

# **RDA-PaNSIG. Workshop on Interoperability**



**Monday, 3 April 2017 - Tuesday, 4 April 2017**

**ALBA Synchrotron**

## **Scientific Programme**

In this workshop, we will explore how to support and enable the interoperability of data within and across user facilities. This would include considering the following aspects:

**Data Discovery:** How do we integrate existing online resources, such as ICAT, the Materials Data Facility, and Globus Online, so that the users are able to locate data to which they have access, no matter where it is stored? This would consider the use of common data and metadata standards and protocols for discovery, access and interpretation of data sets.

**Authentication:** How do we provide a unified authentication scheme that determines whether the users have access to the remote data? This would consider common schemes to identify users and how to accommodate different local authentication schemes.

**Server Queries:** What queries are required so that users can inspect the remote data and selectively download subsets of the data and metadata?

**Transfer Protocols:** How do we handle network requests and how do we serialize the data that is returned? Also, how do we manage the metadata across different repositories (something like... Open Archives Initiative Protocol...)?

**APIs:** What kind of APIs are needed/provided? How would applications need to be modified?

**Reproducibility:** Do we need to capture and store queries to make a) processes reproducible and/or b) facilitate actions repeatedly accessing the same data slab?

**Controlled Vocabulary:** Do we need controlled vocabularies for data and/or resource discovery?

## Data discovery

How do we integrate existing online resources, such as ICAT, the Materials Data Facility, and Globus Online, so that the users are able to locate data to which they have access, no matter where it is stored? This would consider the use of common data and metadata standards and protocols for discovery, access and interpretation of data sets.

## Technical Aspects: Authentication, transfer protocols, Data Formats, Queries and APIs

**Authentication.** How do we provide a unified authentication scheme that determines whether the users have access to the remote data? This would consider common schemes to identify users and how to accommodate different local authentication schemes.

**Data Formats:** Is HDF5 the standard de facto for binary scientific data?, Nexus?. How does it cover the requirements in terms of transfer protocols and throughput? What are the prospects?

Transfer Protocols. How do we handle network requests and how do we serialize the data that is returned?

APIs. What kind of APIs are needed/provided? How would applications need to be modified?

## **Standards: Vocabulary, keywords and methods for queries.**

Controlled Vocabulary. Do we need controlled vocabularies for data and/or resource discovery?

Server Queries. What queries are required so that users can inspect the remote data and selectively download subsets of the data and metadata?

Reproducibility. Do we need to capture and store queries to make a) processes reproducible and/or b) facilitate actions repeatedly accessing the same data slab?

## **PROJECTS AND PROSPECTS IN THE EU, PANDATA/PANDAAS, CALIPSO, LEAPS,**