SPIDER Deliverable D8.2

Data Management Plan

WP	8	**Project and IP Management, Dissemination and Communication activities**
Task	8.1	Strategic and technical coordination with partners and the Commission

Dissemination level	PU	**Due delivery date**	July 31st, 2019
Nature	R	**Actual delivery date**	July 26th, 2019

Lead beneficiary	CEA
Contributing beneficiaries	Wavestone - all partners

Document Version	Date	Author	Comments
0.1	12th July, 2019	N. Deltimple & L. Nicod – Wavestone	Preliminary version for coordinator review.
0.2	17th July, 2019	N. Deltimple & L. Nicod – Wavestone	Completed version for partners' validation.
0.3	25th July, 2019	N. Deltimple & L. Nicod – Wavestone	Completed version with partners edits.
0.3b	4th Nov, 2019	N. Deltimple & L. Nicod – Wavestone	Completed version with partners' approval and track changes accepted for publication on the SPIDER website.

Deliverable abstract

SPIDER is part of the Open research data pilot. This pilot comprises the data needed to validate the results presented in scientific publications and aims at maximizing access to relevant data, taking into account:

- The need to balance openness and protection of scientific and technical knowledge generated in the framework of project's activities;
- Commercialization and IPR;
- Privacy concerns;
- Security;
- Data management and preservation questions.

Therefore, the Data Management Plan (DMP) describes the project's policy concerning any shared data. It will serve to ensure that any shared data is **findable, accessible, interoperable and re-usable (FAIR)**. It covers the handling, curation and preservation of research data during and after the end of the project. It summarises what data will be collected what data will be shared and *via* what methodology and standards.

Deliverable Review

	Reviewer #1: WP Leader		
	Answer	Comments	Type*
1. Is the deliverable in accordance with			
(i) The Description of actions?	☒ Yes ☐ No		☐ M ☐ m ☐ a
2. Is the quality of the deliverable in a status			
(i) That allows it to be sent to European Commission?	☒ Yes ☐ No		☐ M ☐ m ☐ a
(ii) That needs improvement of the writing by the originator of the deliverable?	☐ Yes ☒ No		☐ M ☐ m ☐ a
(iii) That needs further work by the Partners responsible for the deliverable?	☐ Yes ☒ No		☐ M ☐ m ☐ a

* Type of comments: M = Major comment; m = minor comment; a = advice

Table of contents

List of figures:

1. Introduction

A data management plan (DMP) describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), DMPs include information on:

- the handling of research data during & after the end of the project;
- what data will be collected, processed and/or generated;
- which methodology & standards will be applied;
- whether data will be shared/made open access and
- how data will be curated & preserved (including after the end of the project).

The present preliminary data management plan (DMP) has been collectively produced to help the SPIDER teams to comply with the requirements of the Grant Agreement on open access and data management. It will be updated during the project lifetime to catalyse the development of good practices at all levels of the project's data lifecycle (see Figure 1).
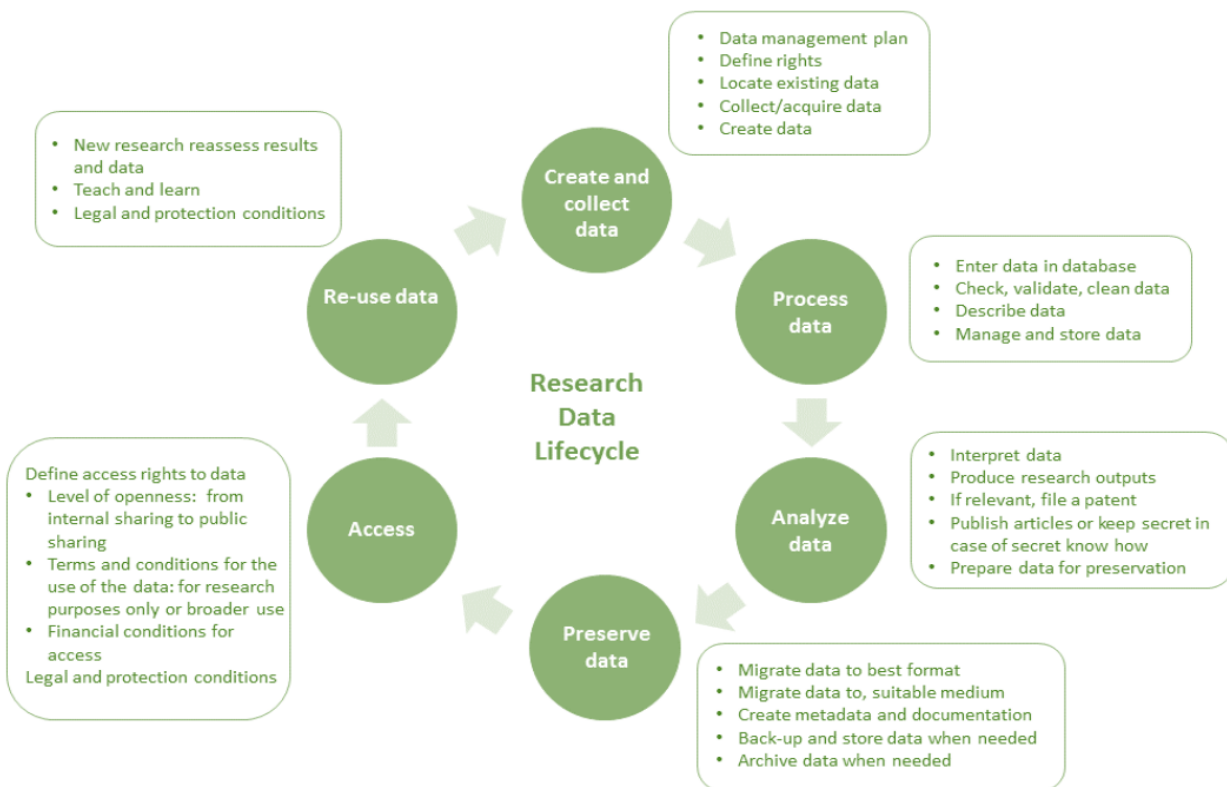
Figure 1 – Research data lifecycle

Research data is subject to a large number of definitions. For the purpose of this deliverable, the SPIDER partners have chosen the one given by Sorbonne-Paris Cité (Cartier et al., 2015):

"Research data: set of factual information recorded on media. They are produced or collected according to various methods during a research process"

Therefore, the data may be of different types: observation, experimental, computational or simulation, derived or compiled data, reference data.

2. Information about the project

Project name*	Safe and Prelithiated hIgh energy DEnsity batteries based on sulphur Rocksalt and silicon chemistries
Project acronym	SPIDER
Project objectives	Generate knowledge-based improvements of Li-ion battery cost, performance, recyclability and safety to enable electric vehicles to rapidly gain market share and reduce CO_2 emissions.
Call number	H2020-NMBP-ST-IND-2018
Funder	European Commission (H2020)
Grant agreement number	814389
Project coordinator and partners	1. COMMISSARIAT A L'ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES (**Coordinator**) 2 SGL CARBON GMBH 3 NANOMAKERS 4 VARTA MICRO INNOVATION GMBH 5 FORSCHUNGSZENTRUM JULICH GMBH 6 TECHNISCHE UNIVERSITAET MUENCHEN 7 RHODIA OPERATIONS 8 FUNDACION CIDETEC 9 VLAAMSE INSTELLING VOOR TECHNOLOGISCH ONDERZOEK N.V. 10 ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS 11 NORGES TEKNISKNATURVITENSKAPELIGE UNIVERSITET 12 WAVESTONE LUXEMBOURG SA 13 CENTRO RICERCHE FIAT SCPA 14 ACCUREC-RECYCLING GMBH
Contact	Cédric HAON, 04 38 78 34 71, cedric.haon@cea.fr CEA
Project duration	01/01/2019 + 44 Months

3. Identified datasets

Currently, at Month 7 of the project, it seems that the following 6 main datasets will emerge from the project :

- **Material synthesis parameters**
- **Electrolyte composition parameters**
- **Cells processing parameters**
- **Recycling process parameters**
- **Characterisation results**
- **Modelling results**

SPIDER data may come from:

- Experiments (synthesis of materials, electrolyte formulation, cell production, prelithiation methods, recycling…)
- Electrochemical characterizations (measurements of electrochemical performances in batteries)
- Physico-chemical characterizations (X-ray diffraction, XPS, BET surface analysis, particle size distribution, elemental analysis…)
- Safety and abuse test (DSC, MMC)
- Activities related to the development and validation of numerical models with or without interaction with the abovementioned R&D actions.

These datasets can have various forms (Raw data, deliverables, reports, peer reviewed publications, presentations) and also various purposes/ impacts as:

- To improve scientific knowledge
- To increase material/prototype/product quality
- To allow process upgrade
- To allow project partners generating patents to protect further exploitation

File formats will mostly be:

- For Raw data: .txt, .jpeg, .tiff, .xls, .mat, .csv, .ppt, .png, and more specifically:
  - X-ray diffraction : .raw, .udf, .udx, .dat
  - Electrochemistry software format raw data (will also be saved as .xlsx): "TE*.DB" (Basytec), .mpr (Biologic), .dfr (Autolab, EIS).
- For Deliverables, reports, publications, presentations: xlsx, docx, pptx, pdf

Total data size should be 30-50Go *a priori per partner*.

Interested audiences may include European battery or components manufacturers, scientific community. Access to third parties for re-use of the processed data published in journals being restricted according to the copy right license agreement of the journals, third parties should seek permission from the journal.

4.   Set-up of first rules

4.1 Data storage

Regarding project data management, all participants will follow best practices for data generation, storage and sharing: document changelog, unified name attribution and appropriate repository information will be kept as clear as possible. The documents will be preferably shared within the consortium through the collaborative document sharing platform https://spider-battery-project.cea.fr/SitePages/Home.aspx (Figure 2). The access to this central repository is regulated by the project coordinator. It is provided for project consortium and to other linked parties upon request from a project team, in accordance with the access provisions in the consortium agreement. This space is password protected. The **platform is hosted by CEA servers** and the security of this platform is guaranteed by the Informatics Systems Direction (DSI) of CEA. Regular back up of all data stored on CEA servers is ensured.

To facilitate document evaluation and review, all SPIDER deliverables and official documents are created using established templates for main MS Office formats (or MS Office-compatible). The overall data produced and/or collected by each consortium partner will be carefully stored and managed by this partner following the agreed rules:

- In a preliminary stage of production (and/or collection) and until full analysis and assessment of significance, a local storage by the authors is not excluded.

- As much as possible, raw data files will be integrated in working documents and stored, during the project life time, on the **SPIDER secure, collaborative electronic document sharing platform** (Figure 2).
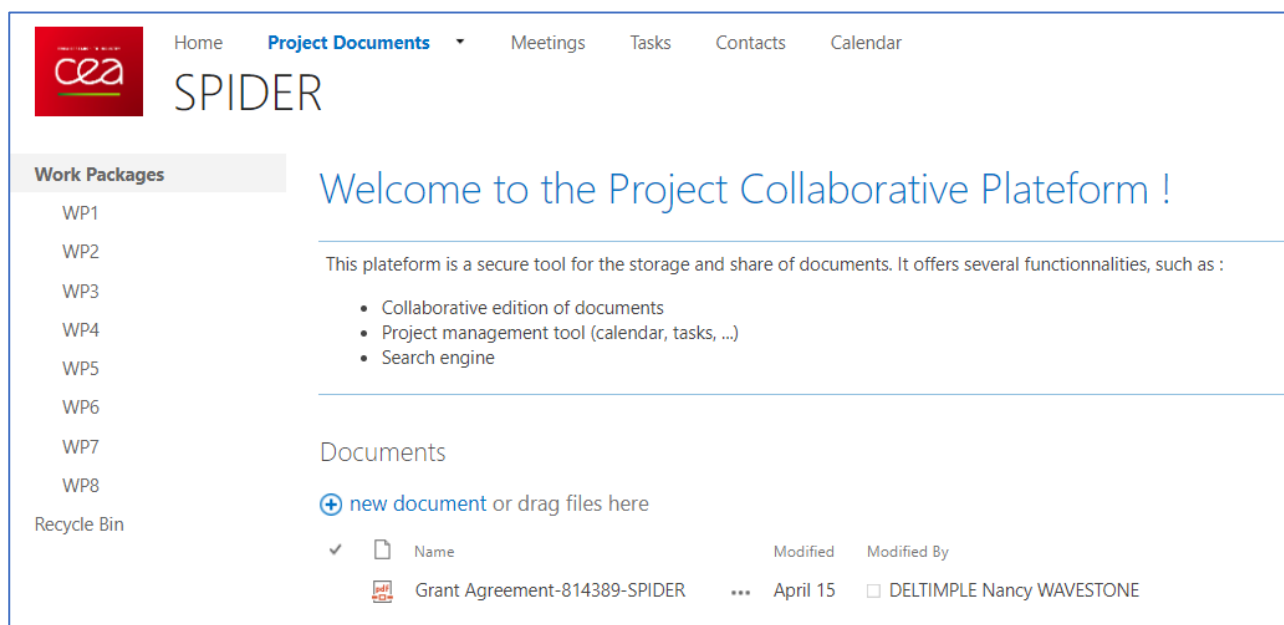

Figure 2 – Directories tree of Spider collaborative platform

Clear versioning will be used, as per the rules defined in the manual of management.

Keywords will also be largely used, both

- Internally, with a clear link to the concerned data,

- **and** during the submission of scientific papers which contain the processed data.

For the sake of inter-disciplinary interoperability, explicit words will be used and eventually definitions can be provided to clearly define terms used in SPIDER Working documents or deliverables, including datasets.

4.2 Making data openly accessible

Open access in H2020 means providing online access to scientific information free of charge to the reader. **Open access** concerns **two categories:**

- **Publications**, i.e. peer-reviewed scientific research articles (primarily published in academic journals or conference proceedings / conference presentations);

- **Research data.**

Since the H2020 requirement for **Open Access publishing is fully embraced by the SPIDER project**, the project will ensure both "**green**" (in addition to publication in subscription journals, the copy of an article is deposited into a repository) and **"gold"** (publications are available directly from the publisher after paying author's fees, envisioned in the project's budget) data publishing.

Research data is information (particularly facts or numbers) collected to be examined and considered, and to serve as a basis for reasoning, discussion or calculation. Open access to research data refers to the right to access and reuse digital research data under the terms and conditions set out in the Grant Agreement (Article 29.3).

The management of SPIDER data requires standardised and timely reporting regarding data which can be shared within the project and with external stakeholders. This will be addressed by all partners under coordination of CEA.

The partners will use standard formats and software to publish data in journals, and prepare the data considering that scientists with physical/chemical or technical background should be able to understand the data vocabulary and relevant terminology.

More specifically, when a collection of data will be **ready to be published** (i.e. authorization received by all concerned parties), the last final version of these data stored on the SPIDER platform will be uploaded on the most pertinent open access public platform, which may be:

- A subject-based repository

- Institutional Repositories

- Data repositories, ie. digital archives collecting and displaying datasets and their metadata, possibly also publications, linking these publications to the underlying data. EU-funded data repositories include OpenAIRE www.openaire.eu and ZENODO https://zenodo.org .

In order to ensure easy availability of key data sets, SPIDER data will be classified into three specific categories as follows.

GOLD DATA CATEGORY

Analysed data that has clear scientific significance, provides interesting new results that have resulted in new understanding in the field, possibly supporting high impact journal publications. Such data will be made available online using a reliable, indexable repository that is compatible with the European Commission's OpenAire platform, such as FigShare.

GREEN DATA CATEGORY

Analysed data that is relevant for the SPIDER partners and can be used to develop new processes and models within the project work plans. Such data will be shared in a searchable repository accessible by all partners, such as the SPIDER platform.

WHITE DATA CATEGORY

Raw data or data that has been analysed but is not thought at that stage to be significant to either the project or wider community. Such data may still have unseen value and will be stored in a local, searchable archive that will be protected against data loss.

Figure 3 – Data management process for new data

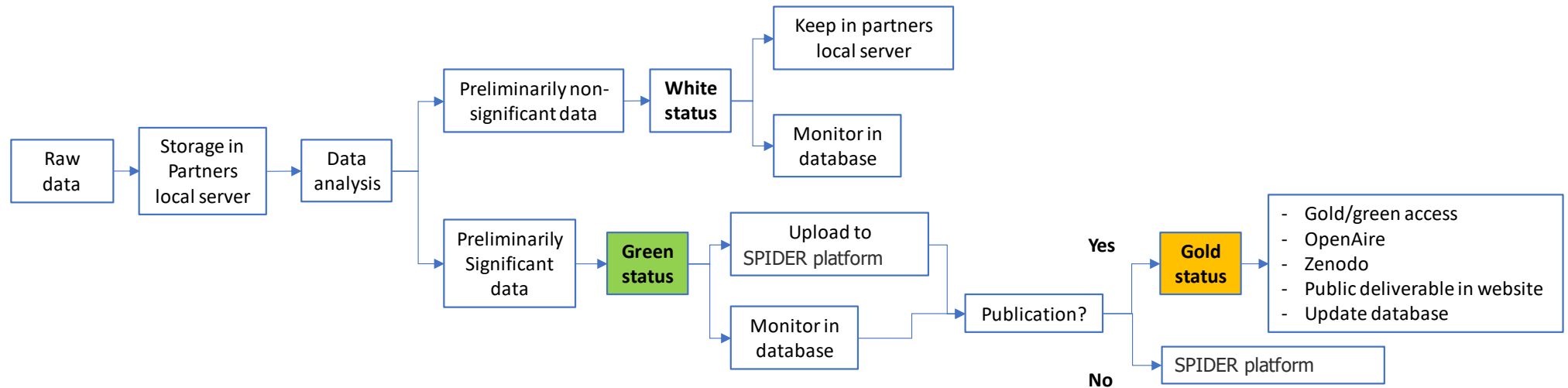The data publication process can therefore be illustrated as follows:

Figure 4–Data management plan for open access data, by category

4.3 Considerations for publishable information

Therefore, the SPIDER consortium, **guided by the Project coordinator and IP manager**, will adopt the following strategy (Figure 5):

If the research results in a ground-breaking innovation, the members of the consortium will consider **two forms of protection**:

- to withhold the data for internal use or

- to apply for a patent in order to commercially exploit the invention and have in return financial gain. In that case, publications will be therefore delayed until the patent filing.

On the contrary, if the technology developments are not going to be withheld or patented, the results (including data) will be published for knowledge-sharing purposes, on the condition they are compatible with the confidentiality of the project.
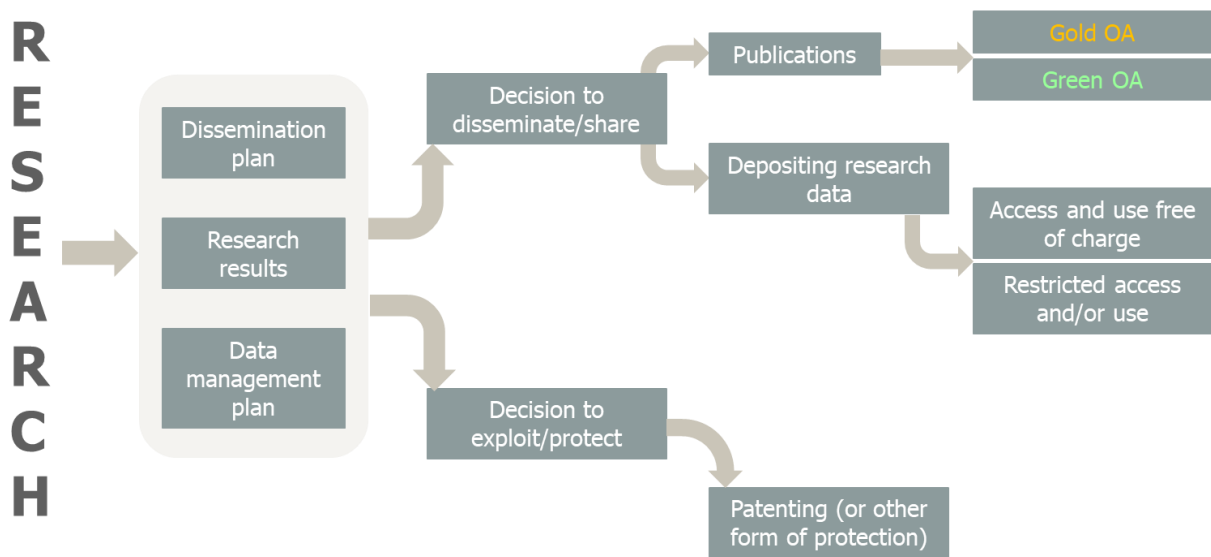

Figure 5 – Process for determining which information is to be made public (from EC's document "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020")

The process and methodology for the validation of publications were described in Deliverable 8.1.

Data will be released as soon as possible but possibly not before patents are fully granted in relevant countries or after a maximum of 6 months embargo for publications.

5. Ethical aspects

According to the Annex 1 of Grant Agreement 814389, the SPIDER Consortium has taken into account all ethical issues listed in the H2020 participant portal for the call topic "Materials for future highly performant electrified vehicle batteries". The consortium is committed to demonstrating research integrity, which means, in particular, avoiding falsification, plagiarism or other research misconduct.

More precisely, all the activities carried out under the SPIDER project comply with ethical principles and relevant national, EU and international legislation, for example the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights. The project does not involve research on humans, animals or cells.

In any case, all the partners will follow their internal protocols to treat any material according to the national law and EU legislation. All chemical waste will be collected and processed by a central facility that respects regulations and environmental and health best practices. All waste will be recycled or appropriately disposed of.

6. Open issues

The partners will discuss further when project progresses, to refine;

- **The list of datasets and their confidentiality level**. This will be assessed considering the re-used pre-existing data, the confidentiality of the process/ object they describe, and the exploitation potential if these datasets are kept and protected by consortium partners. Of course, data related to publications in open access will be open.

- **The datasets naming rules**

- **The classification rules for data access, storage and archiving**

- **The platforms to be used to share datasets** identified as « open », possibly, as mentioned above:

  - « Conventional » platforms for open data storage or data repositories such as Zenodo, Dryad, Figshare, etc. The partners will investigate whether the chosen external repositories are permanent and will ensure that the entity in charge of this repository does not take the ownership of the data. They will also ensure the creation of a DOI (Digital Object Identifier), commonly used to identify journals/data when published in scientific journals. Overviews of repositories can be found on ROAR, OpenDOAR, Re3data, and OpenAIRE provides guidance on proper depositing.

  - The project website, http://www.project-spider.eu/ ,

  - EU-funded data repositories such as OpenAIRE www.openaire.eu and ZENODO https://zenodo.org ,

  - Journals websites[1], etc

- **Data archiving rules and costs**. This will include a risk analysis to determine the relevance of data retention (scientific, strategic, legal, heritage value), including:

  - Legal or contractual obligation to retain certain data for a specified period by making them available on request in certain cases. (*a priori* not applicable for SPIDER)

  - Transparency of research and traceability. Indeed, publishers may request to link data to publications.

  - Data which are non-reproducible or too expensive to reproduce.

  - Political and heritage issues: capitalization of data (*a priori* not applicable for SPIDER).

The type of data to be archived will also be assessed according to the project and its objectives. Data sampling will be considered, to reduce the archive volume, assessing if a selection of data is sufficiently representative of the set.

---

[1] publishers may request to link data to publications.

7.   Annex : References considered by the partners

- The European Commission, Directorate-General for research & innovation (2016a) ***Guidelines on FAIR Data Management in Horizon 2020, V3.0***. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

- The European Commission, Directorate-General for research & innovation (2016b) ***Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, V3.1***. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

- Norms and standards for the metadata :

  - Norm ISO 23081-1 :2006 Metada for records - part 1 principles

  - Norm ISO 15836 :2009 The Dublin Core metadata element set

  - Norm ISO 16684-1:2012 Extensible metadata platform (XMP) specification

- Existing external data repositories

  - http://www.zenodo.org/

  - http://www.re3data.org/

  - http://www.dcc.ac.uk/resources/how-guides