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

Stefano Cozzini

CNR/IOM and eXact lab srl

Trieste, Italy

Istituto Officina dei Materiali

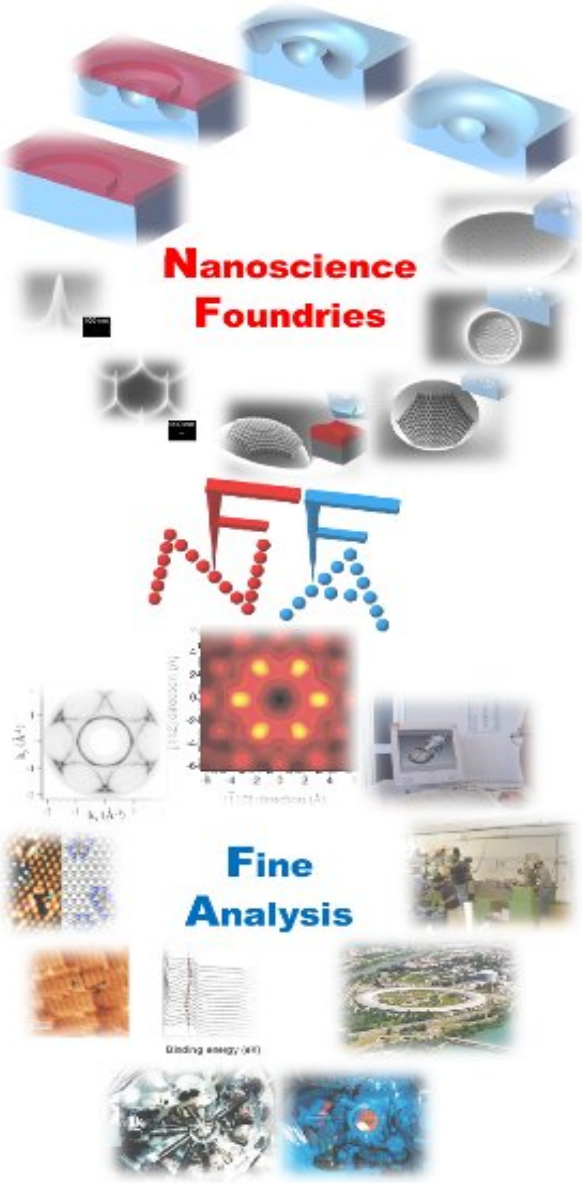
workshop on Big Data and Open Data,
7th -8th May 2014

Agenda

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

The NFFA project:

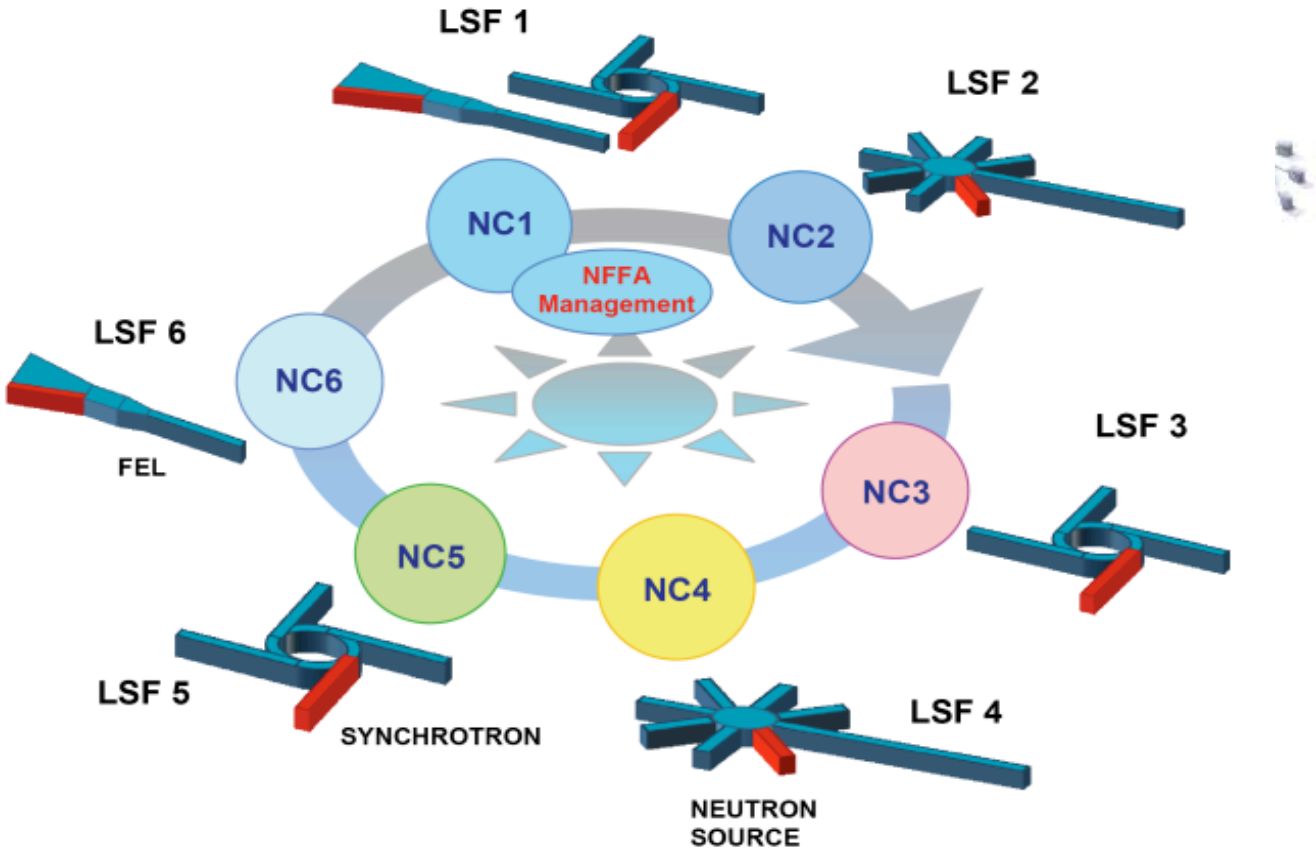
- Started as A EU-funded design study for distribute research infrastructure in nanoscience (2008-2010)
- Leaded by CNR/IOM (Trieste) it involves several large scale facilities in Europe
- Now in a demonstrator phase sponsored by MIUR (2011-2014) for the italian site at CNR/IOM and Elettra
- Currently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

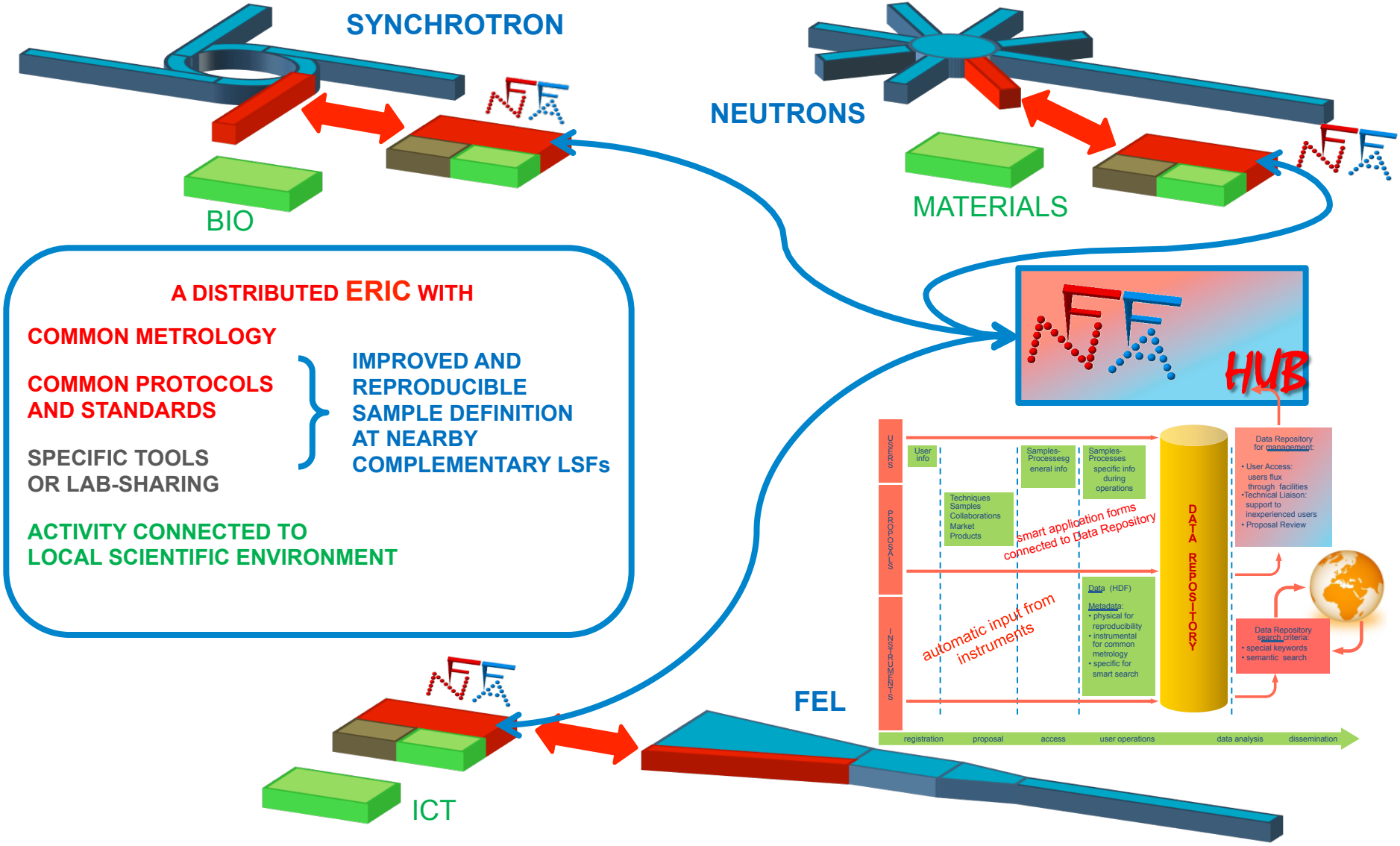


THE NFFA MISSION (www.NFFA.eu)

A DISTRIBUTED INFRASTRUCTURE LINKED TO ANALYTICAL LSFs

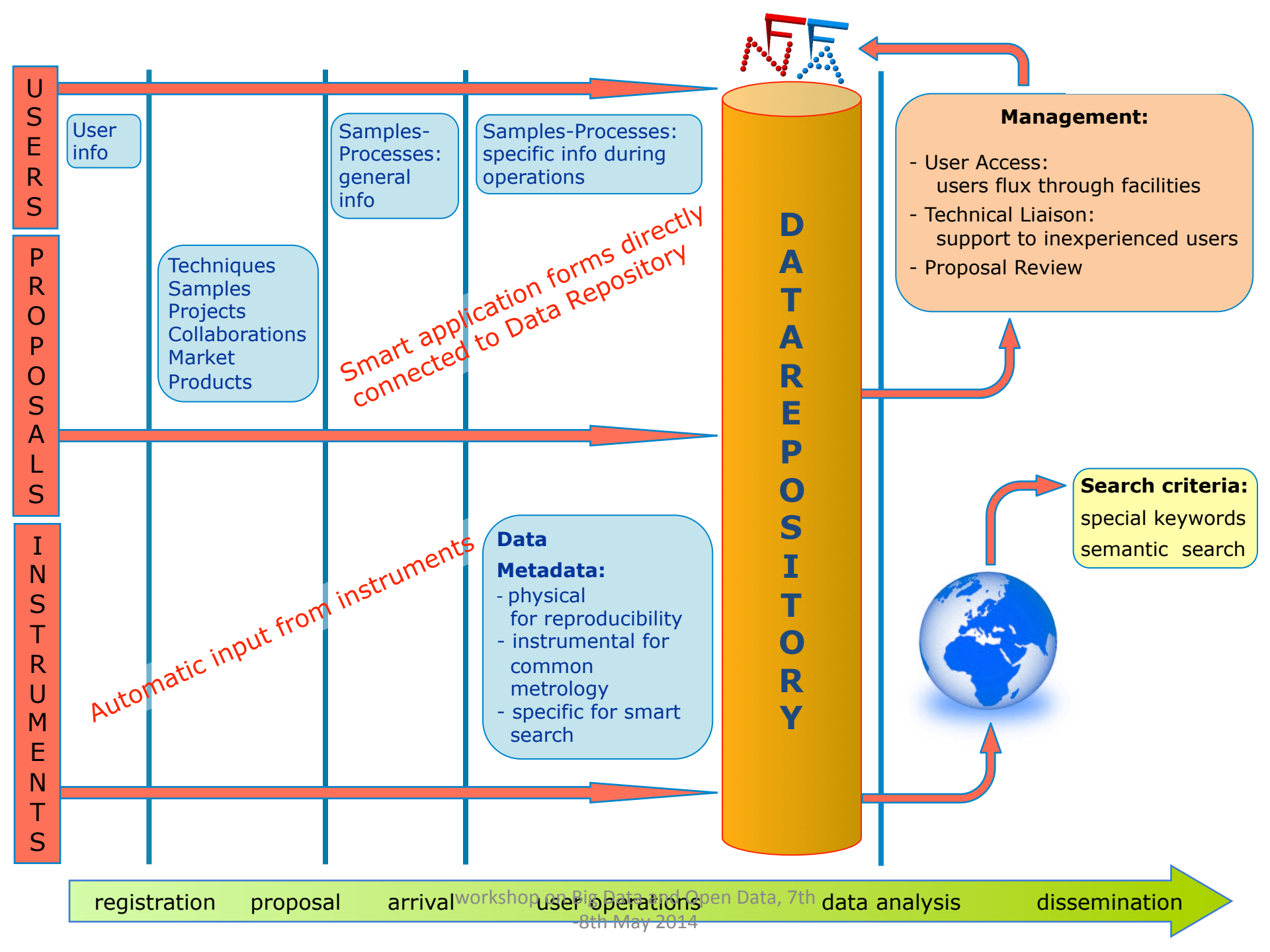
The NFFA aim to support the construction and operation of an ERIC consisting of Nanoscale Science Research Centers at European sites that already host Large Scale Facilities for Fine Analysis of matter.




SYNCHROTRON
NEUTRONS
BIO
MATERIALS
A DISTRIBUTED ERIC WITH
COMMON METROLOGY
COMMON PROTOCOLS AND STANDARDS
SPECIFIC TOOLS OR LAB-SHARING
IMPROVED AND REPRODUCIBLE SAMPLE DEFINITION AT NEARBY COMPLEMENTARY LSFs
ACTIVITY CONNECTED TO LOCAL SCIENTIFIC ENVIRONMENT
HUB
User info
Techniques Samples Collaborations Market Products
Samples-Processesg eneral info
Samples-Processes specific info during operations
Smart application forms connected to Data Repository
DATA REPOSITORY
Data (HDF)
Metadata:
• physical for reproducibility
• instrumental for common metrology
• specific for smart search
automatic input from instruments
Data Repository for management:
• User Access: users flux through facilities
•Technical Liaison: support to inexperienced users
• Proposal Review
Data Repository search criteria:
• special keywords
• semantic search
registration
proposal
access
user operations
data analysis
dissemination
FEL
ICT

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


USERS
PROPOSALS
INSTRUMENTS
User info
Samples-Processes: general info
Samples-Processes: specific info during operations
Techniques
Samples
Projects
Collaborations
Market
Products
Smart application forms directly connected to Data Repository
Automatic input from instruments
Data
Metadata:
- physical for reproducibility
- instrumental for common metrology
- specific for smart search
DATAREPOSITORY
Management:
- User Access: users flux through facilities
- Technical Liaison: support to inexperienced users
- Proposal Review
Search criteria:
special keywords
semantic search
registration
proposal
arrival
user operations
data analysis
dissemination

Not only data…

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

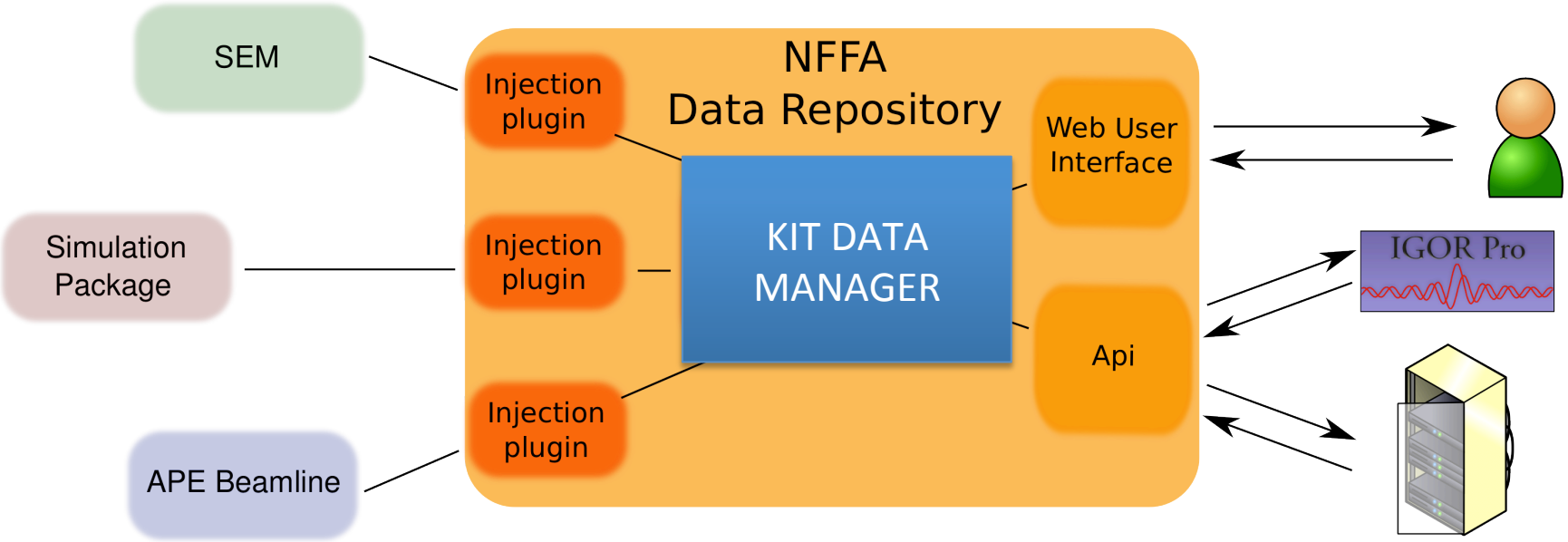
The DATA REPOSITORY should enable the exploitation of truly complementary data as obtained by different methods on true replica samples and environment conditions

WHICH KIND OF METADATA ?

The NFFA DR demonstator

- 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)

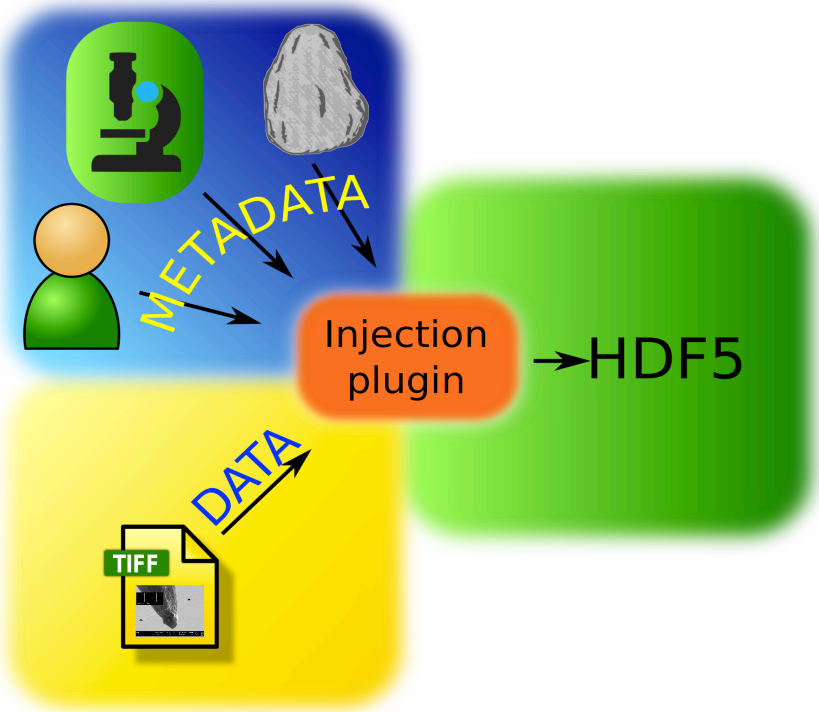
The general schema of the demonstrator

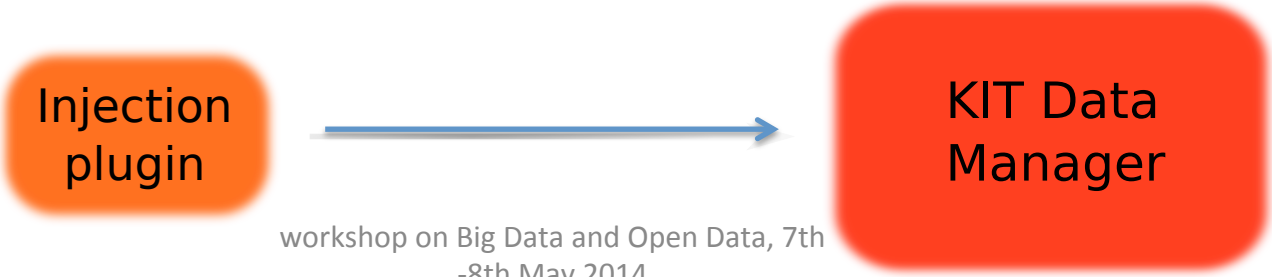


How Plugins works

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

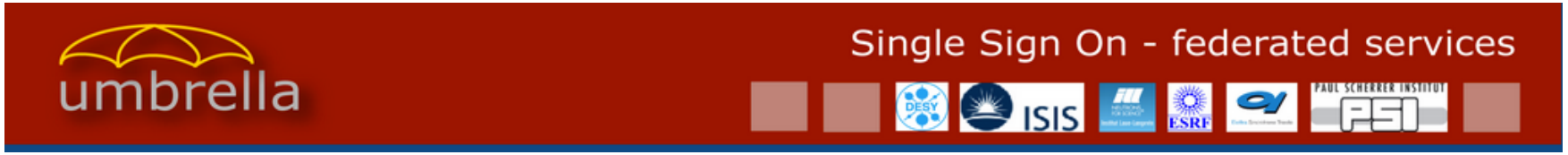
2 . It loads files into the KIT DR via REST interface





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..

NetCDF Climate and Forecast Metadata Convention

Title:	NetCDF Climate and Forecast Metadata Convention
Maintenance Authority:	
Version:	1.6
Date:	2011
Specification:	Original Website
Schema:	
Subjects:	• Weather Forecast

Dublin Core Metadata Element Set

Title:	Dublin Core Metadata Element Set
Maintenance Authority:	Dublin Core Metadata Initiative
Version:	1.1
Date:	2008
Specification:	
Schema:	Original Website
Subjects:	• Metadata

Content Standard for Digital Geospatial Metadata

Title:	Content Standard for Digital Geospatial Metadata
Maintenance Authority:	Metadata Ad Hoc Working Group, Federal Geographic Data Committee
Version:	2.0
Date:	1998
Specification:	Original Website
Schema:	
Subjects:	• Geography • Biology • Oceanography

Astronomy Visualization Metadata Standard

Title:	Astronomy Visualization Metadata Standard
Maintenance Authority:	
Version:	1.10
Date:	2008

Access to Biological Collection Data

Title:	Access to Biological Collection Data
Maintenance Authority:	Biodiversity Information Standards
Version:	2.06
Date:	2005
Specification:	Original Website
Schema:	Original Website
Subjects:	• Speciman data • Observation data • Biodiversity data

SNAPSHOTS taken from http://sdl.syr.edu/

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 nanoscients to:

- Recognize common needs on nano science data
- propose/promote a collaboratation 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