

Grant Agreement No 764879

# EASITRAIN

European Advanced Superconductor Innovation &amp; Training

## DELIVERABLE D1.3

# DATA MANAGEMENT PLAN

---

<b>Document identifier:</b>	EASITrain-WP1-D1.3- EDMS 1946571
<b>Due date:</b>	April, 1 <sup>st</sup> , 2018
<b>Report release date:</b>	March, 23, 2018
<b>Work package:</b>	WP1 Management
<b>Lead beneficiary:</b>	CERN
<b>Document status:</b>	IN WORK
<b>Domain:</b>	Project Management
<b>Keywords:</b>	Data management

---

### Abstract:

This document presents the initial Data Management Plan in agreement with the EC Open Research Data Pilot guidelines, applying the FAIR data management principles by M6 (**D1.3**) using the DMPOnline tool.

Initially, the goal is to make **processed datasets associated with research publications openly available**.

Given the heterogeneity of characterization setups and methods and the different types of analysis for wires and thin films, **this ITN will operate an Open Data working group** that gradually develops the most appropriate approaches and formats of data gathering and publication.

This deliverable is an Initial Data Management Plan.

Copyright notice:

Information provided with this document is subject to the H2020 EASITrain – European Advanced Superconductivity Innovation and Training grant agreement. This Marie Skłodowska-Curie Action (MSCA) Innovative Training Network (ITN) receives funding from the European Union’s H2020 Framework Programme under grant agreement no. 764879. Copyright 2017.

**Delivery Slip**

	<b>Name</b>	<b>Partner</b>	<b>Date</b>
<b>Authored by</b>	Johannes Gutleber	CERN	1 Oct 2017
<b>Edited by</b>	Simon Hopkins	CERN	1 March 2018
<b>Reviewed by</b>	Johannes Gutleber	CERN	20 March 2018
<b>Approved by</b>	Amalia Ballarino, Michael Benedikt	CERN	23 March 2018



## TABLE OF CONTENTS

- 1. DATA SUMMARY ..... 4**
  - 1.1. PURPOSE ..... 4
  - 1.2. RELATION TO THE OBJECTIVES OF THE PROJECT ..... 4
  - 1.3. TYPES AND FORMATS OF DATA ..... 4
  - 1.4. RE-USE OF EXISTING DATA ..... 5
  - 1.5. ORIGIN OF THE DATA ..... 5
  - 1.6. EXPECTED SIZE OF DATA ..... 5
  - 1.7. DATA UTILITY ..... 5
- 2. FAIR DATA ..... 6**
  - 2.1. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA ..... 6
    - 2.1.1. Discoverability ..... 6
    - 2.1.2. Identification ..... 6
    - 2.1.3. Metadata ..... 6
    - 2.1.4. Versioning ..... 8
    - 2.1.5. Naming Conventions ..... 8
    - 2.1.6. General filename and folder naming conventions ..... 10
    - 2.1.7. Data format ..... 10
    - 2.1.8. Keywords ..... 11
    - 2.1.9. Metadata standards ..... 12
    - 2.1.10. Templates ..... 13
    - 2.1.11. Storage administration and access permissions ..... 13
  - 2.2. MAKING DATA OPENLY ACCESSIBLE: ..... 14
  - 2.3. MAKING DATA INTEROPERABLE: ..... 15
  - 2.4. INCREASE DATA RE-USE (THROUGH CLARIFYING LICENSES): ..... 15
    - 2.4.1. Licence ..... 15
    - 2.4.2. Timing ..... 16
    - 2.4.3. Re-Use ..... 16
    - 2.4.4. Quality assurance ..... 16
    - 2.4.5. Validity ..... 16
- 3. ALLOCATION OF RESOURCES ..... 17**
  - 3.1. COST ESTIMATE ..... 17
  - 3.2. DATA MANAGEMENT RESPONSIBILITIES ..... 17
- 4. DATA SECURITY ..... 18**
- 5. ETHICAL ASPECTS ..... 18**
- 6. OTHER ..... 18**
- 7. APPENDIX 1 –THE EASITRAIN CERNBOX/EOS STORAGE AREA ..... 19**
  - 7.1. PREREQUISITES ..... 19
  - 7.2. STEP 1: OBTAIN THE CLIENT SOFTWARE ..... 19
  - 7.3. STEP 2: INSTALL THE CLIENT SOFTWARE ..... 19

## 1. DATA SUMMARY

### 1.1. PURPOSE

Collecting and making available the data of the analysis of superconducting materials to support the credibility and raise the quality of the scientific publications based on those data. Ease the exchange of data within the Consortium and promote the distributed characterization of samples with different methods. Permit follow-up projects and further generations of students continuing the work to build upon existing data sets, to validate the results and to document the improvement of materials and production techniques in a verifiable manner. This approach will ensure a durable impact of this EC funded project beyond the project period.

### 1.2. RELATION TO THE OBJECTIVES OF THE PROJECT

The objective of the project is to advance the performance of superconducting wires and at a later stage thin films by gaining a better understanding of the material behavior, the influence on the production techniques on the performance and to elucidate performance limitations (e.g. quality factor for superconducting thin films on substrate, current limits in wires under high-magnetic field conditions). Managed collection and publication of the data shall help establishing a durable library of results that can help documenting the performance evolution across several years and to permit other researchers validating the results independently.

### 1.3. TYPES AND FORMATS OF DATA

The openly accessible data will be the comprehensive result data sets of characterized samples that are used to create the figures and plots in scientific publications, such that other researchers can compare their results easier and such that further results including historic data can be produced quicker.

The data are value tables in *Open Document Spreadsheet format (.ODS)* for limited amounts of data with typed columns. For larger quantities of numeric data, *UTF-8 encoded, comma separated value in textual format files (.CSV)* with column value and data format description (*FORMAT.TXT*) will be used. In addition, images and raw measurement data files as provided by the measurement instruments will be stored on a project-internal data storage platform (EOS/CERNBOX). Data files and images will be included in the open data sets. Proprietary raw data delivered by the measurement instruments will not be published.

For all published files, a *document record and change track will be included* (author contact information, status, version, change reason and date, description of contents, title, origin of the data including a brief description of the measurement and/or experiment setup) in a separate metadata file for each characterization action called *METADATA.ODS*.

## 1.4. RE-USE OF EXISTING DATA

Existing data from ongoing R&D projects in the scope of the FCC study on superconducting wires and thin films will serve as a basis for the data files.

## 1.5. ORIGIN OF THE DATA

The data stem from experiments and measurement campaigns performed by the ESRs and their colleagues at the beneficiary institutes:

1. Phase A: Superconducting wires and tapes: BRUKER, CERN, COLUMBUS, CNR, TUWIEN
2. Phase B: Superconducting thin films: CERN, HZB, INFN-LNL, USIEGEN

## 1.6. EXPECTED SIZE OF DATA

The size of the data is today not known. Initial experience with storing results from different kind of measurements will permit revising this initial data management plan. The main relevant data sizes will stem from images such as microscopic sample characteristic that are stored in high-resolution bitmap format. However, the total data set size for a single sample characterization is expected to be in the order of tens of MB only.

## 1.7. DATA UTILITY

*Within the Consortium:*

The data sets will be shared within the consortium as the working baseline to produce the scientific publications, to verify and validate the results through repeated experiments at different locations and as a baseline for a comprehensive documentation of the superconducting material performance evaluation in the scope of the world-wide Future Circular Collider (FCC) technology R&D program.

*Beyond the Consortium:*

The data can be used by independent researchers to understand better the contents and conclusions of the scientific publications, which base their findings on the data. Furthermore, independent researchers can use the files to produce figures and publications, showing comparisons of their own results and the EASITRAIN results. Scientists can also use the data files to repeat the experiments and measurements to verify and validate the EASITRAIN research. Finally, the data sets may also be used by scientific writers and the press to produce high-quality infographics, demonstrating the impact potentials of the technology.

## 2. FAIR DATA

### 2.1. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

#### 2.1.1. Discoverability

The *ZENODO.ORG* platform will be used to make the data openly accessible and discoverable. In any case, the data will be made available on the *EASITRAIN collaborative web site* (<http://cern.ch/fcc/easitrain>) together with metadata describing the data sets once they are released on ZENODO. A link will also be provided at the *EASITRAIN public website* <http://easitrain.web.cern.ch>. The data will be indexed using the *EU Open Data Portal* (<http://data.europa.eu>).

Since the open data support the quality and credibility of the open publications, all data are discoverable through the scientific publications. *Each scientific publication will include Digital Object Identifiers (DOI)* that point to the associated open data sets.

#### 2.1.2. Identification

*Each data set will carry a DOI* as unique and persistent identifier. Data sets will be referenced in scientific publications and if the open data platform permits, scientific papers based on the data will be linked on the open data platform. *The DOI is reserved when a ZENODO entry is created* before any data are uploaded to the platform. At this point, the data set is not published and its visibility is classified as "*Closed Access*".

#### 2.1.3. Metadata

The data sets follow the EU Open Data Portal Metadata definitions (<https://ec.europa.eu/digital-single-market/en/news/metadata-specifications-eu-open-data-portal>). From the comprehensive set of fields, a minimum set will be provided. The Dublin Core Metadata Initiative will be followed (<http://dublincore.org>). For all characterizations at least the following metadata will be provided via the ZENODO upload form and the individual metadata files in the data folders:

- **Project identifier** (points to subfolder with the same name)
- **Title** (meaningful name of the sample characterization)
- Alternative title (additional information in concise format)
- Description (description of the dataset)
- **Keywords** (according to the library of congress terms)
- **Identifier (DOI)** reserved for this data set
- URI (link to where data are stored. Usually the path of the folder in the common storage system)
- **Dataset type** (wire, thin film, tape, bulk, power and others and the material)
- Documentation (description of the dataset content)
- Format (file type, usually a compressed archive in ZIP format)
- Issue **date** (date of first issue)
- Last modification date
- Publisher (the EASITrain consortium)
- **Contact point** (first name and last name of main contact organisation)
- Contact full address
- Contact e-mail
- Contact name
- Contact web page
- Version (Major, Minor)

- **Version** description (incremental change record)
- **Licence** (link to licence text)
- **Dataset status** (IN WORK, RELEASED, INVALID)

The top folder of the CERNBOX EASITRAIN project contains a file called *SAMPLES.ODS* that provides the most important metadata, notably the project identifier of the sample characterization that permits pointing to the folder in which the open data are stored:

- *Contact point - the organisation that has performed the characterization*
- *Issue Date - Date at which the data set has been created*
- *Contact name - Person that should be contacted for this data set*
- *Contact e-mail - E-mail address of the contact person*
- **Project identifier** - a local identifier of the dataset in format LLL-YYMMDD
- **OID** - filled if reserved in ZENODO
- **Title** (meaningful name of the sample characterization)
- *Material - type of the material analyzed (e.g. Nb, Nb<sub>3</sub>Sn, MgB<sub>2</sub>)*
- *Sample format - kind of sample (e.g. wire, film, powder)*
- *URI - link to where the data is stored, e.g. a link to the cernbox folder that is shared*

**NOTE:** It is understood that these fields are duplicates of those fields, which are also stored at lower, measurement data set folder level. The repetition at higher level serves creating a simple to use catalogue in the project.

The metadata for each sample characterization are stored in a spreadsheet called *METADATA.ODS*, which is located in the folder of a sample characterization. This file contains the following fields in different columns:

- **Title** (meaningful name of the sample characterisation)
- **Project identifier** - a local identifier of the dataset in format LLL-YYMMDD
- **OID** - filled if reserved in ZENODO
- *Contact point (name of main contact organisation)*
- *Contact full address*
- *Contact e-mail*
- *Contact name*
- *Contact web page*
- *Version (Major, Minor)*
- *Version description (incremental change record)*
- *Licence (link to licence text)*
- *Dataset status (IN WORK, RELEASED, INVALID)*
- *Documentation (description of the dataset content)*
- *Issue date (date of first issue or creation)*
- *Last modification date*
- *Material - type of the material analysed (e.g. Nb, Nb<sub>3</sub>Sn, MgB<sub>2</sub>)*
- *Sample format - kind of sample (e.g. wire, film, powder)*
- *Owner - the person who must authorise the release of marked information (highlighted in the spreadsheet)*

- *Material components* - a list of individual components in the sample
- *Dimensions* - a list of physical dimensions such as mass, diameter, length, width, thickness
- *Synthesis* - a list of steps used to create the sample, including purpose, method and environmental conditions
- *Sample history* - a record of actions at different locations that indicate, how the sample was used and characterised
- *Additional information* - free text additional comments

#### 2.1.4. Versioning

A dataset has a *major* (MM) and a *minor* (mm) version number, separated by a dot (MM.mm).

**Note:** The Zenodo recommended patch number is not used. It should always be “.0”.

If the minor version number is 0, the data set is released. Any minor version number different from zero indicates a data set that is in work. At each release the major version number is incremented by one. For each change in an "in work" version, the minor version number is incremented by one. The first "in work" version starts with a major version number equal to 0. Examples:

V 0.0 - the first draft version

V 0.1 - a second draft version

V 1.0 - the first released version

V 1.1 - an update to Version 1.0, not yet released

V 2.0 - another released version. Version 1.0 is now invalid.

Versions go together with the *dataset status*. It can either be

1. *IN WORK* (not released),
2. *RELEASED* or
3. *INVALID*.

**Note:** *INVALID* and *IN WORK* version must not be published and be referenced in publications.

#### 2.1.5. Naming Conventions

The **project identifier** for a measurement characterisation data set uses the following convention:

***LLL-YYMMDD***

where *LLL* is the three-letter abbreviation of the Consortium member institute at which the data set is created, e.g. TUV for Technische Universität Wien.

*YY* stands for the last two digits of the current year, e.g. 18 for 2018.

*MM* stands for the two digits of the current month, e.g. 04 for April.

*DD* stands for the two digits of the current day of month, e.g. 17 for the seventeenth day.

A complete example for a project identifier is *TUV-180317*.



The three letter abbreviations of an exemplary number of organisations, which typically carry out material characterizations are shown in the table below. This document will be regularly updated. The three-letter abbreviations do not coincide with the typical organisation abbreviations used to identify an organization in EC projects. They merely serve coming to unique project identifiers for sample characterizations.

<b>Organisation</b>	<b>Abbreviation</b>
CERN	<b>CRN</b>
Bruker	<b>BRK</b>
CEA	<b>CEA</b>
CNR SPIN	<b>SPN</b>
Columbus Superconductors	<b>CLB</b>
HZB	<b>HZB</b>
I-CUBE	<b>ICB</b>
INFN LNF	<b>LNF</b>
INFN LNL	<b>LNL</b>
Paul Scherrer Institute	<b>PSI</b>
TU Dresden	<b>DRS</b>
TU Bergakademie Freiberg	<b>TUF</b>
TU Wien	<b>TUW</b>
U. Siegen	<b>SIE</b>
U. Stuttgart	<b>STU</b>
WU Wien	<b>WUW</b>
U. Geneva	<b>GEV</b>
U. Genoa	<b>GOA</b>
TVEL	<b>TVL</b>
Kiswire	<b>KIS</b>
Western Superconductors	<b>WES</b>
Luvata	<b>LVT</b>

For adding additional organization abbreviations, [easitrain-ods@cern.ch](mailto:easitrain-ods@cern.ch) shall be contacted.

### 2.1.6. General filename and folder naming conventions

The filename of a dataset contains a clear, concise and very short name that identifies the contents. Words are in lowercase and separated by underscores ("\_"). CamelCase is discouraged.

Filename extensions are encouraged to ease the understanding of the folder contents.

Individual files are placed in subfolders according to this structure:

*LLL-YYMMDD* is the name of the folder for a sample characterization.

The folder contains at least the following files:

- *README.TXT* - brief description of the folder contents, authors, other useful information
- *LICENCE.TXT* - Text of the Creative Commons CC BY 4.0 licence. Additional note about the data creator if needed including specific clauses on a case-by-case basis
- *METADATA.ODS* - the metadata of the sample characterisation including the change track record

The folder contains subfolders that correspond to different material characteristics investigation methods. The following characterization methods are currently foreseen:

- *./SEM* holds all data that relates to *Scanning Electron Microscopy*
- *./TEM* holds all data that relates to *Transmission Electron Microscopy*
- *./FIB* holds all data that relates to *Focused Ion Beam analysis*
- *./Ic* holds all data that relates to *critical current measurements*

Consortium members shall report needs for further characterization methods to [easitrain-ods@cern.ch](mailto:easitrain-ods@cern.ch) so that the *TEMPLATE* folder at top level can include further examples. The file and folder contents naming convention will evolve according to project needs and with growing experience.

### 2.1.7. Data format

The combined set of sample characterization data is uploaded as a compressed archive file in .ZIP format to ZENODO together with the following separate metadata files:

- *README.TXT*
- *LICENCE.TXT*
- *METADATA.ODS*
- *./{subfolders of selected measurements to be published}*

**Note:** Each subfolder corresponding to a specific measurement action will contain only those files that the Consortium members agree to publish openly. Usually these comprise spreadsheets in open format (.ODS) and comma separated value files (.CSV) as well as high resolution images (.PNG, .TIFF, .JPEG) and high-resolution plots (.EPS, .PDF).

### 2.1.8. Keywords

Each dataset will at least be tagged with the following keywords:

1. *EASITRAIN*
2. *FCC*
3. *H2020*

In addition, appropriate keywords from the Library of Congress Subject Headings will be added (see <http://id.loc.gov/authorities/subjects.html>):

The keywords need at least to include

- *discipline*
- *sample type*
- *material*
- *properties or methods used for characterization*

A selected list of entries from the following keyword terms and the link to the keyword term need to be entered in the two distinct fields foreseen in the ZENODO upload webform:

Term	Identifier link
<b>Disciplines</b>	
Materials Science	<a href="http://id.loc.gov/authorities/subjects/sh85082094">http://id.loc.gov/authorities/subjects/sh85082094</a>
Solid state physics	<a href="http://id.loc.gov/authorities/subjects/sh85124640">http://id.loc.gov/authorities/subjects/sh85124640</a>
Electromagnetism	<a href="http://id.loc.gov/authorities/subjects/sh85042184">http://id.loc.gov/authorities/subjects/sh85042184</a>
Electrical engineering	<a href="http://id.loc.gov/authorities/subjects/sh85041666">http://id.loc.gov/authorities/subjects/sh85041666</a>
<b>Sample type</b>	
Electric conductors	<a href="http://id.loc.gov/authorities/subjects/sh85041622">http://id.loc.gov/authorities/subjects/sh85041622</a>
Electric wire	<a href="http://id.loc.gov/authorities/subjects/sh85042050">http://id.loc.gov/authorities/subjects/sh85042050</a>
Thin films	<a href="http://id.loc.gov/authorities/subjects/sh85134864">http://id.loc.gov/authorities/subjects/sh85134864</a>
Coatings	<a href="http://id.loc.gov/authorities/subjects/sh85027504">http://id.loc.gov/authorities/subjects/sh85027504</a>
Thin film devices	<a href="http://id.loc.gov/authorities/subjects/sh85134863">http://id.loc.gov/authorities/subjects/sh85134863</a>
Powders	<a href="http://id.loc.gov/authorities/subjects/sh85105962">http://id.loc.gov/authorities/subjects/sh85105962</a>
Powder metallurgy	<a href="http://id.loc.gov/authorities/subjects/sh85105959">http://id.loc.gov/authorities/subjects/sh85105959</a>
<b>Materials – general</b>	
Materials	<a href="http://id.loc.gov/authorities/subjects/sh85082065">http://id.loc.gov/authorities/subjects/sh85082065</a>
Electrical engineering--Materials	<a href="http://id.loc.gov/authorities/subjects/sh85041671">http://id.loc.gov/authorities/subjects/sh85041671</a>
Nanostructured materials	<a href="http://id.loc.gov/authorities/subjects/sh93000864">http://id.loc.gov/authorities/subjects/sh93000864</a>
Nanoparticles	<a href="http://id.loc.gov/authorities/subjects/sh85089689">http://id.loc.gov/authorities/subjects/sh85089689</a>
<b>Materials – specific</b>	
Niobium compounds	<a href="http://id.loc.gov/authorities/subjects/sh99000377">http://id.loc.gov/authorities/subjects/sh99000377</a>
Magnesium diboride	<a href="http://id.loc.gov/authorities/subjects/sh2003001055">http://id.loc.gov/authorities/subjects/sh2003001055</a>

High temperature superconductors	<a href="http://id.loc.gov/authorities/subjects/sh88004641">http://id.loc.gov/authorities/subjects/sh88004641</a>
Copper oxide superconductors	<a href="http://id.loc.gov/authorities/subjects/sh88004322">http://id.loc.gov/authorities/subjects/sh88004322</a>
Rare earth metal compounds	<a href="http://id.loc.gov/authorities/subjects/sh85111455">http://id.loc.gov/authorities/subjects/sh85111455</a>
Rare earth oxide thin films	<a href="http://id.loc.gov/authorities/subjects/sh2007000755">http://id.loc.gov/authorities/subjects/sh2007000755</a>
Yttrium compounds	<a href="http://id.loc.gov/authorities/subjects/sh2012003015">http://id.loc.gov/authorities/subjects/sh2012003015</a>
Thallium compounds	<a href="http://id.loc.gov/authorities/subjects/sh85134478">http://id.loc.gov/authorities/subjects/sh85134478</a>
Metal powders	<a href="http://id.loc.gov/authorities/subjects/sh85084071">http://id.loc.gov/authorities/subjects/sh85084071</a>
<b>Properties</b>	
Electric conductivity	<a href="http://id.loc.gov/authorities/subjects/sh85041621">http://id.loc.gov/authorities/subjects/sh85041621</a>
Superconductivity	<a href="http://id.loc.gov/authorities/subjects/sh85130584">http://id.loc.gov/authorities/subjects/sh85130584</a>
Critical currents	<a href="http://id.loc.gov/authorities/subjects/sh92000808">http://id.loc.gov/authorities/subjects/sh92000808</a>
Magnetization	<a href="http://id.loc.gov/authorities/subjects/sh85079774">http://id.loc.gov/authorities/subjects/sh85079774</a>
Magnetic fields	<a href="http://id.loc.gov/authorities/subjects/sh85079703">http://id.loc.gov/authorities/subjects/sh85079703</a>
Electric conductivity	<a href="http://id.loc.gov/authorities/subjects/sh85041621">http://id.loc.gov/authorities/subjects/sh85041621</a>
Microstructure	<a href="http://id.loc.gov/authorities/subjects/sh85084931">http://id.loc.gov/authorities/subjects/sh85084931</a>
<b>Characterisation</b>	
Physical measurements	<a href="http://id.loc.gov/authorities/subjects/sh85101564">http://id.loc.gov/authorities/subjects/sh85101564</a>
Microscopy	<a href="http://id.loc.gov/authorities/subjects/sh92003369">http://id.loc.gov/authorities/subjects/sh92003369</a>
Metallography	<a href="http://id.loc.gov/authorities/subjects/sh85084132">http://id.loc.gov/authorities/subjects/sh85084132</a>
Scanning electron microscopy	<a href="http://id.loc.gov/authorities/subjects/sh91002757">http://id.loc.gov/authorities/subjects/sh91002757</a>
Transmission electron microscopy	<a href="http://id.loc.gov/authorities/subjects/sh93001918">http://id.loc.gov/authorities/subjects/sh93001918</a>
X-ray spectroscopy	<a href="http://id.loc.gov/authorities/subjects/sh85148744">http://id.loc.gov/authorities/subjects/sh85148744</a>
Windowless energy-dispersive X-ray analysis	<a href="http://id.loc.gov/authorities/subjects/sh87006566">http://id.loc.gov/authorities/subjects/sh87006566</a>
X-ray crystallography	<a href="http://id.loc.gov/authorities/subjects/sh85148730">http://id.loc.gov/authorities/subjects/sh85148730</a>
Electric measurements	<a href="http://id.loc.gov/authorities/subjects/sh85041828">http://id.loc.gov/authorities/subjects/sh85041828</a>
Electromagnetic measurements	<a href="http://id.loc.gov/authorities/subjects/sh85042173">http://id.loc.gov/authorities/subjects/sh85042173</a>

### 2.1.9. Metadata standards

The Dublin Core Metadata Initiative will be followed (<http://dublincore.org>) as much as reasonably applicable. Metadata examples concerning the sample measurement campaigns such as <http://icatproject-contrib.github.io/CSMD/csmd-4.0.html> and an existing sample database at CERN have been considered. However, to our best knowledge, no domain-specific metadata standard for those sample characteristics identification campaigns specified in the EASITrain project exist. Therefore, a column-oriented data format with an explanation of the columns will be created in the scope of this project.

This is one tangible outcome of the EASITRAIN Data Management Plan initiative.

### 2.1.10. Templates

Template metadata files can be found in the EASITRAIN cernbox folder

*cernbox\_fcc/easitrain/data/TEMPLATES*

The entire template folder can be copied into the *./data* folder and can be renamed according to the naming convention XXX-YYMMDD for a new sample characterisation. The folder contains exemplary

- *README.TXT*
- *LICENCE.TXT*
- *METADATA.ODS*
- *a list of exemplary folders corresponding to measurement methods*

### 2.1.11. Storage administration and access permissions

Data sets are managed via CERN's scalable storage system EOS and the Cloud storage front-end CERNBOX (<http://www.cernbox.ch>).

Consortium members who need to use this storage platform need to obtain either a CERN IT account or a CERN lightweight account (see <https://account.cern.ch/account/externals/>). Both methods require compliance with CERN's computer security rules and guidelines as well as appropriate training.

The cernbox folder for the EASITrain data sets is located in the EOS scalable storage system. The folder path is */eos/project/fccproject/easitrain*. The folder can be directly reached on the Linux operating filesystem via the *LXPLUS.cern.ch* Linux cluster. 1 TB of storage has originally been allocated. The system is backed up and implements automatic versioning based on file changes.

The owner of the storage is the service user *fccoff*. Administration permissions are managed with CERN's e-group (<http://cern.ch/egroups>) system. The administration e-group is *cernbox-project-fccproject-admins*.

Access permission requests, requests for data set releases and folder administration requests must be set to [easitrain-ods@cern.ch](mailto:easitrain-ods@cern.ch).

Two main groups are used to manage read-only and read-write permissions on the sub-tree:

- *cernbox-project-fccproject-readers* - anyone in this egroup can read in the project space
- *cernbox-project-fccproject-writers* - anyone in this egroup can read, write and delete in the project space. *cernbox-project-fccproject-admins* is also member of this group.

However, in this project, permissions are not granted at top folder level! Finer granularity is *initially* given at the easitrain and its subfolders via the following egroups:

*easitrain-ods-readers* - persons who can create and edit data sets and their metadata files

*easitrain-ods-writers* - persons who can ready the data sets and their metadata files

Membership is decided on a case-by-case basis and is exclusively granted by consensus agreement of the data owners and by requesting access permission to [easitrain.office@cern.ch](mailto:easitrain.office@cern.ch).

## 2.2. MAKING DATA OPENLY ACCESSIBLE:

Data from characterization campaigns will be made openly available only after approval of all persons who were involved in that characterization.

*Note: At this stage of the project, all publishing is carried out by the EASITRAIN project office until the Consortium members have built up sufficient expertise and confidence in the process of open data creation.*

*Note: Persons publishing data on the ZENODO platform need to sign up with an account at that platform at <https://zenodo.org/signup/>.*

To release data for the public, the following steps need to be performed:

1. One person is the *main contact* for the characterization campaign of a sample. The person's contact details are stored in the *METADATA.ODS* file of that characterization.
2. Contact person consolidates all measurement data in the cernbox folder of the data set labelled *XXX-YYMMDD* (see folder naming convention above).
3. Ask [easitrain-ods@cern.ch](mailto:easitrain-ods@cern.ch) to create a data set record on [zenodo.org](https://zenodo.org) and reserve an *OID*. Inform [easitrain-ods@cern.ch](mailto:easitrain-ods@cern.ch) if the data set needs to be anonymized or if it can be published "as is". [easitrain-ods@cern.ch](mailto:easitrain-ods@cern.ch) will only publish fields marked for publication (marked *NO* in the column that indicates if the field shall be public in the metadata spreadsheets) if the publication type is "anonymized".
4. Supply all metadata in cernbox (top level *SAMPLES.ODS* file and data set specific *METADATA.ODS* file), in particular the *OID*
5. Contact person obtains written agreement from all dataset stakeholders (usually the author of the associated publication) and marks the status of the data set as *RELEASED* on *METADATA.ODS*.
6. Request publication of the dataset to [easitrain-ods@cern.ch](mailto:easitrain-ods@cern.ch)
7. Easitrain office assistant will create a publicly sharable, read-only link for the data set and add it in the metadata files
8. Main contact person creates a *ZIP* file of the data set folder that contains only those files that the Consortium partners have agreed to openly publish in the *LLL-YYMMDD* data set folder. The *ZIP* file follows the naming convention *LLL-YYMMDD\_V{MMmm}*, where {*MMmm*} are the major and minor version numbers, e.g. *V0100* for released version 1.0.
9. EASITrain office assistant uploads the data set *ZIP* file to *ZENODO*, filling out the requested *ZENODO* metadata record. The data set is not publicly visible at this point.
10. EASITrain office assistant links the data set on the <http://cern.ch/fcc/easitrain> collaboration site and on the <http://easitrain.web.cern.ch> public project site.
11. EASITrain office assistant notifies all named stakeholders (usually the authors of the publication associated with the data set), providing them the cernbox and *ZENODO* links and asks for review of the information within 5 working days.
12. Upon agreement or when the deadline has expired, the EASITrain office assistant makes the dataset public sends a notification to the entire Consortium mailing list with the link to the data set on cernbox and on *ZENODO*.

There will be different access levels. Sensitive data marked in "light red" color in the metadata file *METADATA.ODS* will be kept on the Consortium's data repository only. In this case, the ZIP file published on ZENODO will contain files in which those fields are cleared and the link to the cernbox folder will not be published on ZENODO. In this case, access for third parties needs to be requested to CERN, the consortium coordinator via [easitrain-ods@cern.ch](mailto:easitrain-ods@cern.ch). Access to production data (details sample composition, origin and other information, which may be subject to IP regulations) needs to be requested to the project management through CERN, the network coordinator. The requestor needs to justify the request and specify the purpose (e.g. academic use, no further dissemination of details). If access is granted by the campaign management and (if applicable) the producer, the specific file version is communicated in electronic form to the requestor. Access may require additional legal procedures such as the creation of a Non-Disclosure Agreement.

### 2.3. MAKING DATA INTEROPERABLE:

The data, metadata and documentation follow open standards and file formats and use controlled vocabularies to ease interoperability and re-use.

The domain-specific meta-data schema will be documented in a final version of the Open Data Management Plan. It is considered to be submitted to the RDA Metadata Standards Directory Working Group (<http://rd-alliance.github.io/metadata-directory/>).

### 2.4. INCREASE DATA RE-USE (THROUGH CLARIFYING LICENSES):

#### 2.4.1. Licence

Anonymized, openly accessible data will be licensed under *Creative Commons CC BY 4.0* (see <https://creativecommons.org/licenses/by/4.0/>).

Users are free to:

**Share** — copy and redistribute the material in any medium or format

**Adapt** — remix, transform, and build upon the material

for any purpose, also commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

**Attribution** — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

**No additional restrictions** — You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable [exception or limitation](#).

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as [publicity, privacy, or moral rights](#) may limit how you use the material.

### 2.4.2. Timing

Data will at latest be made available with the publication of an accompanying scientific publication that references the data sets. All data will at latest be available with the project end, even those data sets, which are not referenced in scientific publications.

### 2.4.3. Re-Use

Anonymized data can be used by other scientists in the fields of superconducting wires, thin-films, magnets, radio frequency cavities, electronics, bulk superconductors and potential downstream application devices.

Original data revealing the sample production technique and/or producer will only be usable by researchers upon explicit request and approval of the characterization campaign manager and the sample producer.

### 2.4.4. Quality assurance

Data sets, metadata and measurement setup and procedure description will be reviewed by at least one peer prior to engaging the release procedure. The author and the reviewer are named in the metadata.

Data sets, metadata and measurement setup and procedure description will be marked **RELEASED** only after approval of the measurement campaign manager (e.g. project supervisor) and one additional reviewer. The approvers are named in the metadata.

Review and release includes a validation of the measured sample, the measurement setup, conditions, procedure and equipment as well as sanity checks against similar studies and control of systematic errors.

In case of data quality uncertainties after release, a new version **IN WORK** is created and the released data set version is marked **INVALID**.

The description of measurement setup (materials and method) are annotated with product references.

The measurement conditions are described.

The measurement location, date and time (periods) are noted.

Any potential and known adverse effects (environmental influences, influences of the measurement equipment) are described in the metadata.

### 2.4.5. Validity

The data will remain usable until the repository withdraws the data or goes out of business.



### 3. ALLOCATION OF RESOURCES

#### 3.1. COST ESTIMATE

A person at CERN will keep the measurement data sets and perform the publication in the open data repository. **The estimated effort is 4 hours per data set, 10 data sets per year, i.e. 40 hours or 1 week per year** over the entire project period. This resource is covered by the project management funds and CERN matching resources.

*Note: the project Coordinator will track the actual efforts and regularly update this estimation.*

Each researcher in the project is responsible to create the data sets using the adopted open data format, providing the metadata files, describing the measurement setup, anonymising the data, reviewing the data sets and performing the release process using the CERN provided storage infrastructure (EOS, cernbox) and the ZENODO platform. **The estimated effort is 20 hours per data set, 10 data sets per year, i.e. five working weeks per year** over the entire project period. This resource is covered by the institutes who carry out the measurement campaigns.

*Note: The participating institutes are strongly encouraged to track the time they are spending to prepare the data sets and to publish them and to report their actual estimates to the Coordinator.*

#### 3.2. DATA MANAGEMENT RESPONSIBILITIES

This data management plan is maintained by Johannes Gutleber (CERN). All work package leaders and deputies commit to cooperate on the establishment of this DMP and to deliver the required information such that the associated deliverables and milestones can be produced in due time with the requirement quality levels:

Data storage and backup responsibilities are covered by the data repository providers. The CERN project repository is managed by CERN IT department. Johannes Gutleber is the site manager. Efrat Tal Hod (Tel Aviv University) provides support for the upload to the CERNBOX/EOS data storage system and Emilie David (CERN) performs a formal (file integrity, naming, metadata completeness) check.

Long-term data preservation will be ensured by CERN at no additional cost.

## 4. DATA SECURITY

All data delivered to the CERN project repository EOS/CERNBOX storage system is backed up by CERN's central IT services. In addition, a copy of released data will be kept on the ZENODO platform. Both services are intended for long-term storage of scientific research data. Upon unintentional loss of data (misuse of the collaborative workspace, accidental removal), Efrat Tal Hod (Tel Aviv University) and Johannes Gutleber (CERN) need to be contacted via email to [easitrain-ods@cern.ch](mailto:easitrain-ods@cern.ch). They will interact with CERN's IT services to restore the latest known copy. No additional costs occur for storage, backup and restore activities.

Nonpublic data sets can be provided by the project members using HTTPS transfer protocol after authentication by sharing a cernbox link by asking the sharing to [easitrain-ods@cern.ch](mailto:easitrain-ods@cern.ch), i.e. in a reasonable secure fashion that counteracts data manipulation.

Sensitive data, i.e. non-anonymized data sets can only be accessed by the author, the measurement campaign management and the project's IT managers (Johannes Gutleber and Efrat Tal Hod). Access to sensitive data may be granted through a request to the network coordinator (CERN) with a justification for the request. Access will be granted on a case-by-case basis in agreement with the measurement campaign manager and, if samples and products from industrial partners are involved, in agreement with the material producers. The data will be communicated in electronic format from the network coordinator to the data requestor in digitally signed and encrypted form. An additional IP access process, such as the establishment of a Non-Disclosure Form may apply.

Every consortium member must inform the network coordinator without delay if a person affiliated (associated or employed) with the institute and who has access to the project data, leaves the institute. In this case, the network coordinator will revoke as soon as technically possible and resources permitting (working hours) the access of the person to the data.

*Note: E-mail is not considered a secure communication channel for data and metadata files. Data can be modified and it is unclear what fields have been modified with respect to the original data source. Therefore only a link to the authentic data source shall be considered reliable information.*

## 5. ETHICAL ASPECTS

Sensitive information will be kept secure. Access to non-anonymized data is managed by the network coordinator in close cooperation with the organisation, who provides the data set. Non-anonymized data will only be communicated in encrypted fashion and digitally signed.

## 6. OTHER

Contact [easitrain-ods@cern.ch](mailto:easitrain-ods@cern.ch) by e-mail for any questions concerning the data sets and their management in the scope of the EASITRAIN H2020 project.

## 7. APPENDIX 1 –THE EASITRAIN CERNBOX/EOS STORAGE AREA

This section gives a step-by-step guideline on how to gain access to the CERNBOX/EOS storage area of the EASITrain sample characterisation data set area.

### 7.1. PREREQUISITES

The described process can only be performed if two prerequisites are met:

1. You have a valid CERN computing account or a valid CERN external account
2. You have been granted read or write access permissions to the EOS/CERNBOX storage area.

Please contact [easitrain-ods@cern.ch](mailto:easitrain-ods@cern.ch) before continuing in order to ensure that the two prerequisites are met!

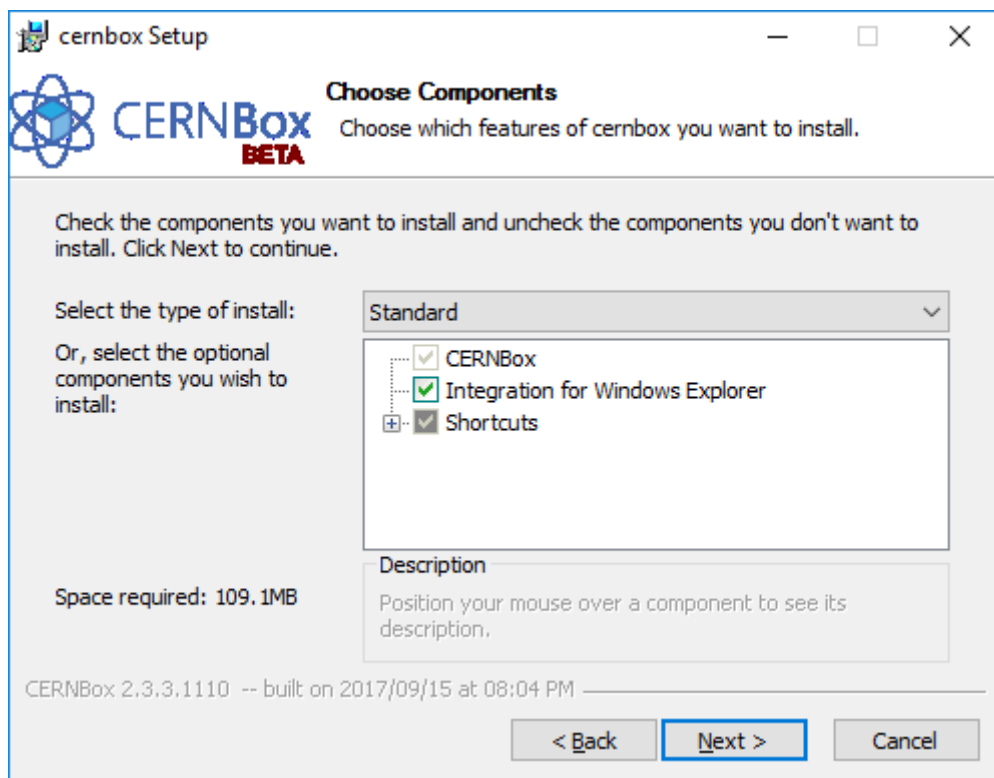
### 7.2. STEP 1: OBTAIN THE CLIENT SOFTWARE

Connect to <http://cernbox.cern.ch> with a Web browser and download the desktop synchronisation client for your local computing platform.

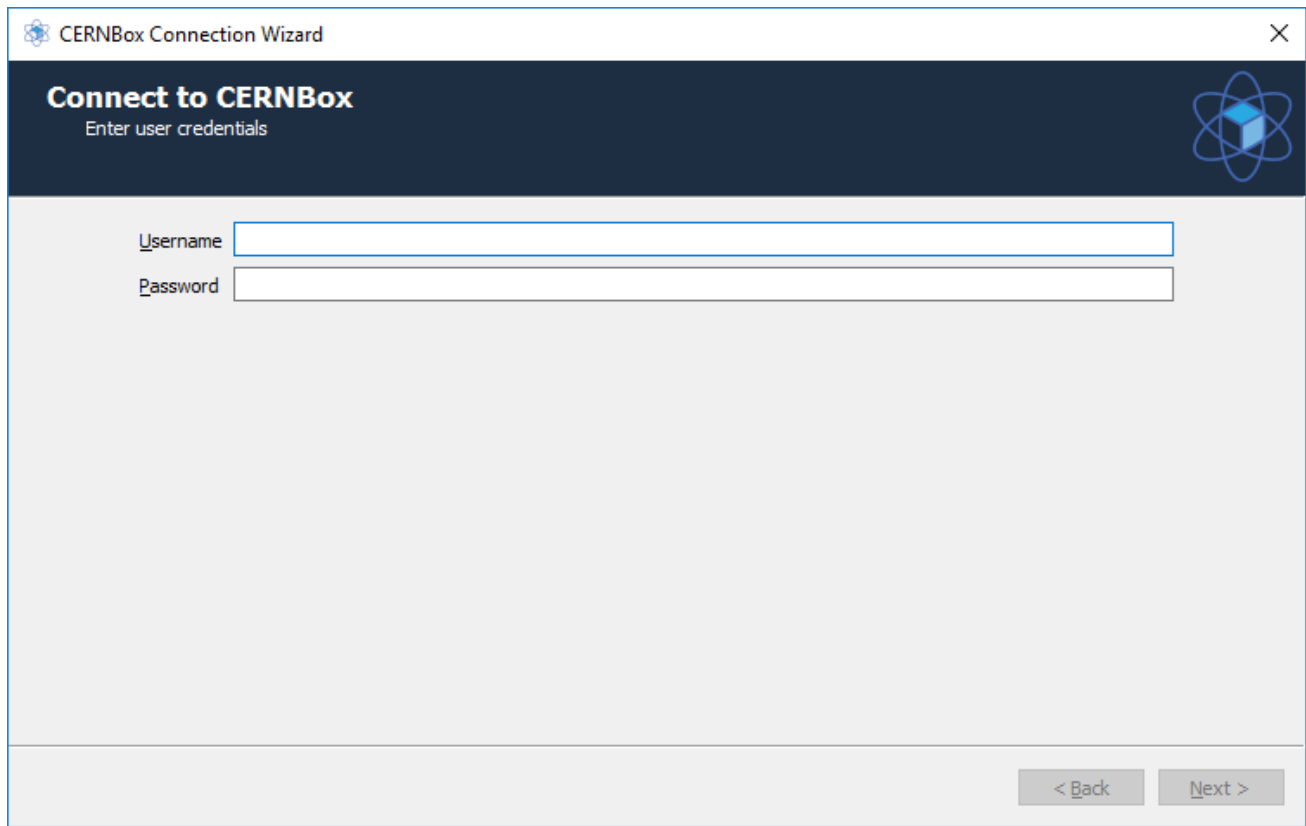
**Note:** Do not use the Web browser access to work on data sets. This method is inefficient for the amount of files and folders to be managed for this project.

### 7.3. STEP 2: INSTALL THE CLIENT SOFTWARE

Open the installer and follow the instructions.



Enter your CERN account or lightweight account credentials when asked:



CERNBox Connection Wizard

**Connect to CERNBox**  
Enter user credentials

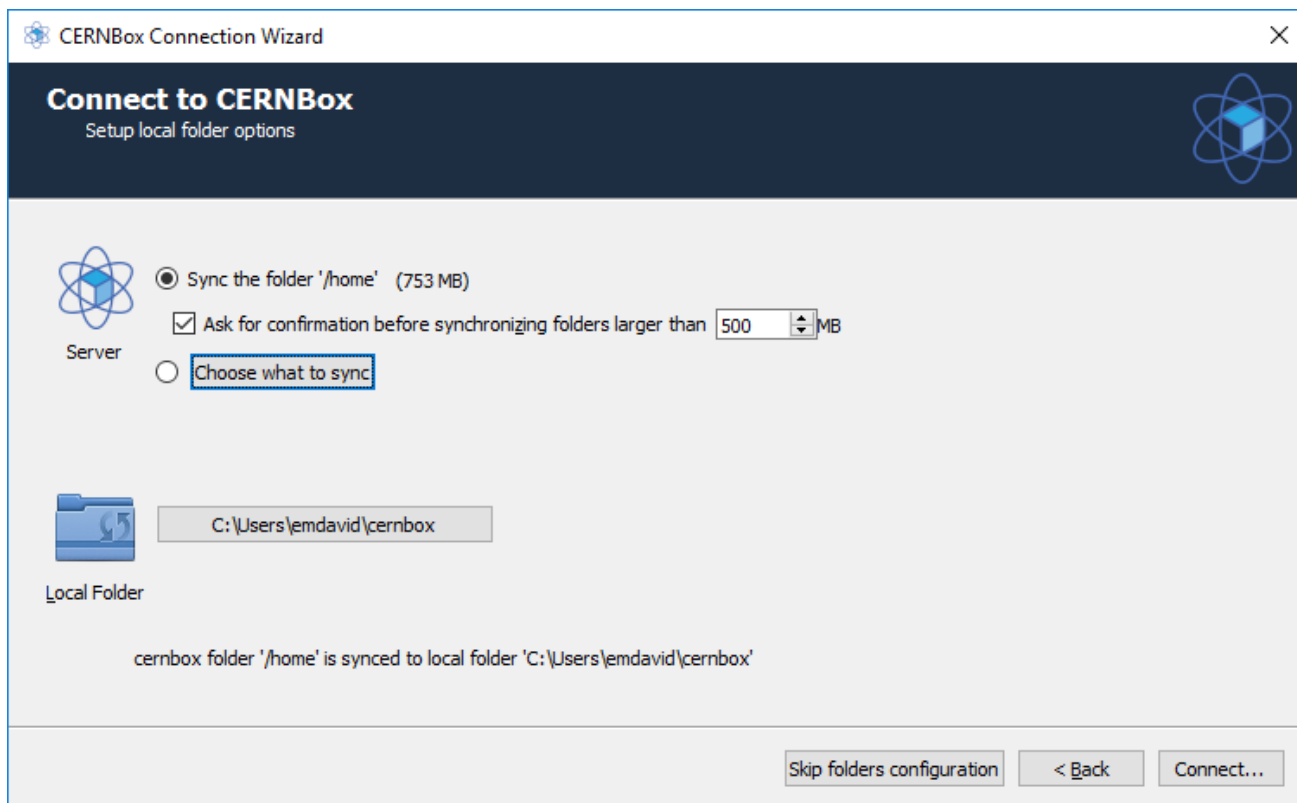
Username

Password

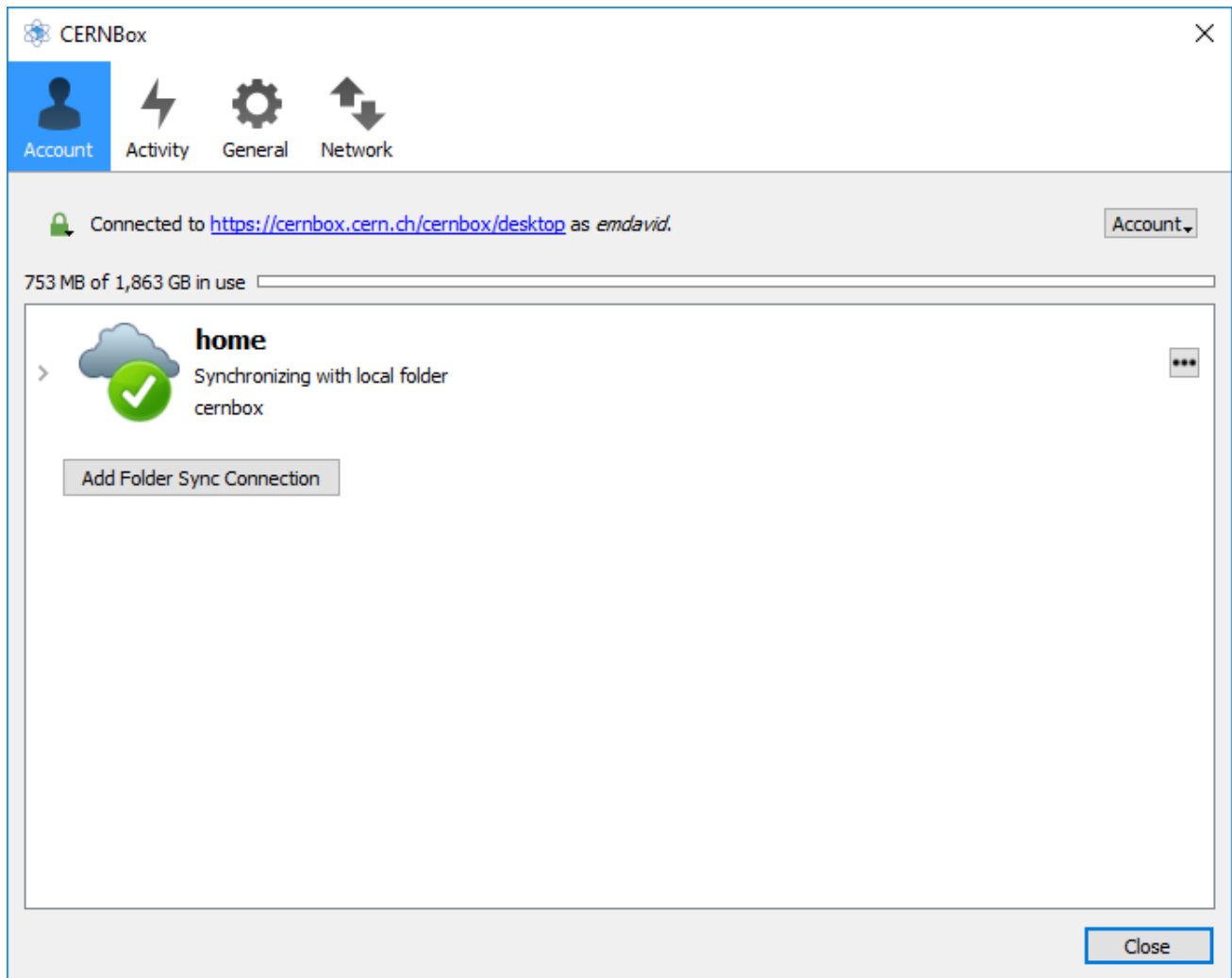
< Back    Next >

Create a folder in your local storage system (harddisk) with a specific name that you want to use for synchronising personal files with cernbox. E.g. call the folder “cernbox”.

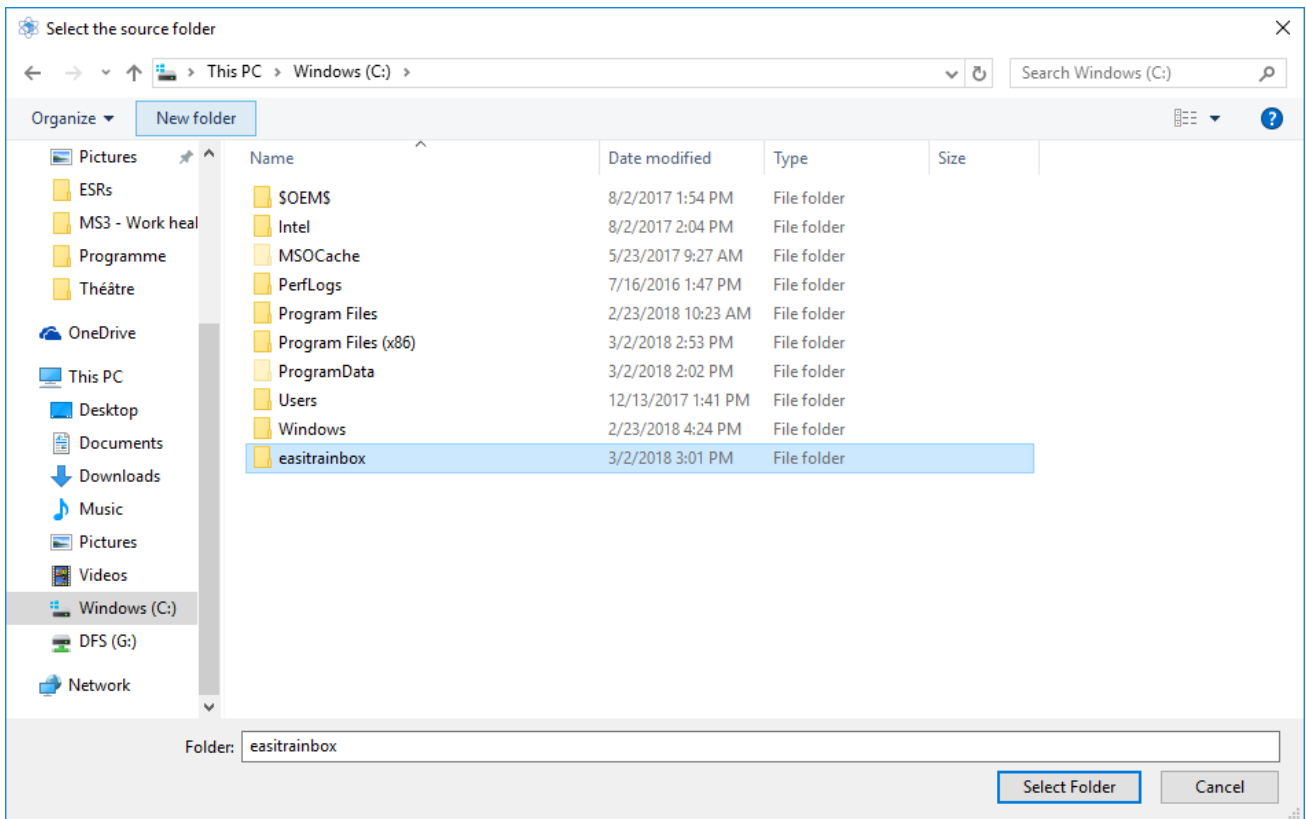
Carefully choose if you want to synchronize folders that are larger than a certain size (e.g. 500 MB) automatically or manually. If you choose to set a limit, do not forget that files are not automatically downloaded to your computer when you are connected or uploaded to the storage system when you make a modification. On the other side, if you choose automatic synchronisation, make sure to pause the cernbox synchronisation when you switch from wired or WIFI network to a mobile-phone network in order to keep the traffic and thus potential additional costs limited!



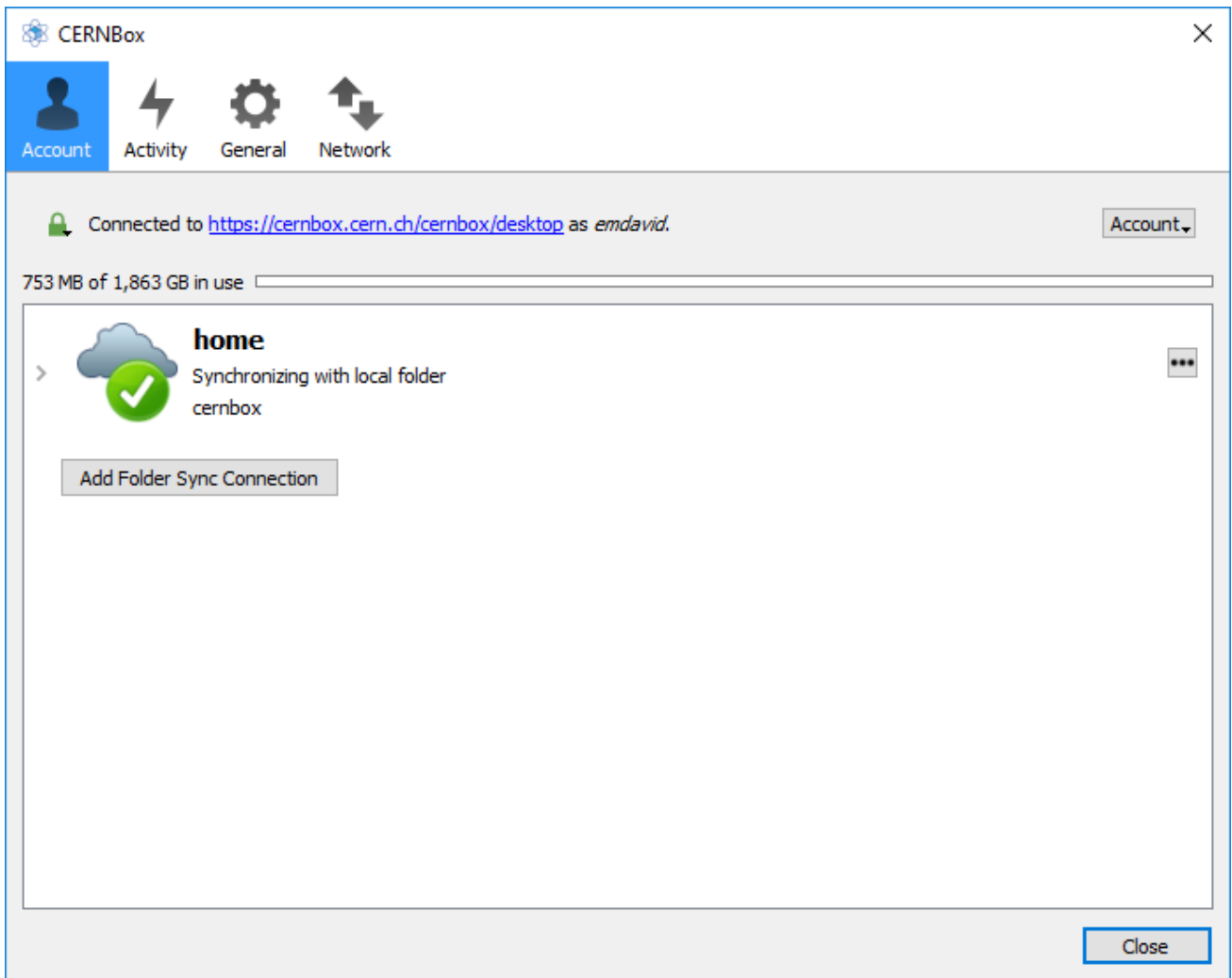
If the installation is successful and your credentials are correct, the client software will perform a first synchronisation of your selected local folder with your personal cernbox cloud storage area. Wait until this synchronisation completes.



Next, create another local folder that you will use to store the data sets in the scope of the project. In this example the folder is called “easitrainbox”.

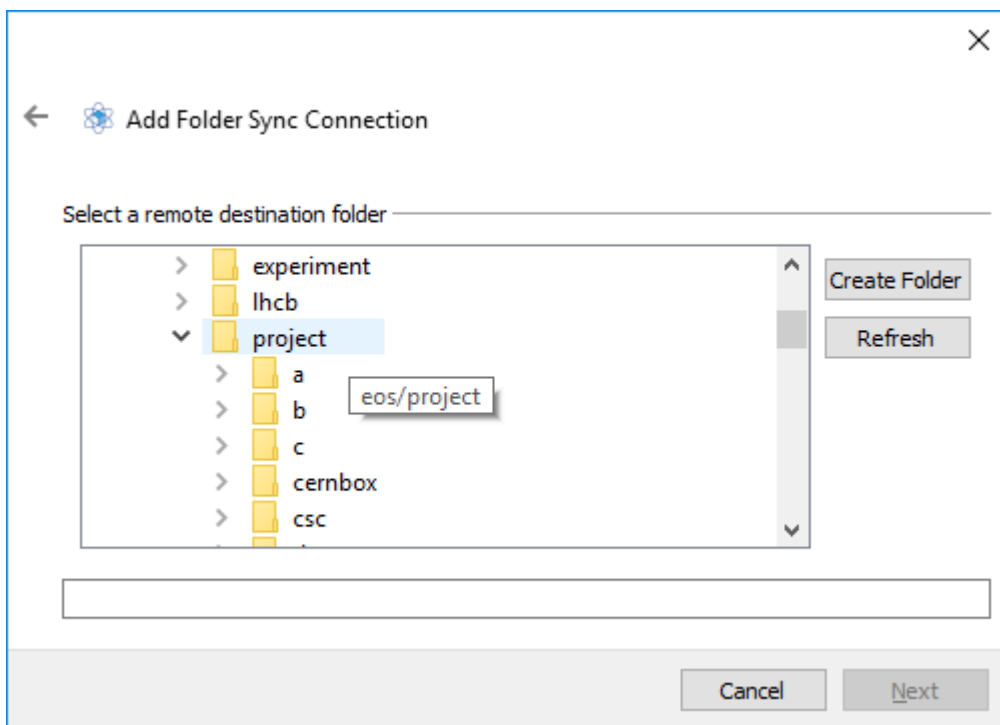
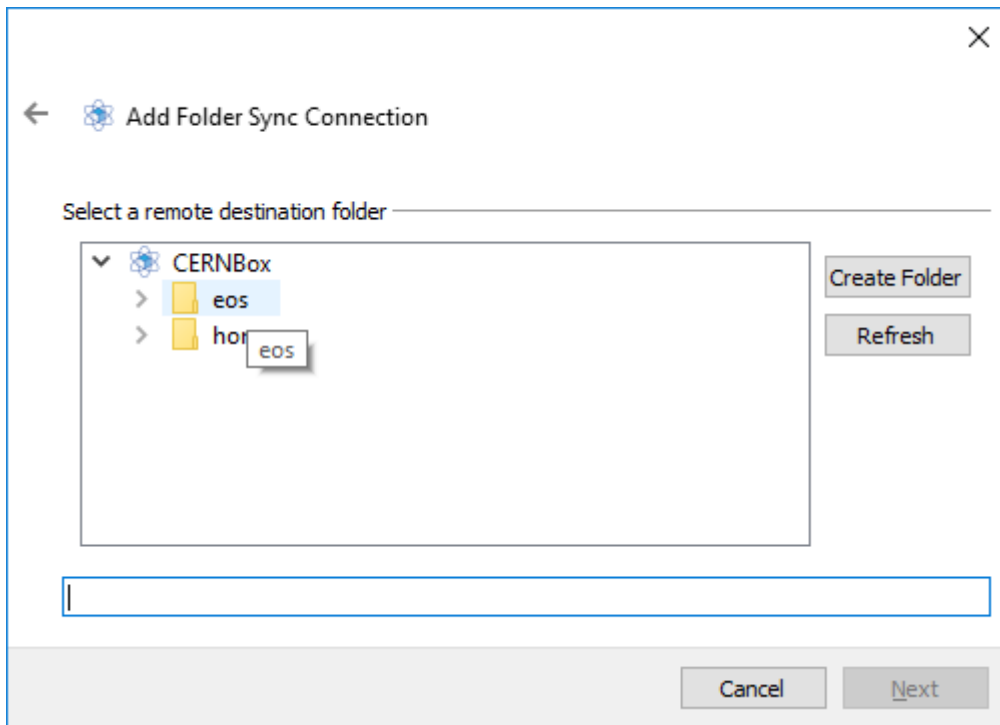


In the client software system, choose “Add Folder Sync Connection” to add another folder synchronization as shown below.

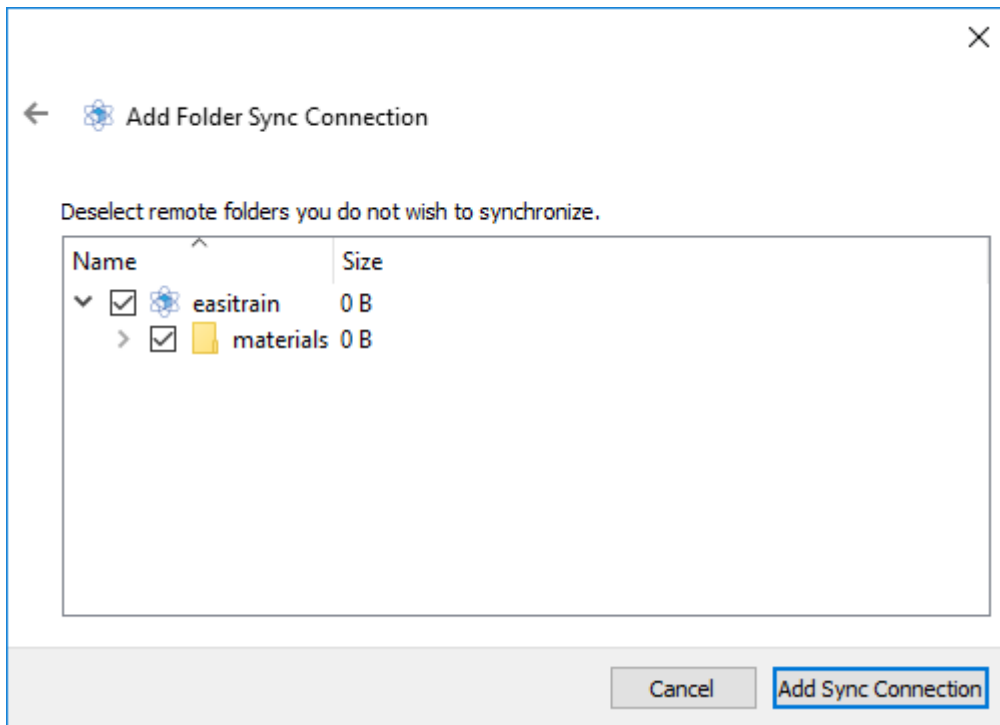


Select your locally created folder “easitrainbox” and continue. Now select the EOS cloud storage path of the easitrain project as shown below (“eos/project/f/fccproject/easitrain”).

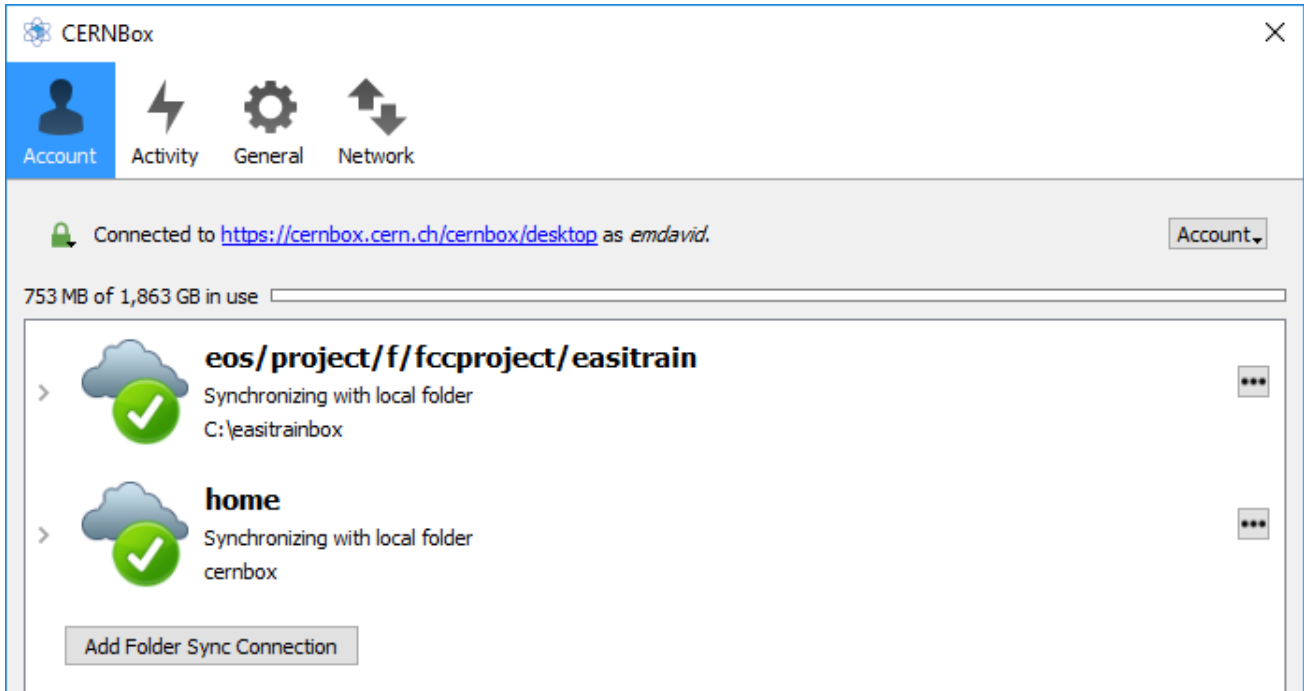




Depending on the current state of the project, the easitrain folder may contain several subfolders such as data, documents, materials, web and others. You may either choose to select the entire easitrain project folder for synchronization or individual subfolders.



If your access permissions are correct, the selection process finishes as outlined below with a first synchronization process.



## 8. APPENDIX 2 – ZENODO

### 8.1. REGISTER FOR AN ACCOUNT

The ZENODO open data platform is reachable at <http://zenodo.org>

All project members working on material characterisation are encouraged to create an account by signing up (see Sign up button on the top right of the platform page).



ZENODO is fully supported by the project coordinator CERN (see logo on sign up page below), ensuring long-term availability of the data sets.

## Research. Shared! Sign up today.

**Citeable. Discoverable.**  
Uploads get a Digital Object Identifier (DOI) to make them easily and uniquely citeable.

**Communities**  
Accept or reject uploads to your own community (e.g workshops, EU projects, institutions or entire disciplines).

**Trusted Research Data Management**  
Built on top of CERN's expertise in managing 100s of petabytes of research data from the Large Hadron Collider.

Sign up with GitHub


Sign up with ORCID

— OR —

Email Address




Username

Password

I'm not a robot  reCAPTCHA  
Privacy - Terms

Sign Up

Already have an account? Log In

## 8.2. EASITRAIN COMMUNITY

An EASITrain community has been created on the ZENODO platform as entry point to all data sets produced in this project. The following URLs exist:

**Collection URL:**

<https://zenodo.org/communities/easitrain/>

Above address links directly to your community collection.

**Upload URL:**

<https://zenodo.org/deposit/new?c=easitrain>

Above address will automatically ensure people who use it will have their record added to your community collection.

**Curation URL:**

<https://zenodo.org/communities/easitrain/curate/>

Above address links to your private curation URL. You will find all uploads pending your curation.

**Harvesting URL:**

[Zenodohttps://zenodo.org/oai2d?verb=ListRecords&set=user-easitrain&metadataPrefix=oai\\_dc](https://zenodo.org/oai2d?verb=ListRecords&set=user-easitrain&metadataPrefix=oai_dc)

Above address links to a OAI-PMH feed, which can be used by other digital repositories to harvest this community.

## 8.3. UPLOAD

To upload a new data set, click on “Communities”, search the “easitrain” community and then click on the “View” button:

## Communities created and curated by Zenodo users

easitrain

Showing 0 to 1 out of 1 communities.

 Sort by ▾

### European Advanced Superconductivity Innovation and Training


View

Curate

The two greatest obstacles to a wide spread adoption of superconductivity remain the limited understanding of its fundamental principles and the yet insufficient capability for large-scale, cost-effective deployment of the technology. Science rather...

Curated by: gutleber

Then, click on “New upload” on the community page.

 New upload

#### Want your upload to appear in this community?

- Click the button above to upload straight to this community.
- The community curator is notified, and will either accept or reject your upload (see community curation policy above).
- If your upload is rejected by the curator, it will still be available on Zenodo, just not in this community.

Community



Complete the individual metadata fields as follows:

Upload type: select **Dataset**

Upload type

Publication
  Poster
  Presentation
  Dataset
  Image
  Video/Audio
  Software
  Lesson
  Other

**Reserve a DOI** by clicking the “Reserve DOI” button and enter that DOI in the METADATA.ODS and SAMPLES.ODS file in the cernbox folders.

Digital Object Identifier

Optional. Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register a new DOI for you. A DOI allows others to easily and unambiguously cite your upload. Please note that it is NOT possible to edit a Zenodo DOI once it has been registered by us, while it is always

Publication date \*

Required. Format: YYYY-MM-DD. In case your upload was already published elsewhere, please use the date of first publication.

Reserve a Digital Object Identifier for your upload. This allows you to know the DOI before you submit your upload, and can thus include it in e.g. publications. The DOI is not finally registered until you submit your upload.

Enter a **meaningful title**.

You must list **all names and affiliations of persons, called “authors” in this form, who have contributed to this sample characterisation campaign. These are the persons who need to approve the publication of the data set.** The list of persons must be taken **from the METADATA.ODS** record in the cernbox folder.

Paste the **description** from the metadata file in the cernbox folder.

The version field must be a **RELEASED version**, i.e. the **minor version must be zero!** Examples 1.0, 2.0, 3.0.

Fill the **keywords** with FCC, EASITRAIN, H2020 and add material type and sample format as separate keywords.

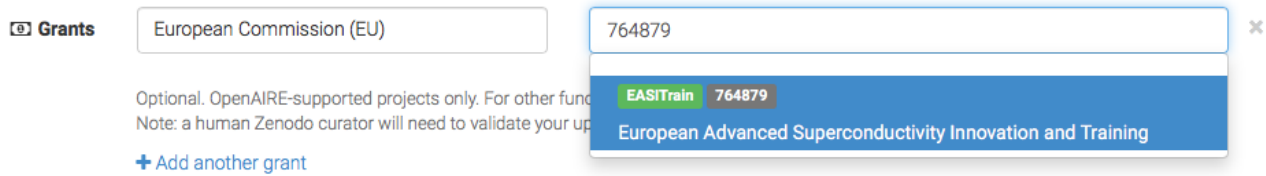
[+ Add another keyword](#)

To create the record, set the “**access rights**” to “Closed Access”.

**Access right \***
 Open Access
  Embargoed Access
  Restricted Access
  Closed Access

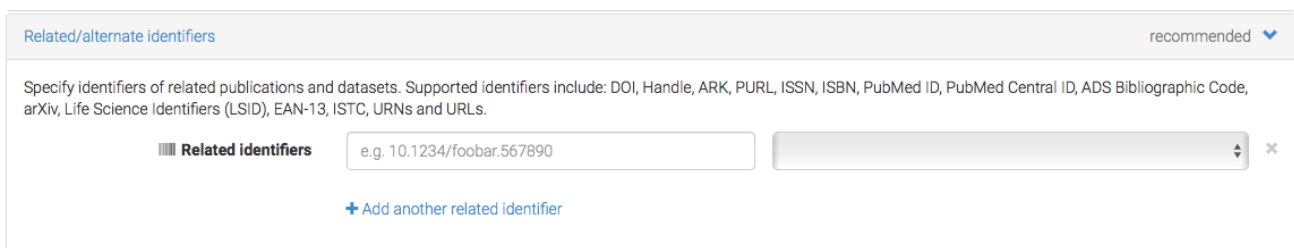
Required. Open access uploads have considerably higher visibility on Zenodo.

Add at least the **EASITRAIN project grant number** (the field has auto completion).



Add any other **grants** that were used in the process of creating, processing and analysing the materials that led to this data set.

In the “Related identifiers” add the additional keywords that are listed in the first part of this document. The keywords consist of a text entry and a URL to the library of congress keyword terms.



The **keywords** need at least to include

- *discipline*
- *sample type*
- *material*
- *properties or methods used for characterization*

Add **any additional persons**, who contributed to the creation of this data set and assign their individual roles.

Add potential references to other data sets and publications, such as materials and methods used to analyse the samples and background publications.

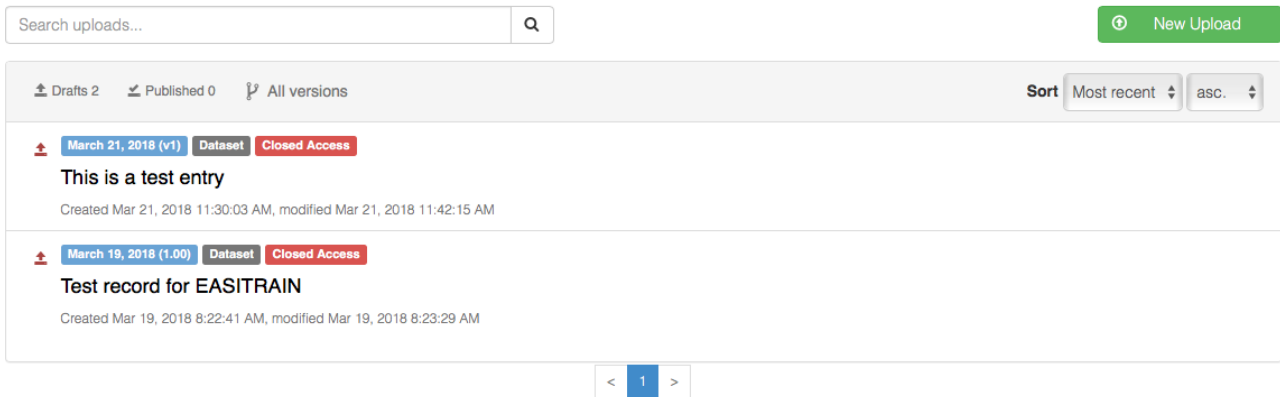
Provide information about the journal in which the corresponding article will be published. This can also be a technical note in the CERN Document Server (cds.cern.ch).

In the subject, add the discipline keywords from the library of congress as outline in this document.

Finally click “Save”. **The record is created, but not published.**

## 8.4. EDIT A RECORD

You can always edit a record by clicking on the “Upload” button at the ZENODO main page. All your records will appear there.



The screenshot shows the ZENODO uploads interface. At the top, there is a search bar labeled "Search uploads..." with a magnifying glass icon and a green "New Upload" button. Below the search bar, there are filters for "Drafts 2", "Published 0", and "All versions". To the right, there is a "Sort" dropdown menu set to "Most recent" and "asc.". The main content area displays two records:

- Record 1:** "March 21, 2018 (v1)" with tags "Dataset" and "Closed Access". The title is "This is a test entry". It was created on Mar 21, 2018 11:30:03 AM and modified on Mar 21, 2018 11:42:15 AM.
- Record 2:** "March 19, 2018 (1.00)" with tags "Dataset" and "Closed Access". The title is "Test record for EASITRAIN". It was created on Mar 19, 2018 8:22:41 AM and modified on Mar 19, 2018 8:23:29 AM.

At the bottom of the records list, there is a pagination control showing "< 1 >".

Then simply click on the record and continue editing until you decide to publish the record.