

PaNdata

*Photon and Neutron
Data Infrastructure*

1st EUDAT User Forum

Barcelona, 7-8 March 2012

Juan Bicarregui
STFC

Overview

The PaNdata Collaboration

The Vision

The PaNdata Europe Project

The PaNdata Open Data Infrastructure Project

Looking Forwards

The PaNdata Collaboration

- Established 2007 with 4 partners
- Expanded since to 11 (now 13) organisations
(see next slide)
- Aims:
 - “...to construct and operate a shared data infrastructure for Neutron and Photon laboratories...”

2007	2008	2009	2010	2011	2012	2013	2014
EDNS (4)		EDNP (10)		PaNdataEurope(11)		Pandata ODI(11)	



PaN-data Partners



PaN-data bring together 11 major European Research Infrastructures

ISIS is the world's leading pulsed spallation neutron source

ILL operates the most intense slow neutron source in the world

PSI operates the Swiss Light Source, SLS, and Neutron Spallation Source, SINQ, and is developing the SwissFEL Free Electron Laser

HZB operates the BER II research reactor the BESSY II synchrotron

CEA/LLB operates neutron scattering spectrometers from the Orphée fission reactor

JCNS Juelich Centre for Neutron Science

ESRF is a third generation synchrotron light source jointly funded by 19 European countries

Diamond is new 3rd generation synchrotron funded by the UK and the Wellcome Trust

DESY operates two synchrotrons, Doris III and Petra III, and the FLASH free electron laser

Soleil is a 2.75 GeV synchrotron radiation facility in operation since 2007

ELETTRA operates a 2-2.4 GeV synchrotron and is building the FERMI Free Electron Laser

ALBA is a new 3 GeV synchrotron facility due to become operational in 2010

MaxLab, Max IV Synchrotron

PaN-data is coordinated by the e-Science Department at the Rutherford Appleton Laboratory, UK



The Science we do - Structure of materials



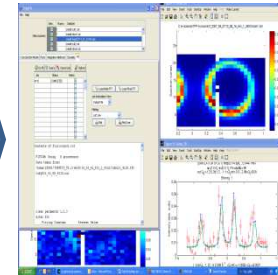
Visit facility on research campus



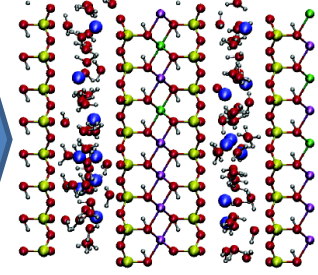
Place sample in beam



Diffraction pattern from sample



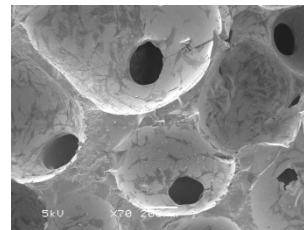
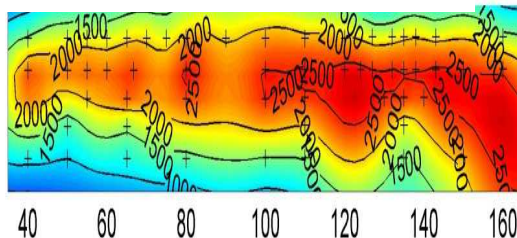
Fitting experimental data to model



Structure of cholesterol in crude oil

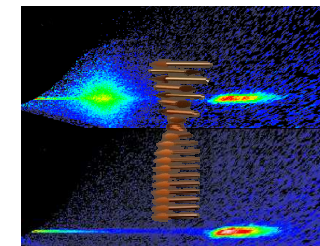
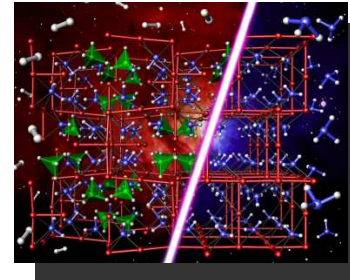
- Over 30,000 user visitors each year:
 - physics, chemistry, biology, medicine,
 - energy, environmental, materials, culture
 - pharmaceuticals, petrochemicals, microelectronics
- Over 5,000 high impact publications per year
 - But so far no integrated data repositories
 - Lacking sustainability & traceability

Longitudinal strain in aircraft wing



Bioactive glass for bone growth

Hydrogen storage for zero emission vehicles



Magnetic moments in electronic storage

PaN-data Applications

The partners operate hundreds of instruments used by over 30,000 scientists each year

These instruments support scientific fields as varied as:

- Physics, Chemistry, Biology, Material sciences, Energy technology, Environmental science, Medical technology and Cultural heritage

Applications include:

- crystallography that reveals the structures of viruses and proteins important for the development of new drugs
- neutron scattering that identifies stresses within engineering components such as turbine blades
- tomography that can image microscopic details of the 3D-structure of the brain

Industrial applications include pharmaceuticals, petrochemicals and microelectronics

PaNdata Facilities



Together represent a capital investment of over 3 Billion €

www.pan-data.eu

Overview

The PaNdata Collaboration

The Vision

The PaNdata Europe Project

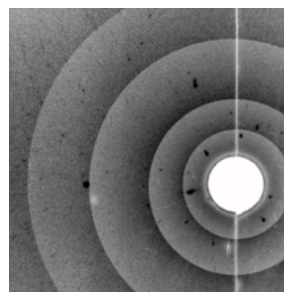
The PaNdata Open Data Infrastructure Project

Looking Forwards

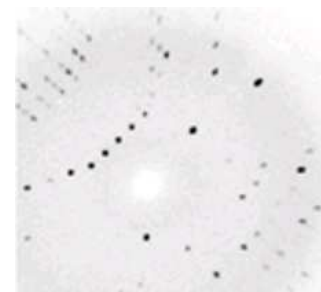
Science driver – Data Integration



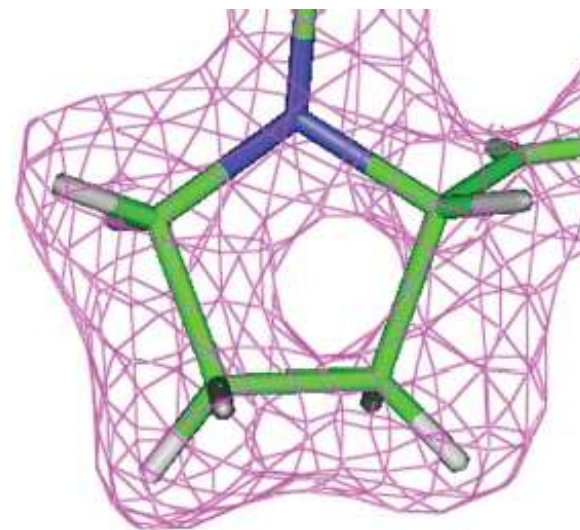
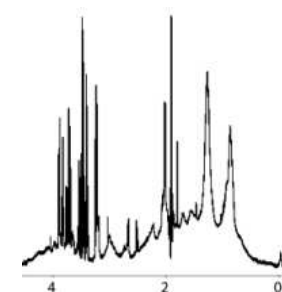
Neutron diffraction



X-ray diffraction

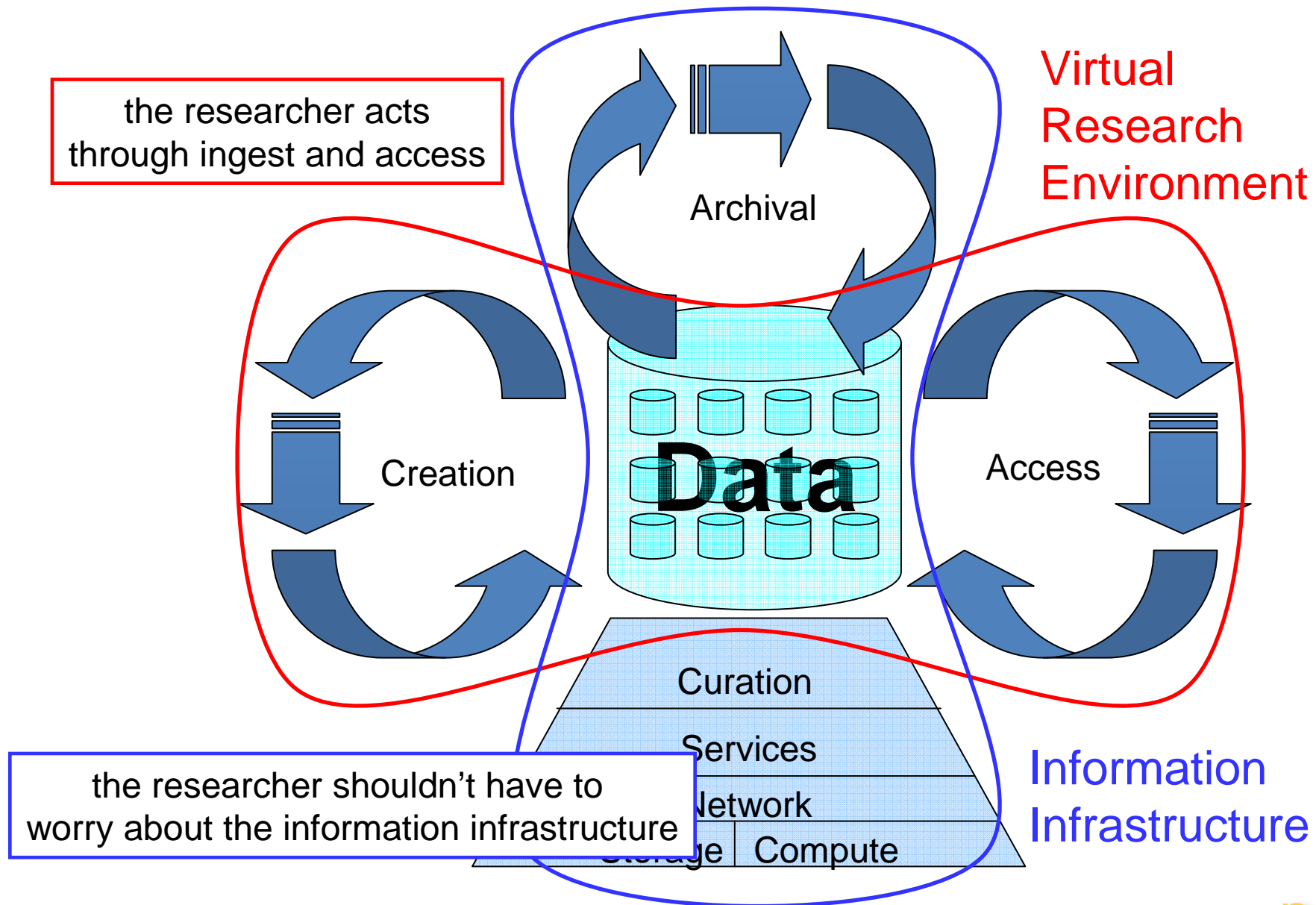


NMR



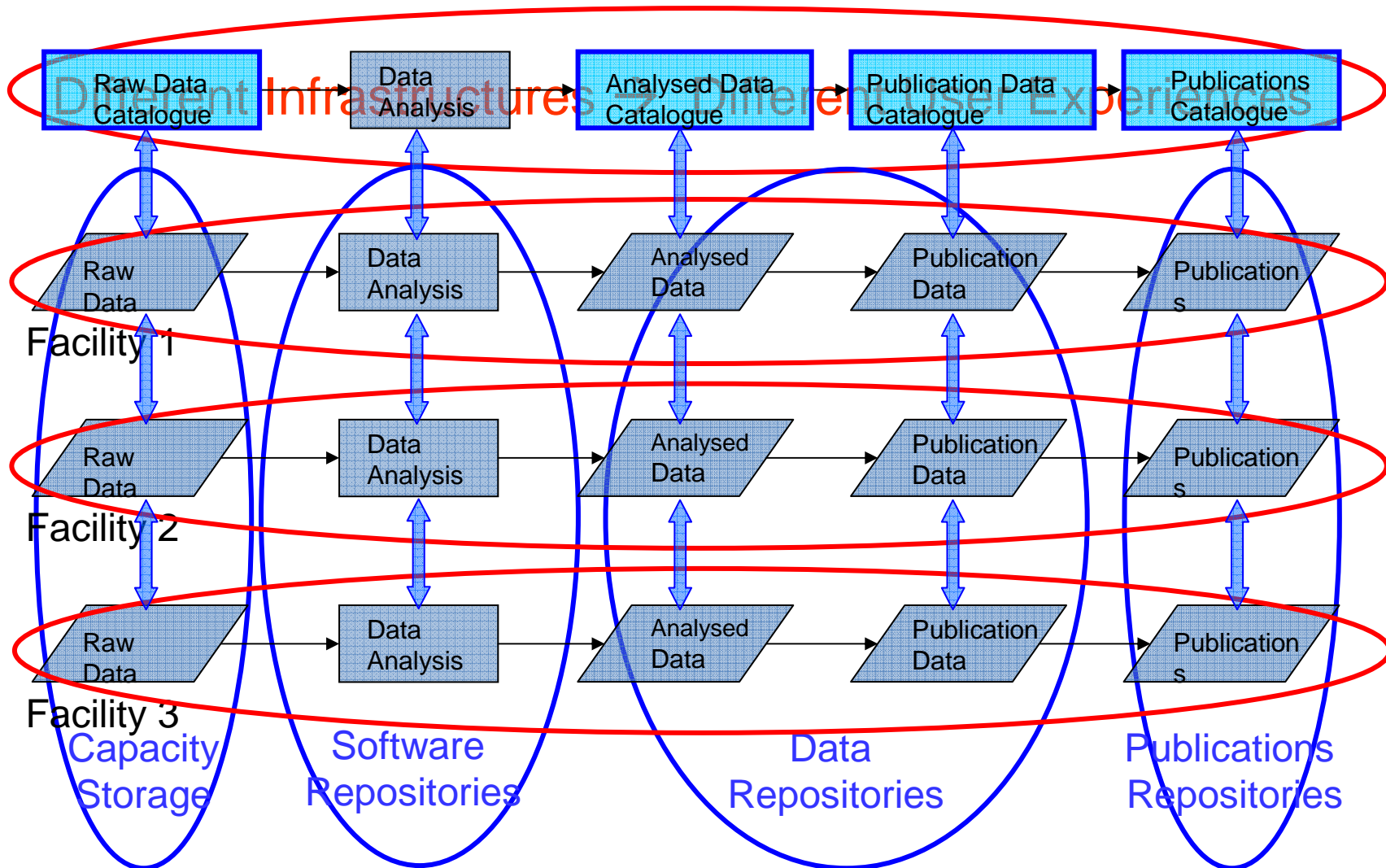
High-quality
structure
refinement

What is e-Infrastructure?



PaNdata Vision

Single Infrastructure → Single User Experience



In words:

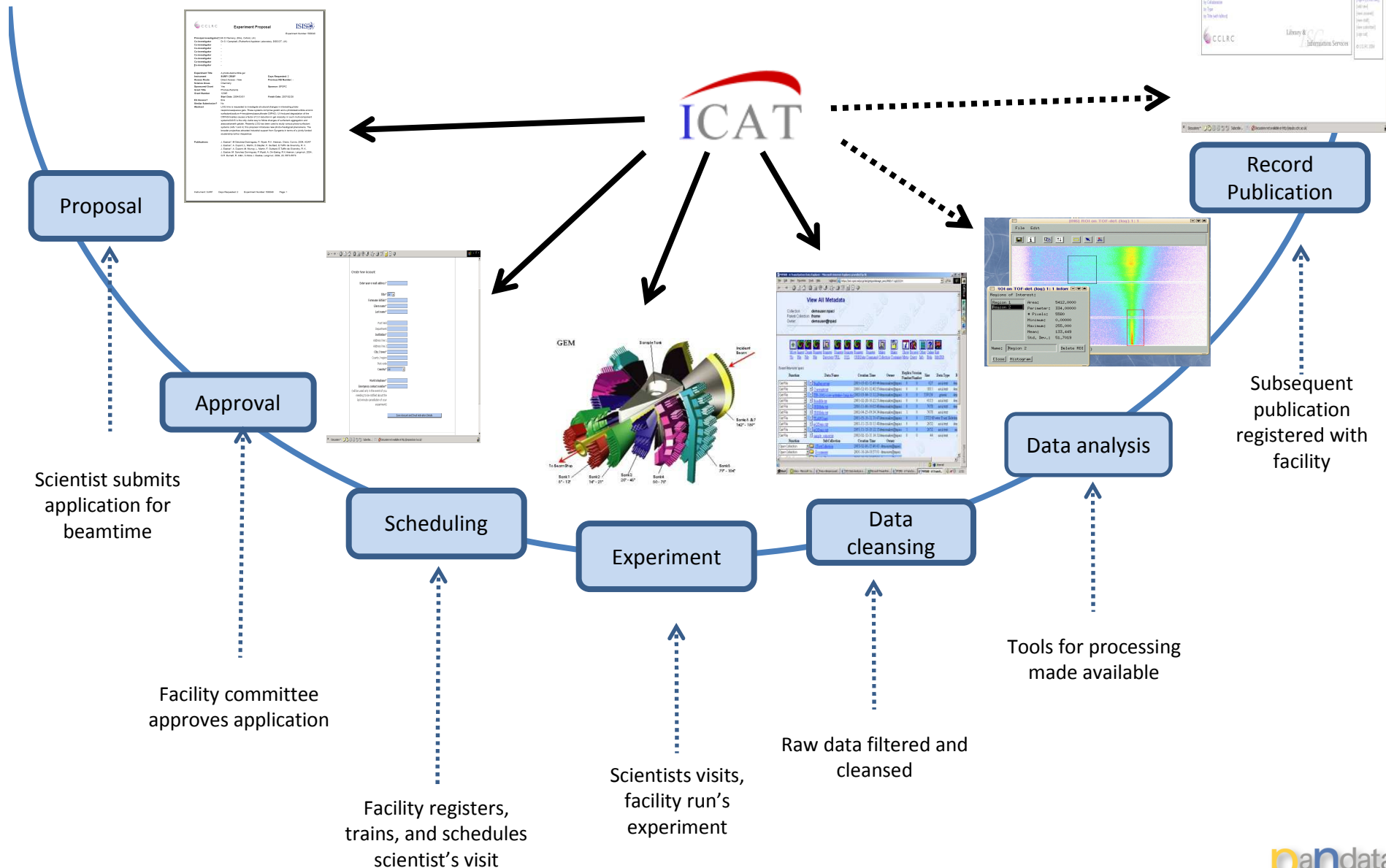
PANdata will provide our user communities with data repositories and data management tools to:

- deal with large sets and large data rates from the experiments,
- enable easy and standardised annotation of data,
- allow transparent and secure remote access to data,
- establish sustainable and compatible data catalogues, allow long-term preservation of data, and
- provide compatible open source data analysis software.

This will have a major impact on our scientific user community because it will offer:

- cross facility and cross discipline data analysis,
- secure access to large data sets over the network instead of using portable media,
- maintaining the records of science by having properly annotated data,
- linking publications to data,
- allowing efficient software developments, and
- efficient scientific collaborations across Europe by providing compatible data formats and analysis software.

Metadata and Digital Curation



Overview

The PaNdata Collaboration

The Vision

The PaNdata Europe Project

The PaNdata Open Data Infrastructure Project

Looking Forwards

PaN-data Standardisation

PaN-data Europe is undertaking 5 standardisation activities:

1. Development of a **common data policy** framework
2. Agreement on protocols for shared **user information exchange**
3. Definition of standards for common **scientific data formats**
4. Strategy for the interoperation of **data analysis software** enabling the most appropriate software to be used independently of where the data is collected
5. **Integration and cross-linking** of research outputs completing the lifecycle of research, linking all information underpinning publications, and supporting the long-term preservation of the research outputs



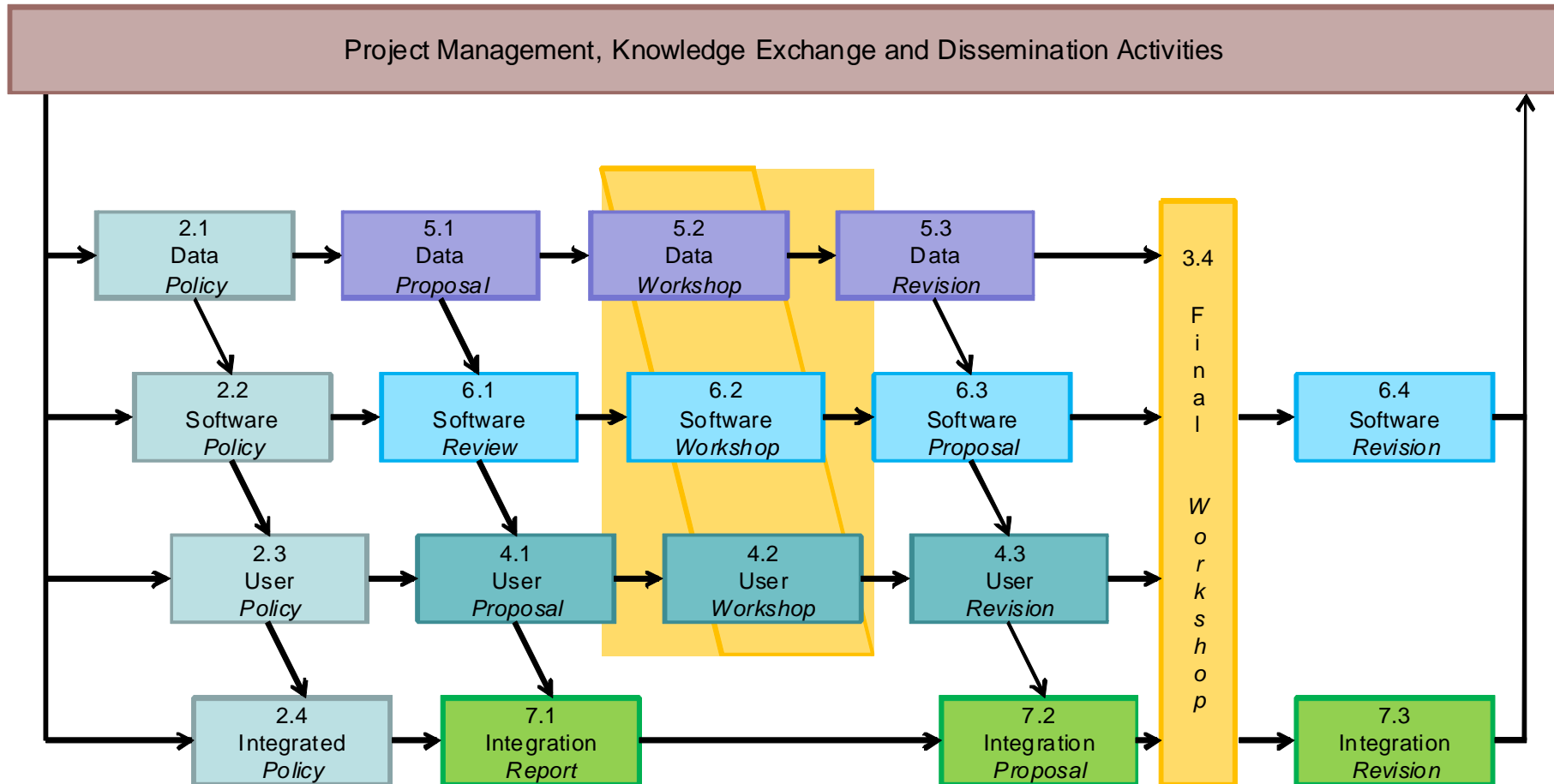
PaN-data Europe – building a sustainable data infrastructure for Neutron and Photon laboratories

PaN-data Europe Timeline

PaN-data Europe runs from June 2010 until December 2011 with workshops in Spring and Autumn 2011.



Dependencies



Dependencies between the major project tasks

Overview

The PaNdata Collaboration

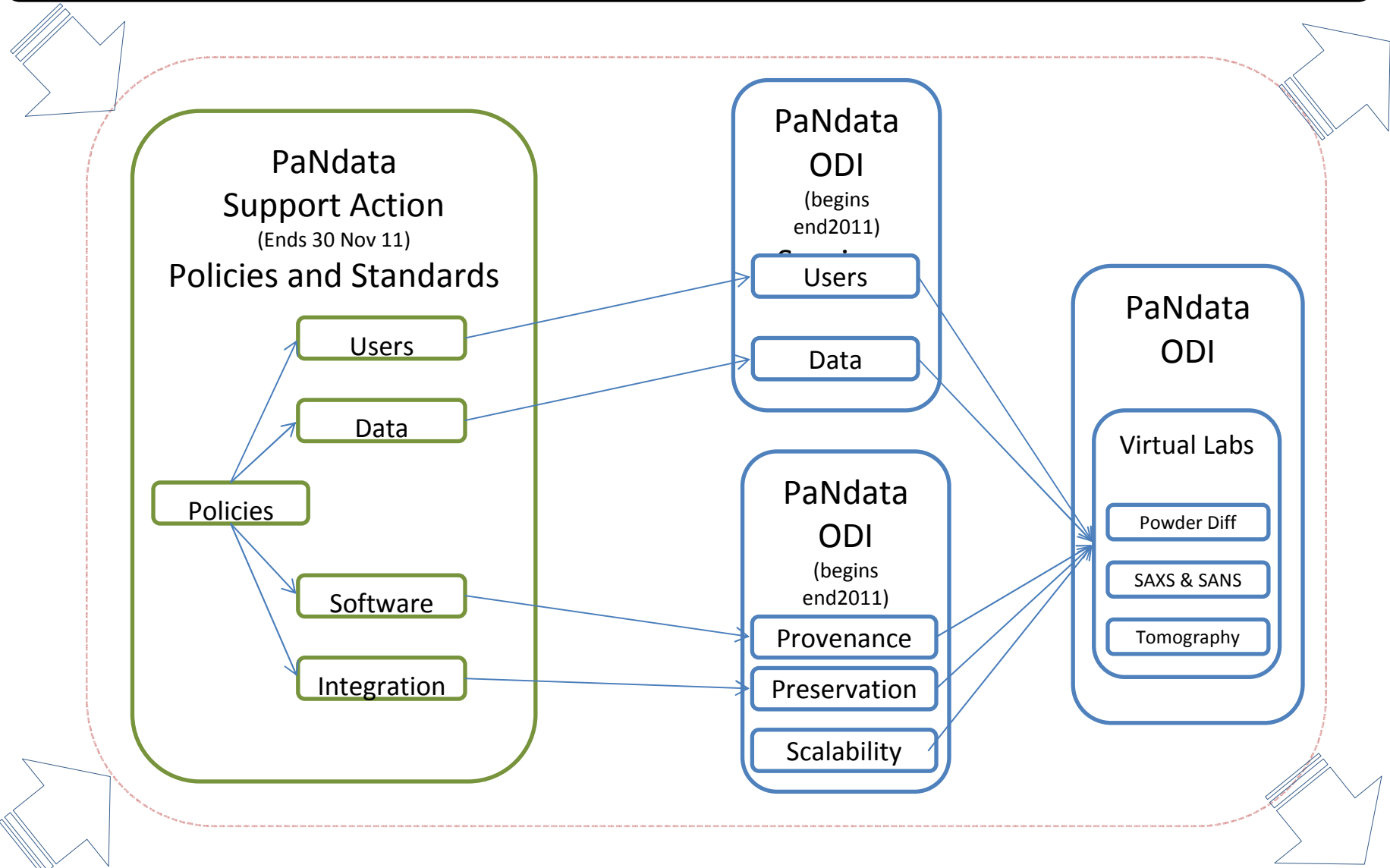
The Vision

The PaNdata Europe Project

The PaNdata Open Data Infrastructure Project

Looking Forwards

ERA Open Access Sharing Initiatives (examples, etc)



ERA Infrastructure Platform Initiatives (EGI, etc)

Objectives

Objective 2 – Users

To deploy, operate and evaluate a system for pan-European **user identification** across the participating facilities and implement common processes for the joint maintenance of that system.

Objective 3 – Data

To deploy, operate and evaluate a generic **catalogue of scientific data** across the participating facilities and promote its integration with other catalogues beyond the project.

Objective 4 – Provenance

To research and develop a conceptual framework, defined as a **metadata model, which can record the analysis process**, and to provide a software infrastructure which implements that model to **record analysis steps** hence enabling the **tracing of the derivation of analysed data outputs**.

Objective 5 – Preservation

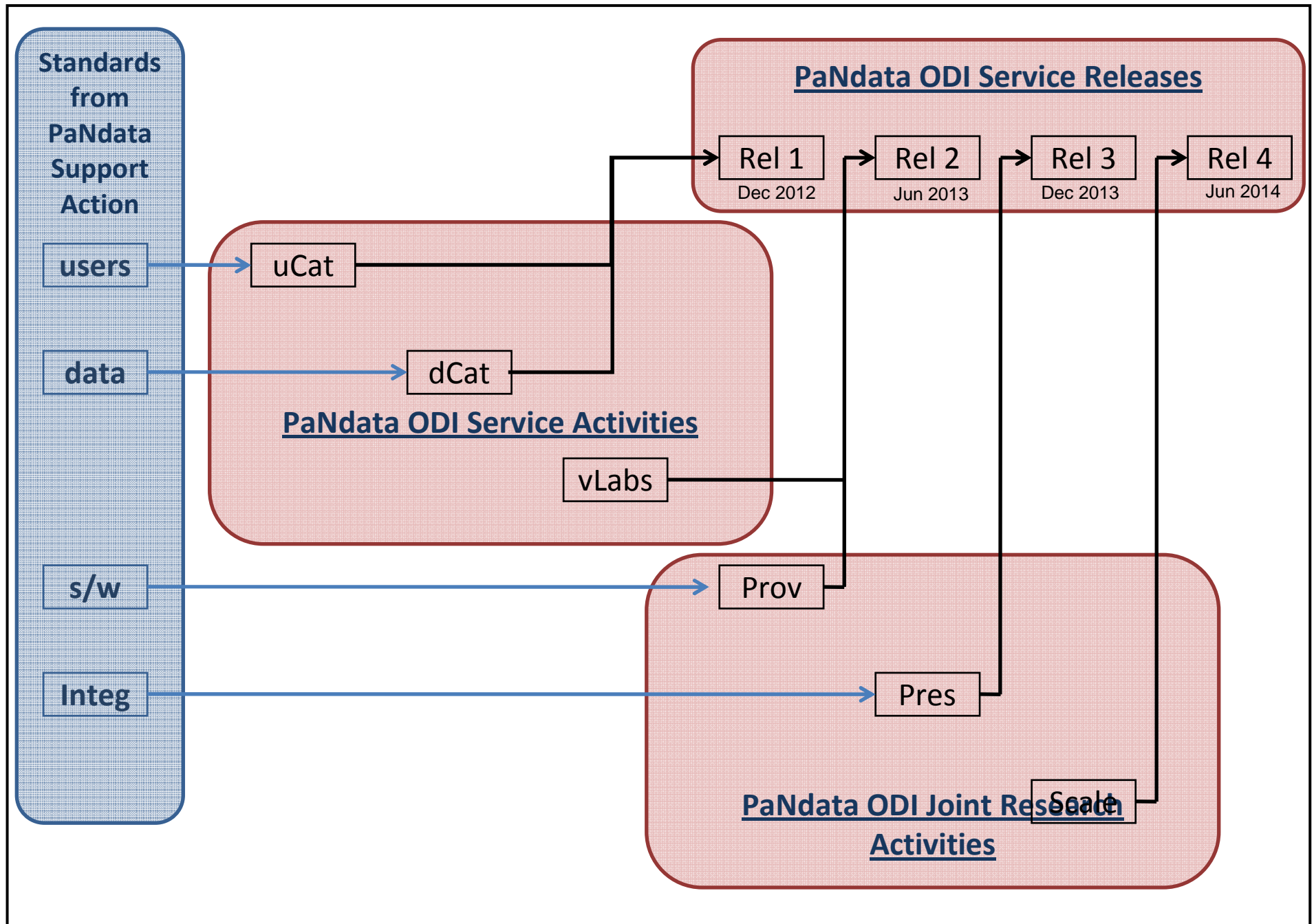
To add to the PaNdata infrastructure extra capabilities oriented towards **long-term preservation** and to integrate these within selected virtual laboratories of the project to demonstrate benefits. These capabilities should, as for the developments in the provenance JRA, be integrated into the normal scientific lifecycle as far as possible. The conceptual foundations will be the **OAIS** standard and the **NeXus** file format.

Objective 6 – Scalability

To develop a **scalable data processing** framework, combining **parallel filesystems** with a parallelized standard data formats (pNexus pHDF5) to permit applications to make most efficient use of dedicated multi-core environments and to permit simultaneous ingest of data from various sources, while maintaining the possibility for real-time data processing.

Objective 7 – Demonstration

To deploy and operate the services and technology developed in the project in **virtual laboratories** for three specific techniques providing a set of integrated end-to-end data services.



Overview

The PaNdata Collaboration

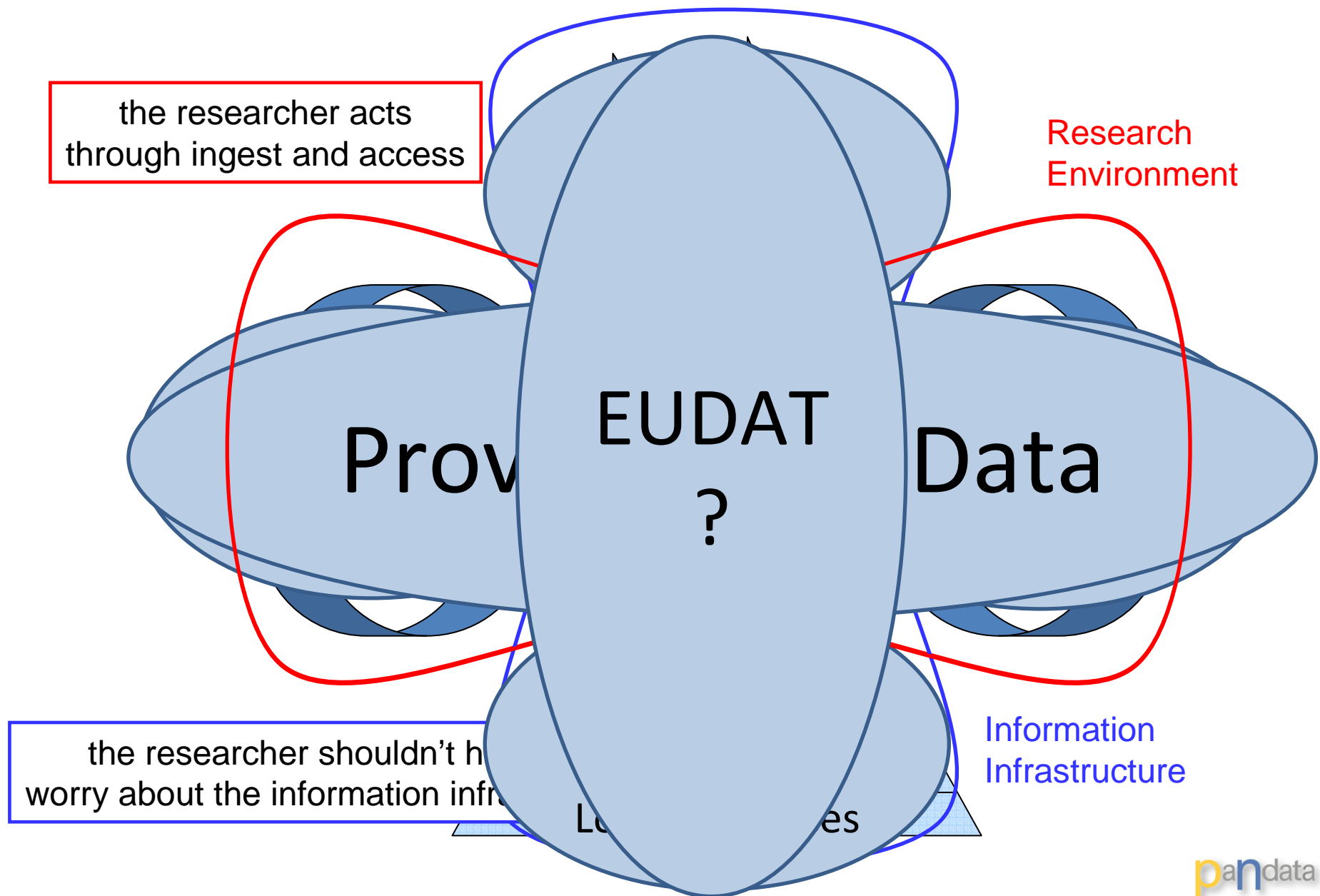
The Vision

The PaNdata Europe Project

The PaNdata Open Data Infrastructure Project

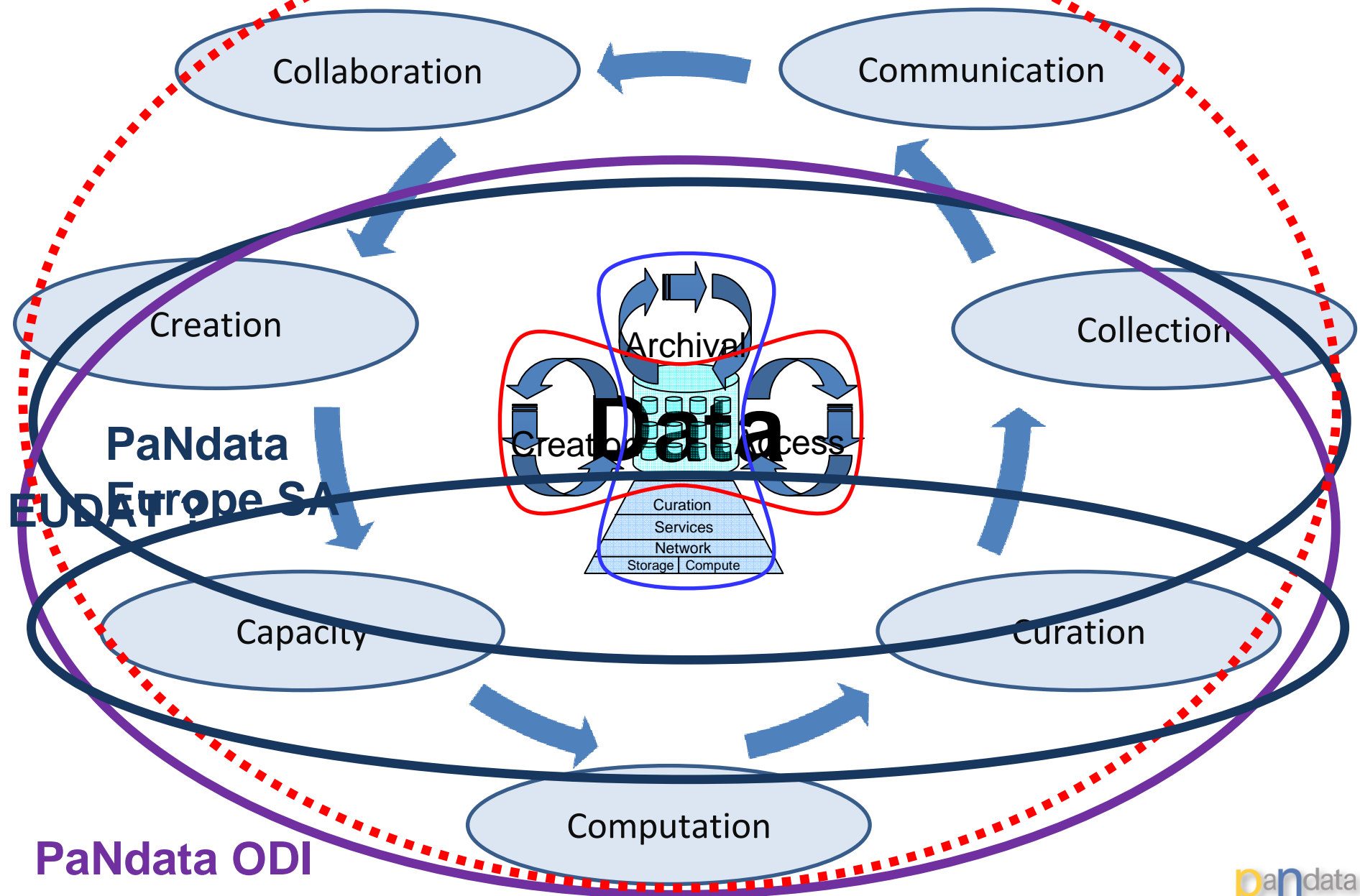
Looking Forwards

The Research Lifecycle



PaNdata VRE

The 7 C's



Overview

The PaNdata Collaboration

The Vision

The PaNdata Europe Project

The PaNdata Open Data Infrastructure Project

Looking Forwards

Thank You



www.pan-data.eu