

BEYOND

D1.2– Data Management Plan

Grant Agreement n°	957020
Project Acronym	BEYOND
Project Title	A reference big data platform implementation and AI analytics toolkit toward innovative data sharing-driven energy service ecosystems for the building sector and beyond
Starting Date	01/12/2020
Duration	36
EU Project Officer	Stavros STAMATOUKOS
Project Coordinator	UBITECH
Consortium Partners	VTT, FVH, CIRCE, Suite5, IGM, KONCAR, ARTELYS, MYTILINEOS, CUERVA, BELIT, URBENER, BEOELEK,
Project Website	beyond-h2020.eu
Cordis	https://cordis.europa.eu/project/id/957020



This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement n° 957020.

D1.2 – Data Management plan

Deliverable No.	D1.2
Deliverable Title	Data Management plan
Work Package	WP 1 - Project Management and Ethics Coordination
WP Leader	UBITECH
Due Date	31/05/2021
Actual Date of submission	30/05/2021
Version	1.0
Status	Submitted
Dissemination Level	Public
Authors	Eleni Tsironi - UBITECH Christina Stratigaki - UBITECH Anna Zoakou - UBITECH
Reviewers	Alexandre Marie - ARTELYS, Milan Vukovic - BELIT

Disclaimer: *The present report reflects only the authors' view. The European Commission is not responsible for any use that may be made of the information it contains.*



Version	Modification(s)	Date	Author(s)
0.1	ToC	10/2/2021	Eleni Tsironi - UBITECH Christina Stratigaki - UBITECH Anna Zoakou - UBITECH
0.3	Section 1,2	16/3/2021	Eleni Tsironi - UBITECH Christina Stratigaki - UBITECH Anna Zoakou - UBITECH
0.5	Section 3	26/3/2021	Eleni Tsironi - UBITECH Christina Stratigaki - UBITECH Anna Zoakou - UBITECH
0.6	Section 4,5	5/4/2021	Eleni Tsironi - UBITECH Christina Stratigaki - UBITECH Anna Zoakou - UBITECH
0.8	Section 6	10/5/2021	Eleni Tsironi - UBITECH Christina Stratigaki - UBITECH Anna Zoakou - UBITECH
0.9	Section 7, Conclusions	24/5/2021	Eleni Tsironi - UBITECH Christina Stratigaki - UBITECH Anna Zoakou - UBITECH
1.0	Updates on reviewers' comments	28/5/2021	Eleni Tsironi - UBITECH Christina Stratigaki - UBITECH Anna Zoakou - UBITECH



EXECUTIVE SUMMARY

Based on European Commission's Guidelines for Horizon 2020, to submit a DMP (*Data Management Plan* as a deliverable) which require project to submit Data Management Plan within the 6 months, the following document represent the DPM of BEYOND project. Any occurring further updates in this respect will be provided in the course of the project duration, on the basis of the actual developments of the technical work. More specifically, in line with the F.A.I.R Principles, this deliverable provides guidelines on how to make the research data collected and/or generated throughout and after the project duration Findable, Accessible, Interoperable and Re-usable. The DMP is, thus, a key document presenting the data management practices to employed by the consortium partners, describing -among other- standards and the methodology that shall be followed for data collection and generation, and whether as well as how such data will be shared. Due to this, all partners were requested to provide input with respect to the artefacts that they will produce/provide in the BEYOND project and the level of dissemination (Public/Private).

It is intended that the present DMP will be updated over the course of the project, whenever significant changes arise, either with respect to the data management practices of individual partners or at project level. Note that a detailed description of the data management practices concerning -in particular- processing of personal data in the context of the piloting activities is provided under WP9 Ethics Requirements deliverables, and also D1.3 - Ethics monitoring report, due in M6 of BEYOND project.



Table of Contents

EXECUTIVE SUMMARY	4
INTRODUCTION.....	8
1. DATA MANAGEMENT IN HORIZON 2020	9
2. PROJECT DATA MANAGEMENT OVERVIEW	12
2.1 Types and formats of artefacts generated/collected	12
2.2 BEYOND Artefacts and Access Rights	12
2.3 Expected size of the data (if known)	16
3. BEYOND ORDP PARTICIPATION	17
3.1 Overview.....	17
3.2 Publishing Infrastructure for Open Access.....	18
3.2.1 Publishing Process.....	19
3.2.2 Publishing Platforms.....	20
3.2.3 Access and Sharing	22
4. FAIR DATA	23
4.1 Making data findable, including provisions for metadata.....	23
4.1.1 Discoverability of the data	24
4.1.2 Data identification mechanism.....	24
4.1.3 Naming conventions used	25
4.1.4 Clear versioning of the documents.....	25
4.1.5 Standards for the metadata creation (if any).....	26
4.2 Making data openly accessible.....	28
4.2.1 Methods or software needed to access the data	29
4.2.2 Deposit of data, associated metadata, documentation and code.....	29
4.3 Making data interoperable.....	31
4.3.1 Interoperability of data assessment	31
4.3.2 Interoperability of data assessment	31
4.4 Making data re-usable.....	31
4.4.1 Increase data re-use through clarifying licenses.....	32
4.4.2 Data quality assurance process.....	32



4.4.3	Length of time for which the data will remain re-usable	32
5.	ALLOCATION OF RESOURCES	33
5.1	Data management responsibilities.....	33
5.2	Cost of potential value of long-term preservation.....	33
6.	DATA SECURITY	34
7.	ETHICAL ASPECTS	35
	CONCLUSIONS.....	36
	ANNEX – ARTEFACT TEMPLATE.....	37

LIST OF TABLES

Table 1	clarifications of terms.....	10
Table 2:	artefacts overview	12
Table 3:	Partners' Research Item provision.....	13
Table 4:	Partners' Software provision	14
Table 5:	Partners' Dataset provision	15
Table 6:	Proposed Document History Table overview.....	25
Table 7:	Document History Template - Example	26
Table 8:	Metadata template for BEYOND datasets.....	26
Table 9:	Making Data Findable Template.....	37
Table 10:	Making data Accessible Template.....	37
Table 11:	Making data Interoperable Template.....	38
Table 12:	Making data Re-usable Template	38



LIST OF FIGURES

Figure 1: BEYOND Artefacts' Types	13
Figure 2: Research Items Access rights	17
Figure 3: Software Artefacts Access rights.....	18
Figure 4: Dataset Artefacts Access rights.....	18
Figure 5: Template to be used for project documentation metadata overview.....	27
Figure 6: Open access to scientific publication and research data in the wider context of dissemination and exploitation	28
Figure 7: Artefacts' Types	29
Figure 8: Beyond repository login page	30
Figure 9: BEYOND Repo Folders Structure	31



INTRODUCTION

This deliverable presents the BEYOND data management plan, as captured in M6 of the project. In accordance with European Commission’s Guidelines for Horizon 2020 Programme¹, to submit a DMP within the first six (6) months of the project, the present report forms the DMP of BEYOND project reflecting the technical progress at the moment of the drafting of the present document. Further updates in this respect will be provided in the course of the project duration, on the basis of the actual developments of the technical work. More specifically, in line with the F.A.I.R Principles, the deliverable provides for how making the research data collected and/or generated throughout and after the project duration Findable, Accessible, Interoperable and Re-usable. To this end, the deliverable outlines -among other- how the research data collected and/or generated will be handled during and after the BEYOND project, describes which standards and methodology for data collection and generation will be followed, and whether and how data will be shared. Note that the document is largely based on the related template provided by the European Commission.

The delivery of the present document falls under WPI activities, that extends until M36 of BEYOND project. In the context of the related activities, it is, thus, intended that the DMP is a living document, subject to updates -to the extent necessary- on the basis of the progress of the project activities.

This Data Management Plan (DMP) outlines how data collected or generated by the BEYOND project will be organized, stored, and shared. It specifies 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.

¹ H2020 Programme Guidelines on FAIR Data Management in Horizon 2020 Version 3.0, 26 July 2016, available at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf



1. Data Management in Horizon 2020

According to the European Commission (EC) all project proposals submitted to "Research and Innovation actions", "Innovation actions" and "Coordination support actions" is required to include a section on research data management which is evaluated under the criterion 'Impact'. Projects participating in the pilot action on open access to research data have to develop a DMP to specify what data will be open².

The DMP is defined as:

“Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. The use of a Data Management Plan is required for projects participating in the Open Research Data Pilot. Other projects are invited to submit a Data Management Plan if relevant for their planned research.”

The purpose of a DMP is to provide a discussion of the main elements of the data management policy that will be used by the applicants with regard to all the datasets that will be generated by the project.

Overall, having taken into account all relevant principles regarding lawful processing of personal data, scientific research data should be easily discoverable, accessible, assessable and intelligible, useable beyond the original purpose for which it was collected and interoperable to specific quality standards.

The BEYOND Data Management also follows the Guidelines on FAIR Data Management in Horizon 2020, released by the European Commission Directorate – General for Research & Innovation. This Horizon 2020 FAIR DMP template³ has been designed to be applicable to any Horizon 2020 project that produces, collects or processes research data. According to these guidelines the management and organization of data should be based on four basic principles, which determine how research outputs should be processed so that they can be more easily accessed, understood, exchanged and reused. This means that data must be findable, accessible, interoperable and re-useable, for example by researchers interested in using the data in further research in the field.

These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution. EC provides a Template

² http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

³ H2020 Programme Guidelines on FAIR Data Management in Horizon 2020 Version 3.0, 26 July 2016, available at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf



D1.2 – Data Management plan

with the FAIR principle. This template is not intended as a strict technical implementation of the FAIR principles, it is rather inspired by FAIR as a general concept. The template represents the set of questions that should be answered with a level of detail appropriate to the project.

TABLE 1 CLARIFICATIONS OF TERMS

Research data	Research data is the evidence that underpins all research conclusions (except those which are purely theoretical) and includes data that have been collected, observed, generated, created or obtained from commercial, government or other sources, for subsequent analysis and synthesis to produce original research results. These results are then used to produce research papers and submitted for publication.
Open research data	Openly accessible research data can typically be accessed, mined, exploited, reproduced and disseminated, free of charge for the user.
Secondary data	Secondary data are data that already exist, regardless of the research to be conducted.
Open access	Open access is understood as the principle that research data should be accessible to relevant users, on equal terms, and at the lowest possible cost. Access should be easy, user-friendly and, if possible, Internet-based.
Metadata	Metadata is data used to describe other data. It summarizes basic information about data, which can make finding and working with instances of data easier.
Research data repositories	Research data repositories are online archives for research data. They can be subject based/thematic, institutional or centralized.

It is possible to develop a single DMP for any project to cover overall approach. However, where there are specific issues for individual datasets (e.g. regarding openness), someone should clearly spell this out.

The template proposes the following issues to be addressed:

- Data Summary
- FAIR data
- Allocation of resources
- Data security
- Ethical aspects
- Other issues
- Further support in developing your DMP

Each of the previously defined has its own set of questions that has to be addressed. The proposed template states that it is not required to provide detailed answers to all the questions of the DMP that needs to be submitted by month 6 of the project,



D1.2 – Data Management plan

subject -also- to potential future updates. Rather, the DMP is intended to be a living document -to the extent necessary- in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur.



2. Project Data Management Overview

As described in the Guidelines on FAIR Data Management in Horizon 2020 a Data Management Plan is a key element to ensure that data is well managed. For this reason, in this section we will firstly identify the type of artefacts that will be generated and collected in the framework of the project. During the lifetime of the BEYOND project, several artefacts will be produced. The artefacts that will be collected/generated are listed below in Section 2.2. As the project evolves, this list may require modifications (addition or removal of artefacts) with respect to the project developments.

2.1 Types and formats of artefacts generated/collected

In order to provide an overview of the different data sets that are currently and will be produced in the BEYOND project, the following table shows the data type, the related WP number and the format, in which the data will be presumably stored.

TABLE 2: ARTEFACTS OVERVIEW

#	Artefact type	Explanation	WP #	Format (indicative)
1	Research Item	Standards, Policies, Questionnaires, Deliverables, Papers	1-8	.xls, .csv, .txt, .docx, .pdf
2	Software	Code, APIs, dashboard	3-6	.xls, .csv, .txt, .docx, .pdf
3	Dataset	BEMS (Building Energy Management Systems) data, IoT sensing data, Metering data, Contract Related data, Demographic data	2,7	.xls, .csv, .txt, .docx, .pdf

2.2 BEYOND Artefacts and Access Rights

In the survey contacted during the first months of the project, the input collected from most of the partners is depicted in the following chart. The types of BEYOND Artefacts



D1.2 – Data Management plan

are distributed as 46% being datasets, 35% research items and the remaining 19% is of a software artefact type.

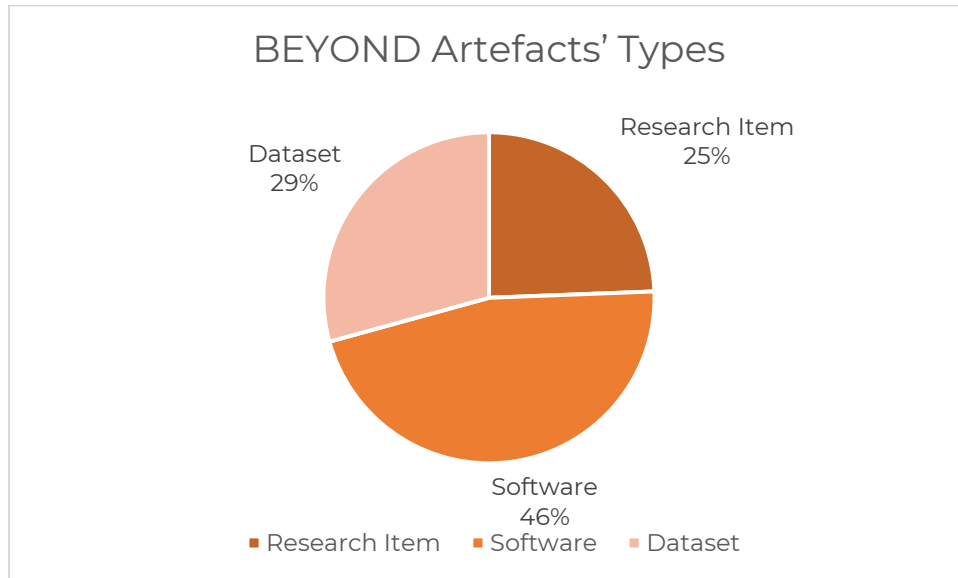


FIGURE 1: BEYOND ARTEFACTS' TYPES

The following tables present the current status and consensus within the Consortium with regards to identified artefacts and their access rights. It is provisioned that those tables are a recurring exercise and all future updates and additions will be documented under the relevant deliverables.

TABLE 3: PARTNERS' RESEARCH ITEM PROVISION

PARTNER	ARTEFACT DESCRIPTIONS	PUBLISHABLE/ NON-PUBLISHABLE
UBITECH	End-user and business requirements (D2.1)	Publishable
UBITECH	End-user and business requirements (D2.2)	Publishable
UBITECH	Pilot asset, data and deployment plans (D2.5)	Publishable
UBITECH	Building portfolio management optimization tool (D6.2)	Publishable
Suite5	Data Exchange Collaboration Activity Planning (D1.5)	Non-publishable
Suite5	BEYOND Framework Architecture-a (D2.6)	Publishable



D1.2 – Data Management plan

PARTNER	ARTEFACT DESCRIPTIONS	PUBLISHABLE/ NON-PUBLISHABLE
Suite5	BEYOND Framework Architecture-b (D2.7)	Publishable
Suite5	BEYOND CIM (D3.2)	Publishable
KONCAR	Flexibility based algorithms for VPP Configurator and DR Strategies Optimization (D6.5)	Publishable
IGM	BEYOND Pretrial validation testing results (D7.1)	Publishable

TABLE 4: PARTNERS' SOFTWARE PROVISION

PARTNER	ARTEFACT DESCRIPTIONS	PUBLISHABLE/ NON-PUBLISHABLE
UBITECH	Portfolio Management Optimization Tool	Non-publishable
UBITECH	Data Security Services Bundle	Non-publishable
BELIT	Energy Performance and Smart Readiness Certification	Publishable
Suite5	Data Collection, Security, Storage, Governance & Management Services Bundles - Beta Release (D3.3)	Publishable
Suite5	Data Collection, Security, Storage, Governance & Management Services Bundles - Release 1.00 (D3.6)	Publishable
Suite5	BEYOND Integrated Platform & Open APIs – Alpha, Mock-ups Release (D3.4)	Publishable
Suite5	BEYOND Integrated Platform & Open APIs – Beta Release (D3.5)	Publishable
Suite5	BEYOND Integrated Platform & Open APIs – Release 1.00 (D3.7)	Publishable
Suite5	BEYOND Integrated Platform & Open APIs – Release 2.00 (D3.8)	Publishable



D1.2 – Data Management plan

PARTNER	ARTEFACT DESCRIPTIONS	PUBLISHABLE/ NON-PUBLISHABLE
Suite5	BEYOND Data Analytics, Sharing & Matchmaking Services Bundles – Beta Release (D4.1)	Publishable
Suite5	BEYOND Data Analytics, Sharing & Matchmaking Services Bundles – Release 1.00 (D4.3)	Publishable
Suite5	BEYOND Baseline Data Analytics – Draft Release (D4.2)	Publishable
Suite5	BEYOND AI Data Analytics Toolkit – Final Release (D4.4)	Publishable
Suite5	Personalized Energy Analytics Tool for Guidance on Energy Performance Optimization and Human-Centric Control Automation (D6.3)	Publishable
KONCAR	Flexibility based VPP Configurator and DR Strategies Optimization Tool (D6.5)	Non-publishable
VTT	Renovation Optimization Decision Support Tool (D5.4)	Non-publishable
VTT	Innovative Energy Services Suite towards Optimized Buildings Energy Performance Management -Digital Twin-based Building Performance Optimization tool- (D6.6)	Non-publishable
ARTELYS	Impact Assessment Tool for Energy Policy Making at urban level (D5.1)	Non-publishable
ARTELYS	District Heating Network Planning and Infrastructure Sizing Tool (D5.2)	Non-publishable

TABLE 5: PARTNERS' DATASET PROVISION

PARTNER	ARTEFACT DESCRIPTIONS	PUBLISHABLE/ NON-PUBLISHABLE
BEOELEK	Pilot asset, data and deployment plans (D2.5)	Non-publishable



D1.2 – Data Management plan

FVH	BEMS data (D2.5)	Non-publishable
FVH	IoT sensing data (D2.5)	Non-publishable
FVH	Helsinki City Open Energy Data https://hri.fi/fi/	Publishable
FVH	Helsinki RES Generation data https://hri.fi/data/en_GB/dataset/auringonsateilyenergian-maarakatoillapaakaupunkiseudulla	Publishable
FVH	Helsinki Air Quality Data	Publishable
FVH	Helsinki Transportation Data	Publishable
MYTILINEOS	BEMS data	Non-publishable
MYTILINEOS	IoT sensing data	Non-publishable
MYTILINEOS	Metering data	Non-publishable
MYTILINEOS	Contract Related data	Non-publishable
MYTILINEOS	Demographic data	Non-publishable

2.3 Expected size of the data (if known)

It is expected that as a research outcome will generate research datasets (i.e. results of the technologies, services of the demos, etc.), publications, new services proposal, dissemination material, etc. Due to size of the project, scope of work and complexity, the expected size cannot be estimated at the moment.



3. BEYOND ORDP Participation

3.1 Overview

The Open Research Data Pilot (ORDP) of the European Commission enables open access and reuse of research data generated by Horizon 2020 projects. There are two main pillars to the Pilot: a) developing a Data Management Plan (DMP) and b) providing open access to research data.

A project that opts-in ORDP have to adhere to the following conditions:

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

The ORDP applies to:

- The data (and metadata) needed to validate results in scientific publications.
- Other curated and/or raw data (and metadata) that are specified in the DMP.

From the current consensus within the consortium some of the BEYOND Artefacts will not be publicly available as depicted in the graphics below

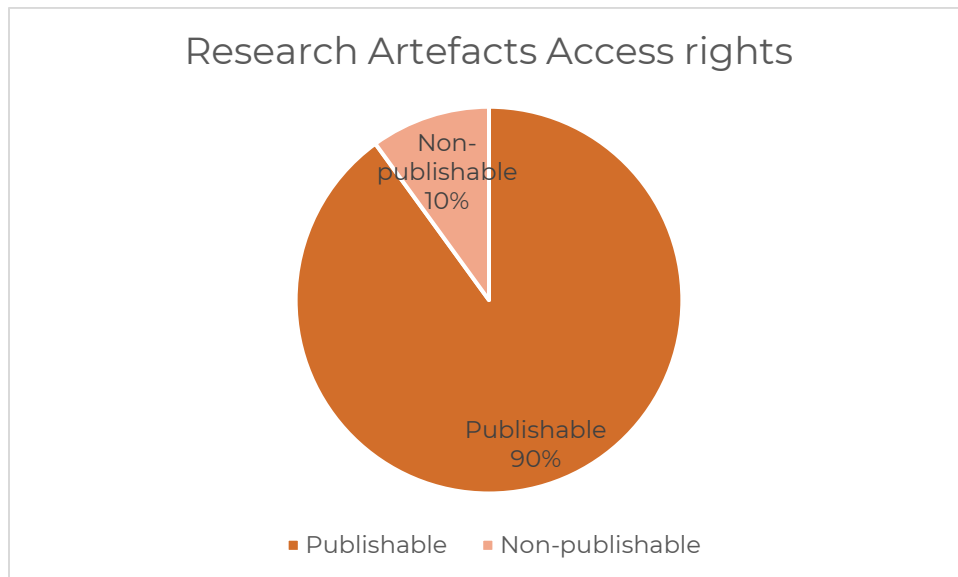


FIGURE 2: RESEARCH ITEMS ACCESS RIGHTS

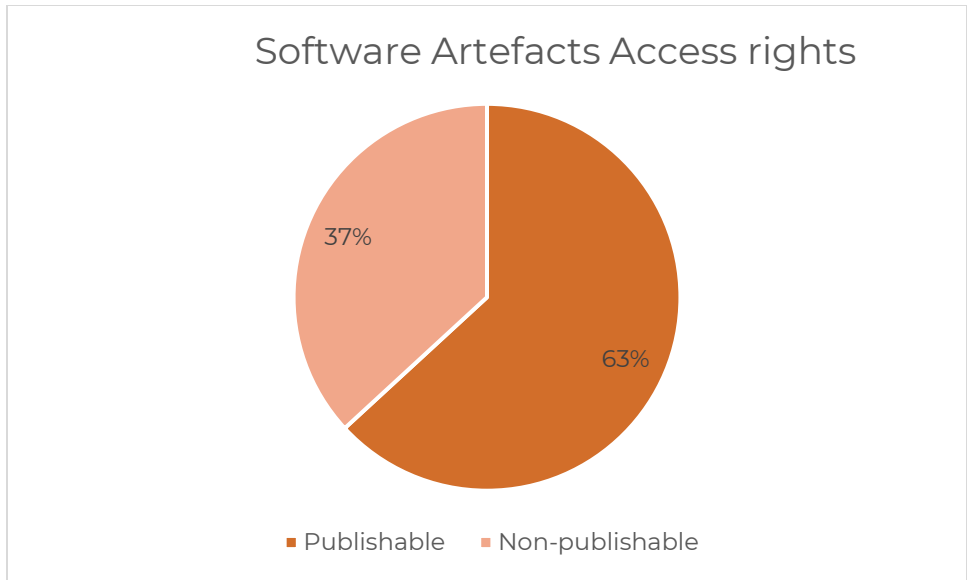


FIGURE 3: SOFTWARE ARTEFACTS ACCESS RIGHTS

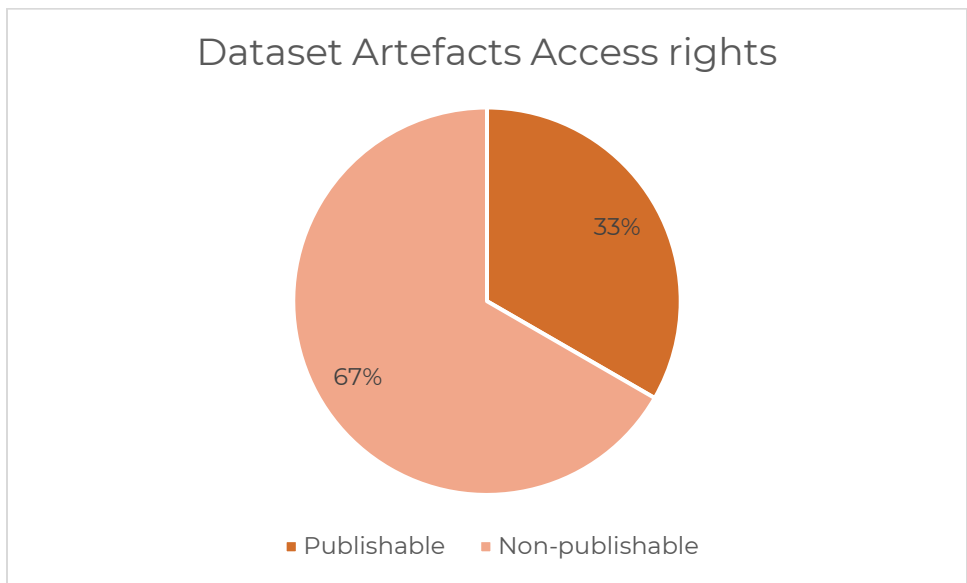


FIGURE 4: DATASET ARTEFACTS ACCESS RIGHTS

3.2 Publishing Infrastructure for Open Access

The BEYOND publication infrastructure consists of a process and several web-based publication platforms that together provide long-term open access to all publishable, generated or collected results of the project. The implementation of the project will be done in accordance with the applicable regulations in national and EU level and, especially, with the General Data Protection Regulation (GDPR) protection of personal data.



D1.2 – Data Management plan

More specifically, cases where personal data information or sensitive information of internet users is collected (IP addresses, email addresses or other personal information) or further processed are not foreseen.

In the potential future case where the BEYOND consortium will collect and/or further process personal data, this will be done in accordance with GDPR. Overall, it is aimed that BEYOND only collects and/or further processes personal data are necessary for the attainment of the project objectives.

Both the process and the used web-based platforms are described in the following subsections.

3.2.1 Publishing Process

BEYOND partners defined a simple, deterministic process that decides if a result in BEYOND must be published or not. The term result is used for all kind of artefacts generated during BEYOND like white papers, scientific publications, and anonymous usage data. By following this process, each result is either classified public or non-public. Public means that the result must be published under the open access policy. Non-public means that it must not be published.

For each result generated or collected during BEYOND runtime, the following questions must be answered to classify it:

Does a result provide significant value to others or is it necessary to understand a scientific conclusion?

If this question is answered with yes, then the result is classified as public. If this question is answered with no, the result is classified as non-public. Such a result could be code that is very specific to BEYOND platform (e.g., a database initialization) which is usually of no scientific interest to anyone, nor does it add any significant contribution.

Does a result include personal information that is not the author's name?

If this question is answered with yes, the result is classified as non-public. Personal information beyond the name must be removed if it should be published. This also bears witness on the repetitive nature of the publishing process, where results which are deemed in the beginning as non-publishable can become publishable once privacy-related information is removed from them.

Does a result allow the identification of individuals even without the name?

If this question is answered with yes, the result is classified as non-public. Sometimes data inference can be used to superimpose different user data and reveal indirectly a single user's identity. As such, in order to make a result publishable, the included information must be reduced to a level where single



individuals cannot be identified. This can be performed by using established anonymization techniques to conceal a single user's identity, e.g., abstraction, dummy users, or non-intersecting features.

Does a result include business or trade secrets of one or more partners of BEYOND?

If this question is answered with yes, the result is classified as non-public, except if the opposite is explicitly stated by the involved partners. Business or trade secrets need to be removed in accordance to all partners' requirements before it can be published.

Does a result name technology that is part of an ongoing, project-related patent application?

If this question is answered with yes, then the result is classified as non-public. Of course, results can be published after patent has been filed.

Can a result be abused for a purpose that is undesired by society in general or contradict with societal norms and BEYOND's ethics?

If this question is answered with yes, the result is classified as non-public.

3.2.2 Publishing Platforms

In BEYOND, we use several platforms to publish our results openly. The following list presents the platforms used during the project and describes their concepts for publishing, storage, and backup.

The project Website

The partners in the project consortium decided early to setup a project-related website <https://beyond-h2020.eu/>. This website describes the mission and the general approach of BEYOND and its development status. A blog informs about news on a regular basis. Later in the project the developed BEYOND platform will be announced. A dedicated area for downloads is used to publish reports and white papers as well as scientific publications (in pre-camera ready form, or through links to the publisher's websites in case these are not open access). All documents are published using the portable document format (PDF)⁴. All downloads are enriched by using simple metadata information, such as the title and the type of the document. The website is hosted by partner UBITECH. All webpage-related data is backed up on a regular basis. All information on the project website can be accessed without creating an account. The website is backed up once per month.

⁴ Note that the site will not host spreadsheets. It exclusively hosts PDFs



Open Access Publication Platforms

Zenodo⁵

Zenodo is a research data archive / online repository which helps researchers to share research results in a wide variety of formats for all fields of science. It was created through EC's OpenAIRE+ project and is now hosted at CERN using one of Europe's most reliably hardware infrastructures. Data is backed nightly and replicated to different locations. Zenodo not only supports the publication of scientific papers or white papers, but also the publication of any structured research data (e.g., using XML). Zenodo provides a connector to GitLab that supports open collaboration for source code and versioning for all kinds of data. All uploaded results are structured by using metadata, like for example the contributors' names, keywords, date, location, kind of document, license, and others. Considering the language of textual metadata items, English is preferred. All metadata is licensed under CC0 license (Creative Commons 'No Rights Reserved'). The property rights or ownership of a result does not change by uploading it to Zenodo.

All public results generated or collected during the project lifetime will be uploaded to Zenodo for long-term storage and open access.

Open Research Europe⁶

Open Research Europe is an open access publishing platform for the publication of research stemming from Horizon 2020 funding across all subject areas. The platform makes it easy for Horizon 2020 beneficiaries to comply with the open access terms of their funding and offers researchers a publishing venue to share their results and insights rapidly and facilitate open, constructive research discussion. Uses an open research publishing model: publication within days of submission, followed by open invited peer review. Includes citations to all supporting data and materials, enabling reanalyzes, replication and reuse. Articles that pass peer review are sent to major indexing databases and repositories.

GitLab

GitLab⁷ is a well-established online repository which supports distributed source code development, management, and revision control. It is primarily used for source code data. It enables world-wide collaboration between developers and provides also some facilities to work on documentation and to track issues. GitLab provides paid and free

⁵ <https://zenodo.org/>

⁶ <https://open-research-europe.ec.europa.eu/>

⁷ <https://about.gitlab.com/>



D1.2 – Data Management plan

service plans. Free service plans can have any number of public, open-access repositories with unlimited collaborators. Private, non-public repositories require a paid service plan. Many open-source projects use GitLab 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 GitLab over provided material. For textual metadata items, English is preferred.

All source-code components that are implemented during this project and decided to be public will be uploaded to an open access GitLab repository.

3.2.3 Access and Sharing

The accessing and sharing of data are ruled by the Description of Action (DoA) which stipulates the dissemination level of each deliverable. Moreover, the project consortium will comply with the FAIR (findable, accessible, interoperable and reusable) (European Commission, 2016) guidelines of the H2020 programme.

The data necessary to successfully complete the project Work Packages (WPs) will be shared without any restrictions amongst the WP partners either via internal repositories or direct communication. Public data will be made available at the project's website or other repositories, as appropriate. Users will be made aware of this data primarily through research publications, patent applications, dissemination activities, invited talks, social networks and the project website. Data will be made available to the project consortium as soon as it is available; to standardization bodies when required; and to the public at the due date of the deliverable, and, in case a research publication is based on that, as soon as the paper is submitted (if submission is anonymous, this will be postponed). If access to confidential data is necessary by the public, restrictive measures will be put in place.



4. FAIR data

BEYOND project supports the reuse of research data and follows FAIR principles⁸. FAIR represents a set of guiding principles to make data **Findable, Accessible, Interoperable, and Reusable**.

The international FAIR Principles have been formulated as a set of guidelines for the reuse of research data. The acronym FAIR stands for findable, accessible, interoperable and reusable. Research data must be of quality that makes them accessible, findable and reusable.

Findable: data has a unique, persistent ID, located in a searchable resource, and documented with meaningful metadata.

Accessible: data is readily and freely retrievable using common methods and protocols, metadata is accessible even if the data is not.

Interoperable: data is presented in broadly recognized standard formats, vocabularies, and languages.

Re-useable: data has clear licenses, and accurate meaningful metadata conformity to relevant community standards and identifying its content and provenance.

4.1 Making data findable, including provisions for metadata

This document launches the data management plan to support the effective collection and integration of the BEYOND data. Storage, processing and sharing (among project participants) will occur via data exchange platforms (such as Nextcloud), whereas interaction with the wider public will be achieved through the official project website. Also, data will be stored at the coordinator's repository and will be kept for minimum 5 years after the end of the project. Where requested, data will be kept for 2 more years.

A naming convention will include a concise description of contents, the host institution collecting the data and the month of publication.

Version numbering will only be an issue if a participant requests withdrawal of their data in which case a version number will be added to the filename.

No specific standards or metadata have been identified for the time being for the proposed datasets.

⁸ Force11 (2016) The FAIR Data Principles, <https://www.force11.org/group/fairgroup/fairprinciples>



D1.2 – Data Management plan

Data will be anonymized meaning that data will not identify any individuals and therefore real names of participants will NOT be distributed.

Data will be shared only in relation to publications (deliverables and papers). As such, the publication will serve as the main piece of metadata for the shared data. When this is not seen as being adequate for the comprehension of the raw data, a report will be shared along with the data explaining their meaning and methods of acquisition.

4.1.1 Discoverability of the data

In order to be able to use the data generated by the project it is essential to integrate data from the participants in the open calls and the activities undertaken by project partners. Taking into account the FAIR data principles (Wilkinson et al., 2016⁹) (meta)data should:

- Be assigned to a globally unique and persistent identifier;
- contain enough metadata to fully interpret the data, and;
- be indexed in a searchable source.

By applying these principles data become retrievable and include their authentication and authorization details.

4.1.2 Data identification mechanism

All documents associated project will be identified with a project name and unique and persistent document type designator and number that will be given to the coordinator for the submission to the EC. Versioning of the document should be part of the document name and title.

As per the documents related to project activities and/or deliverables, the tasks or deliverables number will be used to identify the document followed by a brief title of the activity or deliverable.

Example

BEYOND- D1.2 - Data Management Plan -v1.0.pdf

⁹ Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data3:160018 doi: 10.1038/sdata.2016.18 (2016).



4.1.3 Naming conventions used

Each set of data produced (dataset, deliverables, etc.) will be named in a uniform way and will include a table with a version control.

The recommendations to name documents of the project are as follows¹⁰:

- Choose easily readable identifier names (short and meaningful);
- Do not use acronyms that are not widely accepted;
- Do not use abbreviations or contractions;
- Avoid Language-specific or non-alphanumeric characters;
- Add a two-digit numeric suffix to identify new versions of one document.
- Dates should be included back to front and include the four-digit years: YYYYMMDD.

For deliverables: **BEYOND [Deliverable Code]-[Deliverable Title]_[Partner]-vA.BB**
i.e.: BEYOND_D8.1-Project Management Handbook- *(for submission to the Commission)*

For datasets: **WP [Work Package number] P [Pilot number; pilot activity number] - [description of the activity]** i.e.: WP4 P1.3 Results of demonstration performance.

4.1.4 Clear versioning of the documents

Only documents created by the consortium will be versioned, for this purpose templates include 3 descriptors to identify the versions and status of the documents:

TABLE 6: PROPOSED DOCUMENT HISTORY TABLE OVERVIEW

Version	Date	Author	Notes
1	XX	XX	XX
2			
3			
4			

Moreover, partners, following the recommendations included in section “Naming conventions” will identify the different versions by using a two-digit number following the descriptor Draft. A document reviewed by another partner should be returned to

¹⁰ <https://www.ukdataservice.ac.uk/manage-data/format/organising>



D1.2 – Data Management plan

the principal author by including **rev + acronym** of the organisation. Only the principal author will change the draft number and will add the word FINAL to documents ready to be sent to the EC or those to be used as final versions.

The document history included in the document template should be filled in as follows:

TABLE 7: DOCUMENT HISTORY TEMPLATE - EXAMPLE

Version	Date	Author	Notes
1	XX/XX/2020	ABC	Section 2.1 needs to be completed
2	XX/XX/2021	CDE	Section 2,1 completed. Comments added to the document.
3	XX/XX/2021	ABC	Added suggestions by Artelys
4	XX/XX/2021	XYZ	Included some topics on section 2.1
	XX/XX/2021	ABC	Final version with partners contribution

4.1.5 Standards for the metadata creation (if any)

Basic metadata will be used to facilitate the efficient recall and retrieval of information by project partners and external evaluators and contribute to easily find the information requested. To this end, all documents related to the project have to include in the front-page information about author(s) & editor(s), WP, dissemination level and version.

To support the completeness of metadata, the project provides a metadata template to all stakeholders. The template will be a living document that might be expanded to fit project specific requirements

TABLE 8: METADATA TEMPLATE FOR BEYOND DATASETS.


#	Field	Description
1	Title	A name given to the resource.
2	Creator	An entity primarily responsible for making the resource
3	Subject	The topic of the resource
4	Description	e.g., abstract, table of contents, graphics, ...
5	Publisher	Only for published items.
6	Contributor	Entities that contributed to the making of the resource.
7	Date	The termination of the data collection period.
8	Type	[dataset, article, questionnaire, ...]



#	Field	Description
9	Format	File format of the resource.
10	Identifier	e.g., ISSN if your item has been published
11	Source	Which tools were used to collect the data
12	Language	A language of the resource.
13	Relation	A related resource.
14	Rights	Information about rights held in and over the resource.

In addition to the dataset’s metadata document, dataset providers are compelled to attach additional documents such as:

1. A description of the study
2. Method of research
3. Applied questionnaires
4. Data documentation / usage manual
5. Any other information that might be of interest to a data user



Dx.x - Deliverable Title

Grant Agreement n°	957020
Project Acronym	BEYOND
Project Title	A reference big data platform implementation and AI analytics toolkit toward innovative data sharing-driven energy service ecosystems for the building sector and beyond
Starting Date	01/12/2020
Duration	36
EU Project Officer	Stavros STAMATOUKOS
Project Coordinator	UBITECH
Consortium Partners	VTT, FVH, CIRCE, Suite5, IGM, KONCAR, ARTELYS, MYTILINEOS, CUERVA, BELIT, URBENER, BEOLEK,
Project Website	beyond-h2020.eu
Cordis	https://cordis.europa.eu/project/id/957020

This project has received funding from the European Union’s Horizon 2020 Research and Innovation programme under Grant Agreement n° 957020.

Dxx - Title

Deliverable No.	Dx.x
Deliverable Title	Deliverable Title
Work Package	WP [no] - [title]
WP Leader	[beneficiary]
Due Date	[day, month, year]
Actual Date of submission	[day, month, year]
Version	[no]
Status	[Draft, Peer-reviewed, Submitted, Approved]
Dissemination Level	[confidential or public]
Authors	[name surname] [[beneficiary]], [name surname] [[beneficiary]]
Reviewers	[name surname] [[beneficiary]], [name surname] [[beneficiary]]

Disclaimer: The present report reflects only the authors’ view. The European Commission is not responsible for any use that may be made of the information it contains.

Page 2

FIGURE 5: TEMPLATE TO BE USED FOR PROJECT DOCUMENTATION METADATA OVERVIEW



4.2 Making data openly accessible

Where possible data will be made openly available subject to Ethics and participants’ agreement. However, the personally-identifiable nature of the data collected within BEYOND means that in most instances it would be difficult to release collected data. Where data is made available, we will do so using the Project’s file repository hosted in the coordinator’s premises.

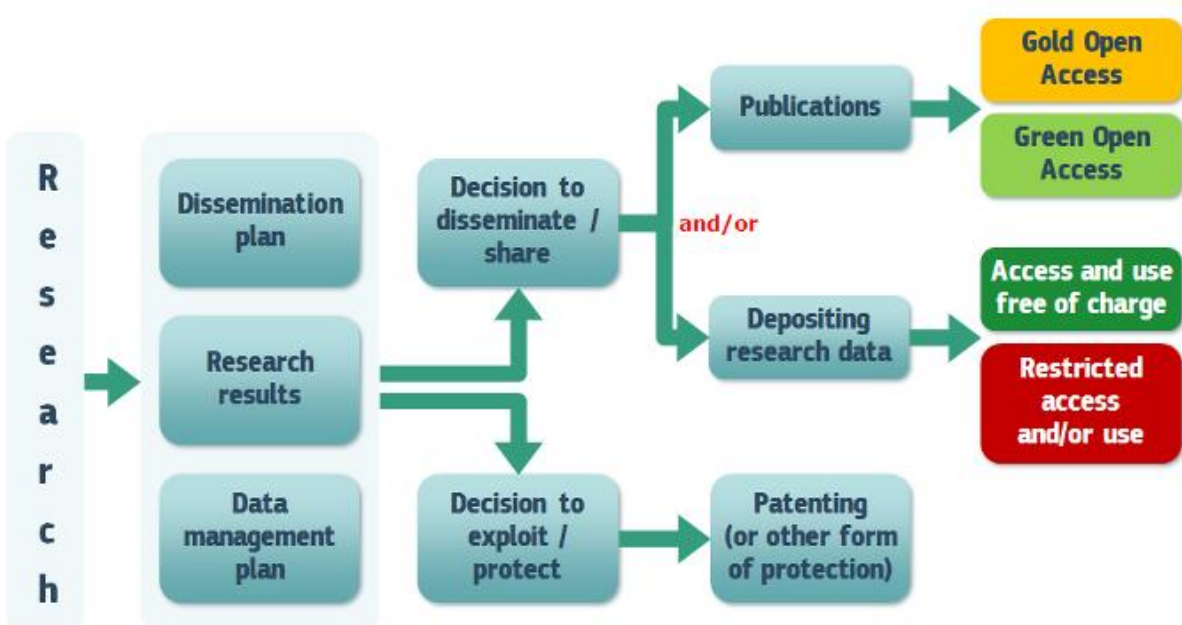


FIGURE 6: OPEN ACCESS TO SCIENTIFIC PUBLICATION AND RESEARCH DATA IN THE WIDER CONTEXT OF DISSEMINATION AND EXPLOITATION¹¹

Prior to release, a requesting party will need to contact the Project Coordinator describing their intended use of a dataset. The Project Coordinator will send a terms and conditions document for them to sign and return. Upon return, the dataset will be released. Documentation will be included with the release of the data.

¹¹ European Commission Directorate-General for Research & Innovation (2017) Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020

D1.2 – Data Management plan

In alignment with the EC Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, BEYOND will also follow a combination of Gold and Green Open Access¹² strategy to its scientific publications, which will be agreed during the first months of the project execution. Gold Access will be encouraged for high-impact journal publications while the self-archiving, Green Access will be granted for the rest of the publications. The availability through the consortium members will be considered while there will also be a relevant repository on the website of the project and in social networking sites for scientists and researchers like ResearchGate.

FIGURE 7: ARTEFACTS' TYPES

#	Artefact Type	Explanation	Accessibility
1	Research Item	Standards, Policies, Questionnaires, Deliverables, Papers	Publishable/ Non-publishable
2	Software	Code, APIs, dashboard	Publishable/ Non-publishable
3	Dataset	BEMS data, IoT sensing data, Metering data, Contract Related data, Demographic data	Publishable/ Non-publishable

4.2.1 Methods or software needed to access the data

No specific software tools will be needed to access the data, since anonymous data sets will be saved and stored in word, pdf or excel to facilitate its exploitation and guarantee their long-term accessibility.

4.2.2 Deposit of data, associated metadata, documentation and code


Data will be deposited and secured on Nextcloud file repository and additional instance of all data on coordinator's account, in the following URL: <https://repository.beyond-h2020.eu>


¹² https://en.wikipedia.org/wiki/Open_access





Username or email

Password 

Log in 

[Forgot password?](#)

[Log in with a device](#)

BEYOND Repository – a safe home for all your data

FIGURE 8: BEYOND REPOSITORY LOGIN PAGE

The screenshot shows a file explorer interface for the BEYOND Repository. The left sidebar contains navigation options: All files, Recent, Favorites, Shares, Tags, Deleted files, 1.8 GB used, and Settings. The main area displays the contents of the 'BEYOND_957020' folder, which includes seven subfolders and a summary row.

Name	Size	Modified
01_Administrative Documents	3.9 MB	a day ago
02_Legal Documents	7.6 MB	2 months ago
03_Work Packages	120.8 MB	10 hours ago
05_Meetings and Telcos	1.6 GB	12 days ago
06_Submitted Deliverables	5.3 MB	a month ago
07_How Tos	10.1 MB	5 days ago
6 folders	1.8 GB	



FIGURE 9: BEYOND REPO FOLDERS STRUCTURE

4.3 Making data interoperable

The concept interoperable demands that both data and metadata must be machine-readable and that a consistent terminology is used.

4.3.1 Interoperability of data assessment

Partners will be responsible of storing the data in a comprehensive format and adapted to the real and current needs of the possible practitioners interested in using, merging or exploiting the data generated throughout the project. The assessment of data interoperability will be updated in future reviews in order to guarantee the BEYOND data fits the needs of a specific scenario (such as data infrastructures, interests or purpose of data).

4.3.2 Interoperability of data assessment

The vocabulary used in the project is a very standard and common language within the business creation culture and the logistics. Vocabulary won't represent any barrier for data interoperability a re-use.

4.4 Making data re-usable

For data to be re-usable, it is -generally- considered that meta(data) have a plurality of accurate and relevant attributes and that they are released with a clear and accessible data usage license. Moreover, it is considered that (meta)data are associated with their provenance and that they meet domain-relevant community standards¹³.

Note that the overall management of knowledge and the provisioning for the establishment of the related Intellectual Property Rights is dictated in detail under BEYOND's Grant Agreement and the consortium agreement stipulating -among other- for the ownership of the background and the foreground knowledge, as well as for the commercial exploitation of the project's results.

¹³ See, also, FAIR data principles (FORCE11 discussion forum) available at: <https://www.force11.org/group/fairgroup/fairprinciples>



4.4.1 Increase data re-use through clarifying licenses

Data will only be available on application/Nextcloud platform/share portal and their use will be restricted to the research use of the licensee and colleagues on a need-to-know basis. This non-commercial licence is renewable after 2 years, data may not be copied or distributed and must be referenced if used in publications. These arrangements will be formalised in a User Access Management license which describes in detail the permitted use of the data.

4.4.2 Data quality assurance process

The project coordinator will be responsible of assuring the quality of the data by making sure dataset follow the FAIR principles included in this plan, and that data is updated.

Personal data processing will be done following the EU, national and international laws taking into account the “data quality” principles listed below¹⁴:

Data processing is adequate, relevant and non-excessive;

- Accurate and kept up to date;
- Processed fairly and lawfully;
- Processed in line with data subjects’ rights;
- Processed in a secure manner;
- Kept for no longer that necessary and for the sole purpose of the project.

Data quality assurance process will be led in accordance with the Regulation (EU) 2016/679 (General Data Protection Regulation) on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

4.4.3 Length of time for which the data will remain re-usable

The Consortium will contribute to maintain data re-usable as long as possible after the end of the project. A first period of 4 years has been established; however, this time can be extended under partners agreement. This period can vary depending on the value of the data after the end of the project.

¹⁴ Wilms, G. Guide on Good Data Protection Practice in Research of the European University Institute. (March 2017). Retrieved from <http://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf>



5. Allocation of resources

5.1 Data management responsibilities

Data will be stored at the Collaboration file repository (Nextcloud), set by the Coordinator as the project's repository, and will be kept for 5 years after the end of the project. Where requested, data will be kept for 2 more years. The handling of the repository on behalf of BEYOND as well as all data management issues related to the project fall in the responsibility of the coordinator.

As for the publications, where the analyses of the empirical research data will be presented, the consortium will publish them in scientific journals that allow open access. The costs related to open access will be claimed as part of the Horizon 2020 grant.

Regarding the data resulting from the activities of the project, each WP leader will be responsible for the storage and compliance of the data and then for uploading in the BEYOND Nextcloud web portal, or other storage systems to share the information of the project.

The BEYOND coordinator assisted by the WP leaders will be responsible for updating this document and develop a strategy to encourage:

- the identification of the most-suitable data-sharing and preservation methods;
- the efficient use of data assuring clear rules on its accessibility;
- the quality of the data stored and
- the storage in a secured in a user-friendly interface.

5.2 Cost of potential value of long-term preservation

As stated in previous section, the costs of data storage and maintenance are not going to require extra funding once the project ends. As per the value of the data, it is important to take into account that the topics covered by the project respond to a current need of the logistics sector and customers' needs. Therefore, data coming out of this project will have a direct impact in the coming years but might not be of relevance as the challenges are being tackled or replaced by other priorities.



6. Data Security

BEYOND data exchange platform (Nextcloud) applies technological and organizational measures to secure processing of personal data against publishing to unauthorized persons, processing in violation of the law and change, loss, damage or destruction.

- **Information security:** SSL (Secure Socket Layer) certificates are applied. In order to ensure the appropriate level of security, the password for the account will exist on the platform only in a coded encrypted form.
- **Options for reading data:** the platform offers the possibility to make data available in a read-only or downloadable format, hindering the access to information by unauthorized users.
- **Back-up policy:** complete and redundant back-ups are done every week. Moreover, every time a modification is done an older version is saved.
- **Accidental deletion or modifications:** in case of a catastrophic event that implies the partial or complete deletion of the data sets, the data from the most recent back up will be automatically restored (back-up won't be older than 60 minutes). In case of accidental deletion or modification only the most recent document will be restored, so in case of accidental changes or deletion data can be easily recovered.
- **Deletion or modification of data by users:** only administrators have the rights to delete or modify the information included in the datasets.
- **Terms and conditions:** the Nextcloud platform have specific terms of use and conditions that have to be accepted by all users of the platform.



7. Ethical Aspects

The BEYOND consortium is aware of the ethical aspects pertinent to the scope of BEYOND, which are addressed under the Work Package 9 on "Ethics Requirements".

In particular, Work Package 9 puts particular emphasis on two aspects pertinent to the scope of BEYOND: a) the involvement of research participants and b) the protection of the personal data to be collected and further processed for the purpose of BEYOND project. Work Package 9, thus, focuses on the Demonstrators' activities by capturing, for instance, the procedures for onboarding research participants and by producing an overview of the technical and organizational measures for the protection of personal data of the research participants.

Note that the content of the earlier stated deliverables on Ethics due in M6 will be updated in the context of the activities under Task 1.3 on "Ethics monitoring, IPR strategies and Responsible Research & Innovation (RRI)", that extends until M36.



Conclusions

This initial version of the data management plan provides the first information on the data to be collected and used throughout the project. The relevance of FAIR data has been described and its application on the used dataset was given. In order to provide for security and the handling of ethical aspects, a first outline of these topics was provided. As mentioned, the DMP is a living document and will be extended and updated through the whole project lifetime by all relevant partners.



Annex – Artefact Template

The following tables try to capture the description of the data that will be produced in the context of BEYOND. Every use case will fill in such a template and subsequently all the templates will be collected with the beginning of WP6, the demonstration applications work package of the project.

TABLE 9: MAKING DATA FINDABLE TEMPLATE

Making data Findable	
Name of data set	<i>Univocal identifier of the considered data [BEYOND_Wx_Tz_01] Please, provide one sentence description.</i>
Data types	<i>[Real time data stream, unstructured like tweets, synthetic data stream, log data of IDS, etc.]</i>
Data generation and/or collection	<i>Description of the type of input used to generate the data and the complete methodology and tools used for data collection</i>
Purpose	<i>What are the data collected/generated specifically used for?</i>
Data origin	<i>[Where applicable, information from applications to be developed by the partner.]</i>
Version control	<i>Version no, Date, Author, Description of change</i>
Search keywords	
Responsible partner of the artefact	

TABLE 10: MAKING DATA ACCESSIBLE TEMPLATE

Making data Accessible	
Accessibility	<i>Open/Confidential</i>
Repository	<i>Description/location of the available data.</i>



Making data Accessible	
Shareability restrictions / related Information	<i>[Where applicable, information from applications to be developed by the partner.]</i>

TABLE 11: MAKING DATA INTEROPERABLE TEMPLATE

Making data Interoperable	
Format	<i>Data format, measuring unit, typical order of magnitude [JSON-like, CSV]</i>
Expected size of the data	<i>[To be defined, 3 TB/Day or 12 GB/day when compressed etc.]</i>
Standards and metadata¹⁵	<i>[The metadata attributes list. The used methodologies.]</i>
Standard software Interfaces	<i>List of the standards used to promote results replicability.</i>
Extensions to standard interfaces	<i>Extensions to the above standards as developed during the project.</i>
Mapping to common ontologies	

TABLE 12: MAKING DATA RE-USABLE TEMPLATE

Making data Reusable	
RE-use of existing data	<i>[No reuse of existing data, for the generation of synthetic datasets, it will be essential to create a recipe, reusing the existing data in logs etc.]</i>
Data backup	<i>Consistent location of the data, including previous releases</i>
Quality Consistency	<i>Constraints determining the quality/currency of the collected data.</i>

¹⁵ Note that the fields pertinent to standards are, also, relevant for reusability purposes.



Making data Reusable	
Emulation tools	<i>Description/location of possible emulation tools useful for replicating the data</i>
Data availability	

