Hestia
Holistic demand response services for European residential communities
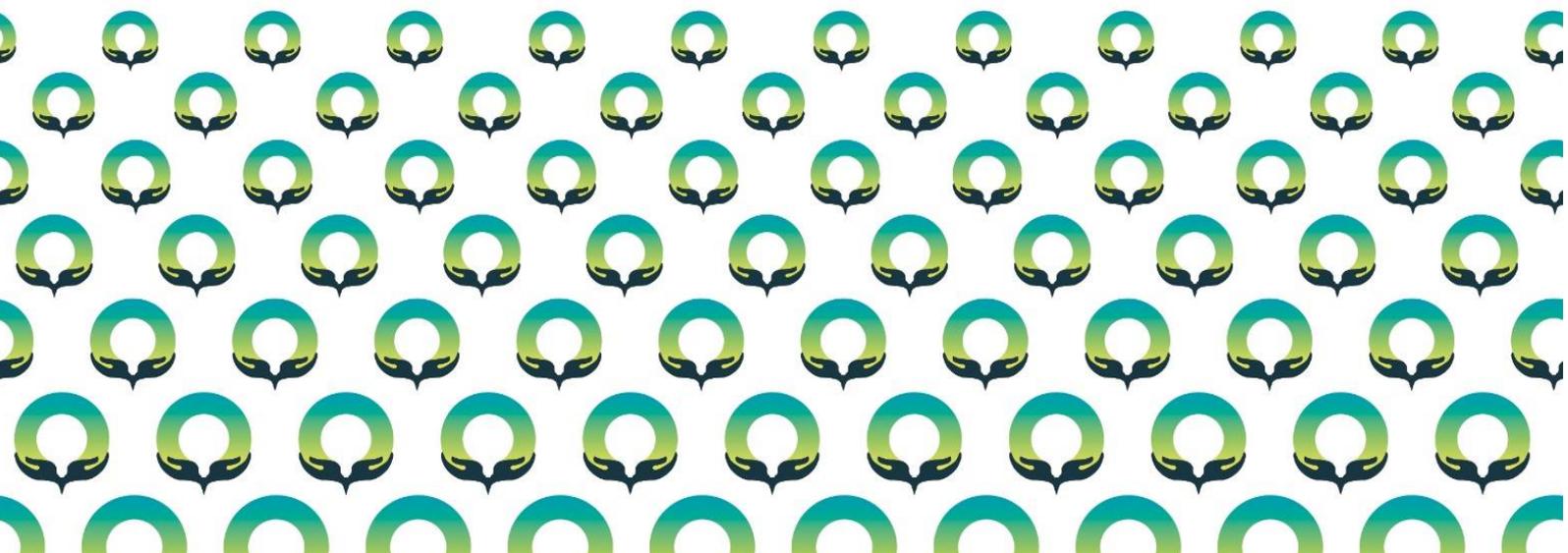
Project Deliverable

D9.3
Data Management Plan

Project Acronym	HESTIA
Project Title	Holistic dEmand response Services for European residenTIAl communities
Project Coordinator	Andrea Martinez SINLOC Andrea.martinez@sinloc.com
Project Duration	01 November 2020 – 31 October 2023
Deliverable No.	D9.3: DATA MANAGEMENT PLAN
Dissemination level*	PUBLIC
Work Package	WP 9 – PROJECT MANAGEMENT
Task	Task 9.4: Data and knowledge management
Lead Partner	CLDI
Contributing beneficiary/ies	ALL
Due date of deliverable	30 April 2021
Actual submission date	27 April 2021

Nature: R = Report, P = Prototype, D = Demonstrator, O = Other

Dissemination level   PU = Public

        PP = Restricted to other programme participants (including the Commission Services)

        RE = Restricted to a group specified by the consortium (including the Commission Services)

        CO = Confidential, only for members of the consortium (including the Commission Services)

DOCUMENT HISTORY

Version	Date	Authors	Description
DMP_v1.0	15.02.2021	CLDI_CMoscardi	First deliverable draft, basic document structure.
DMP_v1.1	22.02.2021	CLDI_CMoscardi	New section incorporated
DMP_v1.2	26.02.2021	CLDI_CMoscardi	Overlapping overview with D10.1
DMP_v1.3	4.03.2021	CLDI_CMoscardi	Added 1st Template for datasets
DMP_v1.4	19.03.2021	CLDI_J.Puate	Added new formats and content
DMP_v1.5	24.03.2021	CLDI_J.Puate	Evaluation of Section 7
DMP_v1.6	01.04.2021	CLDI_J.Puate	Addition of content after internal meeting
DMP_v1.7	13.04.2021	CLDI_J.Puate	Addition of information regarding FAIR and Security sections
DMP_v1.8	21.04.2021	CLDI_J.Puate	Final Draft shared with Internal reviewer
DMP_Final	27.04.2021	CLDI_C.Moscardi	Final Version

AUTHORS

Full Name	Beneficiary Partner	Email	Role
Joan Puaté	CLDI	Joan.puate@clusterdigital.cat	Overall Editor
Christian Moscardi	CLDI	christian.moscardi@clusterdigital.cat	Overall Editor
Laura Perez	R2M	laura.perez@r2msolution.com	Reviewer

Table of Contents

1 Executive Summary

This document is the deliverable **D9.3: Data Management Plan** of the EU-funded H2020 project European project HESTIA - *Holistic dEmand response Services for European residenTIaI communities*.

The purpose of the document is to define the framework and guidelines to be followed by the consortium as a whole and by each individual partner in relation to how data collected, processed and/or generated during the project is managed, both throughout the lifecycle of HESTIA and after the end of the project.

This plan will ensure that HESTIA project's data are managed suitably, in line with the requirements set by the EC in the "Guidelines on FAIR Data Management in Horizon 2020, and also according to the terms and conditions established in the Grant Agreement and its Annexes, as well as on the Consortium Agreement specifications and requirements.

The target audience is consortium members, EU officials and project reviewers.

2 Introduction

In HESTIA, several types of data are going to be collected. This includes data related to the energy consumption of residents, which is extremely sensitive due to privacy considerations; the rationale behind is also addressing current associated problems of the value of data and new exploitation possibilities. This project is an integration project and therefore several consortium members are involved in the data life cycle, the elaboration of the Data Management Plan (DMP) and participation in the '**Pilot on Open Research Data in Horizon 2020'**[1], which comprises defining and managing the following tasks and responsibilities:

(i)   the handling of research data during & after the project;

(ii)   what data will be collected, processed, or generated;

(iii)   what methodology & standards will be applied;

(iv)   whether data will be shared /made open access/; and

(v)   how data will be curated and preserved.

The DMP is implemented across other WPs, based on data delivered. Additionally, the **Open Research Data Pilot (ORDP) of the European Commission** aims to improve and maximise open access and reuse of research data generated by Horizon 2020 projects. There are two main pillars to the Pilot: developing a Data Management Plan (DMP) and providing open access to research data. In order to adhere to Article 29.3 of the Grant Agreement which includes the requirements to comply with the ORD Pilot (except in those cases where it goes against the legitimate interests of the project partners), the research consortia under H2020 will therefor:

- Develop (and keep up to date) a Data Management Plan (DMP).
- Deposit data in a research data repository.
- Ensure third parties can freely access, mine, exploit, reproduce and disseminate your data.
- Provide related information and identify (or provide) the tools needed to use the raw data to validate the research results.

The ORDP applies as well to:

- The data (and metadata) needed to validate results in scientific publications, and associated metadata (i.e., data describing the deposited research data) and
- Other curated and/or raw data (and metadata) as specified by the beneficiaries themselves in their data management plan.

---

[1] https://www.openaire.eu/what-is-the-open-research-data-pilot

An overview of the use of research results is shown in Figure 1. Either the results are exploited which means will be protected by IPR e.g., patents, or they are disseminated. The dissemination can be realized in two ways: first as publication in scientific journals or on conferences, second, the research data can be provided in a data repository. For both dissemination opportunities the EC follows the Open Access principle:
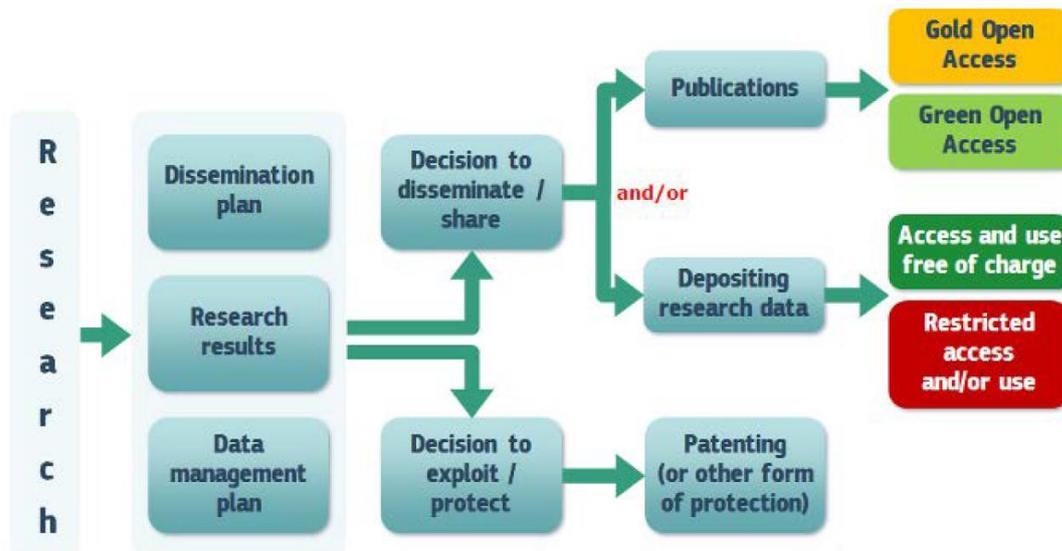

Figure 1. Use of HESTIA research results.

The ORD pilot follows the principle "as open as possible, as closed as necessary", considering the need to balance openness and protection of scientific information. In accordance with this, research data linked to HESTIA's exploitable results will not be put into the open domain if they compromise its commercialisation prospects or have inadequate protection, which is an H2020 obligation. The rest of the research data will be deposited in an open access repository.

3 Scope of this Deliverable

The DMP is a component of WP9 Task 9.4 and, as the leader of this Task, Cluster Digital de Catalunya (CLDI) has drafted and elaborated this document with input from the other partners. All WPs directly contribute to the implementation of the DMP according to their relevant activities. This DMP therefore will describe how research data is managed both throughout the lifecycle of HESTIA and after the end of the project.

It will identify procedures as well and the minimum requirements to collect, store, analyse, and publish data in a consistent way according to the FAIR principles and the H2020 Open Research Data Pilot.

The DMP is, therefore, a living document that will serve as a data management manual for the project activities, considering this a horizontal management activity. It will be regularly updated throughout the project's lifespan according to the timeline specified below. All project partners will be informed of the changes made to this document and, similarly, all partners will be contributing to its development by informing CLDI as relevant information becomes available.

In this regard, a DMP updated version will be released every six (6) months after the scheduled meetings of the Steering Committee, that is to say:

Version	Month	Version	Month
DMP 1st release (D9.4)	M6	DMP 4th release	M24
DMP 2nd release	M12	DMP 5th release	M30
DMP 3rd release	M18	DMP 6th and Final Release	M36

4 Data Summary

4.1 What is the purpose of the data collection/generation and its relation to the objectives of the project?

In general terms, all the information generated within HESTIA's framework will be classified as *Public* -public access- or *Confidential* -only for members of the consortium, including the Commission Services-.

In terms of the type of information, HESTIA will be generating the following:

- On the one hand, private information, that is to say, the one related to the project's internal management, needed to guarantee the achievement of its objectives. Each partner will manage its own private information and share it with other partners whenever this is required by the good accomplishment of the project.
- Secondly, and as a result of the project's development, a series of deliverables will be generated and constitute the project conclusions in its various aspects. Some

deliverables will be subject to public dissemination and some of them are confidential -only for consortium members use, as set out within HESTIA's Grant Agreement (GA).

- Thirdly, the project, through the GA has committed to releasing a minimum number of scientific publications. In these cases, intellectual property policy set out in the Grant Agreement and Consortium Agreement will be strictly observed.

- Finally, technical datasets will be generated throughout the project lifespan, which will constitute the basics of the analytical information on which results, and conclusions of the project will be based. In the framework of HESTIA project, these data will mostly be related to the demonstration activities performed with regard to the communities of households and dwellings that will be monitored. Among others:

  - household/individuals master data: socio-economic characteristics and education,
  - behaviour patterns in terms of energy-saving practices, how frequently user interact with the application, etc.,
  - energy consumption data at dwelling, building or community level (in an aggregated manner),
  - energy production data and other energy assets (e.g., status of storage units) related data,
  - data originating from other community-owned technical assets (common heat pumps, chillers, AHUs, etc.).

4.2 What types and formats of data will the project generate / collect? Will you re-use any existing data and how?

At HESTIA's present stage, there is not enough information for the project partners to precisely describe the datasets involved that will be generated along the project development. In this regard, a first standard template has been shared as a first exercise between partners so minimum categories and information is starting to get defined and gathered. These templates will be properly updated in the different DMP versions as the project advances.

4.3 Dataset DMP Templates

In the upcoming planned updates of this document, this section will include and describe the set of datasets generated throughout the HESTIA's development phases. This will be done by following the first template shown in section 5.1 further below in this document, as well as by creating any additional ones specifically required.

5 Fair Data

The DMP will function as a guiding document to ensure good data management throughout the lifecycle of the project in order to make the information collected and produced Findable, Accessible, Interoperable and Re-usable (FAIR).

F	*Findable* - Discoverable with metadata, identifiable and locatable by means of a standard identification mechanism.
A	*Accessible* - Always available and retrievable online; even if the data is restricted, the metadata is open.
I	*Interoperable* - Both syntactically parseable and semantically understandable, allowing data exchange and reuse between researchers, organisations, or countries.
R	*Re-usable* - Sufficiently described and shared with the least restrictive licences, allowing the widest reuse possible and the least cumbersome integration with other data sources.

Table 5.1. FAIR principles[2].

In this regard, the current version of this document, as well as the successive updates, reflects the current state of concretion in terms of results, publications and datasets and their management. Therefore, in this first version, the document reflects concepts and general strategy regarding the different aspects that make up the Fair Data paradigm.

In this regard, HESTIA adheres to the ORD Pilot launched by the European Commission along with the H2020 program. The members of the consortium embrace the concepts and principles of open science and acknowledge the benefits of reusing and evaluating already produced data for promoting and supporting research and innovation projects at the European level.

Thus, HESTIA will aim to make, all data generated along the project activities (results, publications and datasets), both FAIR and as Open as possible, provided that:

- confidentiality obligations set out in GA and CA are fully observed.
- it is labelled as Open in the framework of HESTIA project.

---

[2] https://www.openaire.eu/what-is-fair-data

- all personal data gets properly anonymized in compliance with the current European and local legislation.

Under these conditions, HESTIA will ensure that by the end of the project all data and information generated throughout the project development will be stored on **ZENODO³** for long term open access -considering embargo periods, if any-, a catch-all repository for EC funded research supported by the OpenAIRE project (Open Access Infrastructure for Research in Europe) that provides the appropriate architecture, procedures and security required for an openly and free access to research data among the research community.

5.1 Making data findable, including provisions for metadata

The Consortium will ensure that all data generated in the frame of HESTIA and according to its categorization -public, confidential- will be findable, that is, searchable and discoverable online. In the case of scientific or industrial publications and datasets, HESTIA will as well ensure a Digital Object Identifier (DOI) is provided, on the one hand, and a set of metadata, on the other.

A set of metadata have been proposed in a common approach for all the publicly available elements -deliverables, publications, datasets- generated. These fields are set out in the table below and their content will describe all elements generated throughout the project evolution and subsequently reported in next DMP versions:

Field	Description
WP and Task	WP1...WP10 Task
Dataset Reference/Name	[WPx] [title] [year] [month] [verx] [.] ext
Dataset Description	Dataset description
Standard used	CIM; Open ADR; etc.
Availability/dissemination level	Private; Consortium; Open
Mandatory Metadata (as per GA)	European Union (EU) Horizon 2020 HESTIA

---

³ ZENODO – online Open Access repository. https://zenodo.org/

	GA957823
	Publication Date
	Length of Embargo Period -if any-
	Persistent Identifier -if applicable-
Specific Metadata	Set of keywords in order to better identify the information
Data Owner	Data Owner
Origin of Data	Origin_of_Data
Format	Data format (.pdf, .doc, .xls, etc.)
	Tools or specific software (if needed)
Expected size	Dataset expected size
Utility	To whom might it be useful, or parts of the datasets will be shared for verification or reuse
Archiving & Preservation	Where will it be stored

5.2 Making data openly accessible

Along the project's execution, most produced data will be openly available as default. Should certain information not be shareable, i.e., because of legal reasons, these will be reported. The datasets will be made available to third parties as soon as they are generated, prepared and reviewed for publication/exploitation and when the conditions of dissemination are decided, and possible protections issues are clarified within the consortium.

However, additional restrictions by setting appropriate embargo periods and/or respecting restrictions from editors of scientific journals and organizers of conferences are also possible. A generally valid statement regarding the embargo periods is not possible at the moment. It can differ from case to case.

In this regard, different approaches will be observed depending on where data is disclosed. Indeed, results, documents, deliverables, datasets, etc., will be managed on several domains, each of which with the appropriate level of privacy and protection for the affected data:

- In the framework of the development of the various WPs, data will be stored on local servers or in cloud environments depending on each partner-specific policy.

- Secondly, the consortium has enabled a SharePoint account, Microsoft Office web-based collaborative platform, with all guarantees offered by this provider. The aim of this environment is to act as a repository of common access for all partners of HESTIA, i.e., work information (data sets, etc.) in the framework of the work packages development and the corresponding deliverables.

- The project website constitutes the first open and public domain. Indeed, in this environment, a space has been set up for project public deliverables, project scientific publications and communication materials. These will be incorporated in accordance with the consortium's strategy that will be defined for this purpose and respecting the provisions of the Consortium Agreement (CA).

- Similarly, scientific publications may be also published -in accordance with the provisions of the CA- on the *Open Research Europe initiative platform*, recently developed by the European Commission and now open for accepting submissions in all fields of research. The publishing platform is accessible through the link https://open-research-europe.ec.europa.eu/ and can count on the following features:

  o   no cost for beneficiaries.
  o   quick publication times, while applying rigorous and open peer review process.
  o   data deposition and sharing are supported by the platform, with no editorial barriers.

- Finally, assuming the requirements of the ORD Pilot, the consortium will publish the information of interest on ZENODO repository, thus allowing its access in the terms and procedures established therein. In this sense, each Data Set or document stored on ZENODO will be assigned a DOI both at the concept level and for each individual version. All data on ZENODO will be fully findable thanks to the included metadata and the search facility available on it.

Most files will be accessible with general use software (Word, Excel, PowerPoint, pdf and their open versions, etc.). Datasets which will be defined in future versions of the DMP will be accessible according to the standard language or format selected by project partner handling the data within their activities. Next versions of the DMP will be more specific for other types of data by specifying format and specific needed software -if any- to make it accessible. In relation to data and documents generated for internal use and communication within the

project, they will be stored within the SharePoint hosted in the project coordinator environment.

The rest of the document categorized as public or accessible according to the Grant Agreement will be made available within the project web portal and listed repositories**.**

As previously mentioned, preference should be given to certified repositories that support open access where possible. In the particular case of scientific and industrial publications, they will be published in journals in gold access with Impact Factor or in self-archiving green access within listed repositories.

5.3 Making data interoperable

At this stage of the project, the information related to data collection, organization and management is still under analysis for further development in order to detail aspects such as:

- standards for formats
- software applications used
- metadata vocabularies or methodologies followed
- standard vocabularies for all data types

Nonetheless, the consortium will ensure the highest level of data interoperability without compromising the quality of the project outcomes. In this regard, in the framework of the project data management, the use of open software and standard vocabularies will always be strongly recommended as well as avoiding project uncommon ontologies.

To this end, specific information on this information will be required as metadata in the different datasets whose description will be provided in future versions of this document as the project evolves.

5.4 Increase data re-use

HESTIA will aim to increase its data re-usability by ensuring the project's research data is well-documented (to help ensure that the data can be correctly interpreted and analysed by others) and has clear licence and provenance information (to govern the terms of the data reuse). As stated before, all public information generated in HESTIA will be uploaded on ZENODO repository. This will ensure open access to the information in a long-term sustainable and safe environment. Indeed, data stored on ZENODO is, as a matter of fact, stored in CERN Data Centre. In this regard, ZENODO guarantees in its website CERNs commitment "*to maintain this data centre over the next 20 years*" and that "*In the highly unlikely event that ZENODO will have to*

close operations", all content would be migrated "*to other suitable repositories, and since all uploads have DOIs, all citations and links to ZENODO resources will not be affected*".

Data published inside HESTIA may be reused by people related to the many areas covered by the project: social, economic, political, environmental, technological. Data will be treated and kept accessible only to project partner team members until the setup, calculations, audit, revision, and other checking phases (included checking any pending third-party intellectual property rights) are completed and disclosure is authorized by the coordinator.

As for the quality control of data, if datasets are updated, the partner that possesses the data has the responsibility to manage the different versions and to make sure that the latest version is available in the case of publicly available data. Quality control of the data is the responsibility of the relevant responsible partner generating the data.

As for the license policy to be applied to the disclosed information, a decision in this regard will be taken in later stages. In this regard, the use of Creative-Commons-Licences (CC) will be discussed within the consortium and reported in upcoming versions of this document.

6 Allocation of resources

Scientific publications produced during the project will be published in journals in gold access with Impact Factor or in self-archiving green access with repositories listed in ZENODO, which will be the main open access repository selected to make data FAIR. However, some papers will be selected to be published under the 'golden' option. For open access publishing, only journals with impact factor will be considered. As ZENODO will be used for long-term preservation, no related costs are foreseen in the project budget.

Also, the *Open Research Europe initiative platform*, a platform recently developed by the European Commission has no costs associated.

Finally, as to *Other Direct Cost' Items* (OGS: Other Goods & Services) declared by partners in the GA, the following budget has been allocated:

- AIT indicates allocation of 3.500€ for Open Access publication fees.
- EDF indicates allocation of 15.000€ for infra costs (servers & data management).
- CIT (MTU) indicates allocation of 2.000 € for 2 publications.
- IMP indicates allocation of 2.000 € for 2 publications.
- CLDI indicates allocation of 2.000€ for Website Hosting & Design.

7 Data Security Aspects

Research data generated, processed, and collected in HESTIA is stored on computers, clusters and servers at each project partner's premises. The facilities are hosted by the partners themselves and are secured according to actual security guidelines. The data are placed in a back-up storage to be able to be restored in case of emergencies. The storage duration is usually 10 years according to the funding rules.

Proper security measures to deal with possible communication threats that HESTIA may encounter in relation to data acquisition, transmission, and access will be put into practice. To accomplish this, complete end-to-end data security measures will be specified and applied. These will be fully described by partner R2M within the activities to be carried in task T5.5 which starts on M10 and final conclusions will be reported on its corresponding Deliverable D5.6 "Data protection and security" at M30.

General project data and documentation is also stored in the project SharePoint. Furthermore, public documents are also provided on the project webpage.

8 Ethical Aspects

8.1 Ethics

The research and development in this proposal involves human participants (community residents) and empirical data related to them. The HESTIA consortium is fully aware that the project's activities may have the potential to generate ethical, fundamental rights, privacy and data protection implications and is fully committed to adhering to the highest ethical, fundamental rights and legal standards, as recognized at the European Union and International level.

All the relevant aspects and concerns have already been presented in the ethics chapter in the Description of the Action (DoA), and in the case of personal data being transferred from the EU to a non-EU country or international organisation, HESTIA's deliverable D10.1 will set out all provisions to guarantee the accordance with Chapter V of the General Data Protection Regulation 2016/679 and full compliance of the GDPR.

9 Literature Resources & Useful Links

- European Commission Open Data & Data Management Information and Guidelines - 2020 – https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm
- Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 - Version 3.2 – 21 March 2017 – https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- The Open Research Europe Initiative Platform, European Commission, 2020 - https://open-research-europe.ec.europa.eu
- ZENODO The Open-Access Data Repository - https://ZENODO.org
- OPEN AIRE Guidelines for Literature Repositories, Data Archives, and Horizon 2020 Open Access requirements: https://guidelines.openaire.eu/en/latest
- Using Identifiers for Open Access- for Authors and Research Materials: http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?action=display&doc_id=4607
- Creative Commons Licensing Types - 2018 - https://creativecommons.org/share
- European Commission (2017) HESTIA Grant Agreement number: 957823 - Funded under Funded under H2020-EU.3.3.1. & EU.3.3.4. - https://cordis.europa.eu/project/id/957823