

PaNdata ODI

*Photon and Neutron
Open Data Infrastructure
ERF meeting
“Big Data and Open Data”
6-7 May 2014
Juan Bicarregui*

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 a 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 the French 2.75 GeV synchrotron

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

ALBA the Spanish 3 GeV synchrotron

Max IV Synchrotron

PaNdata is coordinated by the Scientific Computing Department at Rutherford Appleton Laboratory.

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:

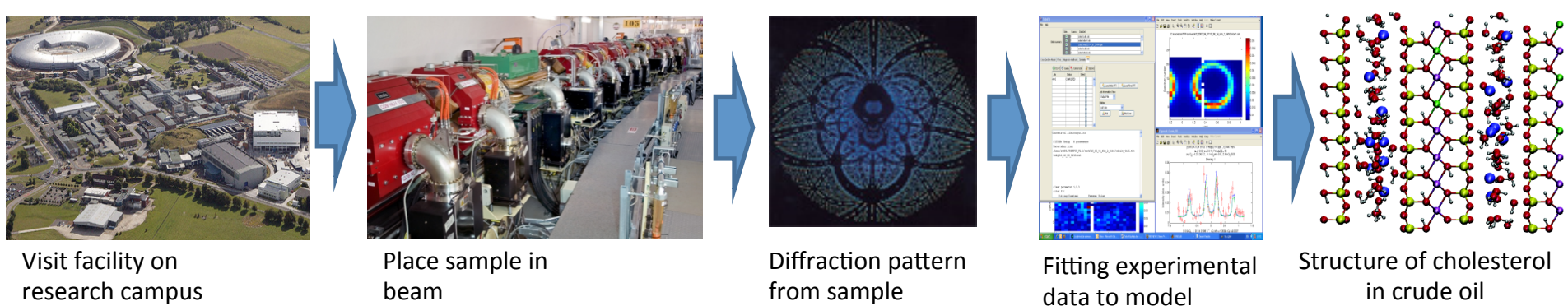
- Physics, Chemistry, Biology, Material sciences, Energy technology, Environmental science, Medical technology and Cultural heritage

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

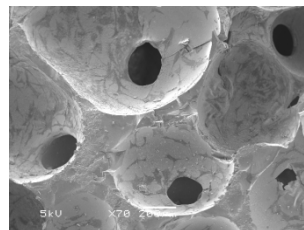
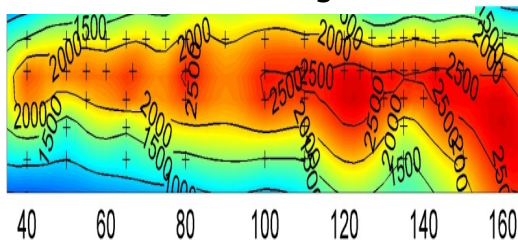
Industrial applications include pharmaceuticals, petrochemicals and microelectronics

The Science we do - Structure of materials



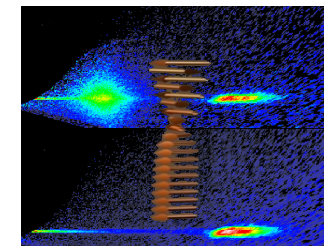
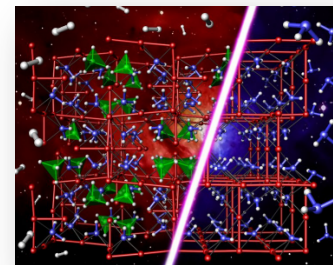
- Over 30,000 user visitors each year
- by small teams - 60,000 short visits
- Most not computing experts
 - Chemistry, biology, medicine, physics
 - energy, environmental, materials, culture
 - pharmaceuticals, petrochemicals, microelectronics
- Over 5,000 high impact publications per year
- In last decade, 4 Nobels and 7 Laureates
 - 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



PaNdata Facilities



Together represent a capital investment of over 3 Billion €

www.pan-data.eu

The PaNdata Collaboration

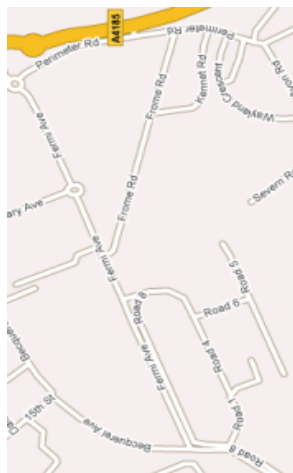
The Vision

The PaNdata Europe Project

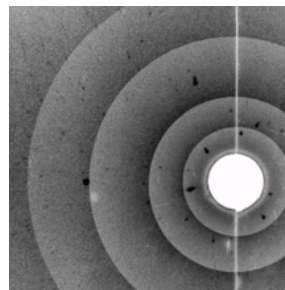
The PaNdata Open Data Infrastructure Project

Looking Forwards

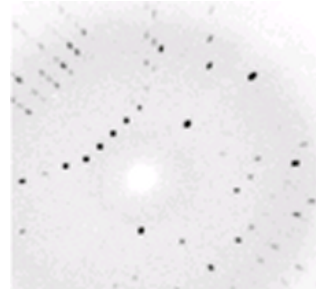
Science driver – Data Integration



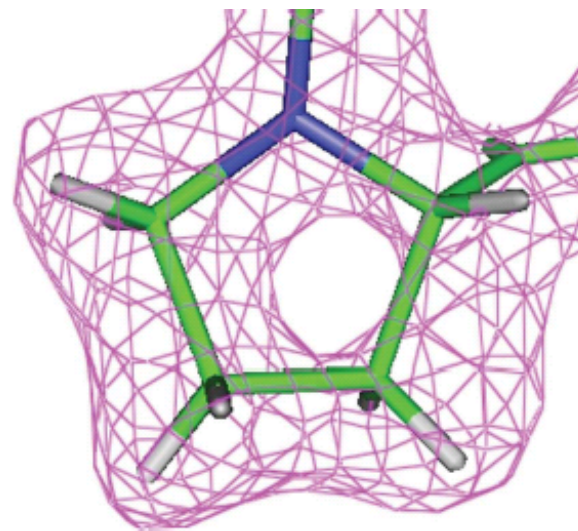
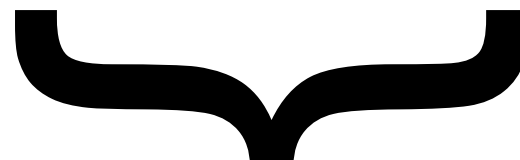
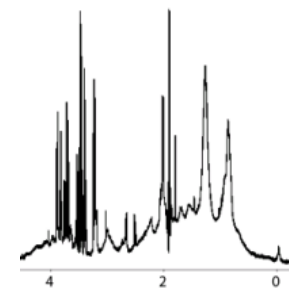
Neutron diffraction



X-ray diffraction



NMR

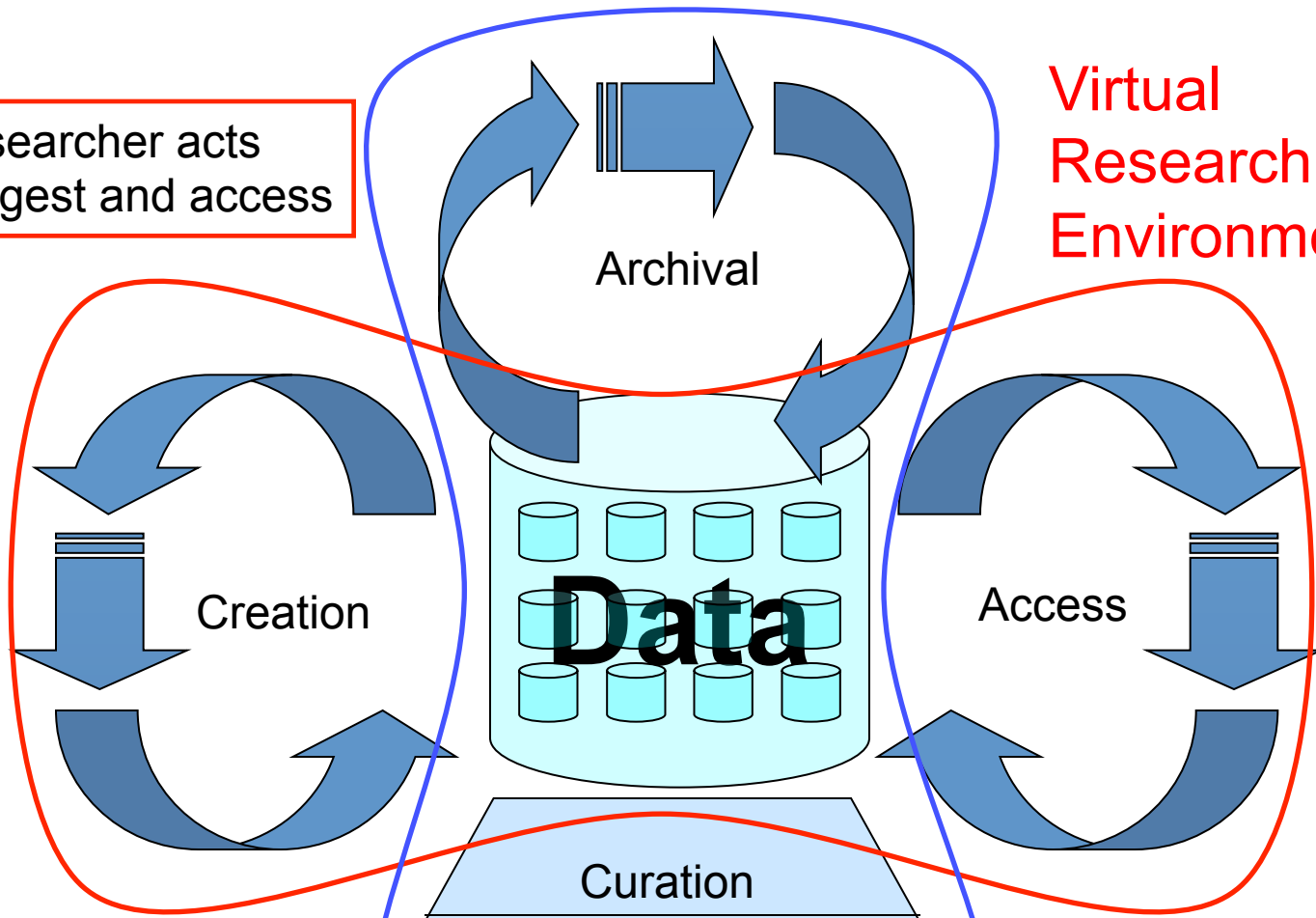


**High-quality
structure
refinement**

What is e-Infrastructure?

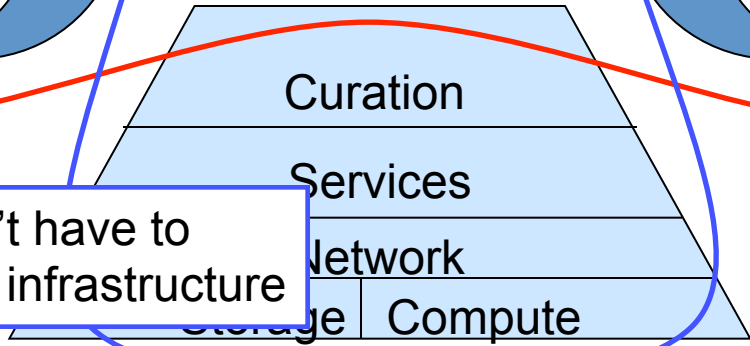
the researcher acts through ingest and access

Virtual Research Environment



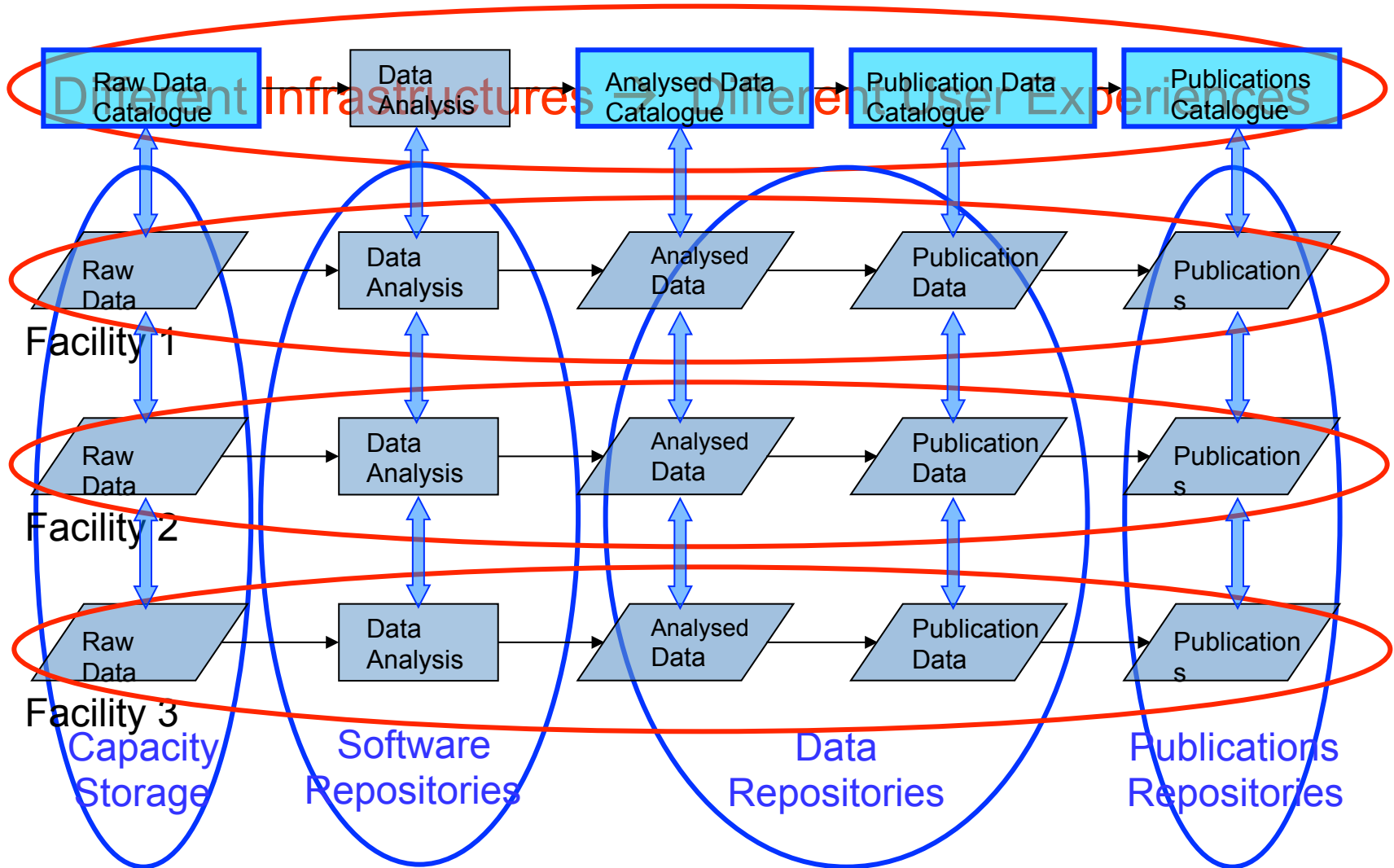
the researcher shouldn't have to worry about the information infrastructure

Information Infrastructure

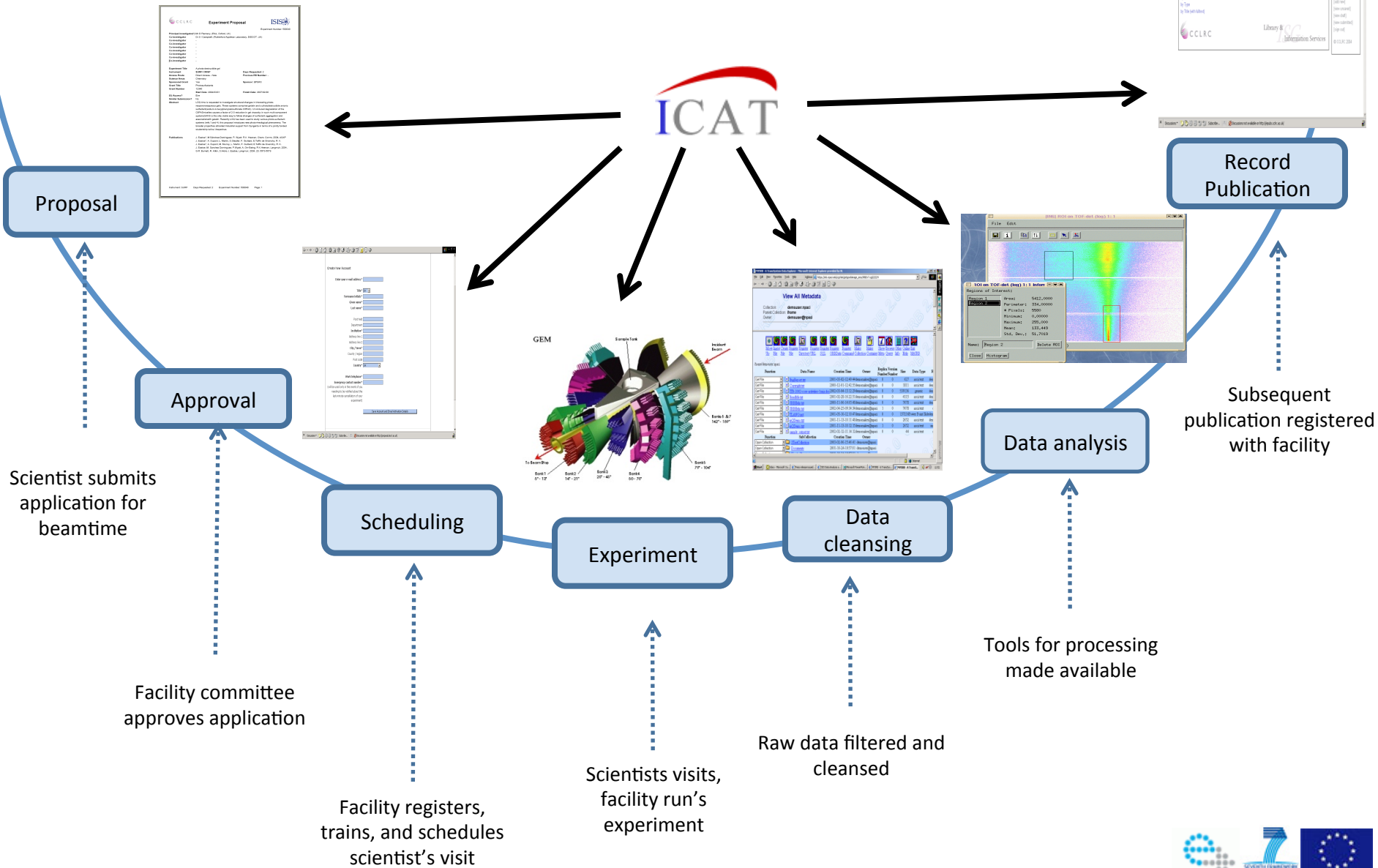


PaNdata Vision

Single Infrastructure → Single User Experience



Metadata



The PaNdata Collaboration

The Vision

The PaNdata Europe Project

The PaNdata Open Data Infrastructure Project

Looking Forwards

PaN-data Standardisation

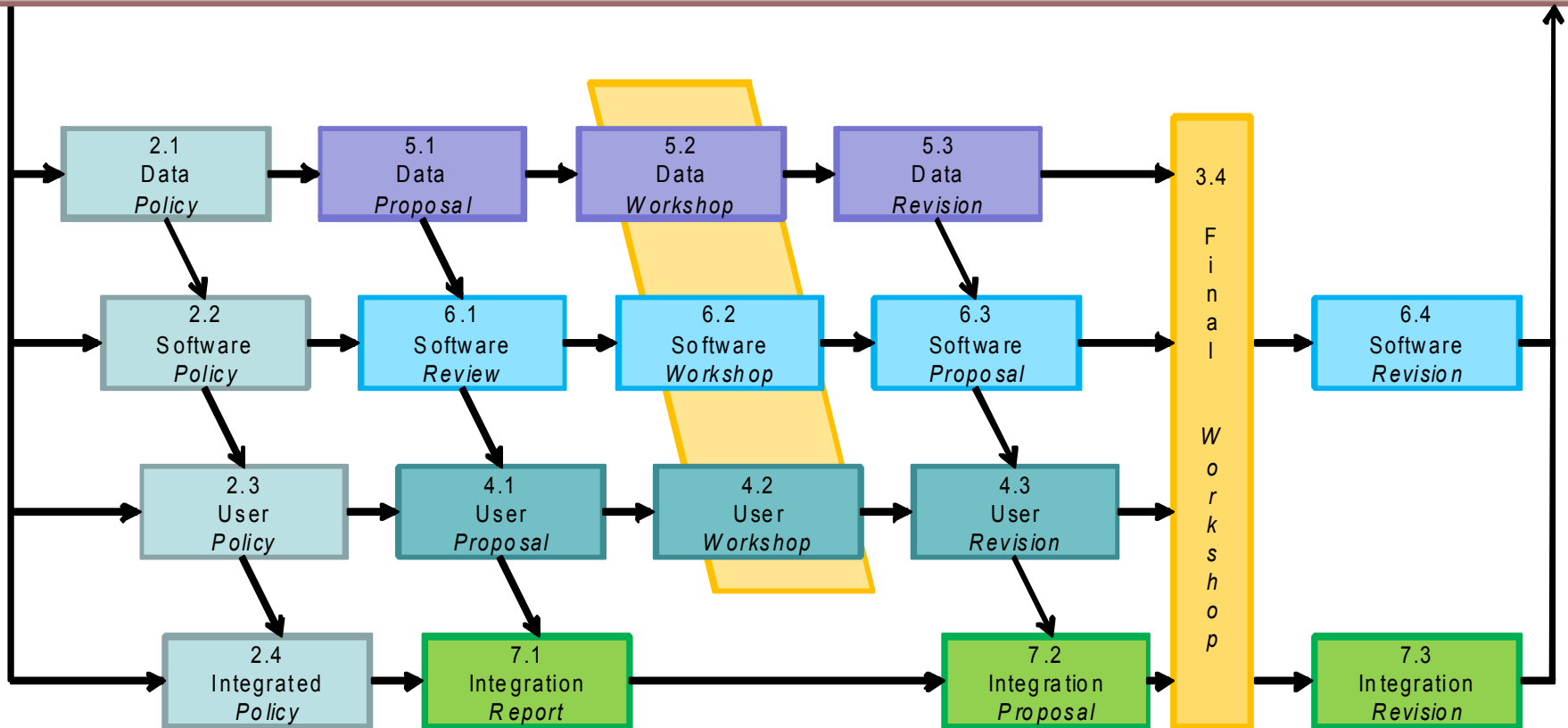
PaN-data Europe is undertaking 5 standardisation activities:

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



Dependencies

Project Management, Knowledge Exchange and Dissemination Activities



Dependencies between the major project tasks

The PaNdata Collaboration

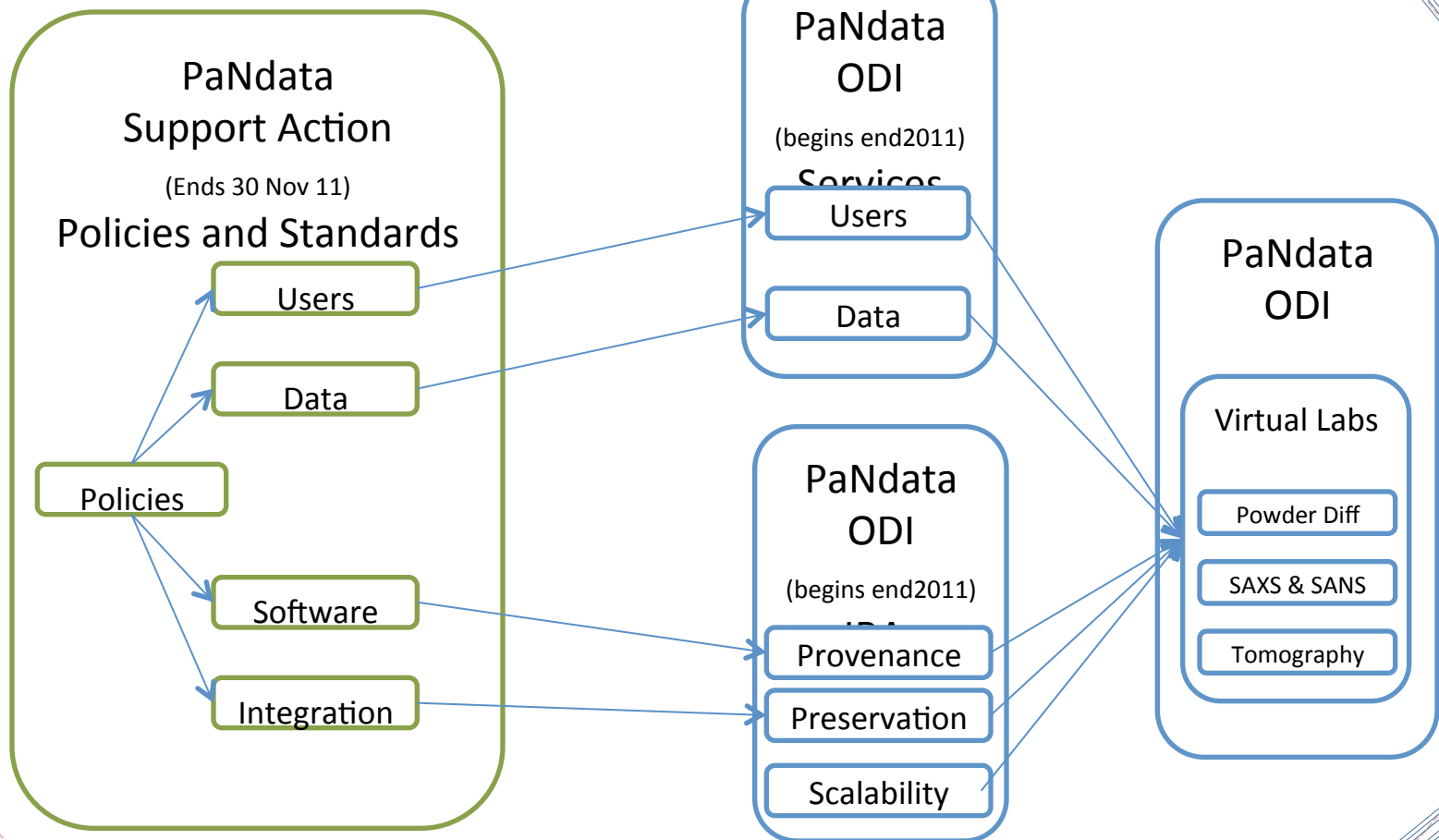
The Vision

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)



Objective – Users (WP3)

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 – Data (WP4)

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 – Demonstration (WP5)

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.

Objective – Provenance (WP6)

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 – Preservation (WP7)

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 – Scalability (WP8)

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.

Engagement (WP2)

Engagement with **other initiatives** and dissemination of project results, in particular to other research infrastructures.

Standards from PaNdata Support Action

users

data

s/w

Integ

uCat

dCat

PaNdata ODI Service Activities

vLabs

PaNdata ODI Service Releases

Rel 1

Rel 2

Rel 3

Rel 4

Prov

Pres

PaNdata ODI Joint Research Activities

Scale

The PaNdata Collaboration

The Vision

The PaNdata Europe Project

The PaNdata Open Data Infrastructure Project

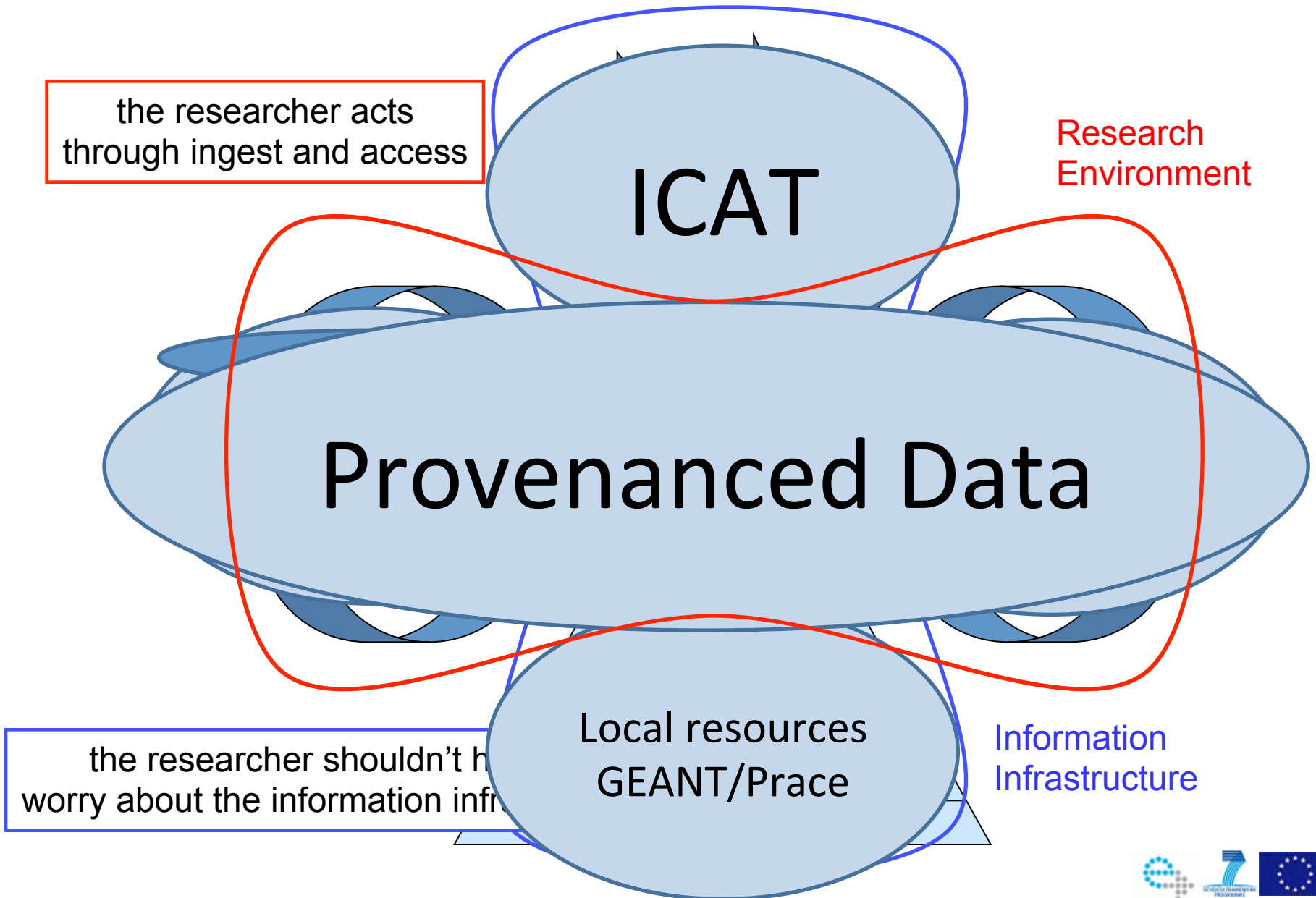
Looking Forwards

Where are we now?

- Service Activities into normal facility operations
 - Umbrella
 - developed primarily by PSI
 - to be deployed by most PaNdata facilities
 - to be operated and supported by ILL
 - ICAT
 - developed primarily by STFC
 - to be deployed at several other facilities
 - supported through an OS project
- Research Activities into future projects
 - Next slide

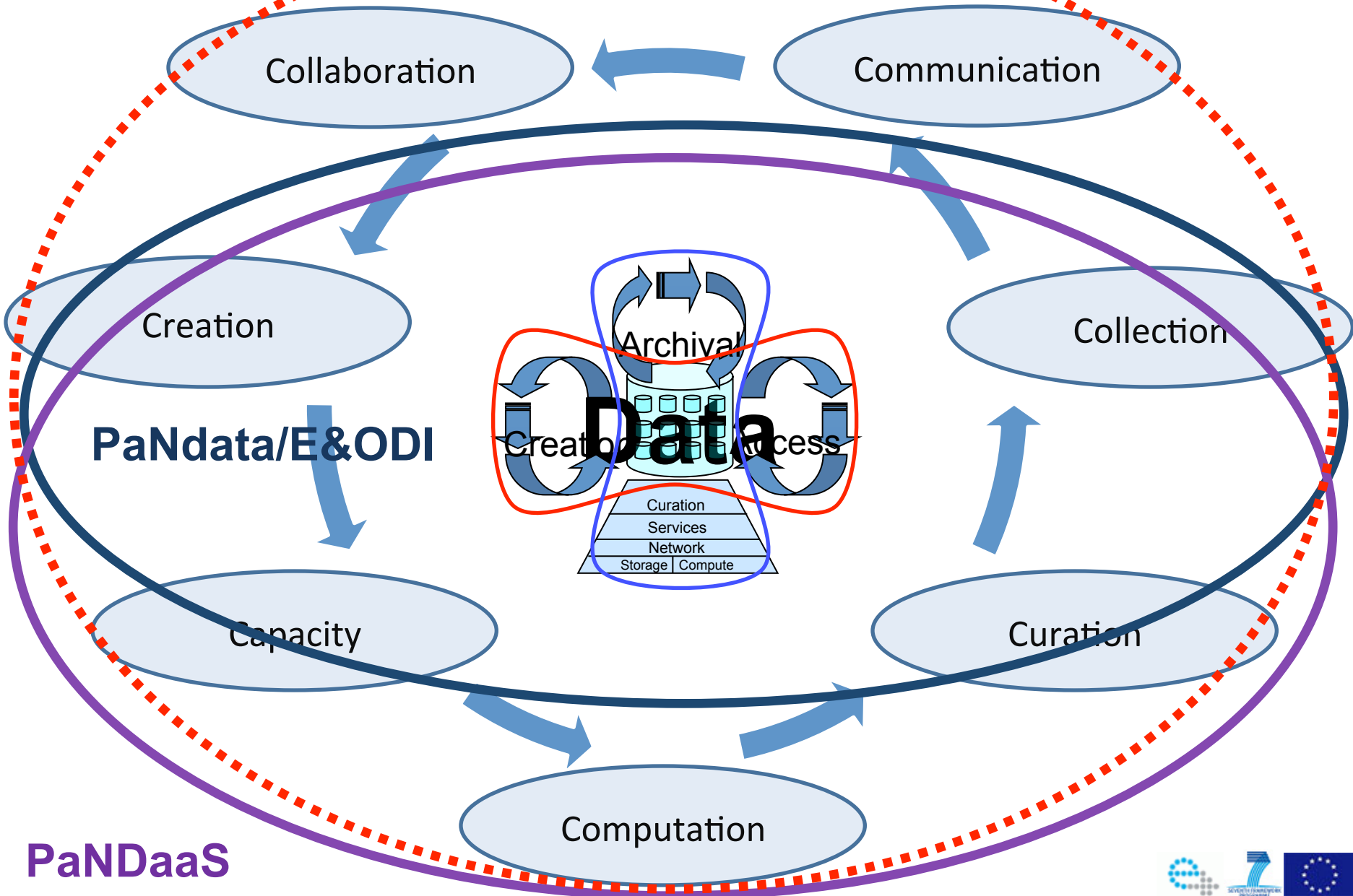
What's next?

- RDA - PaNSIG
 - Photon and Neutron Science Interest Group
- Future Projects
 - PaNDaaS, Photon and Neutron Data as a Service
 - Data as a Service
 - Including data analysis service into the facility provision
 - Led by Rudolf Dimper, ESRF
 - ¿PaN-VRE? completing the life cycle for some techniques



PaNVRE

The 7 C's



The PaNdata Collaboration

The Vision

The PaNdata Europe Project

The PaNdata Open Data Infrastructure Project

Looking Forwards

Thank You



www.pan-data.eu