



Information and Data Management Repository Platform for nanoscience in Europe

The efficient data archiving for nanoscience community is a key challenge, i.e. harvesting from open-access scientific Data Repositories (DR) that could support sample/material preparation protocols with absolute metrology, and adequate metadata for the characterization and scientific investigations. Existing standards, recommendations and evolving best practices of data management should be incorporated, as well as sensible reuse of existing e-infrastructures where applicable rather than building own e-infrastructure for nanoscience from scratch.

The Scientific Challenge

To address the above challenges a very close cooperation with EUDAT and the adoption of their results, whenever possible, is of great significance. We propose to use EUDAT data services to:

- Focus on developing a data service around the NFFA-EUROPE IDRIP rather than developing yet another e-infrastructure;
- Provide data with clear identity as many NFFA partners do not mint persistent identifiers for data and some EUDAT services offer the data identifiers functionality out-of-box;
- Support sharing of long tail experimental data (via B2SHARE);
- Provide easy and flexible discovery of data in a central location (via B2FIND);
- Explore opportunities for scalable and trusted storage and replication of raw experimental data (via B2SAFE).

These services will be integrated into the NFFA-EUROPE IDRIP, with due consideration to the actual data management policies and technology maturity of the project partners.

Who benefits and how?

NFFA-EUROPE relates to the topic “Advanced material research based on large-scale facilities” that aims to further the integration of material science studies, fabrication and analysis (emerging from nanofoundry and characterization research) performed at laboratories connected to state-of-the-art large scale facilities such as neutron sources, synchrotron radiation sources, and free electron lasers (FELs).

Synergies will be realised by NFFA-EUROPE across a range of national and international activities in the fields of metrology, nano-safety, nano-electronics, analysis and conservation of cultural heritage, biomedical nano-imaging and nano-medicine.

Technical Implementation

The overarching goal of NFFA-EUROPE is to implement the first open-access research infrastructure as a platform supporting comprehensive projects for multidisciplinary research at the nanoscale, extending from synthesis to nanocharacterization to theory and numerical simulation.

With its extensive service offer, EUDAT provides a couple of standard services that can be interesting for the NFFA-EUROPE infrastructure. These are for example B2SHARE for publishing datasets, B2FIND to discover datasets and B2ACCESS for user authentication, but also future services that are not in production yet, e.g. the Data Type Registry or the B2NOTE annotation service.

Currently, it is planned to integrate B2SHARE for publishing NFFA datasets on demand. By default, datasets captured in the context of NFFA are stored at the local facility where they have been produced. In addition, the datasets are registered by reference, e.g. using a specific identifier or an URI, at the Information and Data Repository Platform (IDRP) (Figure 1). Afterwards, the access is possible either via the local facility, as long as the scientist is on site, or via a remotely accessible NFFA Portal where datasets can be shared and published on demand, or enforced by policy.

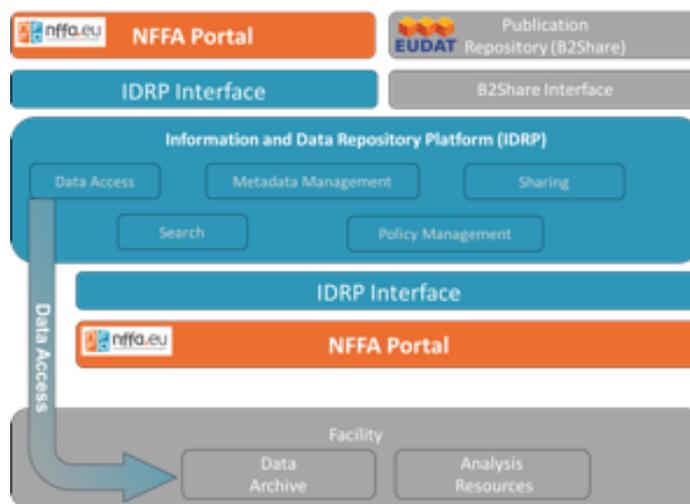


Figure 1 - NFFA Information and Data Repository Platform

The publication of a dataset is performed in several steps. First, the user who wants to publish a dataset has to authorize in B2SHARE using the OAuth2 workflow of B2ACCESS. If publishing is enforced by policy, this step has to be carried out on behalf of a curator, e.g. using preconfigured credentials. After authentication, the data that should be published is transferred by the IDRP from the local archive at the facility to a B2SHARE deposition object. If the local data archive is trusted, it is imaginable to let B2SHARE also reference to the dataset, but for the time being it is assumed that local data archives are not trusted in most cases. As soon as all data have been stored in the deposition object, the object will be committed. In the first prototype, only metadata from the generic domain will be provided, but for the future it is planned to establish a dedicated domain for nanoscience. This effort has already been briefly discussed with the responsible persons and will be deepened in future.

As soon as the deposition object is committed and the obtained persistent identifier supplied by B2SHARE is stored at the IDRP, the dataset is considered published and retrievable via the NFFA Portal and via the EUDAT infrastructure. As there is an automated interface between B2SHARE and B2FIND, the NFFA data record will be present on two EUDAT services: B2SHARE and B2FIND.

Preliminary Results

- Evaluation of B2ACCESS: We evaluated the B2ACCESS service as possible tool for the Authorization and Authentication Infrastructure we are developing for the NFFA IDRP.
- Evaluation of B2SHARE: B2SHARE has been successfully used for deposition of data and seems to be a promising technology for publishing NFFA datasets. NFFA will use an instance located at STFC configured in order to match NFFA requirements.
- Evaluation of B2FIND/NFFA Metadata Domain: Once the nanoscience-specific set of metadata fields will be discussed by NFFA, the metadata profile for the NFFA data provision in B2SHARE will be ready.



Contacts

- Stefano Cozzini, Italian National Research Centre (CNR), stefano.cozzini(at)iom.cnr.it

Further Information

[Learn more about the EUDAT/NFFA Europe Collaboration here](#)
[Scientific Challenges behind the Pilot](#)

[Read more](#)