D1.2 Data Management Plan

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Big Energy Data Value Creation within SYNergetic enERGY-as-a-service Applications through trusted multi party data sharing over an AI big data analytics marketplace

Deliverable nº:	**D1.2**
Deliverable name:	**Data Management Plan**
Version:	**1.0**
Release date:	**30/06/2020**
Dissemination level:	**Public**
Status:	**Submitted**
Author:	**Eva Muñoz (ETRA)**

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Document history:

Version	Date of issue	Content and changes	Edited by
0.1	15/05/2020	First draft version	Eva Muñoz (ETRA)
0.2	08/06/2020	Second draft version	Álvaro Nofuentes (ETRA)
0.3	18/06/2020	Third version, SYNERGY Datasets	Eva Muñoz (ETRA)
1.0	30/06/2020	Final version ready for submission	Eva Muñoz (ETRA)

Peer reviewed by:

Partner	Reviewer
KONCAR	Leila Luttenberger
UBI	Dimitrios Miltiadou
S5	Fenareti Lampathaki

Deliverable beneficiaries:

WP / Task
WP1 / T1.4

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Table of contents

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Abbreviations and Acronyms

Acronym	Description
CERN	European Organization for Nuclear Research
D	Deliverable
DCMI	Dublin Core Metadata Initiative
DMP	Data Management Plan
DoA	Description of Action
DOI	Digital Object Identifier
EC	European Commission
FAIR	Findable, Accessible, Interoperable and Reusable (FAIR).
H2020	Horizon 2020. The EU Framework Programme for Research and Innovation
IPR	Intellectual Property Rights
ISBN	International Standard Book Number
ISRC	International Standard Recording Code
OAI-PMH	Open Archive Initiative-Protocol for Metadata Harvesting
ORD	Open Research Data

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Executive summary

The Data Management Plan (DMP) is a living document that aims at providing an analysis of the main elements of the data management policy that will be used by the SYNERGY Consortium regarding the project research data.

This document describes initially the methodological framework taking into account the H2020 guidelines regarding Open Research Data with a twofold objective: on the one hand, to clearly define how all public data assets issued by the project consortium will be further disseminated to a wider audience and on the other hand, to outline the methodology which can make the research data generated in the context of SYNERGY, findable, accessible, interoperable and reusable (FAIR principles). This DMP will evolve during the implementation of the project, when the project progresses and when significant changes occur, in order to keep an updated version of the guidelines and recommendations and therefore contribute to knowledge discovery and innovation.

The current version is the first iteration in which the envisioned data management strategy is presented and first effort is made to plan the definition of the types of research data that will be generated or collected during the project, the standards that will be used, how the research data will be preserved and what parts of the datasets will be shared for verification or reuse.

This document is open for further iterations and will be updated during the project implementation.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

1 Introduction

1.1 Purpose of the document

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that needs to be considered by the SYNERGY Consortium with regard to the project research data.

The DMP is not a fixed document; on the contrary, it will evolve during the lifespan of the project. This DMP will be a living document in which information will be available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur [1].

This first version of the DMP aims to outline how the SYNERGY project will try to make the research data findable, accessible, interoperable and reusable (FAIR principles) and therefore contribute to knowledge discovery and innovation. Although this first version submitted by month 6 of the project does not provide very detailed information on the specific data sets to be collected, generated and processed during the project, the objective is to lay the foundations for creating an effective data management strategy covering the complete research data life cycle.

1.2 Scope of the document

This report describes the initial data management plan for the project covering the complete research data life cycle. This data management plan, which is expected to be continuously updated, will be used by the SYNERGY consortium as a guideline when handling the research data during and after the end of the project. The current version is the first iteration in which the envisioned data management strategy is presented and it includes a first definition of the types of research data that will be generated or collected during the project, the standards that will be used, how the research data will be preserved and what parts of the datasets will be shared for verification or reuse. The next versions of the DMP will get into more detail and describe the datasets to be produced by the project, the specific conditions that are attached to them and the practical data management procedures to be implemented by the SYNERGY project.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

1.3 Structure of the document

This deliverable follows the template provided by the services of the European Commission (EC) [2] on Data Management Plans. The structure of the template has just been adapted to follow SYNERGY document procedures –e.g. section 1 “Introduction” is mandatory in all the project documents.

On a first section, the data sets that will be used in the project are introduced and briefly described. Then, the document focuses on making SYNERGY research data findable, accessible, interoperable and re-usable (FAIR). The summary of the European Commission guidelines can be found in [3].

Respecting the guidelines provided by the Commission [1], this document will be updated in time with the periodic evaluation/assessment of the project at month 18, 30 and at the end of the project, at month 42. During these reviews and whenever significant changes arise, the current document will be updated with newly available information.

This project has received funding from the European Union’s Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

2 Data Summary

2.1 Purpose of data management and relation to the project

SYNERGY introduces a novel reference big data architecture and platform that leverages data, primary or secondarily related to the electricity domain, coming from diverse sources (APIs, historical data, statistics, sensors/ IoT, weather, energy markets and various other open data sources) to help electricity stakeholders to simultaneously enhance their data reach, improve their internal intelligence on electricity-related optimization functions, while getting involved in novel data (intelligence) sharing/trading models, in order to shift individual decision-making at a collective intelligence level.

SYNERGY participates into the Open Research Data Pilot (ORD Pilot) and therefore open access to research data is applicable (SYNERGY Grant Agreement 872734 [4] – Article 29.3). SYNERGY will comply with the European recommendations regarding Data Management Plans [1], providing a clear procedure for findable, accessible, interoperable and re-usable (FAIR) data and updating the current document along the development of the project.

The purpose of the DMP is to provide an analysis of the main elements of the data management policy that will be used by the Consortium with regard to the project research data. The DMP reflects the consortium's data management policies, systems and procedures - which will be implemented and embedded into research procedures and regularly reviewed throughout the research cycle.

The project will strive to make data open, but cannot overrule limitations that partner institutions put on data that they contribute (as specified in their Background included in the SYNERGY Consortium Agreement). Moreover, an ethical approach will be adopted and maintained throughout the fieldwork process. The responsible partners will assure that the EU standards regarding ethics and Data Management are fulfilled, as stated in D1.3 Ethics Monitoring Report.

2.2 Data set types, formats and standards

Throughout the duration of the project, SYNERGY will gather – and also generate - a variety of data. All of them have to be stored in a way that it is easily accessible by both humans and software, as appropriate.

Broadly, the data falls into two categories:

- Technical and scientific data – this includes raw and processed experimental data, scientific analyses/publications as well as software code and algorithms.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

- Organizational data – e.g. data relevant to the implementation of the Innovation action.

The detailed definition of the different types, formats and standards of the data to be collected, processed and/or generated during the project is taking place in the framework of WP2, where the data assets available from the SYNERGY demonstrators are identified in T2.3, and in WP3, more specifically in task T3.1 "Open Standards, Interoperability and Common Information Model Adaptation".

The results of these tasks will be reported in D2.3 "Ex-Ante Pilot Audits and Pilot Deployment Plan" in M8 and D3.1 "SYNERGY Common Information Model" in M12 respectively. While D2.3 will provide information about relevant data sets involved in the project's demonstrators and their meta-data, D3.1 will detail the energy data modelling landscape and the specific open standards, semantic models and ontologies that are selected for further elaboration depending on their relation to the SYNERGY scope.

In this context, in month 6 the project is not yet in the position to provide a detailed list of the data sets of the project neither the formats nor specific standards to be used. As defined in the DoA and the structure of the project, it was not expected/planned to have the details of this information by Month 6 and therefore it cannot be included in this version of the Data Management Plan. Next version of the DMP (on M18) will include up-to-date decisions on datasets to be used.

2.2.1 SYNERGY Datasets

SYNERGY has started to complete the **[SYNERGY] Continuous Data Collection spreadsheet** in the context of the activities of task T2.3 "Ex-Ante demo surveys, data landscaping and deployment planning". By means of this tool, the demo partners are individually documenting, on the one hand, the demonstrators' data assets that they have available and on the other hand, the data assets they have currently identified as needed.

The different fields that define the data assets can be classified into the following groups (tables will be in dark green when "Existing data assets" and "Requested / Expected Data Asset" have the same fields and there will be a table in light green when "Requested / Expected Data Asset" has a different definition of fields):

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

- General Info (all of them for Existing Data Asset; only some of them for Requested / Expected Data Asset):

Dataset ID	*[Unique identifier following the convention "Country_Partner#no"]*
Demo Case-related ID	*[Unique identifier following the convention "DA_Country_Partner_DC#no"]*
Data Asset Title	*The title of the data asset*
Description	*A brief description of the data asset - At least 2-3 lines to give an overview of the data*
Category	*[1st Data Tier (Primary Energy Data), 2nd Data Tier (Extra-Energy Data)]*
Date of Last Update	*The date of the last update of the specific spreadsheet*

- Data Asset features (all of them for Existing Data Asset; only some of them for Requested / Expected Data Asset):

Volume	*[X GBs / records / transactions per hour / day / month]*
Variety	*[Structured / Unstructured / Semi-structured]*
Type	*[Text / Image / Video / Audio / Other]*
Format	*[csv, xml, json, other]*
Velocity	*[Real-time, Near Real-time, Batch]*
Historical Data Availability	*[Y/N]*
Temporal Coverage	*[From ... To...]*
Spatial Coverage	*[Locations]*
Language	*[e.g. English, Italian, German, Greek, ...]*
Relevant Standards	*[List the international standards to which a data asset complies]*
Veracity	*[Raw, Pre-processed, Processed Data asset]*

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Temporal Resolution	*[The temporal "granularity" of the data, e.g. per minute / hour / day / month]*
Spatial Resolution	*[The spatial "granularity" of the data, e.g. at district / zone / building / area level]*
Dependency / Linking to Other Sources	*[Y/N, If Y, list the other sources or codelists]*

- Data Asset Availability (all of them for Existing Data Asset; only some of them for Requested / Expected Data Asset):

(Potential) Data Asset Owner	*The name of the data asset owner*
Data Asset Available from 3rd Party	*[Y/N]*
Data Asset Provider	*The name of the data asset provider in SYNERGY*
Accessibility Method	*[Through API, As downloadable file, As database extract, Other]*
Frequency of Updates	*[Real-time, Every X minutes / hours, Daily, Weekly, Monthly, Yearly, other]*
Update Strategy	*[Append new data / Replace existing data / other]*
Status **(in case it is a Data Asset Need)**	*(a) Data assets needed but covered by the demonstrator partners, (b) Data assets needed, not covered by the demonstrator partners and arranged with 3rd parties, (c) Data assets needed but still open*

- Data Asset Rights (for Existing Data Asset):

Documentation	*The documentation of the API or data sample (incl. the location and the name of the file in the SYNERGY repository]*
Privacy	*[Confidential (not to be shared at all) / Proprietary (to be shared with appropriate licensing with the demonstrator partners) / Private (to be shared with appropriate licensing within the demonstrator & potentially to be traded with other stakeholders in SYNERGY) / Public (available to all)]*

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

License	*[Exact Licence that is currently applied, e.g. CC Attribution-NonCommercial-ShareAlike (CC BY-NC-SA), or Case-by-Case Bilateral Agreement]*
Sharing Mode	*[Encrypted Data Sharing / Unencrypted Data Sharing / Secure Multi-party Computations (with data always on-premise at providers' side) / Encrypted Intelligence Sharing / Unencrypted Intelligence Sharing]*
Data Asset Consumer(s)	*The list of consumers (in the demonstrator) that are interested in the specific data asset*
Other Stakeholders Potentially Interested in Data Asset Use/Purchase	*[List categories of stakeholders beyong SYNERGY that are potentially interested in the specific data asset]*
Pricing	*[Per Transaction / Subscription / PAYG / N.A.]*
Need for Anonymization	*[Y/N depending on whether the data asset contains sensitive or personal data]*

- Data Asset Rights (for Requested / Expected Data Asset):

Sharing Expectations	*[Confidential (not to be shared at all) / Proprietary (to be shared with appropriate licensing with the demonstrator partners) / Private (to be shared with appropriate licensing within the demonstrator & potentially to be traded with other stakeholders in SYNERGY) / Public (available to all)]*
License	*[e.g. any license allowing commercial use]*
Sharing Mode	*[Data to be downloaded and always available on-premise / Data to be used in the SYNERGY core platform / Data to remain on the data providers' premises but obtain access to use them for an analysis / No preference]*

- Data Analysis:

Types of analysis currently conducted on data	*[e.g. Correlation analysis for..., Predictive analytics for...]*

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

- Data Asset Assessment:

Accuracy	*[Measure of correctness and precision, e.g. whether the dataset is effor-free, Ranked 1 (Low) - 5 (High)]*
Completeness	*[Degree to which a data asset is sufficient in scope, depth, Ranked 1 (Low) - 5 (High)]*
Timeliness	*[How long a data asset remains up-to-date]*
Relevance to specific demonstrator	*[How relevant a data asset is for the specific demonstrator, Ranked 1 (Low) - 5 (High)]*
Relevance to other SYNERGY demonstrators / demo cases (for Existing Data Asset)	*[How relevant a data asset is for other demonstrators, Ranked 1 (Low) - 5 (High), following the convention "D#no_DC#no_Rank X", e.g. D1_DC1_Rank 5 for a very relevant data asset for demo case 1 under demonstrator 1]*
Relevance to other SYNERGY demonstrators / demo cases (for Data Asset Need)	*[How relevant a data asset is for other demonstrators, Ranked 1 (Low) - 5 (High), following the convention "DA_Country_Partner_DC#no_Rank X", e.g. DA_Greece_ELIN_DC1_Rank 5 for a very relevant data asset for a specific demo partner]*
Importance	*[How critical a data asset is for the demonstrator, Ranked 1 (Low) - 5 (High)]*
Rationale	*[Explain the reasons for importance ranking]*

Finally, other energy data sources can be also identified, which will be defined by the same parameters already described in the tables above.

Based upon their privacy attribute, datasets can be divided into four main categories:

- Confidential (not to be shared at all)
- Proprietary (to be shared with appropriate licensing with the demonstrator partners)
- Private (to be shared with appropriate licensing within the demonstrator & potentially to be traded with other stakeholders in SYNERGY)
- Public (available to all)

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

In next versions of the Data Management Plan, and once the [SYNERGY] Continuous Data Collection spreadsheet is analysed, only the information about public dataset will be considered in order to have a full description of them.

Nevertheless, there is already some information collected about **Other Energy Data Sources** which has been tagged as "Public". The following table includes these datasets with their corresponding attributes.

ID	Data Source	Description
OPEN_1	ENTSO E Transparency Platform	Central collection and publication of electricity generation, transportation and consumption data and information for the pan-European market.
OPEN_2	OpenWeatherMap	Weather Meteo Service
OPEN_3	Greek Enex MarketPlace	Central collection and publication of electricity market data for the Greek Market. For the moment day ahead market data are accessible
OPEN_4	Greek IPTO Data Repository	Central collection and publication of electricity network data for the Greek Market.
OPEN_5	Spanish REE Repository	Central collection and publication of electricity network data for the Spanish Market.
OPEN_6	Austrian APG Repository	Central collection and publication of electricity network data for the Austrian Market.
OPEN_7	Nordics Pool Repository	Central collection and publication of electricity network data for the Finnish Market.

2.2.2 SYNERGY Scientific publications

No scientific publications had been produced in the context of the SYNERGY project by the time of writing the deliverable. In the forthcoming versions of this living document, the following template will be used for the management of SYNERGY scientific publications:

Publication reference and name:	The name and the ID of the publication (e.g. DOI).
Publication Abstract:	A short description of the content of the publication.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Standards and metadata:	The type of the document format and any type of metadata associated with the content of the document.
Publication sharing:	Where the final version of the scientific publication is stored.
Archiving and preservation (including storage and backup):.	How the scientific publication will be preserved, archived and preserved

2.3 Re-use of data and software

During the project lifetime, available results from other research activities, publications, and further relevant information available will be analysed. This information will be mainly used for internal project studies and will be provided in respective project deliverables with appropriate references to the origins of the gathered information.

Moreover, SYNERGY will use and exploit some of the preliminary solutions and technologies developed in various H2020 projects: ICARUS [5], BIMERR [6], NOBEL GRID [7], WISEGRID [8], GRIDSOL [9] and FLEXICIENCY [10] projects. These projects are expected to form the basis of some of the SYNERGY Innovation activities, complementing and going beyond the work done in these projects, providing new and more advanced services.

2.4 Timetable for updates of the DMP

As previously indicated, the Data Management Plan is a living document that will be updated over the course of the project whenever significant changes arise, such as new data, modification in consortium policies, changes in consortium composition, external factors, etc.

As a minimum, the Data Management Plan will be updated according to the following timetable:

DMP Version	Delivery Month	Description
V1.0	M6	DMP First version
V2.0	M18	DMP including WP2 and WP3 inputs and data updates
V3.0	M30	DMP including any refinement coming from the first demonstration activities
V4.0	M42	DMP Final version

Table 1 – Timetable for updates of the DMP

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

3 FAIR data

According to [11], the FAIR principles describe four key concepts in research data management. In accordance with the FAIR principles, data should be:

- **Findable** – Easy to find by both humans and computer systems and based on mandatory description of the metadata that allows the discovery of interesting datasets;
- **Accessible** – Long term storage so data can be easily accessed and/or downloaded with well-defined license and access conditions, whether at the level of metadata, or at the level of the actual data content;
- **Interoperable** – Ready to be combined with other datasets by humans, as well as computer systems;
- **Reusable** – Ready to be used for future research and to be processed further using computational methods.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

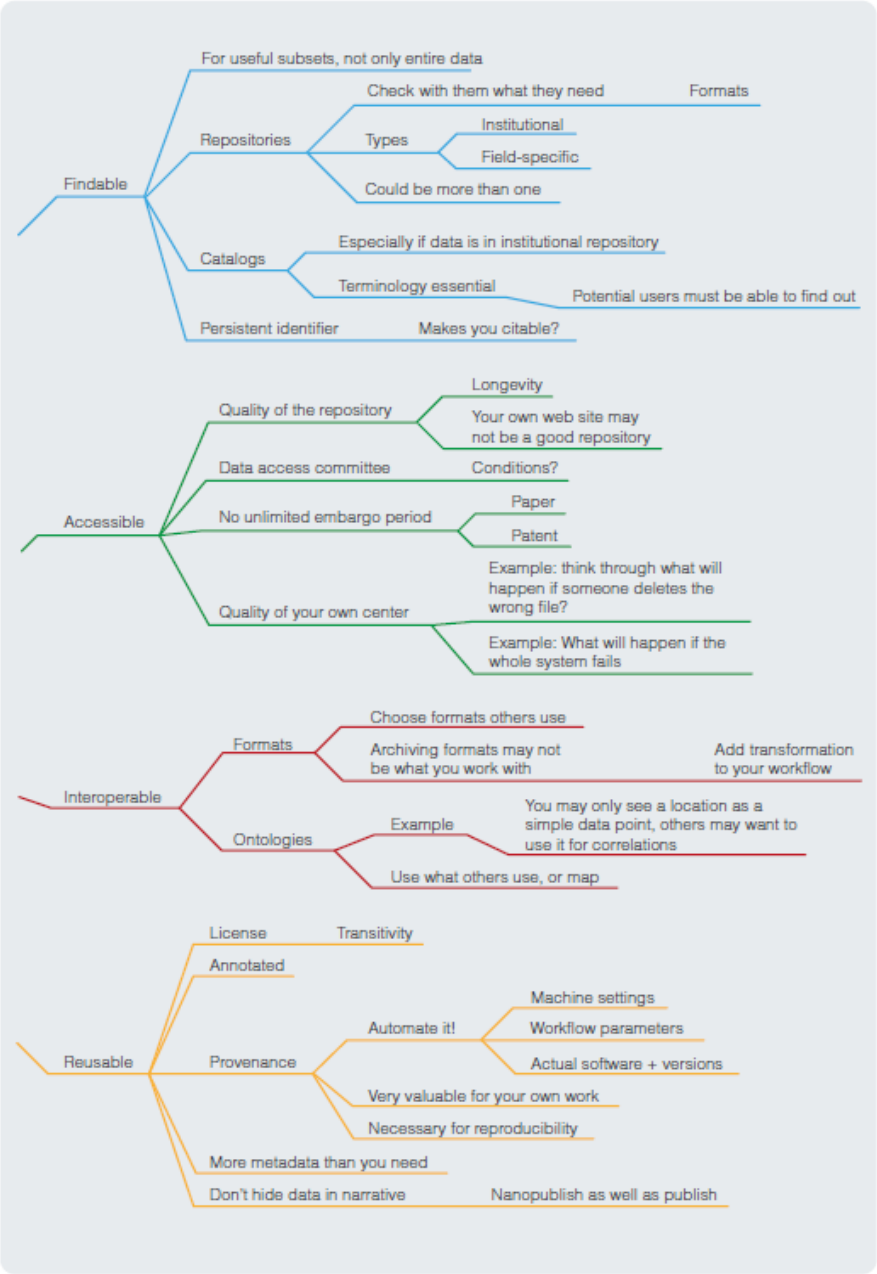

Figure 1 – Research Data Management according to the FAIR principles [12].

3.1 Making data findable, including provisions for metadata

Identification and localisation means will be used for the data to be processed during SYNERGY project. The data to be generated in SYNERGY project will be identifiable and locatable by means of unique identification mechanisms. Files will be uniquely identifiable by using standardised name conventions and clear versioning. These conventions for the documents – and data sets - are already provided in D1.1 Project Handbook [13].

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

SYNERGY project research data will be inventoried and annotated with metadata following the standards promoted by the European Data Portal [14].

> *"Metadata describes the dataset itself (e.g. date of creation, title, content, author, type, size). This information about the data needs to be added to the catalogues to help discover the data. Metadata needs to be both human understandable and machine readable. If it is published as Linked Data, the discoverability of the data is greatly increased. Metadata does not only serve the purposes of description and discovery, but also renders itself as essential for the scope of contextualisation (relevance, quality, restrictions (rights, costs)), as well as for matching users and software to data available on the internet."*
>
> European Data Portal

The European Data Portal strongly recommends the use of the DCAT Application Profile for metadata [14, p. 47] . The Data Category Vocabulary (DCAT) – used for datasets on the internet, is based on the Dublin Core standards.

The Dublin Core metadata standard is a simple yet effective element set for describing a wide range of networked resources. From the perspective of the Dublin Core community, the metadata landscape is currently characterized in terms of four "levels" of interoperability and 15 sections for data description [15] [16]:

Dublin Core Levels of interoperability	
Level 1: Shared term definitions	Shared vocabularies defined in natural language
Level 2: Formal semantic interoperability	Shared vocabularies based on formal semantics
Level 3: Description Set syntactic interoperability	Shared formal vocabularies in exchangeable records
Level 4: Description Set Profile interoperability	Shared formal vocabularies and constraints in records

Table 2 – Interoperability Levels (Source: DCMI)

Contributor	An entity responsible for making contributions to the resource
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant
Creator	An entity primarily responsible for making the resource

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Date	A point or period of time associated with an event in the lifecycle of the resource
Description	An account of the resource
Format	The file format, physical medium, or dimensions of the resource
Identifier	An unambiguous reference to the resource within a given context
Language	A language of the resource
Publisher	An entity responsible for making the resource available
Relation	A related resource
Rights	Information about rights held in and over the resource
Source	A related resource from which the described resource is derive
Subject	The topic of the resource
Title	A name given to the resource
Type	The nature or genre of the resource

Table 3 – Dublin Core Metadata Element Set – 15 Elements Overview

The fifteen element "Dublin Core" described in this standard is part of a larger set of metadata vocabularies and technical specifications maintained by the Dublin Core Metadata Initiative (DCMI). The fifteen element descriptions have been formally endorsed in the following standards:

- ISO Standard 15836:2009 of February 2009 (confirmed in 2014)
- ANSI/NISO Standard Z39.85-2012 of February 2013
- IETF RFC 5013 of August 2007

The DCAT Application Profile is based on Dublin Core standards but is not a vocabulary. It is a simple specification for metadata descriptions of EU governmental data and portals.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

The metadata will be published with the data using a machine-readable format and standard terms to define the metadata; the overall features of the dataset will be described with information about local parameters, licence, origin and quality.

The figure below summaries the best practices outlined by the European Data Portal which will be followed by the SYNERGY partners when using the DCAT-AP.

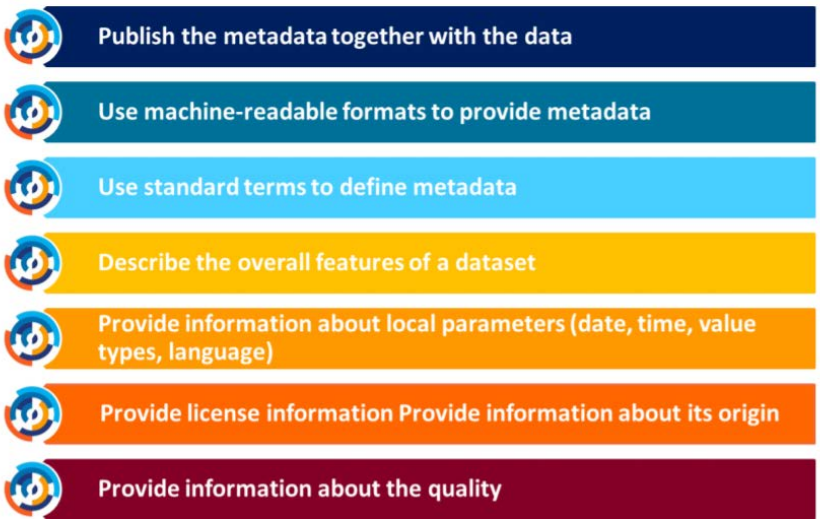

Figure 2 – Summary of Metadata Best Practices (Source DCAT-AP)

3.2 Making data openly accessible

An analysis of which SYNERGY research data will be made openly accessible and which data will be kept closed will be done in a later stage of the project. As indicated in section 2.2, the starting point is the definition of all types of research data to be handled and generated during and after the end of the project and this activity will be done in the framework of the ongoing tasks in WP2 and WP3. Once this action is completed, the data accessibility analysis will be carried out including the specification of software tools required to access the data. The outcome shall be implemented in the coming versions of the DMP. At the current stage, such an analysis would not make much sense since the data sets are not defined yet.

During the project lifetime, information on the following aspects will be elaborated for all data sets on a case-by-case base, before making consortium decision on handling of the particular data generated or collected:

- Nature and scale of the data in consideration,
- To whom it could be useful / targeted audience and its size / level of interest,

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

- Information on the existence of similar data and possible synergies,
- Possibility for integration and reuse of the provided data by external users / researchers, and
- Any further related issue

In general terms, the SYNERGY research data will be made available when possible without compromising privacy, ethical or commercial sustainability, to parties with a legitimate research interest. In the case of SYNERGY, certain research data – especially considering the demo sites – is sensitive due to security issues and therefore will be kept confidential.

3.2.1 The SYNERGY Platform

SYNERGY will deliver a flagship industrial data intelligence platform in the energy domain, acting as a trusted hub for sharing, trading, analysing and acquiring proprietary energy data and their derivative energy data intelligence. The SYNERGY integrated platform will build on the state-of-the-art, reusing and integrating into its solution a plethora of enabling technologies, tools and methods coming from open source initiatives, as well as from highly innovative R&D breakthroughs of the consortium. It will be scalable and interoperable-by-design as any energy stakeholder (esp. SMEs) and legacy system will be able to connect to its Open APIs and leverage its data-driven services bundles, infrastructures and documentation provided to embark on their data analytics-driven digital transformation journey.

The SYNERGY reference architecture to be developed in the context of Task 2.4 is open-by-design and, together with the energy Common Information Model (built on existing energy standards and models), the Open APIs and the reference demonstrators' adaptors shall be delivered as open specifications under appropriate Creative Commons licenses.

The SYNERGY Platform will facilitate the execution of the 5 large-scale demonstrators (and 21 demo cases), that are characterized by enhanced data synergies/ exchanges, real-time data flows and management of big data volumes and velocities.

Delivering this platform is the main objective of the SYNERGY project and will be realised in the context of WP3 and WP4.

The SYNERGY big data platform will be hosted in the cloud and shall be accompanied by on-premise deployments by the different electricity value chain stakeholders. Details about the hosting of the platform will be available in due time once the initial platform releases are available.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

3.2.2 Open Access publications

The rules and principles of the European Commission's Horizon 2020 Framework Programme clearly defines that scientific results generated within H2020 projects should be made available as open access publications, i.e. freely available online to any user. Following these rules, enforced in the SYNERGY Grant Agreement – Article 29.2, open access will be ensured to all peer-reviewed scientific publications related to SYNERGY and its composite solutions. Furthermore, other project outputs and results may be provided in the Open Access upon respective decision of the project consortium.

According to the "Open access to publications and data in Horizon 2020 Fact sheet" [17], two main routes exist for open access to scientific peer-reviewed publications:

- Self-archiving (also called 'Green' open access) means that the published article or the final peer-reviewed manuscript is archived by the researcher in an online repository before, after or alongside its publication. Access to the article is often – but not necessarily - delayed ('embargo period' of six months of publication) as some scientific publishers may wish to recoup their investment by selling subscriptions and charging pay-per-download view fees during an exclusivity period
- Open access publishing (also called 'Gold' open access) means that an article is immediately provided in open access mode by the scientific publisher. The associated costs are shifted away from readers, and instead charged to the research institute to which the researcher is affiliated, or to the funding agency supporting the research

Since these two routes described above are not mutually exclusive in an EU-funded action, within the SYNERGY project each beneficiary will be able to choose the most suitable approach for each publication concerned. In any case, this specific aspect will be further discussed and described in an updated version of the Data Management Plan.

For the moment, parallel publishing has been chosen as the primary strategy for providing open access. This allows consortium members to publish their results in the scientific fora and journals of their choice for maximum impact and still ensure optimal dissemination of the results by open access. It is worth to mention that the overwhelming majority of academic journals support either the gold, the green or a hybrid open access route. This means that SYNERGY beneficiaries have the freedom to publish where they feel it is the most appropriate. All publications will have a DOI making them easily findable and citable.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

It is important to note that the open access requirement does not mean an obligation to publish results. As stated by the guidelines: "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 graph below illustrates the decision related to research results and possible path for publication or other options.

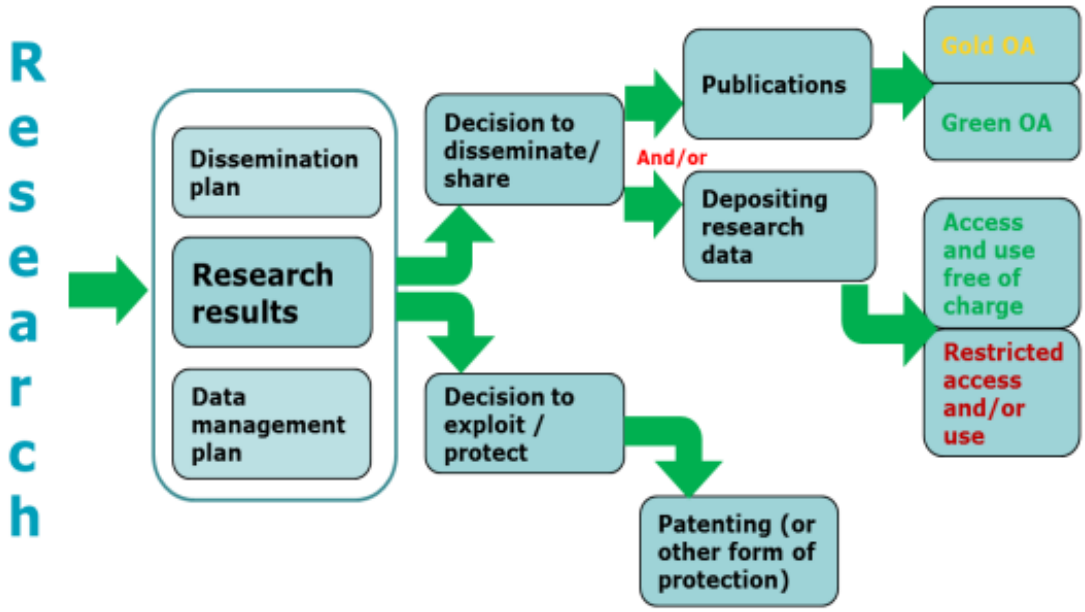

Figure 3 - Open access to scientific publication and research data in the wider context of dissemination and exploitation **[3]**.

SYENRGY will use two main tools as the platforms for accessing the project scientific results: Researchgate and Zenodo.

3.2.2.1 Open research publications: Researchgate

SYNERGY will create a space in Researchgate [18] to create exposure for the work produced and to interact with audience and peers. This space will be found under the Researchgate Projects section and will allow promoting all scientific publications of the project. All SYNERGY partners can follow the SYNERGY Project and upload periodically the material produced within the project. Each project collaborator will be able to ask questions and get feedback, as well as to share all the new research with the project followers.

Open Access documents will be published using the PDF format. Provided the type of research that has been added supports adding a DOI (also including data), the platform allows to generate one. Researchgate also supports harvesting of all content via the OAI-PMH protocol [19]. All downloads are

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

enriched by using simple metadata information like the, title, a short description and the type of the document.

ResearchGate is the professional network for scientists and researchers. It helps researchers connect and make it easy for them to share and access scientific output, knowledge, and expertise. It allows publicly recommending a research item, just sharing it privately with a select person or group, or sharing a project with people that may be interested in it.

Researchgate indicates the different versions of the publication that can be handled as well as copyright terms and conditions in https://explore.researchgate.net/display/support/Copyright.

3.2.2.2 Open research data repository: Zenodo

For implementing open access to the research data sets, SYNERGY will use Zenodo [20], an established online European scientific repository that is fully integrated with OpenAIRE [21]. The decision on whether a research data set will be uploaded to and opened for access in Zenodo will be made on a case-by-case basis between the Project Coordinator (ETRA), the Technical Coordinator (SUITE5) and the partner(s) that have ownership of the data.

When a data package study has been marked as public, it will be made openly available. Data gathered by partners outside of the project work plan and protected by IPR, or inside the work plan but containing confidential information (e.g. related to personal data), will be kept closed for privacy reasons.

Zenodo [20] offers a simple online service that enables researchers, scientists, EU projects and institutions to share, preserve and showcase multidisciplinary research results (data and publications), that are not part of the existing institutional or subject-based repositories of the research communities. It provides service hosting according to industry best practices in CERN's professional data centres. A detailed description of Zenodo's policies regarding the handling of the data and usage of the service is found at https://zenodo.org/policies

As previously indicated, SYNERGY intends to share datasets in the publicly accessible disciplinary repository Zenodo using descriptive metadata as required/provided by that repository. Zenodo assigns all publicly available uploads a Digital Object Identifier (DOI) to make the upload easily and uniquely

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

citeable. (It is NOT possible to edit a Zenodo DOI once it has been registered). Zenodo further supports harvesting of all content via the OAI-PMH protocol [19].

3.2.3 SYNERGY Website and public deliverables

The SYNERGY website describes the mission and the general approach of the project and its development status, as well as provide a short description of the project's objective and its methodology, news, events and updates that are relevant to the project's activities.

The project public deliverables will be available for download on the website after their submission and approval from the EC. Confidential deliverables will be stored in the Alfresco repository only, which is accessible for restricted users. Confidential deliverables might be requested by external parties, in which case the Consortium might make decision to disseminate corresponding deliverables or specific parts of the deliverables to particular external parties. The project deliverables on the website will be provided in the widely adopted PDF format.

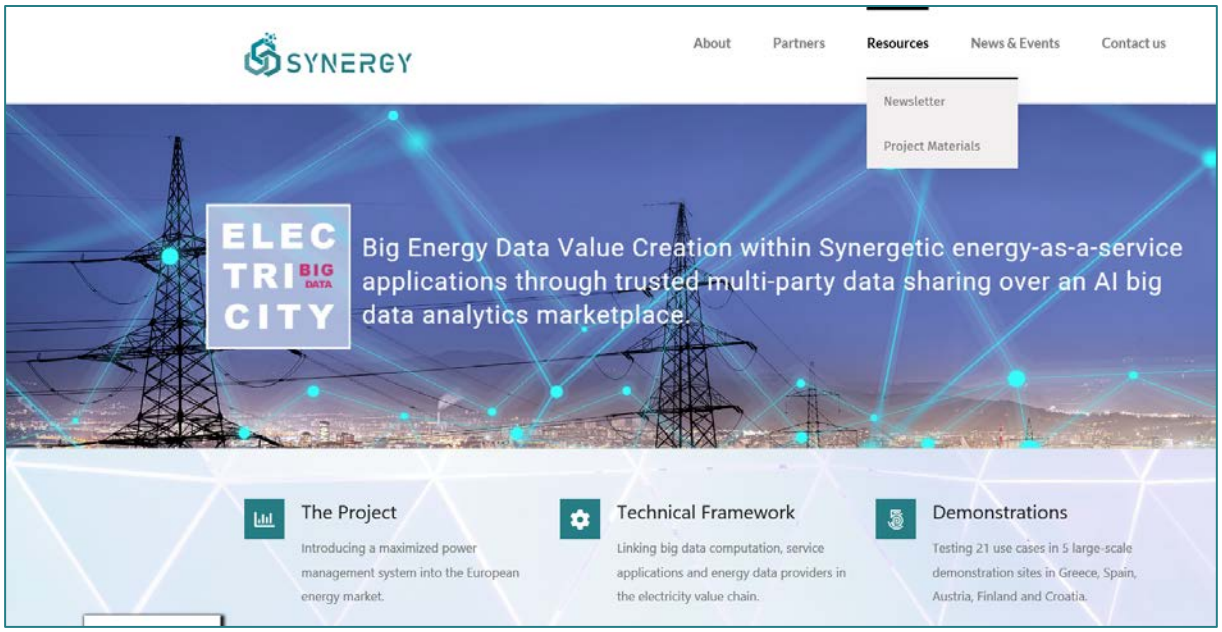

Figure 4 SYNERGY Website, Resources

3.2.4 SYNERGY internal repository

During the life cycle of SYNERGY, data collected or generated by the project will be stored and systematically organised in the official project repository on Alfresco.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Alfresco [22] is a flexible content management web application developed using Java technology. In the framework of the SYNERGY project, it is used mainly as a repository to securely store and share files, making them available to the whole Consortium. The Documents Library consists of a project internal area, that is not possible to be accessed by external users.

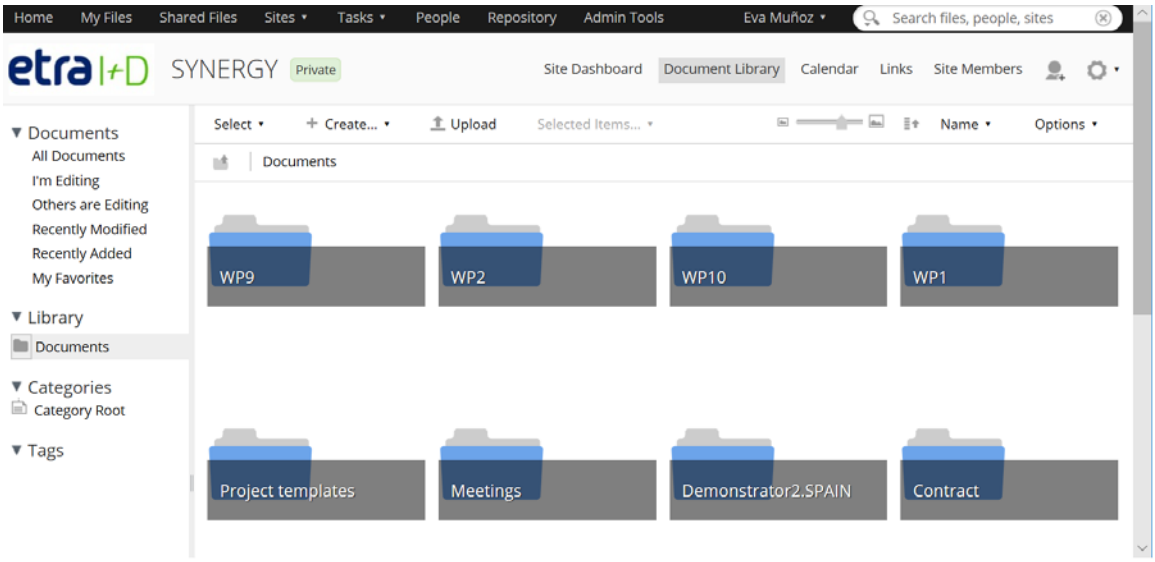

Figure 5 – SYNERGY repository overview

As it is depicted in the snapshots below, folders are organized in a hierarchical and clear structure and files are uniquely identifiable and versioned using a name convention.

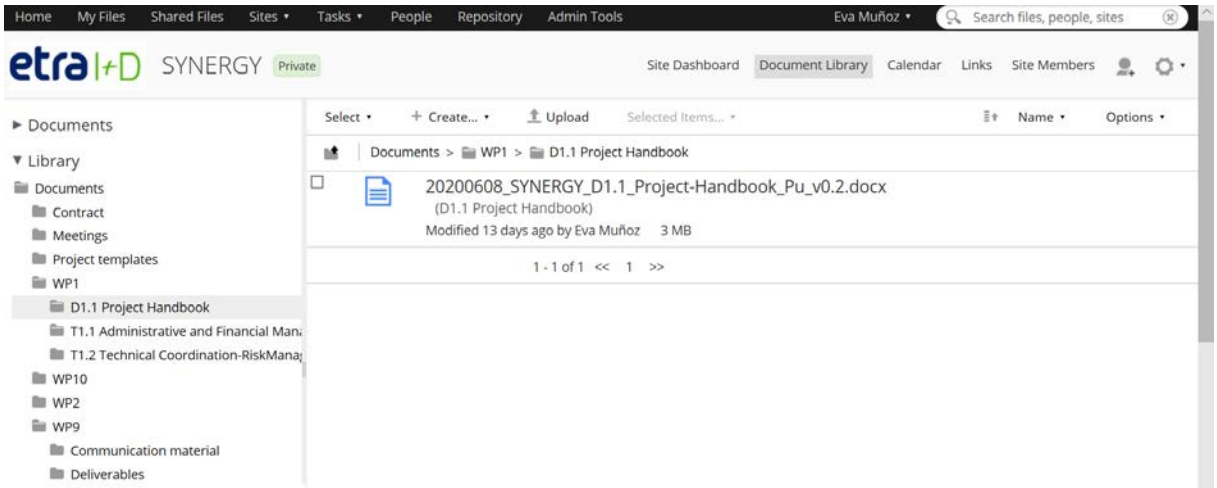

Figure 6 – Example of folders and files in SYNERGY repository

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

3.2.5 Github / Gitlab

SYNERGY open source code will be available in GitHub. All source-code components that are implemented during the project and decided to be public will be uploaded to an open access GitHub repository, linked also with the Zenodo account of the project.

GitHub [23] is an online repository which supports distributed source code development, management, and revision control. It is primarily used for source code data. It enables worldwide collaboration between developers and provides also some facilities to work on documentation and to track issues.

GitHub provides paid and free service plans. Private, non-public repositories require a paid service plan while free service plans can have any number of public, open access repositories with unlimited collaborators. Many open-source projects use GitHub to share their results for free. The platform uses metadata like contributors' nicknames, keywords, time, and data file types to structure the projects and their results. The terms of service state that no intellectual property rights are claimed by the GitHub Inc. over the provided material. For textual metadata items, English is preferred.

3.3 Making data interoperable

An assessment of the data interoperability, specifying what data and metadata vocabularies, standards or methodologies will be followed in WP3 to facilitate interoperability, will be carried out in a later stage of the project. The assessment will allow to determine whether standard vocabulary is being used for all data types present in the certain data set in order to allow inter-disciplinary interoperability. Again, for the development of this task, the starting point is the definition of all types of research data to be handled and generated during and after the end of the project and the definition of the components or actors-components that will be communicating in the scope of the SYNERGY platform as part of the SYNERGY platform architecture.

Concerning standards and interoperable data models within the project, the following actions are planned for the upcoming months as part of the work that is taking place in WP3:

- Identification of the structure/model of the data that are to be handled or exchanged between data value chain stakeholders in the scope of SYNERGY project;
- For each dataset and in general, assess the available standards, data models and ontologies;
- Evaluate the applicability of the standards and data models identified in the SYNERGY project;
- Develop the SYNERGY Common Information Model.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

The outcomes of the actions described above will be integrated in the coming versions of the DMP.

3.4 Increase data re-use (through clarifying licences)

As previously indicated, data will be treated on a case study basis during the project. Once a data set is marked as **public**, and, therefore, made publicly available on Zenodo, it will be fully reusable (with the possibility of specifying an embargo period or with controlled access to whitelist of persons; see Zenodo policies in [24].

When possible, and as recommended by European Commission's guidelines [3], data will be made available with Creative Commons Licences (CC BY or CC0). The data sets may be given different licenses according to their specificities[1].

The Zenodo repository ensures sustainable archiving of the final research data. Items deposited in Zenodo will be retained for the lifetime of the repository, which is currently the lifetime of the host laboratory CERN and has an experimental programme defined for at least the next 20 years. All publicly available uploads on Zenodo will be stored safely for the future in the same cloud infrastructure as research data from CERN's Large Hadron Collider and using CERN's battle-tested repository software INVENIO, which is used by some of the world's largest repositories such as INSPIRE HEP and CERN Document Server.

The data will remain re-usable at least until Zenodo discontinues the dataset(s) (i.e. warrantied for a minimum of 20 years).

The project envisages adopting the "data pedigree" concept, which assure that each piece of relevant information is traceable back to the original data sources. This data lineage along with metadata allows for quality audit and sensitivity analyses of the outputs.

Data re-use be also ensured for **confidential and public datasets** through the SYNERGY data marketplace. SYNERGY is designed to address a wide variety of data in motion and at rest that will be ingested, curated, linked, safeguarded, analysed (in 3 different modalities: edge analytics and secure multi-party computations in the on-premise environment, and big data analytics in the secure experimentation playgrounds in the SYNERGY cloud infrastructure) and shared in a trusted, controlled manner.

[1] The EUDAT B2SHARE tool includes a built-in license wizard that facilitates the selection of an adequate license for research data (https://eudat.eu/services/userdoc/license-selector)

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

4 Responsibilities and decision making

As indicated in previous sections, the Data Management Plan presented in this deliverable is just the first version, and the related Consortium discussions will be continuously carried out, to identify the relevant project outputs as well as to decide on way and means of their open access (if applicable). To ensure it, a dedicated time slot will be reserved at each project plenary meeting and, if needed, at selected Consortium audio conferences. The EC and the project reviewers will be informed about related work done and publications provided in the project management reports.

Individual responsibilities on data management in the project consortium are:

- Project Coordinator (ETRA) – to prepare and lead related discussions at the relevant project meetings and to maintain the project document repository Alfresco;
- Technical Coordinator (SUITE5) – to identify data collected by the project and technical project outcomes eventually suitable for publication; moreover, to ensure dataset integrity and compatibility for its use during the project lifetime by different partners.
- Dissemination Manager (GECO) – to identify publications suitable for publication in the considered repositories and maintain SYNERGY inputs for the Open Access;
- Each individual partner – to identify own project results suitable for publication and to share the published scientific articles in advance with project coordinator and dissemination manager.

The Project Coordinator and the Dissemination Manager have a particular responsibility to ensure that data shared through the SYNERGY website are easily available, but also that backups are performed and that proprietary data are secured.

Moreover, each SYNERGY partner has to respect the policies set out in this DMP. Data sets have to be created, managed and stored appropriately and in line with applicable legislation. Validation and registration of datasets and metadata is the responsibility of the partner that generates the data in the Work Package (WP). Metadata constitute an underlying definition or description of the datasets, and facilitate finding and working with particular instances of data. Additional responsibilities undertaken by the SYNERGY project partners include:

- Backing up data assets for sharing through Open Access repositories. It is the responsibility of the partner possessing these data assets.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

- Quality control of the data assets. It is the responsibility of the demo partner providing the data.
- Managing different versions in case the data assets are updated, and making sure that the latest version is available in the case of publicly available data.
- Consulting the concerned partner(s) before publishing data in the open domain that can be associated with an exploitable result. It is the responsibility of all project partners involved in this activity.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

5 Data security

The Zenodo and Alfresco repositories will ensure secure and safe storage of both public and non-public data respectively.

Zenodo provides clear security guaranties. All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files and metadata are backed up on a nightly basis. Files are regularly checked against their checksums (using MD5 algorithm) to assure that file content remains constant. In case of closure of the repository, Zenodo ensures that efforts will be made to integrate all content into suitable alternatives [24].

Alfresco is hosted on a private internal server with local backup mechanism (managed by the project coordinator ETRA).

The servers hosting the research data will be accessible only by authorized system administrators. Files containing confidential data should be protected by owners using local encryption tools (i.e. password-protected archives) before being uploaded to shared repositories. Interaction through web user interfaces will use https protocol (i.e. secure). Also, a secure file transfer protocol (sftp) will be provided as the need arises.

To assure data security and privacy, the SYNERGY platform will support advanced data anonymization and encryption mechanisms, as well as cloud and on-premise storage on servers to which only the relevant staff have access. More specifically the servers onto which the data will be stored will have server side encryption. This means that the server's administration personnel will be able to generate public keys for specific personnel who will have access to the data but will not be able to access the data themselves (since the private keys required for this access will be generated on the machine of the person with access to the data). This means that only the required personnel (and stakeholders that have active data contracts) will have access to the data and, even in the remote case of a possible data leak or server hack, the data stolen will be fully encrypted and thus virtually fully non-accessible.

Finally, and after a retention period (to be defined), a secure deletion software will be used to destroy data, i.e. using Gutman algorithm (35-pass overwrite technique).

If deemed necessary, a full format can be used in conjunction with overwriting, to provide further assurance that data cannot be recovered, guaranteeing the destruction of the project personal data.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

The following guidelines will be used in order to ensure the security of the data:

- use anonymised and aggregated data instead of individual data;
- encrypt data by the local researchers and not allowing the data to leave their premises unencrypted;;
- store data in at least two separate locations to avoid loss of data;
- limit the use of USB flash drives;
- label files in a systematically structured way in order to ensure the coherence of the final dataset.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

6 Compliance with GDPR

The current section undertakes the initial approach towards safeguarding compliance with GDPR [25] principles. As indicated along the whole document, this is the first version of the SYNERGY Data Management Plan, and thus many issues are still open and will be resolved as the project progresses. The updates will be documented in the future versions of the current document.

In order to check compliance with GDPR, the following items must be identified:

- the data subjects for which information will be collected
- the personally identifiable information per data subject
- the roles of the stakeholders associated with data handling activities
- the processing activities associated with the personally identifiable information per data subject and per stakeholder role.

6.1 Data subjects

According to the GDPR, a Data Subject is a physical person or a group of physical persons for which information is gathered. Towards identifying the data subjects, the data sources that will be used for the data collection have to be examined. Since this is work in progress in the T2.3, Data Subjects will be fully described in future versions of the DMP.

6.2 Personally Identifiable Information

Based upon the data subjects that can be identified, the next step will be to assess the personally identifiable information per data subject.

As a general principle, any user-related data that may lead to single user identification will not be made available to the consortium, and should not be stored on the SYNERGY platform. On the contrary, these data assets, (which are either proprietary or confidential and not public), will be anonymised anonymised and encrypted by the data owners prior to being stored on the SYNERGY platform.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

6.3 Roles

The current section undertakes the analysis of the legal entities involved in the collection, access and processing of the Personally Identifiable Information included in the previous section, according to Article 4 of the EU GDPR, along with their mapping to the SYNERGY project partners and external stakeholders.

Thus, according to Article 4 of the EU GDPR, we can identify three main legal entities: 1) the Data Controllers, 2) the Data Processors and 3) the Data Recipients.

Data Controller

Controller – "means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data".

In SYNERGY, the Data Controllers are actually the owners of the Data Assets that will be made available to the consortium, either in a public, in a proprietary or a confidential manner. They provide the information, either generated or collected internally, or collected and/or compiled from external stakeholders that will be managed and analysed in the context of the project. These organizations are the demo partners that are part of any of the five SYNERGY demonstrator countries: HEDNO, FVH, CUE, COBRA, URB, CAV, IPTO, EPA, KRK, VERD, GUS, ENES.

Data Processor

Processor – "means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller".

In SYNERGY, the Data Processors are the technical partners who are developing the SYNERGY Platform, who provide the technical infrastructure and the storage facilities for the data handling and analysis, and for the data storage activities respectively, and who design and implement the data access, security and privacy policies. These partners are those that participate in technical WPs: WP3, WP4, WP5, WP6 and WP7: UBI, ETRA, HEDNO, ICCS, FVH, VTT, CUE, COBRA, S5, CIRCE, CAV, IPTO, EPA, KRK, MAG, KBZ, KONCAR, UCY, VERD, ENES.

Data Recipient

Recipient – "means a natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not".

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

In SYNERGY, the Data Recipients are the (external) stakeholders to which the raw data assets along with the outcomes of the data analytics are transferred by the Data Controller or are made available by the Data Processors to be used by means of the different Energy Applications developed on top of the SYNERGY Platform. Such stakeholders include all the actors of the electricity value chain and become data recipients once data contracts between the involved stakeholders have been signed and enforced in the SYNERGY platform.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

7 Allocation of resources

As this preliminary DMP is currently based on the use of free resources and open source software, the only costs that will be incurred are related to the server(s) (hardware) required to run them and the working time needed to setup, maintain and evolve the different tools (efforts measured by person-months).

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

8 Ethical aspects

Ethical issues are covered in WP11 and also reported in D1.3 Ethics Monitoring Report. In this report, the procedure for handling SYNERGY ethical issues is presented. The SYNERGY consortium has to comply with all European and national legislation and directives relevant to the country where the data collection is taking place. The collection, processing and transmission of personal data will be analysed under principles of (a) The Universal Declaration of Human Rights and the Convention 108 for the Protection of Individuals with Regard to Automatic Processing of Personal Data, (b) The **General Data Protection Regulation (GDPR)** (Regulation (EU) 2016/679) [25], and (c) The national laws applying its provisions. Any additional regulations at national level that do not fall under the GDPR and apply to data protection or any other sensitive information will also be taken into account.

Data managed during the project will be processed only under the following preconditions which need to be met: (a) When the data subject has given her/his consent; (b) When the processing is necessary for the performance of or the entering into a contract; (c) When processing is necessary for compliance with a legal obligation; and (d) When processing is necessary in order to protect the vital interests of the data subject.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

9 Conclusions

This document sets the guidelines and recommendations to be followed in order to make the project research data Findable, Accessible, Interoperable and Reusable (FAIR) and therefore contribute to knowledge discovery and innovation.

The main elements of the data management policy that is used and will be used by the SYNERGY project are analysed and studied in this deliverable.

As it has been stated in the document, the Data Management Plan is a living document that needs to be updated all along the project implementation, in order to cover all changes or progresses that might occur during the project lifetime.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

10 References

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

[2] European Commission, "TEMPLATE HORIZON 2020 DATA MANAGEMENT PLAN (DMP)," [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx.

[3] EC, "H2020 Programme Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020," [Online]. Available: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

[4] SYNERGY Project, "Grant Agreement number: 872734 — SYNERGY," 2019.

[5] ICARUS Consortium, "ICARUS H2020 project," [Online]. Available: https://icarus2020.eu/.

[6] BIMERR Consortium, "BIMERR H2020 project," [Online]. Available: https://bimerr.eu/.

[7] NOBELGRID Consortium, [Online]. Available: www.nobelgrid.eu.

[8] WiseGrid Consortium, "WiseGrid H2020 project," [Online]. Available: https://www.wisegrid.eu/.

[9] GRIDSOL Consortium, "GRIDSOL H2020 project," [Online]. Available: https://www.gridsolproject.eu/.

[10] CORDIS, "FLEXICIENCY H2020 project," [Online]. Available: https://cordis.europa.eu/project/id/646482/es.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

[11] "Research Data Management plans. Elseiver.," [Online]. Available: https://www.publishingcampus.elsevier.com/websites/elsevier_publishingcampus/files/Guides/Research%20Data%20Management%20Plan.pdf.

[12] K. Russell, "FAIR data principles and Data Management Plans," 31 October 2017. [Online]. Available: https://www.slideshare.net/AustralianNationalDataService/fair-data-principles-and-data-management-plans-31-oct-2017.

[13] SYNERGY Consortium, "D1.1 Project Handbook," June 2020.

[14] European Data Portal,, "Open Data Goldbook for Data Managers and Data Holders,," [Online]. Available: https://www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf/..

[15] "Dublin Core Metadata Initiative," [Online]. Available: http://dublincore.org/metadata-basics/.

[16] Dublin Core Metadata Element Set, [Online]. Available: http://dublincore.org/documents/dces/.

[17] "Fact sheet open access to publications and data in Horizon 2020 - Fact sheet," [Online]. Available: https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/Open_Access_in_H2020.pdf.

[18] Researchgate, [Online]. Available: https://www.researchgate.net/. [Accessed June 2020].

[19] Open Archives Initiative, "Protocol for Metadata Harvesting - Tools," [Online]. Available: http://www.openarchives.org/pmh/tools/. [Accessed June 2020].

[20] Zenodo - Data Repository, [Online]. Available: https://zenodo.org/.

[21] [Online]. Available: https://www.openaire.eu/.

[22] ALFRESCO, [Online]. Available: https://www.alfresco.com/.

[23] GitHub, [Online]. Available: https://github.com/. [Accessed June 2020].

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

[24] "Zenodo policies," [Online]. Available: https://www.zenodo.org/policies.

[25] European Union, *REGULATION (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data,* 2016.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

ANNEXES

I. ANNEX A

This table provides a summary of the Data Management Plan (DMP) issues to be addressed, as outlined in the Horizon 2020 Data Management Plan Template [3].

DMP component	Issues to be addressed
1. Data summary	• 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
2. FAIR Data **2.1. Making data findable, including provisions for metadata**	• 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 • Outline the approach towards search keyword • Outline the approach for clear versioning • Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
2.2 Making data openly accessible	• 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
2.3. Making data interoperable	• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

	• 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?
2.4. Increase data re-use (through clarifying licences)	• 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
3. Allocation of resources	• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs • Clearly identify responsibilities for data management in your project • Describe costs and potential value of long-term preservation
4. Data security	• Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	• To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Table 4 - Data Management Plan Scope Summary

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.