance to the "Guidelines on Data Management in Horizon 2020". The H2020 DMP describes, among others:

- the handling of research data during and after the project
- the type of data that will be collected, processed, or gathered
- what methodology and standards will be applied
- whether and how the data will be made (openly) accessible
- how the data is stored

The first version of the data management plan (DMP) of CIRCULAR AGRONOMICS (**CA**) project is presented as **Deliverable 6.4**, and will be updated as work progresses. An updated version will be presented in **M18**.

2. DATA SUMMARY

<u>Points to be addressed</u>: Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

The CA data summary has been conceived and designed for three main work package groups:
- **GROUP 1. Experimental and Technical (WP1, 2 and 3)**,
- **GROUP 2. Socio-economic and Environmental Assessment (WP4 and 5)**
- **GROUP 3. Exploitation, Communication, Dissemination and Coordination (WP6 and 7**)

The purpose is to collect and harmonize European agro-industrial data referring to agricultural practices, soil characteristics and livestock farming at regional level, food industry, waste management, technical, socio-economic, environmental and other external information data related to the performance of the sector, as well as data generated during the project. The data collection will comply with all national and EU ethics and legal requirements. Access to use these data is needed to address the different CA objectives as specified in the grant agreement and summarized below.

- **Objective 1.** Increase the understanding of C, N, P flows and the related potential to reduce environmental impacts at farm and regional level under different bio-geographical conditions.
- **Objective 2.** Closing loops within cropland farming, from livestock to cropland farming and to increase the reuse of waste/wastewater from food-industry to improve soil fertility and to increase nutrient use efficiency.
- **Objective 3.** Highlight the performance of different prototypes of agro-ecological systems and increase sustainability of food production in the EU.
- **Objective 4.** To contribute to the improvement of the European Agricultural Policies by providing evidence based, farmer led and consumer relevant recommendations for the agri-food chain.

Project Number: *773649*
Project Acronym: *CIRCULAR AGRONOMICS*
Deliverable name   *D6.4. Data Management Plan*

Table 1. Data Summary

	GROUP 1 (WP1, WP2 and WP3)	GROUP 2 (WP4 and WP5)	GROUP 3 (WP6 and WP7)
Collected/generated data	Generated and collected data	Generated and Collected data	Generated and collected data
Origin(s)	Most of them based on field and experimental studies which will be performed during the project: ▪ Data from soils characterization ▪ Data from laboratory experiments ▪ Data on C, N and P stocks, flows and emissions in crop farming and livestock farming. Data from respiration chambers and precision feeding systems: gas flows and measurements of gaseous emissions from dairy cows ▪ Data of fertilizer (organic and inorganic) dosages ▪ Data of above- and belowground biodiversity ▪ Data from manure and residues characterization and emissions ▪ Data from energy and mass balances as well as efficiencies	Several, mainly provided by internal surveys and external and official databases as well as data collected from the project's case studies: ▪ Data from energy and mass balances/efficiencies ▪ Data from surveys ▪ Data from cost assessment /economic data ▪ Data from environmental and economic impact and life cycle assessment ▪ Data of exports/imports EU/Africa	Several, mainly contacts of partners, stakeholders, participants in events and training activities and those generated in dissemination and communication activities (workshops, conferences, reports): ▪ Data from dissemination and communication events ▪ Partners contact information ▪ Data from internal and external reports
Nature	Experimental data, laboratory data. Scientific reports	Surveys. Structured data (processed datasets) related to consumers and farmers acceptance/behavior, cost and life cycle assessment and scientific and technological structured data	Minutes from meetings, workshops and conference. Specific dissemination data. Reports
Purpose of the data	Objectives 1, 2 and 3	Objective 3 and 4	Objective 4
Size	To be evaluated during the course of the project. The expected size depends on the extent and the nature of the data that are made available.	To be evaluated during the course of the project. The expected size depends on the extent and the nature of the data that are made available (data from surveys: at least 300 farmers and 5000 consumers)	To be evaluated during the course of the project. The expected size depends on the extent and the nature of the data that are made available.
Data Utility	▪ CA consortium; ▪ European Commission services and European Agencies; ▪ EU National Bodies; ▪ The general public including the broader scientific community; ▪ Internal/external reports, deliverables, scientific report and publications, etc.	▪ CA consortium; ▪ European Commission services and European Agencies; ▪ EU National Bodies; ▪ The general public including the broader scientific community; ▪ Internal/external reports, deliverables, scientific report and publications, etc.	▪ CA consortium; ▪ European Commission services and European Agencies; ▪ EU National Bodies; ▪ The general public including the broader scientific community; ▪ Internal/external reports, deliverables, scientific report and publications, etc.

3. FAIR DATA

Points to be addressed:

In general terms, your research 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.

3.1. Making data FINDABLE, including provisions for metadata:

CA participates in the Open Research Data Pilot and aims to make data open whenever possible, but as closed as necessary when taking into consideration personal data and privacy. If it affects personal confidentiality and privacy, data will not be shared publicly. Moreover, for some specific research data e.g. resulting from the investigations accomplished in the pilot plants, the participation in the Open Research Data Pilot of the EC is considered incompatible with the Horizon 2020 obligation to protect results that can be reasonably expected to be commercially or industrially exploited. However, each Circular Agronomics beneficiary must ensure open access to all peer-reviewed scientific publications relating to its results. In addition to the peer-reviewed scientific publications, the Circular Agronomics consortium will open the access to the project data as much as compatible with the exploitation plan

Points to be addressed**:**
- ▪ Outline the discoverability of data (metadata provision)
- ▪ Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- ▪ Outline naming conventions used
- ▪ Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

▪ **Metadata provision and standards:**

The metadata are a series of structured information, common to all the single measurement data in one data collection, which facilitate understanding, tracing and working with the data. In particular, metadata means a set of descriptive elements providing information on the "container of the data" (i.e. the title of a data collection, the abstract describing this data collection, etc.). For aggregated data, the metadata includes also information about the statistical method used for the aggregation. Metadata will be use to describe, discover and trace existing data collected by the CA project and the data that will be generated by it over the next years.

Once the data are generated or collected they will be made publicly available through uploading them in a data repository and made identifiable trough a PID (Persistent Identifier for Data, e.g. DOI) for each dataset and the associated metadata. The data collected will include the PID, origin of the data, description of the data, purpose of the experiment, duration and location, and the size and scale (number of raw and processed/calculated parameters and number of values in each parameter including missing values).

All metadata will be stored internally in JSON-format in line with a defined JSON schema. According to the OpenAIRE Guidelines (see section 4.2. regarding ZENODO generic repository), metadata is exported in several standard formats such as MARCXML, Dublin Core, and DataCite Metadata Schema

▪ **Naming convention:**

For metadata, dataset and template names we will define naming convention consisting in the following parts:
- • A root composed by:
  - - the short and meaningful name of the dataset/template
  - - the acronym/short name of the data provider organisation(s)
- • A suffix indicating the date of the last upload into the Repository in YYYYMMDD format.
- • Each of these elements are separated by an underscore: _

Project Number:
Project Acronym:
Deliverable name

773649
CIRCULAR AGRONOMICS
D6.4. Data Management Plan

3.2. Making data opening ACCESSIBLE

Points to be addressed**:**

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available
- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

Data, apart from that which includes sensitive data and general personal data affected by GDPR, or raising any ethical concerns (unless consented), will be shared. After publication, project's data will be deposited in a specific community created for Circular Agronomics in the ZENODO repository (https://zenodo.org/communities/circagro/) (Figure 1) (to be updated if needed with the revision of the DMP on M18). Partners may deposit restricted files with the ability to share access with others if certain requirements are met. These files will not be made publicly available and sharing will be made possible only by the approval of depositor of the original file. The ZENODO's repository system does not require any specific software and provides open access according to national and EU regulation. Only CA's partners will have permissions to upload datasets to the community. Go to https://zenodo.org/deposit/new?c=circagro, to upload information.
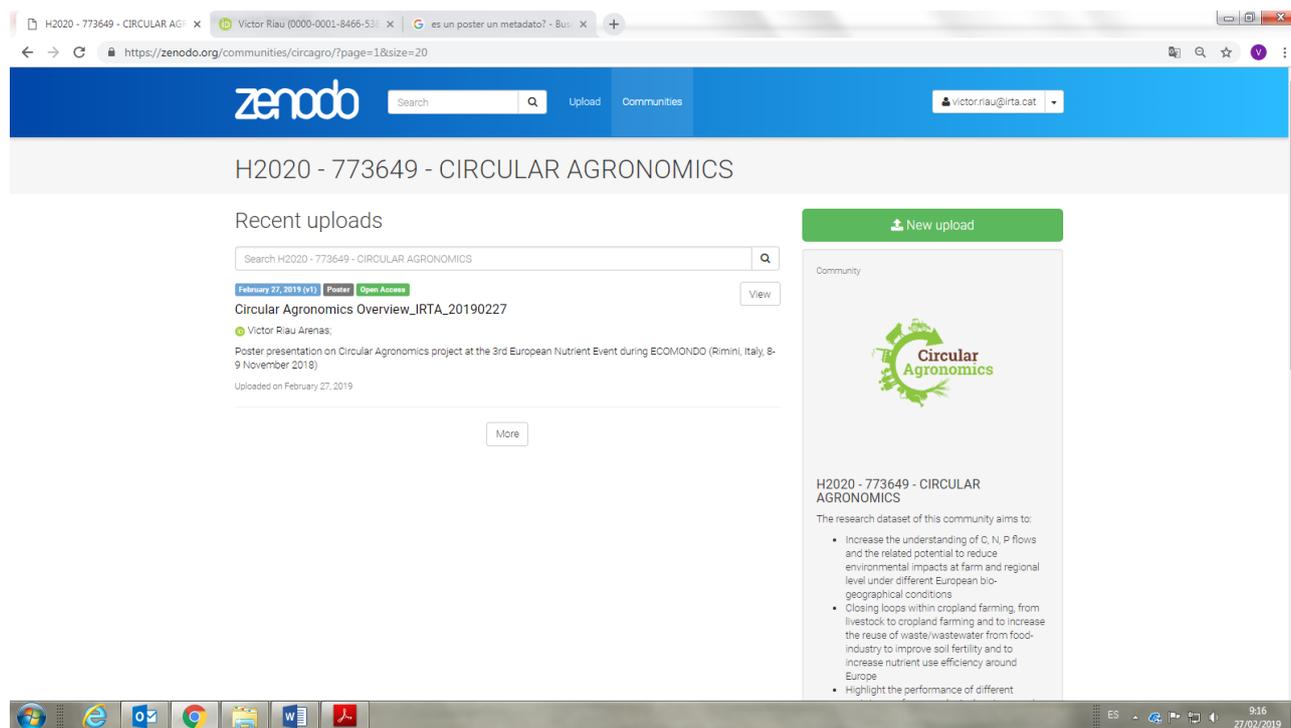

Figure 1. Circular Agronomics Community in ZENODO. Example of the deposition of the poster presented at ENE3 (Rimini, Italy November 2018) in the repository

For ease of access for partners, a direct link will be provided with metadata and direct link to all those datasets on One Drive. Metadata will be located on a dedicated page of the 773649_CIRCULAR AGRONOMICS_Data folder which will remain active for a minimum of 5 years after the completion of the project.

In accordance with the Grant Agreement all research related data (excluding personal and sensitive data) will be stored for at least five years after the end of the research project (in case there is a high interest in the datasets or due to different national legislation, data may be stored for a longer period, which will be transparently discussed and approved within the consortium and relevant parties). Data that are used for publication will be stored at least five years after publication.

3.3. Making data INTEROPERABLE

<u>Points to be addressed</u>**:**

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

CA will put efforts in producing interoperable data, allowing data exchange and re-use between researchers, institutions, organisations, countries. It means CA is adhering to general standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins. The data and metadata, standards and methodologies the CA project follows are with the aim to make the project data interoperable and standard vocabularies for all data types present in the project will allow inter-disciplinary and trans-disciplinary interoperability. AGROVOC or GEMET common vocabularies are suggested since they compile most of the vocabulary regarding the agri-food sector and the related environmental aspects. In the unlikely case that uncommon or project-unique specific ontologies or vocabularies have to be used, CA will provide mappings to more commonly used standard vocabularies for all uncommon data types to allow maximum interoperability. Any assumptions made, and the mapping steps to more commonly used vocabularies will be documented in a readme file and stored on One Drive.

Data produced will use standard formats depending of the type of data (e.g. txt, .doc, .xlsx, .xlsm, .xml, .csv, .pdf, .jpg, .gif, .docx, .png, .svg, .json) and made compliant with available (open) software applications, facilitating the recombination with various datasets from different origins. It is important to note that, when needed, data should be encoded in UTF8.

The following formats are suggested:
- .txt/.csv/.xml/.json (for data)
- .odf/.odc/.pdf/.ppt (for reports/presentations)
- .jpg/.png/.gif/.svg/ (for graphics)

3.4. Increase Data RE-USE (through clarifying licences)

<u>Points to be addressed</u>**:**

- Specify how the data will be licenced to permit the widest reuse possible
- Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
- Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
- Describe data quality assurance processes
- Specify the length of time for which the data will remain re-usable

This section will be compiled throughout the course of the project, when we get more information on the datasets that are made available for CIRCULAR AGRONOMICS. However, whenever suitable, data will be Open Access licensed data (after considering of personal data, intellectual property rights and any additional legal and ethical requirements). Based on Open Access regulation, CA aims to allow data to be re-used by third parties, but with restrictions if IPR or other rights demanding such restrictions. Copyright of the data are based on EU H2020 guidelines and Digital Curation Centre (DCC), an internationally-recognised centre of expertise in digital curation with a focus on building capability and skills for research data management. CA data licensing[1] is still under revision and will be updated throughout the course of the project

Access to the research data will be dependent on any agreed 'embargo period' based on national and EU regulations. The 'embargo period' is applied to give time to publish the work or seek patents, were applicable and this will be as short as possible until the work is accepted for publication or patent, bearing in mind that research data should be made available

---

[1] Types of Licences applied: the type of permit assigned from the dataset owner to exploit it, e.g. Creative Commons licences: https://creativecommons.org/share-your-work/licensing-types-examples/

Project Number: *773649*
Project Acronym: *CIRCULAR AGRONOMICS*
Deliverable name   *D6.4. Data Management Plan*

as soon as possible. Data will be stored based on the contractual terms, until which it can be re-usable (all research related data will be stored at least for five years after the end of the research project). Because of the combined natural and social science nature of the data there is no time limit for its reusability.

▪ <u>Restrictions on data sharing</u>

The principle of data sharing will be to maximize information flow through open access approach, in which all beneficiaries will have grant free access to results produced during the implementation of the Circular Agronomics project. The open access approach will be extended to the scientific community only for non-confidential data. Nevertheless, confidential information should not be disclosed outside the Consortium, to avoid loss of IP protection rights and damage to disclosing beneficiary. In this sense Sect. 8, Sect.9 and Sec.10 of Consortium Agreement will regulate data management and restrictions on sharing should be strictly observed by Circular Agronomics beneficiaries.

As confidential and non-confidential data will be stored in the cloud system, parties that are not signatory of the consortium agreement will not be allowed to access data, unless signing separate Non-Disclosure Agreement (NDA). Therefore, separate NDA could be used to minimize restriction to data sharing with third parties

4. ALLOCATION OF RESOURCES (STORAGE AND PRESERVATION)

Data storage and preservation is the responsibility of each partner and they have to provide the funds for it. Any costs for making CIRCULAR AGRONOMICS generated data FAIR and ensuring Open Access for publications is covered by the EU Grant funding for the H2020 project and should be allocated in each partner budget.

Regarding publication of results, two main routes to open access[2] are suggested and recommended in Horizon2020 projects by the European Commission and in the context of CIRCULAR AGRONOMICS:

1) 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.

2) 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.

In the context of research funding, open access requirements do not imply an obligation to publish results. The decision to publish is entirely up to the grant beneficiaries. Open access becomes an issue *only if* publication is chosen as a means of dissemination.

The DMP applies to all research of all consortium partners (universities, multi-actors, SME partners) and individually each researcher or research team employed or subcontracted will be responsible for managing their data adequately. Where CIRCULAR AGRONOMICS researchers plan to publish with co-authors outside of the project (e.g. with the brother H2020 project Nutri2Cycle or with partners outside of H2020), they will make them aware of the CIRCULAR AGRONOMICS DMP requirements and data collection procedures and make sure that primary research data are stored to the same standard as required for H2020 projects.

5. RESPONSIBILITIES FOR DATA MANAGEMENT

IRTA leads the Data Management, which is the responsibility of WP 7 leader and the project coordinator Dr. Victor Riau (victor.riau@irta.cat). However, It is the responsibility of the workpackage leaders to coordinate the specific data collection in accordance with this data management plan and each partner is responsible for the data generated in their specific WP and their management. Moreover, **IRTA's Technology Watch Unit,** specialized in knowledge management, intellectual property rights and technology watch, will help the WP leaders to identify how to generate and manage appropriately.

---

[2] Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

6. LONG-TERM PRESERVATION ISSUES

For an efficient curation, preservation and provision of access to the data collected or produced under the project, as is defined in the data management plan, during the project each WP leader is responsible for the data generated in their specific tasks and their management, storage and preservation. These data will may be also stored and accessed by everyone on the project intranet to ensure that these data can survive long-term changes in storage media, devices and data formats.

At the end of the project, the data without restrictions will be deposited in Open Science Data repositories or published in open access journals.

7. DATA SECURITY

Data security is the responsibility of each partner but general recommendations are given:

- Back up your files/data regularly and use different media for it (external hard drives, computer hard drives, departmental server, etc.)
- Enable computer firewalls and keep antimalware software up-to-date and operational.
- Users must have access to the computers and/or servers via individual user accounts and not shared accounts.
- Collaborative networks/platforms/Intranets: Permission-controlled files so that users, depending on their status, can "read only", "write", or "execute" files. Computers connected networks should not store sensitive data, unless data is encrypted, so as to minimize network vulnerabilities.
- Cloud-based storage is useful as a secondary or tertiary storage location for your files. Research Data repositories (e.g. Zenodo) are recommended in this sense. However, this via is not recommended for sensitive or confidential data that will be encrypted or treated anonymously.

In CIRCULAR AGRONOMICS, there will be little sensitive data collected. In case of personal data (for instance in WP4 and WP5), these will be stored in an internal database located in the local institutional servers of the corresponding leaders.

8. ETHICAL ASPECTS

A specific WP8 and the associated deliverables regarding Ethics requirements on POPD, NEC, A, EPQ and H has been set up. Deliverables are confidential, only for members of the consortium, including the Commission Services.

9. OTHER ISSUES

Another important procedure in the DMP is to maintain a quality protocol, for this reason to ensure the "data quality" on the FAIR Data Management plan, each WP leader JOINTLY with the Data Management Plan LEAD, will follow and review the application of the DMP.

Without data standards and data quality, the data interoperability (data exchange and re-use between researchers and institutions) is not possible. Data fields and the content of those fields need to be standardized, and agreements on representations, formats and definitions of common data are needed. Therefore, in case metadata standards did not exist in a discipline, the WP leader ("data management task") will specify what kind of metadata will be created[3]/adapted. If needed, the coordinator can provide alternative solutions, such as implementing general Metadata Standards and recommendations, Research Data Alliance providing a Metadata Standards Directory.

---

[3] In the absence of standards, rules must be created to describe how the data is recorded to ensure consistency across multiple sources.

BIBLIOGRAPHY

1. DMP online
2. Guidelines on FAIR Data Management in Horizon 2020. Url: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
3. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
4. How to create a DMP Plan. OpenAIRE. URL: https://www.openaire.eu/opendatapilot-dmp
5. Metadata Standards Directory. URL: http://rd-alliance.github.io/metadata-directory/
6. Metadata Standards. The Science Data Literacy Project. URL: http://sdl.syr.edu/?page_id=32
7. OpenAIRE. Data Managemenet Plan. URL: https://www.openaire.eu/opendatapilot-dmp
8. RDA Metadata Standards Directory. URL: http://rd-alliance.github.io/metadata-directory/
9. Registry of Research Data Repositories. URL: http://www.re3data.org/
10. Zenodo repository: URL: https://zenodo.org/