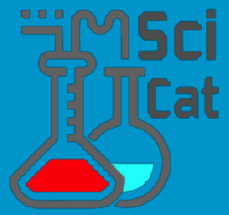
EUROPEAN SPALLATION SOURCE

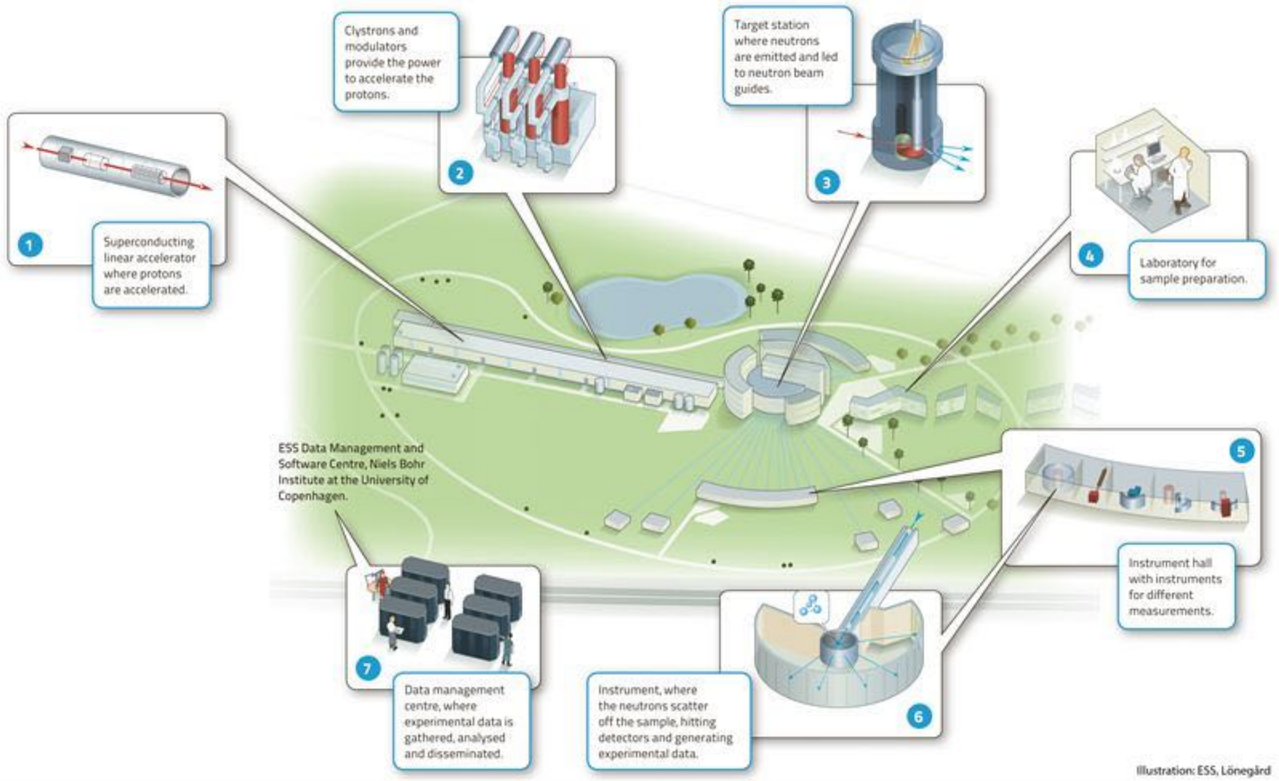
ESS Publication Workflow

Gareth Murphy
Experimental Control and Data Curation
2018-07-03


brightness


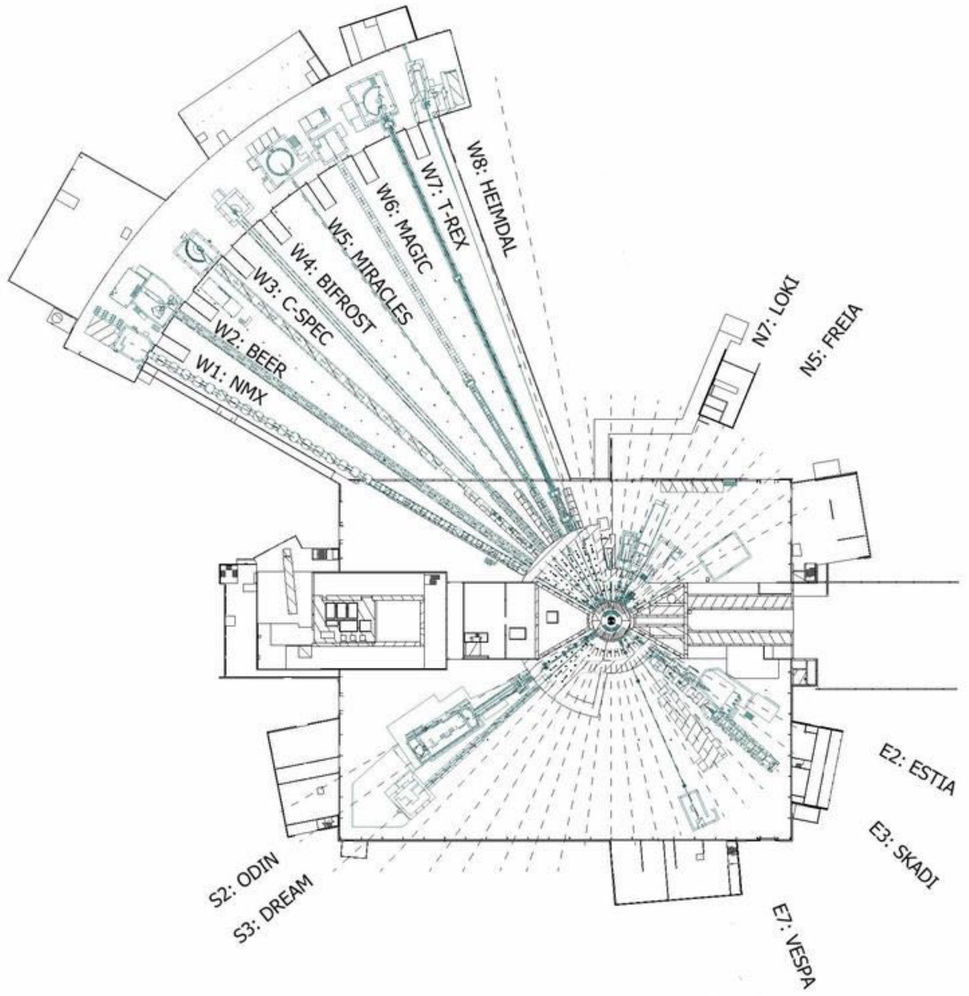
Powerful pulsed neutron source

- 17 Partner countries
- Construction work in progress

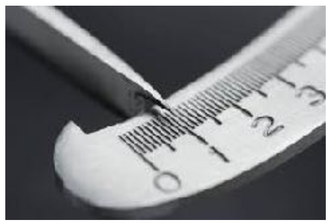


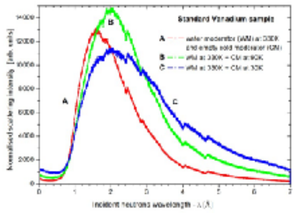
European Spallation Source

- European Spallation Source Scandinavia
- 15 instruments/beamlines
- Imaging, spectroscopy, diffraction
- Each instrument has different data requirements
- Traditionally, communities have had different data types, formats, analysis and reduction methods, standards - problem for data management
- By standardizing across instruments, we can make this process simpler and quicker



Raw, reduced and derived data

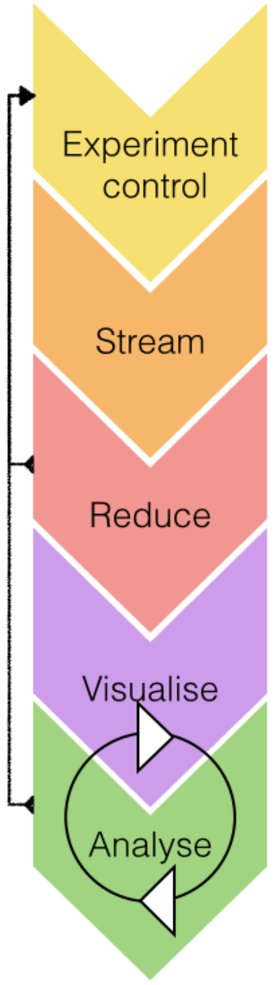




- Raw data - unprocessed data at full resolution, with communications artifacts removed (e.g. frame headers)
- Reduced - transformed and corrected from instrument units to physical units,
- Derived data - images, plots, statistics
- NASA define several processing levels raw = level 0, reduced = level 1, derived = level 2

Data Management & Software Centre (DMSC)

- DMSC - one team to rule the data
- Create uniform file writer for every beam line
- Connect data acquisition to data reduction and analysis
- Create/acquire metadata and send to data catalogue
- Owner + ORCiD, time, wavelength, license, type



brightness

EUROPEAN SPALLATION SOURCE

- BrightnESS is a Horizon2020 program to support ESS
- Years of archives of legacy brightnESS data need to be supported
- 250,000 files, many different formats and types
- Metadata need to be preserved and made accessible
- Test case for SciCat

ESS Data Policy

data and results.

3. RAW AND META DATA COLLECTED BY ESS

3.1. Raw data and associated meta data

3.1.1. DMSC will act as the custodian of raw data and associated meta data.

3.1.2. ESS shall provide a persistent identifier for containers of raw data and meta data at the point of creation of said container.

3.2. Security, curation, archival of raw and meta data

3.2.1. Raw data will be read-only for the duration of their life time.

3.2.2. Raw and meta data will be stored in a defined format.

3.2.3. ESS will curate a means to read the raw data and meta data.

3.2.4. Two (2) copies of the raw data and meta data will be kept in physically different locations for at least first five (5) years commencing at the end of an ESS operations cycle when the raw data was taken.

3.2.5. ESS will move raw data and meta data to archival facilities for long term curation for a minimum of 10 years after the restricted access period.

3.3. Access to Raw data and associated meta data by ESS/PT

3.3.1. Access to the raw data and meta data shall be restricted to the original PT and ESS staff for the first three (3) years after the end of the ESS operations cycle when the raw data was taken. After this period, the raw and meta data will be open access.

3.3.2. ESS staff shall respect the confidentiality of the scientific research data for the restricted access period but may exercise the right to make public thereafter.

3.3.3. The PT shall be able to obtain an electronic copy of the scientific research data they collected.

3.3.4. It is the responsibility of the PT to curate their electronic copy.

3.3.5. Publications including data collected at ESS will cite the persistent identifier of the data.

3.4. Open Access to Raw and meta data

EUROPEAN SPALLATION SOURCE

3.4.1. Three (3) years after the end of an ESS operations cycle when the raw data was generated, ESS shall make public an index of the raw data and meta data collected in an open on-line catalogue and notify PT of this event.

3.4.2. The PT may request an extension of the restricted access period of up to three years, by submitting a written request specifying the reasons for the extension to the ESS director of science who will decide following the advice of the chair of the science advisory committee.

3.4.3. The PT may request their data be made open before the expiration of the restricted access period.

3.4.4. The online index shall catalogue raw data and associated meta data for example: sample names/compositions, instrument names and basic parameters, number of runs, volume of data, sample environment equipment used, sample data itself, temperature ranges, etc., and the title from the original proposal with the name of the PI).

3.4.5. Access to the online catalogue of open data will be given to a user, not associated with the PT, providing said user registers with ESS, fulfils the ESS requirements for access as defined in the ESS Access Policy, and accepts the terms of this scientific data policy.

3.4.6. Publications or other generated intellectual property resulting from access to open data from ESS will cite the original persistent identifier of the data, the PT and the proposal number.

4. RESULTS

4.1. Ownership of all results (intellectual property) derived from the analysis of the raw data shall be determined by the specific user agreement, subject to the European Spallation Source ERIC Intellectual Property Rights and Inventions Policy.

Landing pages

- Digital Object Identifier (DOI) must connect to accessible landing page, which displays metadata
- https://doi.org/10.17199/BRIGHTNESS.D5.1
- Need landing page server
- Users should be able to make their data public and acquire a DOI and landing page

Secure | https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1161

EARTHDATA Other DAACs Feedback

ORNL DAAC
DISTRIBUTED ACTIVE ARCHIVE CENTER FOR BIOGEOCHEMICAL DYNAMICS

About Us | Get Data | Submit Data | Tools | Resources | Sign in

Search ORNL DAAC | Search

DAAC Home > Get Data > Regional/Global > North American Carbon Program (NACP) > Landing page

NACP Aboveground Biomass and Carbon Baseline Data, V.2 (NBCD 2000), U.S.A., 2000

Overview

DOI	https://doi.org/10.3334/ORNLDAAC/1161
Version	2
Project	NACP
Release Date	2013-05-29
Usage	2356 downloads
Citations	18 publications cited this service

Download Data 14.1GB | User Guide

Spatial Coverage

Bounding rectangle

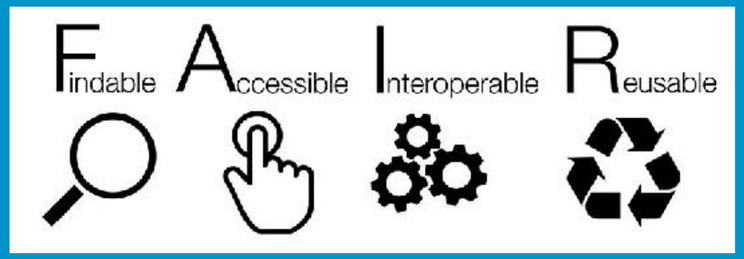
N: 49.79 | S: 26.52 | E: -67.96 | W: -126.46

Temporal Coverage

1999-01-01 to 2002-12-31

Description

The NBCD 2000 (National Biomass and Carbon Dataset for the Year 2000) data set provides a high-resolution (30 m) map of year-2000 baseline estimates of basal area-weighted canopy height, aboveground live dry biomass, and standing carbon stock for the conterminous United States. This data set distributes, for each of 66 map zones, a set of six raster files in GeoTIFF format. There is a detailed README companion file for each map zone. There is also an ArcGIS shapefile (mapping_zone_shapefile.shp) with the boundaries of all the map zones. A mosaic image of biomass at 240 m resolution for the whole conterminous U.S. is also included. Please read this important note regarding the differences of Version 2 from Version 1 of the NBCD 2000 data. With Version 1, in some mapping zones, certain land cover types (in particular Shrubs, NLCD Type 52) were missing from and unaccounted for in modeled estimates because of a lack of reference data. In Version 1, when landcover types were missing in the models, the model for the deciduous tree cover type was applied. While more woody vegetation was mapped, the authors think this had little effect on model performance as in most cases NLCD version 1 cover type was not a strong



EUROPEAN SPALLATION SOURCE

To be Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. (meta)data are registered or indexed in a searchable resource.
F4. metadata specify the data identifier.

TO BE ACCESSIBLE:

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2 metadata are accessible, even when the data are no longer available.

TO BE INTEROPERABLE:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

TO BE RE-USABLE:

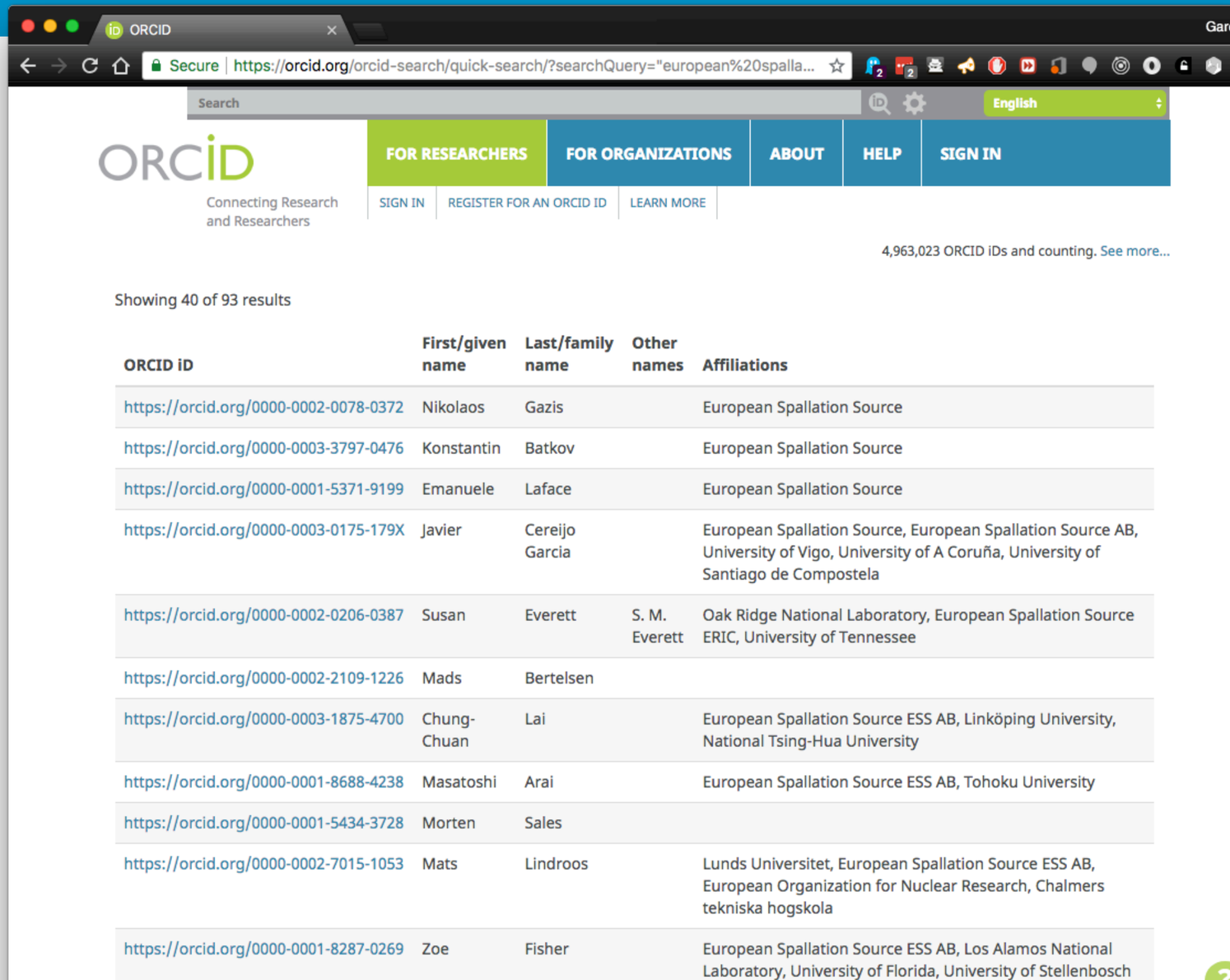
R1. meta(data) have a plurality of accurate and relevant attributes.
R1.1. (meta)data are released with a clear and accessible data usage license.
R1.2. (meta)data are associated with their provenance.
R1.3. (meta)data meet domain-relevant community standards

Photon and Neutron Open Science Cloud (PANOSC)

- FAIR - PaNOSC will comply with the FAIR principles in the following ways:
- Findable - all data will have a DOI, rich metadata, common api for federated search
- Accessible - api will support open protocol, metadata accessible even without data
- Inter-operable - metadata to follow community standards (Nexus), register metadata
- Reusable - follow community standardise metadata, clear licence (CC-BY)

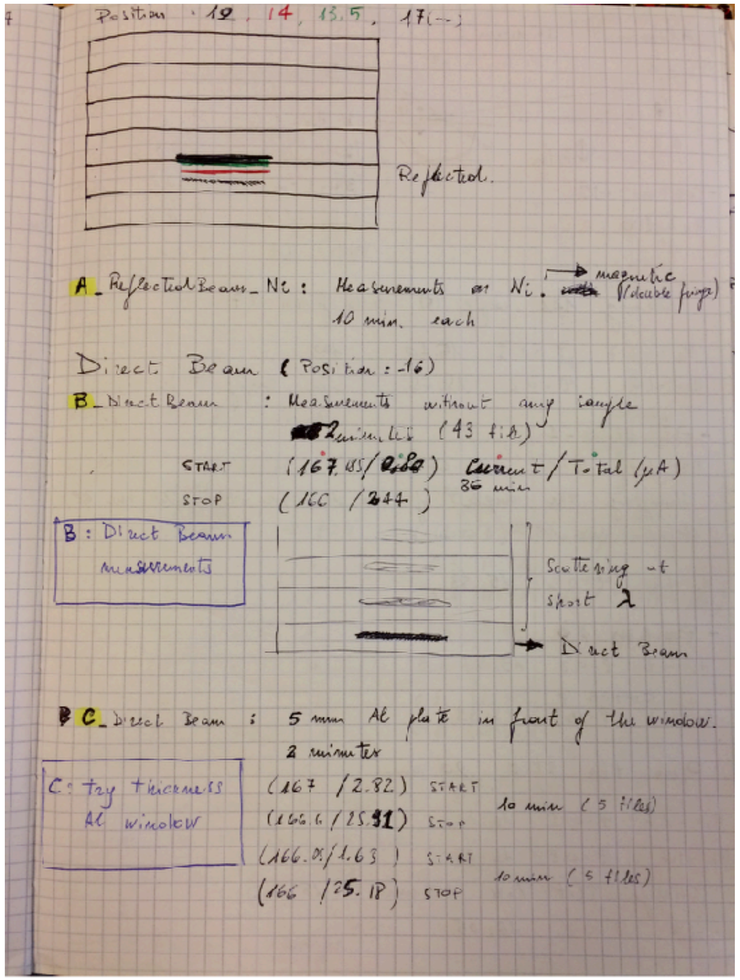
Researcher persistent identifier

- ORCID
- Can uniquely identify researcher using instruments
- Can follow data use and citations
- Data creator/steward can be identified uniquely

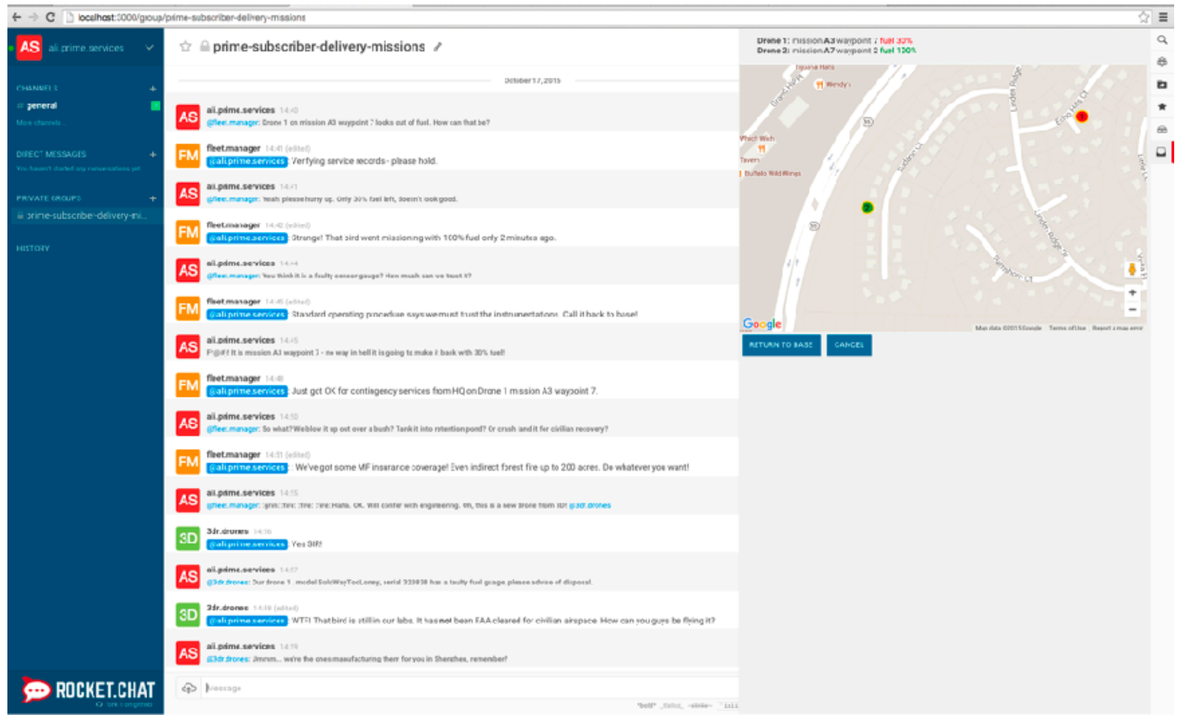


ORCID iD	First/given name	Last/family name	Other names	Affiliations
https://orcid.org/0000-0002-0078-0372	Nikolaos	Gazis		European Spallation Source
https://orcid.org/0000-0003-3797-0476	Konstantin	Batkov		European Spallation Source
https://orcid.org/0000-0001-5371-9199	Emanuele	Laface		European Spallation Source
https://orcid.org/0000-0003-0175-179X	Javier	Cereijo Garcia		European Spallation Source, European Spallation Source AB, University of Vigo, University of A Coruña, University of Santiago de Compostela
https://orcid.org/0000-0002-0206-0387	Susan	Everett	S. M. Everett	Oak Ridge National Laboratory, European Spallation Source ERIC, University of Tennessee
https://orcid.org/0000-0002-2109-1226	Mads	Bertelsen		
https://orcid.org/0000-0003-1875-4700	Chung-Chuan	Lai		European Spallation Source ESS AB, Linköping University, National Tsing-Hua University
https://orcid.org/0000-0001-8688-4238	Masatoshi	Arai		European Spallation Source ESS AB, Tohoku University
https://orcid.org/0000-0001-5434-3728	Morten	Sales		
https://orcid.org/0000-0002-7015-1053	Mats	Lindroos		Lunds Universitet, European Spallation Source ESS AB, European Organization for Nuclear Research, Chalmers tekniska hogskola
https://orcid.org/0000-0001-8287-0269	Zoe	Fisher		European Spallation Source ESS AB, Los Alamos National Laboratory, University of Florida, University of Stellenbosch

Researcher logbooks



- How to capture researcher logs
- Use a web chat application like Rocket Chat to capture collaboration



Summary

- ESS requires Open Access metadata and data
- SciCat already provides PIDs
- We still need DOI integration and landing pages
- BrightnESS data can be a good test case for SciCat publication workflow

- Open Access data
- Persistent Identifiers (PIDs)
- Digital Object Identifiers (DOIs)
- Landing page