Data and Metadata for Nanoscience: Nano foundries and Fine Analysis (NFFA) Data Repository:

Stefano Cozzini
CNR/IOM and eXact lab srl
Trieste, Italy
Agenda

• Introducing the NFFA project
• The NFFA DATA repository efforts
• Data and Metadata for Nanoscience
• Conclusions
The NFFA project:

• Started as A EU-funded design study for distribute research infrastructure in nanoscience (2008-2010)
• Led by CNR/IOM (Trieste) it involves several large scale facilities in Europe
• Now in a demonstrator phase sponsored by MIUR (2011-2014) for the Italian site at CNR/IOM and Elettra
• Currently a new EU proposal is under preparation
THE NFFA MISSION (www.NFFA.eu)

A DISTRIBUTED INFRASTRUCTURE LINKED TO ANALYTICAL LSFs

The NFFA aim to support the construction and operation of an ERIC consisting of Nanoscale Science Research Centers at European sites that already host Large Scale Facilities for Fine Analysis of matter.

workshop on Big Data and Open Data, 7th -8th May 2014
A DISTRIBUTED ERIC WITH
COMMON METROLOGY
COMMON PROTOCOLS
AND STANDARDS
SPECIFIC TOOLS
OR LAB-SHARING
ACTIVITY CONNECTED TO
LOCAL SCIENTIFIC ENVIRONMENT

IMPROVED AND
REPRODUCIBLE
SAMPLE DEFINITION
AT NEARBY
COMPLEMENTARY LSFs

BIO
SYNCHROTRON
NEUTRONS
MATERIALS
FEL
ICT

workshop on Big Data and Open Data, 7th -8th May 2014
Data Repository for Nanoscience

- NFFA ha address the creation of the first Data Repository (DR) in nanoscience.

- Goals:
  - Store all the data produced in the NFFA centers
  - Make data accessible/ searchable
  - Less invasive as possible for the final users
Smart application forms directly connected to Data Repository.

**Data Repository**

**Management:**
- User Access: users flux through facilities
- Technical Liaison: support to inexperienced users
- Proposal Review

**Search criteria:**
特殊关键词 semantic search

**Data Metadata:**
- physical for reproducibility
- instrumental for common metrology
- specific for smart search

**Data**

**Instruments**

**Proposals**

**Users**

**Registration** proposal arrival user operations data analysis dissemination

Workshop on Big Data and Open Data, 7th-8th May 2014
Not only data...

- The DR needs to include the relevant information
  - for data analysis
  - for the full reproducibility of preparations and experiments

The DATA REPOSITORY should enable the exploitation of truly complementary data as obtained by different methods on true replica samples and environment conditions

**WHICH KIND OF METADATA ?**
The NFFA DR demonstrator

• A prototype implemented in the design study developed in collaboration with Elettra
• Currently enhancing it for further expansion
  – Active collaboration with Karlsruhe IT IPE
• The demonstrator collects data from three different instruments:
  – A Scanning Electron Microscope (SEM) instrument,
  – a synchrotron radiation beamline for spectroscopy (APE)
  – an open package for first principle quantum simulation (Quantum Espresso)
The general schema of the demonstrator
How Plugins works

1. Injection plugin converts all data in HDF5 files and all metadata in HDF5 attributes transparently to the users

2. It loads files into the KIT DR via REST interface
Open Issues for NFFA DR

• Authentication/Authorization policies for users and data in a distribute environment

• How to enable smart data discovery?
  complex problem addressed in many ways in different scientific enviroment..
Several conventions/standards around..
What about nanoscience?

At present the data generated within each of the many techniques are handled using different formats and with different underlying data models thus effectively preventing reuse/interoperability at the experimental data level.

This makes extremely difficult to discover and integrate nanoscience data

Are we missing metadata standards for nanoscience?
NFFA proposal

The project envisions a shared effort driven by nanoscientists to:

• Recognize common needs on nano science data
• propose/promote a collaboration to address the challenge
Conclusions

• Data repository needs for NFFA identified
• Demonstrator DR built with the scope of validating the approach by user communities
• A new enhanced version of the prototype is under development
• A nanoscience community effort to identify/define metadata standards for nanoscience should be promoted within appropriate organizations