

**BrightnESS<sup>2</sup>**

**Bringing Together a Neutron Ecosystem for Sustainable Science with ESS**

**H2020-INFRADEV-3-2018-1**

**Grant Agreement Number: 823867**

brightness<sup>2</sup>

**Deliverable Report**  
**D1.2 Data Management Plan**



## 1 Project Deliverable Information Sheet

BrightnESS <sup>2</sup> Project	Project Ref. No. 823867	
	Project Title: BrightnESS <sup>2</sup> – Bringing Together a Neutron Ecosystem for Sustainable Science with ESS	
	Project Website: brightness.esss.se	
	Deliverable No.: D1.2 Data Management Plan	
	Deliverable Type: Report	
	Dissemination Level:	Contractual Delivery Date: June 30 <sup>th</sup> 2019
		Actual Delivery Date: June 30 <sup>th</sup> 2019
		Revised Delivery Date: January 28 <sup>th</sup> 2021
EC Project Officer: René Martins		

## 2 Document Control Sheet

Document	Title: BrightnESS <sup>2</sup> D1.2 Data Management Plan	
	Version: 2	
	Available at: TBD	
	Files: 1	
Authorship	Written by	Erika Lundqvist
	Contributors	Tobias Richter, Gareth Murphy, Margaret Armstrong
	Reviewed by	Ute Gunsenheimer
	Approved	BrightnESS <sup>2</sup> Steering Board

## 3 List of Abbreviations and Acronyms

CHES	Collaboration Home for ESS
DEUNET	European Deuteration Network
DMP	Data Management Plan
DOI	Digital Object Identifier
EC	European Commission
ERIC	European Research Infrastructure Consortium
ESS	European Spallation Source ERIC
EU	European Union
FAIR Principles	Findable, Accessible, Interoperable, and Reusable
GDPR	General Data Protection Regulation
HDF5	Hierarchical Data Format 5
IKC	In-kind Contribution
ILO	Industrial Liaison Officer
IT	Information Technology
OA	Open Access
PI	Principal Investigator
PLM	Project Lifecycle Management
POPD	Processing of Personal Data
WP	Work Package

## 4 List of Tables

TABLE 1 PROJECT TIMELINES FOR BRIGHTNESS <sup>2</sup> AND ESS .....	4
TABLE 2 OVERVIEW OF THE MAIN BRIGHTNESS <sup>2</sup> DATA RELEASE CHANNELS .....	7
TABLE 3 NQL DATA LOCATIONS AND REPOSITORY REFERENCES.....	8
TABLE 4 DATA MANAGEMENT ROLES .....	10



## Table of Content

1	Project Deliverable Information Sheet .....	1
2	Document Control Sheet .....	1
3	List of Abbreviations and Acronyms .....	2
4	List of Tables .....	2
5	Executive Summary .....	4
6	Report on Implementation Process .....	4
6.1	Scope of BrightnESS <sup>2</sup> Data Management Plan .....	4
6.1.1	Purpose of the data collection/generation and its relation to the objectives of the project	4
6.1.2	Types and formats of data the project will generate/collect .....	5
6.1.3	What is the origin of the data? .....	5
6.1.4	Expected size of the data .....	6
6.1.5	Data Utility .....	6
6.2	FAIR data .....	6
6.2.1	Naming conventions .....	7
6.2.2	Keywords to optimize possibilities for re-use .....	7
6.2.3	Version numbers .....	7
6.2.4	Metadata .....	7
6.3	Making data openly accessible .....	7
6.3.1	Data Accessibility .....	8
6.3.2	Location of data and associated metadata, documentation and code .....	8
6.3.3	Making data interoperable .....	9
6.3.4	Increase data re-use (through clarifying licenses) .....	9
6.4	Allocation of resources .....	9
6.4.1	Responsibility structure for data management in BrightnESS <sup>2</sup> .....	10
6.4.2	Resources for long term preservation .....	10
6.5	Data security .....	10
6.5.1	Is the data safely stored in certified repositories for long term preservation and curation?	10
6.6	Ethical aspects .....	10
6.6.1	Informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data .....	11



## 5 Executive Summary

This document is Deliverable 1.2 and concerns the Data Management Plan (DMP) for the H2020 BrightnESS<sup>2</sup> project at the European Spallation Source (ESS) European Research Infrastructure Consortium (ERIC) in Lund Sweden. The Plan follows the H2020 FAIR Data Management Plan (DMP) template (version 26 July 2016). The purpose of this deliverable is to support the data management life cycle for all data that will be collected, processed or generated by the project. It will provide a description of the data types the project will generate and how the data will be collected and stored and made available for validation, exploitation and re-use by others. The information in this DMP will be updated during the course of the project. All data will be stored on the ESS data catalogue and published with open access. The data management plan is a living document and will be updated with changes to metadata format and requirements.

## 6 Report on Implementation Process

The procedures that will be implemented will be in line with the *European Spallation Source ERIC Policy for Scientific Data*, *ESS Rules for Personal Data Protection*, *Privacy Notice for internal ESS IT Systems*, *Website privacy policy and cookies policy*, the *ESS IT Security Plan*, *ESS Rules for archiving of documented information*, along with national legislation of each consortium partner and in line with the European Union (EU) standards, including the General Data Protection Regulation (GDPR).

Project name	Project timeline
BrightnESS <sup>2</sup>	January 1 <sup>st</sup> , 2019 to December 31 <sup>st</sup> , 2021
ESS	Construction: 2013 - 2019 Initial Operations: 2019 - 2025 Operations: 2025 - 2065

Table 1 Project timelines for BrightnESS<sup>2</sup> and ESS

### 6.1 Scope of BrightnESS<sup>2</sup> Data Management Plan

#### 6.1.1 Purpose of the data collection/generation and its relation to the objectives of the project

**Scientific data:** Research results can be obtained more reliably, faster and cost efficiently by linking sample preparation, experiment validation and analysis tools, and recognizing that methodological specialisation is required to cater adequately for various user communities.

In WP2, Task 2.3 aims to create a Neutron Quality Label as internal certification within the participating neutron facilities in Europe. In order to guarantee the sustainability of the Label and to ensure continuous development of the technique, the collaboration between neutron facilities is prerequisite and will be strengthened by the proposed activities, and contribute to standardisation. The second subtask focuses on a pilot project aiming to improve neutron usage for soft matter and life sciences by increasing access to high quality deuterated samples. This directly profits ESS when the first instruments for small angle neutron scattering and neutron reflectometry are starting user operation and at the same time extends the range of deuterated molecules available to the European user community.

**IKC data:** WP3 uses a software tool (XRM+) which has enhanced the information transfer between the IKC partners and ESS, that was set up under the Project BrightnESS, funded by the European Union Framework Programme for Research and Innovation Horizon 2020, under grant agreement 676548,



and will be maintained during BrightnESS<sup>2</sup>. It is the single repository where all the “vital signs” of each in-kind contribution is stored and already records the details of €550M worth of work. XRM+ is being integrated into the ESS professional project management tool Primavera P6 so that planners at ESS are instantly aware of changes made by the partners. This helps the partners to keep track of the in-kind for which they are responsible, and provides vital management information to ESS to allow it to track progress on the IKC both in overview and at a detailed level. In turn, through the Field Coordinators, ESS is able to quickly trace changes and understand what impact the work at each of the partner institutes will have on what is happening as the enormous ESS facility is built.

**Industrial data:** The tasks of WP4 are designed to build on the foundation established during BrightnESS (G.A No. 676548) and to help ESS to prepare for the changes anticipated in innovation activities and industrial engagement as the facility moves closer to operations. Currently, industry predominantly fulfils the role of suppliers to ESS and co-creators with ESS, and is hence an active beneficiary of what the ESFRI Working Group on Innovation calls an “upstream business model”. Once the ESS user programme starts, a new dimension of industrial engagement will emerge in the form of users. This will result in a two-way business model that encompasses both upstream and downstream activities. The establishment and maintenance of good collaboration with industry in all directions and at all levels represents a critical link in the innovation chain.

**Socio-economic impact data:** The ability to measure and evaluate scientific as well as socio-economic impact is key in making the ESS facility sustainable in the long-term. BrightnESS (G.A No. 676548) played a critical role in laying down the foundation early in the life-cycle of ESS by defining clear objectives and strategy, and setting up an overall approach to systematically assess its socio-economic impact. Within the framework of WP5, data will be collected to demonstrate the socio-economic impact of ESS generated during the construction phase.

**Personal data:** The collection of personal data will be kept to a minimum, only as necessary and with the consent of the person in question, for example, in the shape of event participation signature lists, photographs from events, etc.

## 6.1.2 Types and formats of data the project will generate/collect

The data set types can be summarized as the following:

- Scientific data: mostly NeXus/HDF5 formatted data files – see <https://www.nexusformat.org/> (WP2)
- IKC data: In-Kind (WP3)
- Industrial data: Surveys, interviews, workshops (WP4)
- Deliverables (All WPs), documents

## 6.1.3 What is the origin of the data?

**Scientific data:** In the development and practice of the Neutron Quality Label in WP2 Task 2.3, standard samples and procedures will be developed, allowing the evaluation of physical parameters of the specific measuring configuration such as lateral-, strain- and time-resolution, reproducibility, accuracy and precision of strain values. These values shall be reported for every measurement. A copy of each standard sample will be distributed to participating neutron facilities and common developed procedures will apply on a daily basis. Specifically, data will be collected at the ISIS ENGIN-X instrument, at the Neca MPISI diffraction station, the SALSA instrument at ILL and STRESS-SPEC at the MLZ in Garching.

The soft-matter pilot also in WP2 Task 2.3, aims to combine the existing expertise of the different DEUNET partners to produce selectively or perdeuterated samples. Moreover, the pilot project aims



to demonstrate the way forward to an efficient experimental workflow between support labs in the area of soft matter and life sciences at the participating neutron facilities.

**IKC data:** The IKC process can be separated into three fundamental phases: preparation (managing expressions of interest and contracting with Partner institutions), implementation (tracking the delivery of the contracted solutions, services, materials etc.) and conclusion (IKC value crediting process is approved, applied and the closeout of a contribution is conducted). This process is recorded in XRM+, a tool developed under BrightnESS (G.A No. 676548).

**Industrial data:** The innovation and collaboration potential of the ILO Network will be assessed through an online survey and expert interviews with ILOs. Building on the surveys completed during BrightnESS (G.A No. 676548), the type of data to be sought concerns the ILOs individual expertise and long-term strategies. The findings will feed into the discussions during interactive workshops. Two interactive workshops will be organised for ILOs to jointly develop a long-term strategy on how to best utilise the ILO Network to fuel innovation at ESS. The strategy will be presented in Deliverable D4.1 Report on the evolution potential of the ESS ILO Network.

**Socio-economic impact data:** Data will be collected using internal ESS systems as well as external databases. The data will cover a wide variety of areas and will include scientific data, technical data, in-kind data, economic and financial data, data related to international collaborations, projects funded by national and international agencies, media and press, human capital, environment etc. The list presented here is not exhaustive and will be expanded and updated as the work on socio-economic impact progresses

#### 6.1.4 Expected size of the data

The expected size is approximately 100 GB of data files. Documents will be a small fraction, e.g. 100 MB. This may change over the lifetime of the project and will be accordingly updated in the data management plan.

#### 6.1.5 Data Utility

The standardized Neutron Quality label produced through BrightnESS<sup>2</sup> will be of value to collaborators and researchers in the community. In addition to the data supporting development projects, the scientifically valuable experimental data produced through BrightnESS<sup>2</sup> will demonstrate the use of the new standards to the same research communities as the general ESS experimental data. It is expected that the primary users will be colleagues from the respective research communities – where, e.g. the value of easy data repository access to world class experimental data for, especially, young researchers and in connection with research training, is important.

## 6.2 FAIR data

Publications prepared within BrightnESS<sup>2</sup> will be public to the extent possible. The aim is to make all publications and the data collected publicly available (gold open access, OA), unless this is in conflict with privacy issues or potential further exploitation with the project outcomes. All the public deliverables will be accessible from the project web site. In general, all published datasets and publications from BrightnESS<sup>2</sup>/ESS will be mandated to support persistent citation – i.e. to use e.g. DOI for all datasets and publications.

Datasets on SciCat will use a persistent identifier and a DOI for all datasets publicly available and findable.



### 6.2.1 Naming conventions

Researchers using the resources offered in the BrightnESS<sup>2</sup> and ESS data repositories, will be requested to include a bibliographic citation to all BrightnESS<sup>2</sup>/ESS products that they use in their publications. Such citations will help others find the products and see how they have been used.

The content of a data product citation should include as much of the following information as appropriate:

- contributing PI/investigators/authors
- year of release
- product title/proposal ID
- publisher [BrightnESS<sup>2</sup> – a program funded under the H2020 programme for Research (grant no. 823867; ESS ERIC)]
- date/version accessed
- digital object identifier

In general, all published datasets and publications from BrightnESS<sup>2</sup>/ESS will be mandated to support persistent citation – i.e. to use e.g. DOI for all datasets and publications.

### 6.2.2 Keywords to optimize possibilities for re-use

Keywords identifying the project and data – e.g. science discipline, sample type, methodology used, etc. – will be provided when data is published within certain formats, such as in journal publications.

### 6.2.3 Version numbers

For the standard sample datasets, only one version is expected. If new data is regenerated it will be tagged, version 1.1, version 2.1 etc.

### 6.2.4 Metadata

The metadata associated with the data products can be divided into three general categories:

- **Repository-level metadata** – repository/archive context related information (e.g. upload date, embargo information)
- **Project-level/Dublin Core metadata** – detailing the context in which the data was produced (e.g. project/proposal info), the owner/custodian of the data (e.g. PI), how it is, and can be, used (e.g. persistent identifier, Digital Object Identifier, associated publications)
- **Data-level metadata** – detailed technical information on the data and the dataset (e.g. origin, type, acquisition, processing history). This will be as shown in the NeXus Class Definitions. <https://www.nexusformat.org/> stored using HDF5 standard format.

## 6.3 Making data openly accessible

Dissemination of data to different stakeholder communities is an integral part of WP5 It is through the data that BrightnESS<sup>2</sup> hopes to generate further interest into the research possibilities and capacity of ESS. BrightnESS<sup>2</sup> is committed to distributing results and publications via Open Access as a means for widening the ESS target audience.

Data product	Release channel
Data files	SciCat ( <a href="https://scicat.esss.se">https://scicat.esss.se</a> )
Scientific publications	OpenAire

Table 2 Overview of the main BrightnESS<sup>2</sup> data release channels





Parts of the project deliverables will be made publicly available through the BrightnESS<sup>2</sup> website. WP5 will actively promote the fact that this information is available.

### 6.3.1 Data Accessibility

Access to the data products will depend on the repository. Openly accessible data products hosted on the ESS SciCat repository will be accessible for anonymous download. Scientific publications will be made available/referenced through the OpenAire platform.

### 6.3.2 Location of data and associated metadata, documentation and code

**Raw and processed data:** For raw and processed data, the SciCat, the ESS data repository <https://scicat.esss.se>, will preserve all links to the data, allowing transparent access to the archived data. BrightnESS<sup>2</sup> partners are responsible for ensuring that DMSC is informed of new datasets and metadata to be deposited in the SciCat catalogue.

The NQL data collected at several different neutron facilities will remain at the facilities and be subject to their data and access policies. The raw data measured on the participating instruments are stored in the respective archives under a unique experiment number, along with the metadata to ensure the ease of finding the data. The data is available for public access within 3 years and will eventually be made public. All participating facilities are committed to store the data for the long term (minimum of 10 years) and the data can be easily accessed and downloaded, with access restriction according to the data policy of each facility (Table 3).

Instrument	Facility	Measurement	Data policy & availability
ENGIN-X	ISIS	RB2030003	<ul style="list-style-type: none"> <li>DOI: <a href="https://doi.org/10.5286/ISIS.E.RB203003">https://doi.org/10.5286/ISIS.E.RB203003</a></li> <li>Data portal: <a href="https://data.isis.stfc.ac.uk/">https://data.isis.stfc.ac.uk/</a> <ul style="list-style-type: none"> <li>First 3 years: Limited to users registered on proposal</li> <li>After 3 years: Available for general ISIS registered user</li> </ul> </li> </ul> <a href="https://www.isis.stfc.ac.uk/Pages/Data-Policy.aspx">https://www.isis.stfc.ac.uk/Pages/Data-Policy.aspx</a>
MPISI	Necsa	2019/12/20-6201	<ul style="list-style-type: none"> <li>Data can be made on-line to collaborators upon request (<a href="mailto:communication@necsa.co.za">communication@necsa.co.za</a>) through Necsa managed FTP server</li> <li>Data is accessible by public after 3 years</li> </ul>
SALSA	ILL	INTER-468	<ul style="list-style-type: none"> <li>DOI: <a href="https://doi.ill.fr/10.5291/ILL-DATA.INTER-468">https://doi.ill.fr/10.5291/ILL-DATA.INTER-468</a></li> <li>Data portal: <a href="https://data.ill.fr">https://data.ill.fr</a> <ul style="list-style-type: none"> <li>First 3 years: Limited to users registered on proposal</li> <li>After 3 years: Available for general users registered in the ILL user club</li> </ul> </li> </ul> <a href="https://www.ill.eu/users/user-guide/after-your-experiment/data-management">https://www.ill.eu/users/user-guide/after-your-experiment/data-management</a>
STRESS-SPEC	MLZ	16405	<ul style="list-style-type: none"> <li>Unique identifier will be available in the future, web-based access mechanism subject to PI permission is currently available</li> <li>After publication, PI must make the data publicly available</li> </ul> <a href="https://www01.mlz-garching.de/englisch/user-office/terms-of-reference.html#Publications">https://www01.mlz-garching.de/englisch/user-office/terms-of-reference.html#Publications</a>

Table 3 NQL data locations and repository references

**BrightnESS<sup>2</sup> deliverables:** The BrightnESS<sup>2</sup> website, where all public deliverables can be found, will remain open and will be actively maintained by ESS and BrightnESS<sup>2</sup> partners for at least five years after the project has been successfully concluded.

**Scientific publications:** The scientific data generated by BrightnESS<sup>2</sup> during the construction of ESS will result in the publication of multiple articles. BrightnESS<sup>2</sup> will commit to making these publications accessible through the open access infrastructure OpenAire. Scientific publications produced by ESS will follow the ESS Publication, Communication and Affiliation Policy.

### 6.3.3 Making data interoperable

The standard formats used will be NeXus (<https://www.nexusformat.org/>) and HDF5 (<https://www.hdfgroup.org>). These are widely known and used within the research community. Other standards as they arise will be added to the data management plan.

### 6.3.4 Increase data re-use (through clarifying licenses)

Data will be licensed as open data which ensures data can be reused for scientific purposes. Data in the BrightnESS<sup>2</sup> and ESS repositories may be subject to embargo. These would typically be:

- **Researcher proprietary period** – a 3-year period after data has been created, in which the data will only be accessible by the PI and collaborators. After the end of the proprietary period, the dataset will be made openly accessible to the community. Project-level metadata may be generally accessible as soon as the data enters the repository.
- **Journal embargo** – occasionally scientific journals (e.g. Nature) may impose an embargo on scientific data/results until the results have been published. After the embargo period has ended, the data may be released.
- **Confidentiality considerations** – a subset of ESS experimental projects will involve industry involvement, which may impose certain confidentiality considerations to respect commercial interests.

Project deliverables will remain on the website, available for anonymous download. Scientific publications produced by ESS will follow the *ESS Publication, Communication and Affiliation Policy*; as well as comply with *OpenAire* obligations.

## 6.4 Allocation of resources

Any costs considering the small amounts of research data envisaged from BrightnESS<sup>2</sup> will be met from the BrightnESS<sup>2</sup> budget. Costs of making ESS Data FAIR will be borne by ESS. Data curation and retention for data subject to other facilities' data policies (see table 3 for the NQL data) will be funded through the normal operations budget, as that data is being treated under the standard workflow for users at the respective facility.



#### 6.4.1 Responsibility structure for data management in BrightnESS<sup>2</sup>

Agent	Class	Roles
BrightnESS <sup>2</sup>	Primary stakeholder	Data creator Data owner Data user Data hosting
ESS	Primary stakeholder	Data creator Data owner Data user Data hosting
BrightnESS <sup>2</sup> Collaborators	Primary stakeholder	Data creator Data owner Data user
Community	Secondary stakeholder	Data user

Table 4 Data Management Roles

#### 6.4.2 Resources for long term preservation

All project-related data will be kept for the required number of years after the closing of the project, after which data, if no longer of value to ESS, can be destroyed. All costs associated to data preservation or maintenance after the project closing will be borne by ESS.

#### 6.5 Data security

During the lifetime of the BrightnESS<sup>2</sup> project, relevant documents will be stored in a Product Lifecycle Management (PLM) system known as CNESS (Collaboration Home for ESS). CNESS is structured in a way such that it can be easily accessed both by humans and by software as appropriate. Information Technology (IT) has developed a number of project portals for accessing relevant data stored in CNESS. Although IT will provide the technical infrastructure for this it is the central ESS project office's responsibility that the PLM data for BrightnESS<sup>2</sup> is available and can be accessed by the organisations delivering or wanting to use the technologies.

SciCat uses a secure https server and all data is protected by passwords.

All physical formats of data will be scanned, uploaded to CNESS, with the physical documents archived according to the ESS Rules for archiving of documented information. This process will take place periodically.

##### 6.5.1 Is the data safely stored in certified repositories for long term preservation and curation?

Yes. CNESS is maintained by the IT division at ESS and encrypted in accordance with the IT Security Plan. Physical forms of data are stored in a locked container for the required duration, then, if deemed no longer of value to ESS, destroyed.

ESS SciCat will store data in the long term for preservation and curation purposes.

#### 6.6 Ethical aspects

BrightnESS<sup>2</sup> beneficiaries are aware of the ethical issues that may be involved when gathering and storing research data, and the requirement to ensure that there are measures in place to monitor all

activities taken during the term of the project, and ensure their compliance with the appropriate regulations. BrightnESS<sup>2</sup> has two Ethics requirements as deliverables *D6.1 Requirement No.1 - Templates of the information sheets (in language and terms intelligible to the participants) and the informed consent forms (if applicable) must be kept on file* and *D6.2 Requirement No 2 – POPD*, these will outline the ethical considerations which will apply to the data management in the BrightnESS<sup>2</sup> Project.

## **6.6.1 Informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data**

Personal user information will be protected in accordance with the *ESS Rules for Personal Data Protection, EU Data Protection Directive (Directive 95/46/EC)* and, where applicable, the *EU General Data Protection Regulation*.

