



BrightnESS

**Building a research infrastructure and synergies for highest
scientific impact on ESS**

H2020-INFRADEV-1-2015-1

Grant Agreement Number: 676548

brightness

Deliverable Report: D1.2 Data Management Plan

1 Project Deliverable Information Sheet

BrightnESS Project	Project Ref. No. 676548	
	Project Title: BrightnESS - Building a research infrastructure and synergies for highest scientific impact on ESS	
	Project Website: brightness.esss.se	
	Deliverable No.: 1.2	
	Deliverable Type: Report	
	Dissemination Level: Public	Contractual Delivery Date: 01.03.2016
		Actual Delivery Date: 20.4.2016
	EC Project Officer: Bernhard Fabianek	

2 Document Control Sheet

Document	Title: BrightnESS_Deliverable 1.2	
	Version: 2.0	
	Available at: https://brightness.esss.se	
	Files: 1	
Authorship	Written by	Roy Pennings (WP1 leader), Raquel Costa (WP1)
	Contributors	Tobias Richter (WP5 leader), Richard Hall-Wilton (WP4 leader), Judith Freita Ramos (WP4), Brian Lindgren Jensen (WP5), Sune Rastad Bahn (WP5)
	Reviewed by	Ute Gunsenheimer (WP6 leader)
	Approved by	BrightnESS Steering Board (18-04-2016)

3 List of Abbreviations

DDT	Directors Discretionary Time
DMP	Data Management Plan
DMSC	Data Management and Software Centre
ERIC	European Research Infrastructure Consortium
ESS	European Spallation Source ERIC
EPICS	Experimental Physics & Industrial Control System
IK	In-Kind
IKC	In-Kind Contribution
MIS	Management Information System
OSS	Open Source Software
PI	Principal Investigator
RI	Research Infrastructure
TAC	Time Allocation Committee



4 Table of Contents

1	Project Deliverable Information Sheet	2
2	Document Control Sheet	2
3	List of Abbreviations	2
4	Table of Contents	3
5	Executive Summary	4
6	Report on implantation Process and Status of Deliverable	5
6.1	Scope of BrightnESS DMP	5
6.2	Data set types, standards and metadata	6
6.2.1	Data relevant to the setting up and running of an ERIC (WP2, WP3, WP6).....	6
6.2.2	Data Acquisition Software	7
6.2.2.1	Data Processing Software	7
6.2.2.2	Standards	8
6.2.3	Metadata	8
6.2.3.1	Metadata overview.....	9
6.3	Data sharing	10
6.3.1	Data Sharing and Dissemination	10
6.3.1.1	Stakeholder Identification	11
6.3.1.2	Data Sharing Policy	11
6.3.1.3	Ethical and Privacy Issues.....	12
6.3.1.4	Data Copyright Issues	12
6.3.1.5	Embargoes	12
6.3.1.6	Intended Future Data Users	13
6.3.1.7	Data Citation Policy	13
6.3.1.8	Publication Policy.....	14
6.3.1.9	Data management responsibilities.....	14
6.3.2	Archiving and Preservation	14
6.3.2.1	Archiving Policy and Data Lifecycle Management.....	14
7	References	15



5 Executive Summary

This document is Deliverable 1.2 and concerns the Data Management Plan (DMP) for the H2020 BrightnESS project at the European Spallation Source (ESS) in Lund Sweden. The Plan follows the H2020 Work Programme 2014-2015 Guidelines for DMP and Art. 29.3 of the Grant Agreement. 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.



6 Report on implantation Process and Status of Deliverable

This document describes the Data Management Plan (DMP) as deliverable 1.2 for the BrightnESS project. The deliverable outlines how research data collected or generated will be handled during and after the BrightnESS action. It is important to note that the DMP will evolve and further develop during the project’s lifecycle. This document follows the template “Guidelines on Data Management in Horizon 2020” provided by the European Commission (EC).

6.1 Scope of BrightnESS DMP

The BrightnESS project will develop groundbreaking and innovative neutron detector solutions for use at the European Spallation Source (ESS). As part of this development project, a large quantity of technical and research data-products will be created within the areas of modeled and experimental data, as well as within software development. While some of these data-products will be in the form of test data of immediate value, but with limited re-usability, the bulk part of the software and data products will have considerable re-usability value for collaborators and third-party researchers – either in the form of software and algorithms that can be used as-is, or even re-purposed for adjacent uses – or, in the form of modeling and experimental data that has technical and scientific value, both for new, original research, and for verifying and reproducing existing results.

The procedures that will be implemented will be in line with national legislation of each consortium partner and in line with the European Union (EU) standards.

This Data Management Plan will apply to all data under control by the project. We strive to make data open, but cannot overrule limitations that partner institutions put on data that they contribute (as specified in their Background grant agreement for example).

As the BrightnESS products are developed for use in the ESS project, all pertinent data products will ultimately be transferred to, stored in, and made accessible from, a general ESS data repository. As the general ESS data repository is first expected to be online from 2018 (see project timelines in Table 6-1), earlier BrightnESS data products will initially be made available through dedicated channels (e.g. the BrightnESS website).

To incorporate the transitory nature of the data management, i.e. taking into account the requirements of migrating the data from BrightnESS repositories to the ESS repositories, subsequent iterations of the BrightnESS data management plan might be influenced by elements of a future ESS data management plan. This broad scope (encompassing elements of a general ESS data management plan) can cause a certain degree of generalization in some areas of the plan. As the BrightnESS and ESS DMPs both are living documents, they will be regularly updated as areas in the DMPs are further developed and characterized.

Project name	Project timeline
BrightnESS	Sep. 1, 2015 to Aug 31, 2018
ESS	Construction: 2014-2019 Initial Operations: 2019-2025 Operations: 2025-2065

Table 6-1: Project timelines for BrightnESS and ESS



6.2 Data set types, standards and metadata

During the duration of the project, BrightnESS will generate a variety of data. All of it has to be stored in a way such that it is easily accessible by both humans and software, as appropriate.

Broadly, the data falls into two categories:

- **Organizational data** – e.g. data relevant to the implementation of an ERIC
- **Technical and scientific data** – this includes simulation and modeling, raw and processed experimental data, scientific analyses/publications as well as software code and algorithms

Excluding Section 6.2.1 below, this DMP will only describe the treatment of technical and scientific data.

6.2.1 Data relevant to the setting up and running of an ERIC (WP2, WP3, WP6)

- a. Processes and procedures for large scale In-Kind Contribution management from multi-country partners
- b. Processes and procedures for dealing as an ERIC with various taxation matters
- c. Processes for In-Kind accounting
- d. Processes and procedures for managing the enlargement of an ERIC

Technically, this type of data does not fall into the definition of the European Commission on research data. However, the BrightnESS consortium feels that the data will be directly relevant to other research communities wishing to create their own Research Infrastructure (RI). The costs of building such RI are becoming prohibitive for a single national science community, company or even national state. In-Kind contributions from multi-country partners are one of the most promising ways forward, but this approach is as yet uncommon, technically very complex and subject to a wide variety of legal, legislative and financial interpretations and constraints. The ERIC status created by the European Commission to overcome some of these constraints is helpful, but EU Member States have limited experience with this model and can still include additional national rules. In that perspective, the data from the mentioned work packages can be exploited as best practice catalysts for future RIs, and will accordingly be made accessible for verification and re-use.

Most data generated by work packages 2, 3 and 6 will be in the form of documents that describe processes, which will - where relevant - be supported by flow charts, monitoring tools, online platforms etc. Documents will not only show end-results of a negotiation (for example with a view to solving VAT-issues, taxation related to posted or seconded staff, legal aspects on transfer of ownership of in-kind contributions or best-practice accounting of in-kind) but also the steps toward reaching this end-result. This approach will allow others who are interested in setting up a research infrastructure through large-scale in-kind contributions to understand the pitfalls and processes involved. The production of data will follow the overall GANTT chart for each of the work packages; as specific items are resolved the approved documents will be published immediately.

Storage of data created by the mentioned work packages will be on the BrightnESS web portal. Access is open to anyone, with the exception of specific ESS-related internal business



information. Publication of the availability of the data will take place through WP6 dissemination activities.

Within work packages 2, 3 and 6 data will also be collected on the future market for materials research using neutrons as well as technical innovation potential for industry, both during the construction phase of ESS as well as thereafter when ESS is in standard operations mode. Data from organizations in the academic sector will be obtained through at least one online survey questionnaire. Innovation ecosystems data will be collected from mapping via desk research and qualitative interviews. Finally, industry data will be collected from regional focus groups with 5-10 industrial participants for each of the six BrightnESS Regional Hubs located in the different partner countries. Specific data sets within each of these groups will be further specified during the course of the project. Data obtained will not be identified by individual, but will be anonymized, compiled and analyzed as a group. Third parties will have full access to the anonymized compiled data. The data will be stored in the public section of the BrightnESS website (pdf-format) and promoted as part of WP6 dissemination activities.

6.2.2 Data Acquisition Software

BrightnESS includes work packages in the area of detector event processing software. Detector event processing is a task traditionally done via electronics, but ESS’s software-based processing will allow scientist to be more flexible and selective, and potentially more precise, in discriminating and locating neutron events on detectors. These software products will be of potential general use for third-party detector development, and will be developed using open source software principles on community accessible platforms.

6.2.2.1 Data Processing Software

A wide range of scientific software frameworks, environments and applications is needed to handle the modeling, processing, analysis and visualization of the data from BrightnESS and ESS (see Table 6-2 for a sample listing of software products used). The software packages needed for re-exploitation of repository data are Open Source Software projects developed by the scientific community in open development projects on GitHub.

ESS is a strong and active partner in these scientific software development projects, and will continually support their continued development and distribution.

Software project	Project type	Project description and location	Location
Mantid	OSS	Framework for manipulating neutron scattering data – used for data reduction of experimental data at ESS. Community standard software – strong community development support.	http://www.mantidproject.org
McStas	OSS	General tool for simulating neutron scattering instruments and experiments – strong community support.	http://www.mcstas.org
SasView	OSS	Small angle scattering analysis software – managed by an international collaboration of neutron research facilities.	http://www.sasview.org



ParaView	OSS	Data analysis and visualization application – strong development support.	http://www.paraview.org
----------	-----	---	---

Table 6-2: Sample software packages used for data processing, modeling and analysis

6.2.2.2 Standards

The full ESS policy on standards has not yet been finalized. The tasks within BrightnESS, especially task 4.4, will be used to integrate this process and finalize the documents. These will feature in the next update of this DMP. Current standards can be seen in Table 6-3.

The following generalized data set description format will be used:

Resource name	Title of the resource
Resource type	Choose value: document, source code, data file
Media type	The medium through the content is provided
Language(s)	Language(s) of the content
License	Licensing conditions and terms for use of the data and/or tool(s)
Distribution medium	Channel for delivery or access to the data and/or tool(s)
Usage	Foreseen use of the resource for which it has been produced
Size	Size of the resource (numerical unit)
Description	Brief description of the data and/or the tool(s) and their use

Standard	Type	Description
CIF	File format	Crystallographic Information File – standardized text file format
EPICS	Control system	Software environment for distributed control systems
HDF5	Data format	HDF5 is a community standard data model, library, and file format for storing and managing data.
JSON	Data format	JavaScript Object Notation – lightweight data-interchange format
Markdown	Markup language	Lightweight markup language for documentation
NeXus	Data format	An international scientific standard for neutron data format
ODF	File format	Open Document format for structured information (documents, spreadsheets etc.)
OpenXML	File format	Office Open XML format for structured information (documents, spreadsheets etc.)
PDF	File format	Portable Document Format – open standard
RST	File format	reStructuredText – file format for technical documentation

Table 6-3: Selected standards used in the BrightnESS project

6.2.3 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 metadata** – detailing the context in which the data was produced (e.g. project/proposal info), who is responsible (e.g. PI), how it is, and can be, used (e.g. Digital Object Identifier, associated publications)
- **Data-level metadata** – detailed technical information on the data and the dataset (e.g. origin, type, acquisition, processing history)

6.2.3.1 Metadata overview

Overview of the three types of metadata associated with project-data is presented in Table 6-4 below:

Repository-level metadata	
Key type	Key description
Unique dataset identifier	ESS dataset identifier
Upload date	Timestamp for the addition of the data to the repository
Version history	Revision log for changes to the attached metadata
Embargo restrictions	Data release restrictions due to e.g. proprietary periods (e.g. program-specific or imposed by publication journals) or confidentiality terms (e.g. commercial/industry restrictions)
Project-level metadata	
Key type	Key description
PI and collaborators	Unique researcher identifiers: ORCID (and possibly e.g. ResearcherID)
Proposal ID	Unique ESS proposal identifier for the proposal which has been granted beamtime for this project
Proposal Title	Title of the proposal/project
Project Type	Type of project (e.g. standard TAC proposal, DDT project, simulated data etc.)
Project Keywords	Keywords identifying the project and data – e.g. science discipline, sample type, methodology used etc.
Unique universal identifier	DOI for dataset
Experiment logbook identifier	A reference to the relevant beamline user logbook for the experiment
Associated publications	References to the DOI of publications using the data set
Data-level metadata	
Data-origin	
Key type	Key description



Data class	experimental, simulated
Data state	raw, processed
Data source	instrument/experiment
Data timestamps	experiment timestamps
Data-class	
Key type	Key description
Data type	eventfile, image, spectrum, hypercube
File type	NeXus, XML (cf standards)
Processing record	
Key type	Key description
Processing software versions	Versions/identifiers of software packages used in the processing history
Calibration data used	Identifiers for the calibration data sets used for data calibration
Processing history	All processing steps applied to the data, processing timestamps, parameters etc.

Table 6-4: Sample overview of functional project-data metadata

6.3 Data sharing

6.3.1 Data Sharing and Dissemination

Dissemination of data to different scientific communities is an integral part of WP6. It is through the data that BrightnESS hopes to generate further interest into the research possibilities and capacity of ESS.

BrightnESS is committed to distributing results and publications via Open Access as a means for widening the ESS target audience. Data files will be released through Zenodo or directly by the Data Management and Software Centre (DMSC). Source codes and software will be made publicly available through public repositories (Github and Bitbucket). Where specific software analytical tools have been developed by ESS, those tools will also be made publicly accessible for re-use. Scientific publications are already being made available/referenced through the OpenAire platform (see also Section 3.6.1.9 and Table 6-5 for an overview of the BrightnESS data release channels).

Data product	Release channel
Data files	Zenodo, BrightnESS website, DMSC
Source code/software products	Public repositories (GitHub, Bitbucket)
Scientific publications	OpenAire

Table 6-5: Overview of BrightnESS data release channels

Parts of the project deliverables (in particular related to WP2, WP3 and WP6) will be made publicly available through the BrightnESS website. WP6 will actively promote the fact that this information is available.



6.3.1.1 Stakeholder Identification

Agent	Class	Roles
BrightnESS	Primary stakeholder	Data creator Data owner Data user Data hosting
ESS	Primary stakeholder	Data creator Data owner Data user Data hosting
BrightnESS Collaborators	Primary stakeholder	Data creator Data owner Data user
Community	Secondary stakeholder	Data user

6.3.1.2 Data Sharing Policy

The data sharing policy depends on which type of data is to be shared.

6.3.1.2.1 Accessing data

Access to the data products will depend on the repository. Openly accessible data products hosted on the BrightnESS web portal will be accessible for anonymous download from the BrightnESS website.

6.3.1.2.2 Raw and processed BrightnESS measurement data and developed software

The data retention policy for the raw data measurements on detectors done at ESS in the context of BrightnESS for the purpose of construction, QA and calibration of the beam lines and detectors, is:

Raw data will be centrally stored and will be made available to third parties within 4 weeks after final approval of the technical reports by the ESS Detector Work Package Manager.

The data retention policy for the processed BrightnESS developed data - meaning the analysis, model development and statistics derived from the raw data sets – is:

All processed data will be centrally stored and will be made available to third parties within 4 weeks after final approval of the technical reports by the ESS Detector Work Package Manager, together with any ESS-developed tool which allows validation of the analysis.

The processed data and analysis itself will flow into ESS technical reports on measurements and test of detector prototypes and simulations for QA approval. These reports are already made public by ESS, so BrightnESS will follow standard procedure.

Computer source code (mainly algorithms and related frameworks) will mainly be created and maintained at ESS’ Data Management and Software Centre (DMSC) in Copenhagen. Once algorithms are validated and published (in a publication or through project deliverable), the source code will be put on GitHub as the central repository.



6.3.1.2.3 *Raw experimental data*

The raw data of a project will generally be openly available after a proprietary period of six months after its creation, and will remain available in the ESS repository for 5 years after the project has finished.

6.3.1.2.4 *Data product enhancement*

To enhance the value of the data products in the repository, the data will – through metadata information - contain:

- Links to relevant experiment logbooks
- Contain contact info (through ORCID) on the project PI
- Contain data processing history, allowing for validation and reproduction
- Processing software version information with links to tools and software necessary for data use and analysis, as well as links to general documentation detailing how the data can be understood and re-analyzed (if needed).

6.3.1.3 *Data Obligations*

During the lifetime of BrightnESS, ESS will not produce data-sets for industry. ESS recognizes that the open accessibility of certain data-sets may be limited by future industry contracts or programmes, but the general accessibility policy already in place at ESS is that a third party is allowed commercial re-exploitation of repository data, as long as fair use guidelines are observed (e.g. full and clear credit and acknowledgement is given to originating PI/collaborators and BrightnESS/ESS, and no claim is made to ownership/copyright of the data-products). Through its Industry Focus Group meetings across Europe (part of work package 6), BrightnESS is supporting ESS in understanding the requirements, parameters and exceptions for use of industry data-sets once it is in standard operations mode. This knowledge will flow into further decision-making at ESS Council level.

6.3.1.4 *Ethical and Privacy Issues*

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

6.3.1.5 *Data Copyright Issues*

Experimental and modeling data created as part of BrightnESS will, to the extent possible, be released under a non-restrictive license (Creative Commons or similar). Software created and developed as part of BrightnESS will be released under an Open Source Software license (e.g. GNU General Public License (GPL)). For some data products, institutional and/or funding agency policies regarding copyright may need to be taken into account.

6.3.1.6 *Embargoes*

Data in the BrightnESS and ESS repositories may be subject to embargo. These would typically be:



- **Researcher proprietary period** – a limited, fixed, 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 will typically 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.

See section 6.3.1.2 above, for current policies and embargo durations.

6.3.1.7 Intended Future Data Users

As described in section 6.1 above, the datasets produced through BrightnESS will be of value to collaborators and researchers in the community for e.g. detector development. In addition to the data supporting development projects, the scientifically valuable experimental data produced through BrightnESS will be of interest to the same research communities as the general ESS experimental data. It is expected that the primary repository users likely 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 difficult to underestimate.

To ensure that data can be used by other users and combined with other datasets, the data in the repositories will have a very high level of interoperability due to the widely accepted community standards for data formats used – as described in section 6.2.2.2 above.

6.3.1.8 Data Citation Policy

Researchers using the resources offered in the BrightnESS and ESS data repositories, will be requested to include a bibliographic citation to all BrightnESS/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 – a program funded under the H2020 programme for Research (grant no. 676548; ESS ERIC)
- date/version accessed
- digital object identifier



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

6.3.1.9 Publication Policy

The scientific data generated by BrightnESS during the construction of ESS will result in the publication of multiple articles. The first one has recently been published in a peer-reviewed journal. BrightnESS 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.

6.3.1.10 Data management responsibilities

The table below provides information on the management of BrightnESS data in relation to registration of raw data, data integrity and security, storage and access:

Activity	Responsible
Registration of data sets	WP leader or delegated person
Verify rights to use the data	WP leader
Catalogue or version control for files	WP leader or DMSC (for data located at DMSC from 2018)
Backing up data	WP leader can be delegated to DMSC
Security and protection of data	WP leader can be delegated to DMSC

6.3.2 Archiving and Preservation

Archiving and preservation of raw and processed data, technical documents and standards being developed as part of BrightnESS is important for the ESS, as it will allow re-tracing the technical development for future research activities. This archiving will be performed by ESS as well as the ESS and BrightnESS collaborators. The collected and enhanced data will be stored in the ESS repository and at the sites of relevant BrightnESS project partners.

6.3.2.1 Archiving Policy and Data Lifecycle Management

The archiving policy depends on the type of data.

6.3.2.1.1 Raw and processed data

For raw and processed data, at least two copies will be kept in at least two different physical locations. One of the physical locations for the data will be the DMSC in Denmark, the other will be at one or more external scientific archive providers which are still to be selected. The ESS data repository will preserve all links to the data, allowing transparent access to the archived data.

6.3.2.1.2 BrightnESS deliverables

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



7 References

- Atlassian. (n.d.). *Bitbucket*. Retrieved from <https://bitbucket.org>
- BrightnESS project (EC H2020). (n.d.). *BrightnESS website*. Retrieved from <https://brightness.esss.se>
- EUDAT. (n.d.). Retrieved from <https://www.eudat.eu>
- European Parliament. (n.d.). *EU General Data Protection Regulation*. Retrieved from <http://ec.europa.eu/justice/data-protection/>
- European Parliament. (n.d.). *EU Protection of personal data*. Retrieved from <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=URISERV%3AI14012>
- European Spallation Source ERIC. (n.d.). *ESS website*. Retrieved from <https://www.esss.se>
- GitHub, Inc. (n.d.). *GitHub*. Retrieved from <https://github.com>
- *GNU General Public License*. (n.d.). Retrieved from <http://www.gnu.org/licenses/gpl-3.0.en.html>
- *HDF5*. (n.d.). Retrieved from <https://www.hdfgroup.org/HDF5/>
- *moonshot*. (n.d.). Retrieved from <https://wiki.moonshot.ja.net>
- *NeXus homepage*. (n.d.). Retrieved from <http://www.nexusformat.org>
- *OpenAire*. (n.d.). Retrieved from <https://www.openaire.eu>
- *ParaView Homepage*. (n.d.). Retrieved from <http://www.paraview.org>
- *UmbrellaID*. (n.d.). Retrieved from <http://pan-data.eu/Umbrella>
- *Zenodo*. (n.d.). Retrieved from <http://zenodo.org>