
Communities adoption of EUDAT services and Lessons Learnt



FINAL PROGRAMME | REGISTRATION | POSTERS & DEMOS | VENUE & ACCOMMODATION

Communities adoption of EUDAT services and Lessons Learnt

Let's meet at: 13:30 - 17:00, Room: Lagos | 24 Jan 2018

Session Chair: Daan Broeder, Meertens Institute

Objectives:

- Present how diverse communities have integrated EUDAT services with their own architecture
- Live demos
- Analyse what can be learned from communities adoption of EUDAT services

More on the session:

A series of presentations and demos that illustrate the uptake of EUDAT services by the communities. This work presented and demonstrated has been executed in the context of uptake plans by the EUDAT core communities or within the (extended) EUDAT Data Pilots. At the end of the demos a "Lessons Learnt" discussion will take place focusing on what has been EUDAT's contribution to the data and compute landscape from the communities perspective and what we have learned.

Agenda:

Demos of the EUDAT data pilots:

- **FAIR Data Pilot, Luiz Bonino:** The FAIR Data Point is an approach developed by the Dutch Techcentre for Life Sciences to expose metadata and data in a way compliant with the FAIR data principles. This approach includes a RESTful API as well as a metadata layered structure that allows data repositories and data owners to define the rich metadata required by the FAIR principles. In this EUDAT pilot we extended the B2SHARE service to align with the FAIR Data Point specifications and, therefore, increasing the B2SHARE FAIRness, which improves the findability, accessibility, interoperability and reuse of the data deposited in this service. In this demo we will demonstrate the extensions made on B2SHARE to comply with the APIs and metadata definitions of the FAIR Data Point and how these extensions can improve findability, accessibility, interoperability and reuse. ([Download presentation](#))
 - **Porto University Data Pilot, Christina Robeiro, Joao Rocha:** The DataPublication@ U.Porto pilot uses Dendro (github.com/feup-infolab/dendro), a Research Data Management platform, for data description and preparation, and B2SHARE as a repository for datasets in the long tail of science. Dendro is built on open-source tools and its ontology-based data model provides a flexible metadata infrastructure that can be extended via domain-specific ontologies. The interface of Dendro and B2SHARE performs dataset deposits on demand, transparently loading data and metadata into B2SHARE. Descriptors present in the B2SHARE model are automatically filled in, and a full RDF metadata record is associated with the dataset, for interoperability with other systems. The [DataPublication@U.Porto](#) demo provides a compact view of the data management workflow in a project developed by a research team. ([Download Presentation](#))
 - **PAIRQURS Data Pilot, Javier Arino:** PAIRQURS is the data publication component of LIFE+RESPIRA, a citizen science project that models and maps at fine scale the air pollution levels in the city of Pamplona and that can be extended to other cities. Two years worth of air pollution data (gas concentrations and particles) collected by roaming portable sensors carried around by volunteers allowed production of high-resolution maps adapted to various climatic and calendar conditions. The maps and datasets are deposited in B2SHARE under a detailed metadata schema allowing retrieval of relevant data for further analysis and for ingestion to a route planner that selects clean routes. This demo describes how users may invoke an app to calculate a bike or pedestrian route that accounts for expected pollution levels. The app looks up the current or forecast weather and calendar conditions and queries B2SHARE for the relevant pollution map layer. The app then calculates the cleanest, fastest, or easier route based on the retrieved layer. ([Download Presentation](#))
 - **ENES core community, Hannes Thiemann, Heiner Widmann, Asela Rajapakse:** This demo will showcase how the requirements of the ENES partners involved in EUDAT have been incorporated into the relevant EUDAT services. These are B2FIND, B2HANDLE, B2SHARE and GEF. The current and intended integration of these services into the ENES landscape will be illustrated by some examples: ([Download presentation](#))
 - The metadata publication of quality-checked data in B2FIND
 - The role of the B2HANDLE Library in the ESGF software stack
 - The intended task of the B2SHARE repository in the ECAS workflow within the framework of EOSC-hub
 - The uptake of GEF against the background of the ENES use case including the GEF prototype for EGI integration as a first step towards integration of the GEF with the climate4impact platform
 - **CLARIN core community, Dieter van Uytvanck, Josef Mitsuka, Claus Zinn:** This demo will showcase how EUDAT is used by CLARIN as a safe backup solution, the B2SAFE uptake within CLARIN and the how B2DROP has been bridged with CLARIN's Language Switch Board ([Download presentation](#))
 - **ILTER core community, Johannes Peterseil, Christoph Wohner:** The documentation of datasets and the related observation facilities are important requisites for reproducible science. This enhances both the discoverability and reproducibility of the analysis through standardised metadata. Depositing research data objects, linked to persistent identifiers, in a managed and curated repository is another important aspect. The development focused on two aspects: a) the generic documentation of observation facilities and b) the linkage of B2SHARE as a trusted repository to a community metadata catalogue. This resulted in the adaption of [DEIMS-SDR](#) including both aspects. The demonstration shows the features of DEIMS-SDR and the integration of B2SHARE as trusted repository. ([Download presentation](#))
 - **EISCAT Data Pilot, Claudia Martens:** The European Incoherent Scatter Scientific Association (EISCAT) operates three large incoherent scatter radars, which are instruments capable of probing the Earth's ionosphere as well as near-Earth objects such as meteors. These instruments will soon be replaced by the next generation incoherent scatter radar, EISCAT 3D, expected to start in 2021. It will consist of arrays of almost 10,000 antennas which will provide three-dimensional measurements of the ionospheric incoherent scatter parameters as well as other targets. The data volumes will be so large that it is practically impossible for users to download and process data at their offices. Therefore a suitable user portal and workflow system, interfacing to storage and HPC/HTC systems, will be necessary. The purpose of this EUDAT pilot, Unified Access to EISCAT radar data, has been to evaluate the use of EUDAT B2* services by establishing a unified archival and search system for the existing EISCAT incoherent scatter radars. In this demo, it will be shown how to search for data from an EISCAT experiment by utilizing different facets. B2FIND will be used to find experiment level B2SHARE records in order to identify an event of interest, assigned level 2 data may then be accessible through the links from the corresponding level 2 B2SHARE record. The outcomes of the pilot will be used to explore how to customize EUDAT services for data storage, user authentication, data discovery and data access for EISCAT 3D. ([Download presentation](#))
 - **EPOS core community, Alberto Michelini, Peter Danecek:** We present the activities carried out by the EPOS community group in EUDAT2020. These stem from the implementation of the following services all instrumental for the realization of the EPOS thematic core services. They include: Data access control (AAI); Data management, replication and preservation; Data accessibility; Remote processing (data intensive); Data Provenance; Data Versioning; ArrayDB (data intensive). The data demo will include the following "production-ready" services: Data access control (AAI) to the European Integrated seismic Data Archive (EIDA); Integration through B2ACCESS (GFZ); Management, replication and preservation; B2SAFE integration through node implementation (INGV); Data accessibility through HTTP-API integration; EPOS API (INGV), etc. ([Download presentation](#))
 - **ELIXIR core community, Jinny Chien:** The ELIXIR Compute Platform is being established to provide a pool of computational and storage resources to enable life-science analysis. Key to many of the analysis activities is the local availability of reference data sets. These reference data sets vary in size, complexity and release frequency but are sources of information essential for downstream analysis. Given the wide availability of local analysis resources, a service that can simplify the process of maintaining local replicas of reference data sets with minimal local intervention are highly desired. The Reference Data Set Distribution Service (RSDSDS) addresses this use case by providing a mechanism for storing the structure of a reference data set, and then building and making releases to subscribed consumers of the service. This use case is one that we see occurring beyond the life-science community. The demonstration will take the audience through the registration of a reference data set and the creation of a release by the data owner. A new data user (e.g. researcher or site admin) will select a data set and a version for replication on to their site. We will also show how a new data set release is replicated to existing subscribers across multiple sites. ([Download presentation](#))
 - **Discussion on lessons learnt**
- Presentations:**
- **FAIR Data Pilot, Luiz Bonino** ([Download presentation](#))
 - **Porto University Data Pilot, Christina Robeiro, Joao Rocha; Yulia Karimova** ([Download Presentation](#))
 - **PAIRQURS Data Pilot, Javier Arino** ([Download Presentation](#))
 - **ENES core community, Hannes Thiemann, Heiner Widmann, Asela Rajapakse** ([Download presentation](#))
 - **CLARIN core community, Dieter van Uytvanck, Josef Mitsuka, Claus Zinn** ([Download presentation](#))
 - **ILTER core community, Johannes Peterseil, Christoph Wohner** ([Download presentation](#))
 - **EISCAT Data Pilot, Claudia Martens** ([Download presentation](#))
 - **EPOS core community, Alberto Michelini, Peter Danecek** ([Download presentation](#))
 - **ELIXIR core community, Jinny Chien** ([Download presentation](#))
 - **ICOS, Maggie Hellstrom** ([Download presentation](#))



[Read more](#)