

Grant Agreement N°763990

UPWARDS

Deliverable D1.4

Data Management Plan

WP	1	Consortium management
Task	1.5	Data management

Dissemination level¹	PU	Due delivery date	30/09/2018
Nature²	R	Actual delivery date	21/09/2018

Lead beneficiary	Fraunhofer
Contributing beneficiaries	Wavestone, SINTEF

Document Version	Date	Author	Comments ³
VF	05/11/2019	Benjamin Adrian – Fraunhofer	revised version with updated container repository
VF	07/09/2018	Christian Salzig – Fraunhofer	Final version for evaluation
VF_GG	11/09/2018	Gary Grima – Wavestone	Review
	13/09/2018	Jens Kjaer Jorgesen – SINTEF	Review
VF_JKJ_AW	14/09/2018	Andreas Wirsén – Fraunhofer	Review

¹ Dissemination level: **PU** = Public, **PP** = Restricted to other programme participants (including the JU), **RE** = Restricted to a group specified by the consortium (including the JU), **CO** = Confidential, only for members of the consortium (including the JU)

² Nature of the deliverable: **R** = Report, **P** = Prototype, **D** = Demonstrator, **O** = Other

³ Creation, modification, final version for evaluation, revised version following evaluation, final

VR	14/09/2018	Christian Salzig – Fraunhofer	revised version following evaluation
VR_CS	18/09/2018	Christian Salzig – Fraunhofer	Evaluated at plenary meeting
VF	21/09/2018	Alejandro Simon - Wavestone	Final

Deliverable abstract

The scope of present document is to comply with Consortium Commitment Deliverable D1.4, namely “UPWARDS Data Management Plan”. It describes the initial strategy for data collection, storage, maintenance, and publication. This includes database architecture, data versioning, and metadata management.

As described in the UPWARDS GA, it will take into account the requirements for data integration and processing of UPWARDS virtual wind turbine. It will include information on how research data will be handled during and after the end of the project; what data will be collected, processed and/or generated; which methodology and standards will be applied; whether data will be shared/made open access and how data will be organized and accessible during and after the project.

Table of Contents

1. Data Summary	4
1.1. Purpose of the Data Collection / Relation to the Project Objectives	4
1.2. Types, Formats, and Utility of the collected Data	5
1.3. UPWARDS data management methodology	6
Architecture	6
Process	7
Reporting	8
2. FAIR Data	9
2.1. Making data findable, including provisions for metadata	9
Naming conventions	9
Digital Object Identifiers (DOI)	9
Metadata	9
2.2. Making data openly accessible	10
2.3. Making data interoperable	10
2.4. Increase data re-use	10
3. Allocation of Resources	11
4. Data Security	11
5. Ethical Aspects	12
6. Other Procedures	12
7. Annex A: Principles of Zenodo	13
7.1. Best Effort Principles	13
7.2. FAIR Principles	13
8. Annex B: Zenodo.org General Policies	16
8.1. Content	16
8.2. Access and Reuse	16
8.3. Removal	17
8.4. Longevity	17
9. Annex C – Data Questionnaire	18

1. Data Summary

1.1. Purpose of the Data Collection / Relation to the Project Objectives

The data collected by UPWARDS is technical data of new wind turbines required to parameterize integrated simulation model as well as social data.

Major objective 1	The UPWARDS project will establish a high-fidelity multi-physics, mechatronic and multi-scale simulation framework for wind turbines that enables integrated modelling of wind flow, mechanical movements, structural/control dynamics and stresses with a level of detail that today only is achievable in a sequential fashion meaning a comprehensive holistic vision is not possible. The collected technical data will be required to parametrize the model.
Major objective 2	UPWARDS will define a virtual prototype of a 15MW horizontal axis wind turbine including descriptions of aerodynamic design, structural design, transmission and generator, and control system. The purpose of the virtual prototype is to serve as a study case to ensure that the developed simulation tools performs as required, and enable generation of realistic and relevant simulation result for knowledge extraction and further exploitation.
Major objective 3	Using the innovative resources that will be developed as well as feedback from the public and stakeholder opinions and needs that will be gathered, the UPWARDS project will perform high-fidelity simulations of important wind turbine related phenomena to exploit and increase the understanding of their physical behavior and interaction. State of the art data mining methods will be used to extract and structure relevant information from the data that will inform new wind turbine designs.

To fulfill the above-mentioned objectives, a clear data management strategy and methodology, also enabling the ability to easily share data between partners are both necessary. The large volume of the data that will be generated during this project poses significant challenges for the implementation of such a methodology.

In order to determine the needs and preferences, a questionnaire based on the "Guidelines on FAIR Data Management in Horizon 2020" of the European Commission was prepared and distributed to all partners (cf. Annex C).

To ensure easy availability of key data sets to key stakeholders and audiences, UPWARDS data will be classified into three specific categories as follows.

GOLD	GREEN	WHITE
<p>Analysed data that has clear scientific significance, provides interesting results that have resulted in new understanding in the field, possibly supporting high impact journal publications.</p> <p>Such data will be made available online using a reliable, indexable repository that is compatible with the European Commission's OpenAIRE platform i.e. Zenodo.</p>	<p>Analysed data that is relevant for the UPWARDS partners and can be used to develop new models within the project work plans.</p> <p>Such data will be shared in a searchable repository accessible by all partners, the UPWARDS website intranet.</p>	<p>Raw data or data that has been analysed but is not thought at this stage to be significant to either the project or wider community.</p> <p>Such data may still have unseen value and will be stored in a local, searchable archive</p>

All data generated during the project will be tracked and recorded in a project database. The stored information will describe the category of the data, the type of data, a brief description of the data and where it can be located.

1.2. Types, Formats, and Utility of the collected Data

The type of data and formats of data are described and agreed in the Grant Agreement pp 37-38. The update is given in the table below:

Type of data	WP	Standards	Accessibility	Curation/preservation
Progress, interim and final reports	1	EC templates	Restricted to the project partners and the EC	Website's intranet
Raw model output data	2	Model dependent	Restricted to the project partners	Depending on data status (c.f. section 1.3)
Processed model output data (adapted to interface with the following models in the chain)	2	Model dependent	Restricted to the project partners	Depending on data status (c.f. section 1.3)
Flow results around wind turbine, strains, stresses, velocities, accelerations, local forces & moments in mechanisms, sensors and actuators data	3	Model dependent, curves, 3D graphical views and realistic animations	Restricted to the project partners	Depending on data status (c.f. section 1.3)
Wind turbine CFD database	4	Model dependent	Restricted to the project partners	Depending on data status (c.f. section 1.3)
Wind turbine near-field noise database	4	Model dependent	Restricted to the project partners	Depending on data status (c.f. section 1.3)
Report and data on the effect of fatigue loading history on damage development	5	Templates Established by partners	Public domain	website's intranet
Simulation model and data of blade substructure	5	Model dependent	Public domain	website's intranet
Experimental fatigue material	5	Templates Established by partners	Restricted to the project partners and the EC	website's intranet
Integrated system simulation	6	As in WP 2-5	As in WP 2-5	Raw data: local server Results: Depending on data status (c.f. section 1.3)
Business cases/market studies	7, 8	Templates established by partners	Restricted to the project partners and the EC. All subject of publications and related to the open	Dedicated project data management system

			pilot strategy will be made available	Open data for publications and open pilot strategies will be followed.
Dissemination data	7, 8	Defined in dissemination plan	Public	Public domain (website and intranet)

In WP6 an integrated system simulation model is generated. The data is depending on the models of WP2-WP5 and the data accessibility is defined in the corresponding WPs. Raw data and intermediate results of the simulations will be stored at partners' premises.

The impact of the UPWARDS project outcomes within and beyond its lifespan will be maximized through a systematic set of communication, dissemination, market analyses and business planning actions, and transmit the project results to the relevant stakeholders including policy makers, industry, and society.

1.3. UPWARDS data management methodology

The roadmap for data management in UPWARDS is shown in Figure 1.1. There are three key aspects that will be considered:

1. Architecture. The structure and function of the data archiving system. The physical locations of the data storage repositories and the tools required to manage and recall data.
2. Process. The process for adding to the UPWARDS data and ensuring that data is in a standardized format and easily searchable.
3. Reporting. The process of monitoring the data archives, and the ability to extract Key Performance Indicators that can inform decision making for further improvements in the Data Management plan.

Architecture

Due to the high volume of data that is expected to be generated during the project, all of the data will be required to be stored locally at each partner's site, but results of significance will be shared internally using a common UPWARDS repository and externally using standard online repositories, as shown in Figure 1.1.

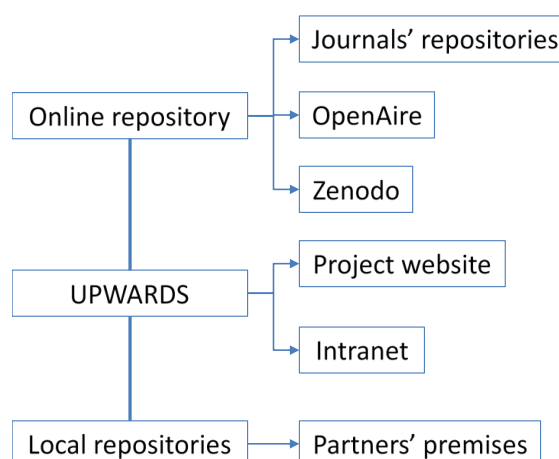


Figure 1.1 Architecture

The database with the data available will be stored in the UPWARDS intranet and kept updated.

A contribution of Upwards is to provide the integration system with all components as Open Accessible Framework. Hence, we specified to package each simulation step as software container in Docker⁴. Further, we defined a concrete format each container has to follow in order to be used in Upwards and integrated within the workflow.

- The container can be created by executing a provided build script (i.e., Dockerfile)
- The container executes the encapsulated software by passing a single command line script (here, a Bash script named runAll.sh). This script takes input data from files and writes simulation results into files. These files are stored at the host system and mounted into the container runtime.
- Licenses of proprietary or commercial software programs (e.g., StarCCM+, Samcef) and databases (e.g., ERA5) are injected via configurable license server paths by the use of environment variables.
- Commercial software binaries are injected by placing the files in dedicated directories, which are then applied and integrated by the build script.
- All resources and documentation is managed in a web-based software managing and version control system GITLAB⁵, which is hosted and provided by Fraunhofer ITWM. It is accessible to each consortium partner⁶.

Furthermore, for large data sets with green status (e.g. simulation results with significant relevance) Fraunhofer ITWM provides access to a large-scale file sharing server.

Process

The management of UPWARDS data requires standardized and timely reporting of data which can be shared within the project and with external stakeholders.

The preliminary standard process for handling new data generated during the project is defined in Figure 1.2 and will be the backbone of the definitive data management plan. This process will ensure that all data generated is captured and can be made available to both internal and external stakeholders if required. UPWARDS' partners will also be able to see that data has been recorded in advance of processing and analysis and can prepare activities in advance of the publication of results.

⁴ <https://docker.com>

⁵ <https://about.gitlab.com/>

⁶ <https://gitlab.itwm.fraunhofer.de/Adrian/upwards>

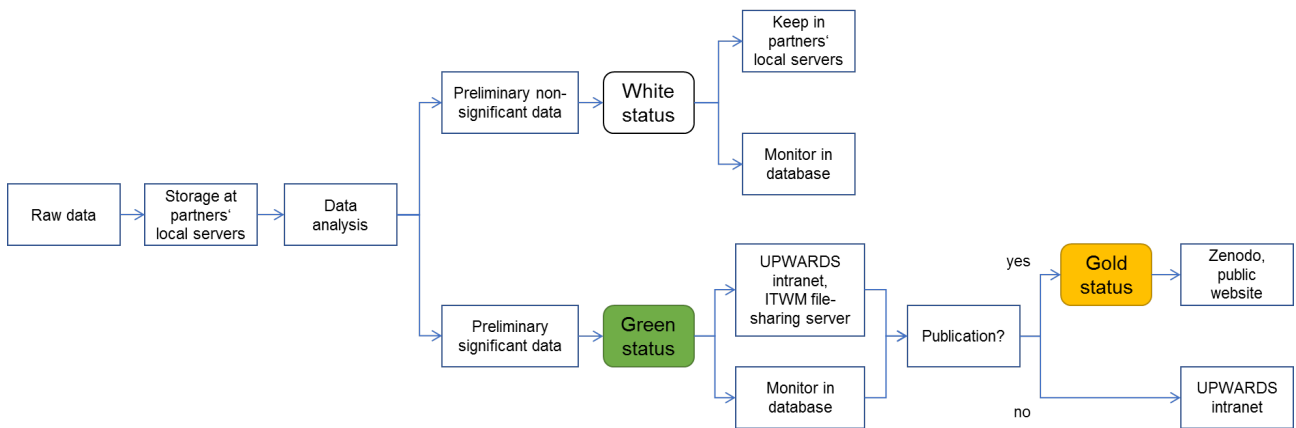


Figure 1.2 Process

Reporting

Database monitoring tools will be used for extracting key KPIs from the extensive data sets in the archive.

2. FAIR Data

UPWARDS will use the Zenodo depository (zenodo.org) to store all data which are released for free access. Zenodo is a free large capacity platform for the exchange and curation of research data managed by CERN established as a result of the OpenAIRE project. Zenodo has built in functionality to meet most FAIR criteria and to generate searchable metadata (see section 7.2).

In Zenodo data is stored in records with associated metadata. All data must be associated with a community. For that purpose, a "UPWARDS H2020 Project" community has been established that all data from UPWARDS will be associated to. In addition, data will be associated to other communities as the "Wind Energy" community.

2.1. Making data findable, including provisions for metadata

To make the open access data findable they will be assigned unique Digital Object Identifiers, descriptive names keywords and metadata. Relevant search keywords will be assigned to all data sets and included in the metadata.

Naming conventions

Data will be named using the following naming conventions:

Deliverables: [DT] UPWARDS_[DN]_[UDN].[VN]

Publications: [DT] UPWARDS_[PN]_[UDN].[VN]

[DT] Descriptive Text

[DN] Deliverable Number

[PN] Publication Number

[UDN] Unique Data Number

[VN] Version Number

Digital Object Identifiers (DOI)

DOIs for all datasets will be reserved and assigned with the DOI functionality provided by Zenodo. DOI versioning will be used to assign unique identifiers to updated versions of the data records.

Metadata

Metadata associated to each published dataset will by default be

- Digital Object Identifiers and version numbers
- Bibliographic information
- Keywords
- Abstract/description
- Associated project and communities
- Associated publications and reports
- Grant information
- Access and licensing info
- Language

2.2. Making data openly accessible

As described above, the open data is exchanged via the Zenodo depository. Metadata, including licenses for individual records and data collections, can be harvested using the OAI-PMH protocol by the record identifier and the name of the collection. Metadata can also be retrieved via the public REST API. The data can be called up on the Internet at www.zenodo.org and is therefore accessible via any web browser application. The data is freely searchable and the identity of the persons accessing the data is not determined.

2.3. Making data interoperable

All files use standard scientific notations as S.I. units and vocabulary as in ISO test standards.

The depository Zenodo, where shared data will be stored, uses JSON schemes as internal representation of metadata and offers export to other popular formats such as Dublin Core, MARCXML, BibTeX, CSL, DataCite and export to Mendeley. The data record metadata will utilize the vocabularies applied by Zenodo. For certain terms these refer to open, external vocabularies, e.g.: license (Open Definition), funders (FundRef) and grants (OpenAIRE). Reference to any external metadata is done with a resolvable URL.

2.4. Increase data re-use

The data will be licensed under different levels of creative commons licenses (<https://creativecommons.org/licenses/>). As default, the CC-BY-SA license will be applied for open UPWARDS data. This license lets others remix, tweak, and build upon the data even for commercial purposes, as long as they credit UPWARDS and license their new creations under the identical terms. This license is often compared to “copyleft” free and open source software licenses. All new works based on CC-BY-SA licensed data will carry the same license, so any derivatives will also allow commercial use. This does not preclude use of less restrictive licenses as CC-BY or more restrictive licenses as CC-BY-NC not allowing commercial usage. This will be assessed in each case.

For data published in scientific journals, the data are made available at the same time as Open Access is granted for the paper or preprint. The data will accompany the paper. For data associated with public deliverables data will be shared after approval of the deliverable by the EC.

Open data will be reusable as defined by their licenses. Data defined as confidential will not be reusable as default due to commercial exploitation. See table in section 1.2.

The data re-usability is only limited by the lifetime of the Zenodo depository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental program defined for the next 20 years at least. In cases where Zenodo expires, its policy is to transfer data and metadata to other appropriate depositories.

3. Allocation of Resources

The cost is only the PM cost required to organize and upload the data and will be covered by the project grants. Fraunhofer ITWM will be responsible for the data management with Dr. Andreas Wirsén as technical manager.

Wavestone will be in charge of updating the project related databases i.e. Zenodo, monitoring database in UPWARDS intranet and the continuous reporting feature in the EC's participant portal.

Gold Data stored on Zenodo depository: Since an externally freely usable, already financed depository is used, there are no costs for long-term archiving for the project. The longevity of the data curation is only limited by the lifetime of the Zenodo depository, which currently has an experimental program defined for the next 20 years at least

Self-archiving or so-called 'green' Open Access will be also applied through the developed Zenodo repository. As also required, open access to the publication will be ensured in a maximal delay of 6 months. The difference between gold and green open access (including related fees) can be found in the table below.

	Gold open access	Green open access
Definition	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 to (for example) the university or research institute to which the researcher is affiliated, or to the funding agency supporting the research.	Self-archiving (also called 'Green' open access) means that the published article or the final peer-reviewed manuscript is archived by the researcher – or a representative - in an online repository before, after or alongside its publication. Access to the article is often – but not necessarily - delayed ('embargo period') as some scientific publishers may wish to recoup their investment by selling subscriptions and charging pay-per-download:view fees during an exclusivity period.
Options	<ul style="list-style-type: none"> • Publish in an open access journal • Or in a journal which supports open access 	<ul style="list-style-type: none"> • Link to the article • Select a journal that features an open archive • Self-archive a version of the article
Access	<ul style="list-style-type: none"> • Public access is to the final published article • Access is immediate 	<ul style="list-style-type: none"> • Free access to a version of the article • Time delay may apply (embargo period)
Fees	<ul style="list-style-type: none"> • Open access fee is paid by the author • Fees range between \$500 and \$5,000 USD depending on the journal 	<ul style="list-style-type: none"> • No fee is payable by the author as publishing costs are covered by library subscriptions
Use	<ul style="list-style-type: none"> • Authors can choose between a commercial and noncommercial user license 	<ul style="list-style-type: none"> • Accepted manuscripts should attach a Creative Common Licence • Authors retain the right to reuse their articles for a wide range of purposes

4. Data Security

The data security is as specified by the Zenodo depository (see section 8.4):

1. **Versions:** Data files are versioned. Records are not versioned. The uploaded data is archived as Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and record are preserved.
2. **Replicas:** All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
3. **Retention period:** Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental program defined for the next 20 years at least.

4. **Functional preservation:** Zenodo makes no promises of usability and understandability of deposited objects over time.
5. **File preservation:** Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
6. **Fixity and authenticity:** All data files are stored along with a MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant.
7. **Succession plans:** In case of closure of the repository, best efforts will be made to integrate all content into suitable alternative institutional and/or subject based repositories.

If the file-sharing server for large-scale data with green status expires, ITWM will inform all partners at an early stage in order to secure the data with significant relevance.

5. Ethical Aspects

No sensitive personal data will be collected (see Grant Agreement, Annex 1 (part B), pp. 97-98).

6. Other Procedures

No other national, funder, sectorial, or departmental procedures for data management are planned.

7. Annex A: Principles of Zenodo

Source (Date: August 2018): <http://about.zenodo.org/principles/>

7.1. Best Effort Principles

Zenodo does not sign SLAs (service-level agreements). This is not a weakness, it is by design and marks a philosophy that we believe is most appropriate for Science. Instead, Zenodo is run by leading practitioners according to best practices.

What Science needs is inherent reliability, or more accurately demonstrated reliability based on open best practices. Furthermore the users should be able to influence these best practices. In the long-term, a service which is trusted is much more valuable than one for which assurances must be bought.

Service failure can never be undone. Enforcing an SLA means being prepared to litigate against the contract, which means compensation, frequently assessed on the basis of loss of revenue... but none of these concepts have any place or relevance in the free exchange of research results!

Living by these principles, Zenodo strives to make available architecture, implementation, practices and statistics. Please see for example the infrastructure page. We are also aiming to have these certified.

7.2. FAIR Principles

FAIR Principles definition as referenced from: Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).

To be Findable:

- F1: (meta)data are assigned a globally unique and persistent identifier
 - A DOI is issued to every published record on Zenodo.
- F2: data are described with rich metadata (defined by R1 below)
 - Zenodo's metadata is compliant with DataCite's Metadata Schema minimum and recommended terms, with a few additional enrichments.
- F3: metadata clearly and explicitly include the identifier of the data it describes
 - The DOI is a top-level and a mandatory field in the metadata of each record.
- F4: (meta)data are registered or indexed in a searchable resource
 - Metadata of each record is indexed and searchable directly in Zenodo's search engine immediately after publishing.
 - Metadata of each record is sent to DataCite servers during DOI registration and indexed there.

To be Accessible:

- A1: (meta)data are retrievable by their identifier using a standardized communications protocol
 - Metadata for individual records as well as record collections are harvestable using the OAI-PMH protocol by the record identifier and the collection name.

- Metadata is also retrievable through the public REST API.

A1.1: the protocol is open, free, and universally implementable

- See point A1. OAI-PMH and REST are open, free and universal protocols for information retrieval on the web.

A1.2: the protocol allows for an authentication and authorization procedure, where necessary

- Metadata are publicly accessible and licensed under public domain. No authorization is ever necessary to retrieve it.

A2: metadata are accessible, even when the data are no longer available

- Data and metadata will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.
- Metadata are stored in high-availability database servers at CERN, which are separate to the data itself.

To be Interoperable:

I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

- Zenodo uses JSON Schema as internal representation of metadata and offers export to other popular formats such as Dublin Core or MARCXML.

I2: (meta)data use vocabularies that follow FAIR principles

- For certain terms we refer to open, external vocabularies, e.g.: license (Open Definition), funders (FundRef) and grants (OpenAIRE).

I3: (meta)data include qualified references to other (meta)data

- Each referenced external piece of metadata is qualified by a resolvable URL.

To be Reusable:

R1: (meta)data are richly described with a plurality of accurate and relevant attributes

- Each record contains a minimum of DataCite's mandatory terms, with optionally additional DataCite recommended terms and Zenodo's enrichments.

R1.1: (meta)data are released with a clear and accessible data usage license

- License is one of the mandatory terms in Zenodo's metadata, and is referring to an Open Definition license.
- Data downloaded by the users is subject to the license specified in the metadata by the uploader.

R1.2: (meta)data are associated with detailed provenance

- All data and metadata uploaded is traceable to a registered Zenodo user.
- Metadata can optionally describe the original authors of the published work.

R1.3: (meta)data meet domain-relevant community standards

- Zenodo is not a domain-specific repository, yet through compliance with DataCite's Metadata Schema, metadata meets one of the broadest cross-domain standards available.

8. Annex B: Zenodo.org General Policies

Source (Date: August 2018): <http://about.zenodo.org/policies/>

8.1. Content

- **Scope:** All fields of research. All types of research artifacts. Content must not violate privacy or copyright, or breach confidentiality or non-disclosure agreements for data collected from human subjects.
- **Status of research data:** Any status is accepted, from any stage of the research lifecycle.
- **Eligible depositors:** Anyone may register as user of Zenodo. All users are allowed to deposit content for which they possess the appropriate rights.
- **Ownership:** By uploading content, no change of ownership is implied and no property rights are transferred to CERN. All uploaded content remains the property of the parties prior to submission.
- **Data file formats:** All formats are allowed - even preservation unfriendly. We are working on guidelines and features that will help people deposit in preservation friendly formats.
- **Volume and size limitations:** Total files size limit per record is 50GB. Higher quotas can be requested and granted on a case-by-case basis.
- **Data quality:** All information is provided “as-is”, and the user shall hold Zenodo and information providers supplying data to Zenodo free and harmless in connection with the use of such information.
- **Metadata types and sources:** All metadata is stored internally in JSON-format according to a defined JSON schema. Metadata is exported in several standard formats such as MARCXML, Dublin Core, and DataCite Metadata Schema (according to the OpenAIRE Guidelines).
- **Language:** For textual items, English is preferred but all languages are accepted.
- **Licenses:** Users must specify a license for all publicly available files. Licenses for closed access files may be specified in the description field.

8.2. Access and Reuse

- **Access to data objects:** Files may be deposited under closed, open, or embargoed access. Files deposited under closed access are protected against unauthorized access at all levels. Access to metadata and data files is provided over standard protocols such as HTTP and OAI-PMH.
- **Use and re-use of data objects:** Use and re-use is subject to the license under which the data objects were deposited.
- **Embargo status:** Users may deposit content under an embargo status and provide and end date for the embargo. The repository will restrict access to the data until the end of the embargo period; at which time, the content will become publically available automatically.
- **Restricted Access:** Users may deposit restricted files with the ability to share access with others if certain requirements are met. These files will not be made

publicly available and sharing will be made possible only by the approval of depositor of the original file.

- **Metadata access and reuse:** Metadata is licensed under CC0, except for email addresses. All metadata is exported via OAI-PMH and can be harvested.

8.3. Removal

- **Revocation:** Content not considered to fall under the scope of the repository will be removed and associated DOIs issued by Zenodo revoked. Please signal promptly, ideally no later than 24 hours from upload, any suspected policy violation. Alternatively, content found to already have an external DOI will have the Zenodo DOI invalidated and the record updated to indicate the original external DOI. User access may be revoked on violation of Terms of Use.
- **Withdrawal:** If the uploaded research object must later be withdrawn, the reason for the withdrawal will be indicated on a tombstone page, which will henceforth be served in its place. Withdrawal is considered an exceptional action, which normally should be requested and fully justified by the original uploader. In any other circumstance reasonable attempts will be made to contact the original uploader to obtain consent. The DOI and the URL of the original object are retained.

8.4. Longevity

- **Versions:** Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and record are preserved.
- **Replicas:** All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- **Retention period:** Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.
- **Functional preservation:** Zenodo makes no promises of usability and understandability of deposited objects over time.
- **File preservation:** Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
- **Fixity and authenticity:** All data files are stored along with a MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant.
- **Succession plans:** In case of closure of the repository, best efforts will be made to integrate all content into suitable alternative institutional and/or subject based repositories.

9. Annex C – Data Questionnaire

Partner

Name	Company/Institute/University Acronym	Date

Data Summary

Data Type – short description, its origin (method, software, ...) and relation to the project objectives
Data Format – File type? Expected file size range? Expected number of files?
Data Utility – Will the data be re-used? How? To whom might it be useful? Partners? External?

FAIR Data Principles

Making Data Findable – Is metadata generated? Keywords? Naming conventions? Version numbers?
Making Data Accessible – Which methods, software or documentation are needed to access the data? Is it possible to provide the relevant method/software/documentation? Also for external?
Access restrictions – Can the data be shared? To Partners? To External? Specify contractual, legal and voluntary restrictions! If data cannot be shared with external, are there alternatives (similar data, reduced resolution, ...)?

Making Data Interoperable – Is data exchange and re-use between institutions, countries, etc. easy? Access by open software tools? Common metadata vocabularies, standards? If not, do you provide mappings to common standards?

Increase Re-Use – When will the data be made available? Delay, e.g. to seek patents? If yes, why and how long

Data storage

Storage properties – Version control? Access control? Password? Data recovery? Secure transfer?

Storage lifetime – how long is it intended that the data remains available/re-usable after project end?

Storage costs – Are there resources available for data long-term preservation? (eligible H2020 costs)

Other Issues

Ethical aspects – Any ethical or legal aspects that can have an impact on data sharing

Comments – Anything, I've forgotten to ask.

