CINECA

CINECA SCAI SuperComputing Applications and Innovation

SuperComputing Applications and Innovation

Open Science 2014

Giuseppe Fiameni

g.fiameni@cineca.it

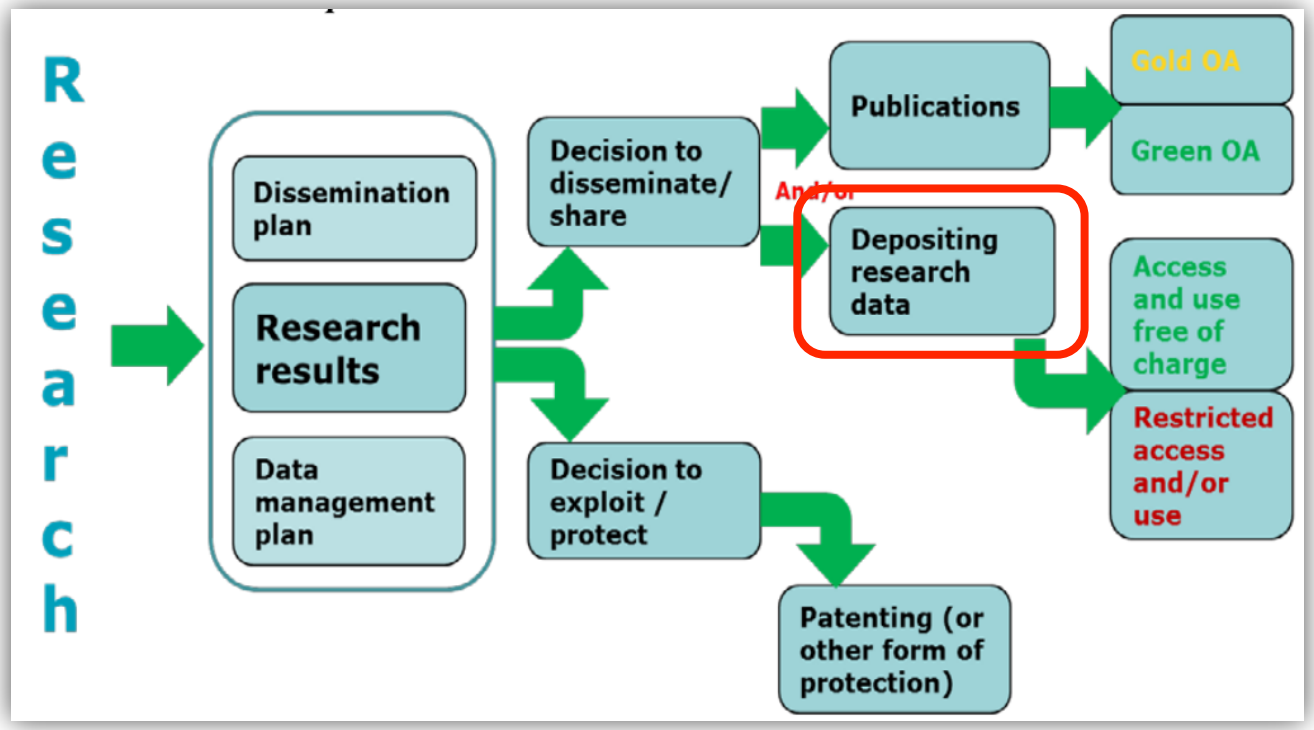
CINECA

Casalecchio di Reno (BO)

Via Magnanelli 6/3, 40033 Casalecchio di Reno | 051 6171411 | www.cineca.it

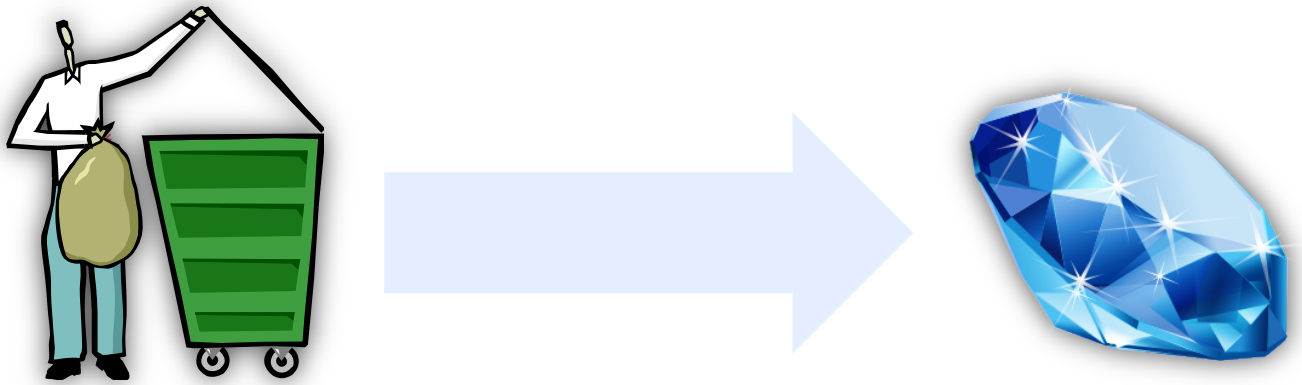
Outline

- Reference data lifecycle model
- G8 + O5 principles
- The EUDAT project
- The CINECA perspective

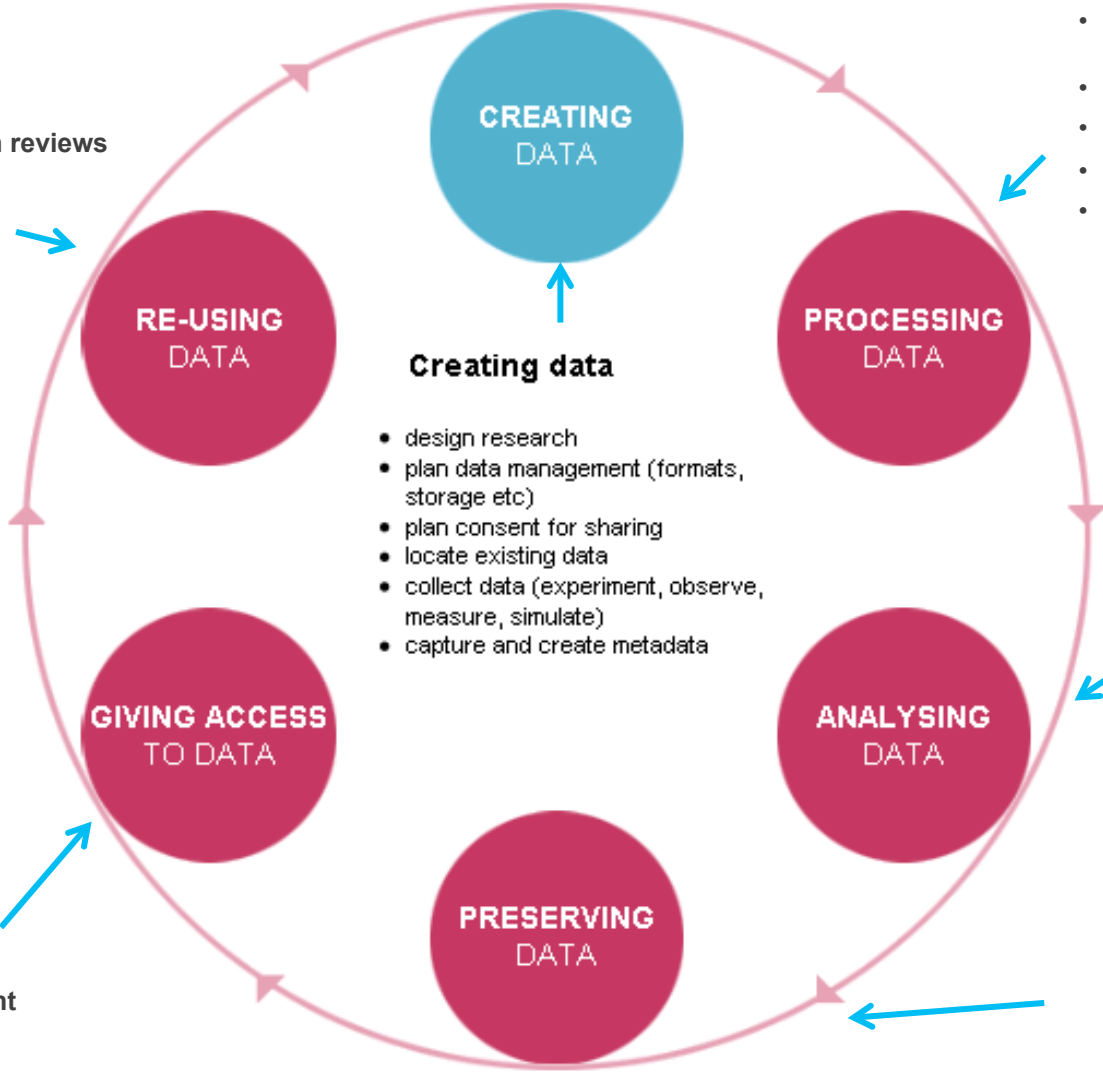

Research
Dissemination plan
Research results
Data management plan
Decision to disseminate/ share
Decision to exploit / protect
Publications
Depositing research data
And/or
Gold OA
Green OA
Access and use free of charge
Restricted access and/or use
Patenting (or other form of protection)

Data as an asset

"One of the most significant changes of the past decade has been the widespread recognition of data as an asset rather than the refuse of research."



Research model data lifecycle



- Follow-up research
- New research
- Undertake research reviews
- Scrutinize findings
- Teach and learn

- Enter data, digitize, transcribe, translate
- Check, validate, clean data
- Anonymise data where necessary
- Describe data
- Manage and store data

- Interpret data
- Derive data
- Produce research outputs
- Author publications
- Prepare data for preservation

- Migrate data to best format
- Migrate data to suitable medium
- Backup and store data
- Create metadata and documentation
- Archive data

- Distribute data
- Share data
- Control access
- Establish copyright
- Promote data

http://www.data-archive.ac.uk/create-manage/life-cycle

G8 + O5 White Paper Draft v1.0
5 Principles for an Open Data Infrastructure

- **1. Discoverability**
  - Implementation of appropriate **persistent identifier** frameworks, adoption of descriptive metadata standards, and the use of appropriate data formats and taxonomies.
- **2. Accessibility**
  - **Publicly funded scientific research data should be made openly available** with as few restrictions as possible. Ethical, legal or commercial constraints may be imposed on the use of research data to ensure that the research process is not damaged by inappropriate release of data.
- **3. Understandability**
  - Scientific data sets must be understandable in order to be effectively used. A set of numbers, texts, pictures or even videos alone cannot be understandable without **additional context, semantics, data analysis tools, and algorithms**.
- **4. Manageability**
  - Data management policies and plans must make it clear **who is responsible for maintaining the availability of data and how the associated costs** are to be met including issues associated with curation, storage and services.
- **5. People**
  - A global approach to research data infrastructure requires a **highly skilled and adaptable workforce** and culture that is able to capture the available data and make it available to those that are able to use it appropriately.

G8+O5 Data working group and GSO Report Recommendations 9 & 10

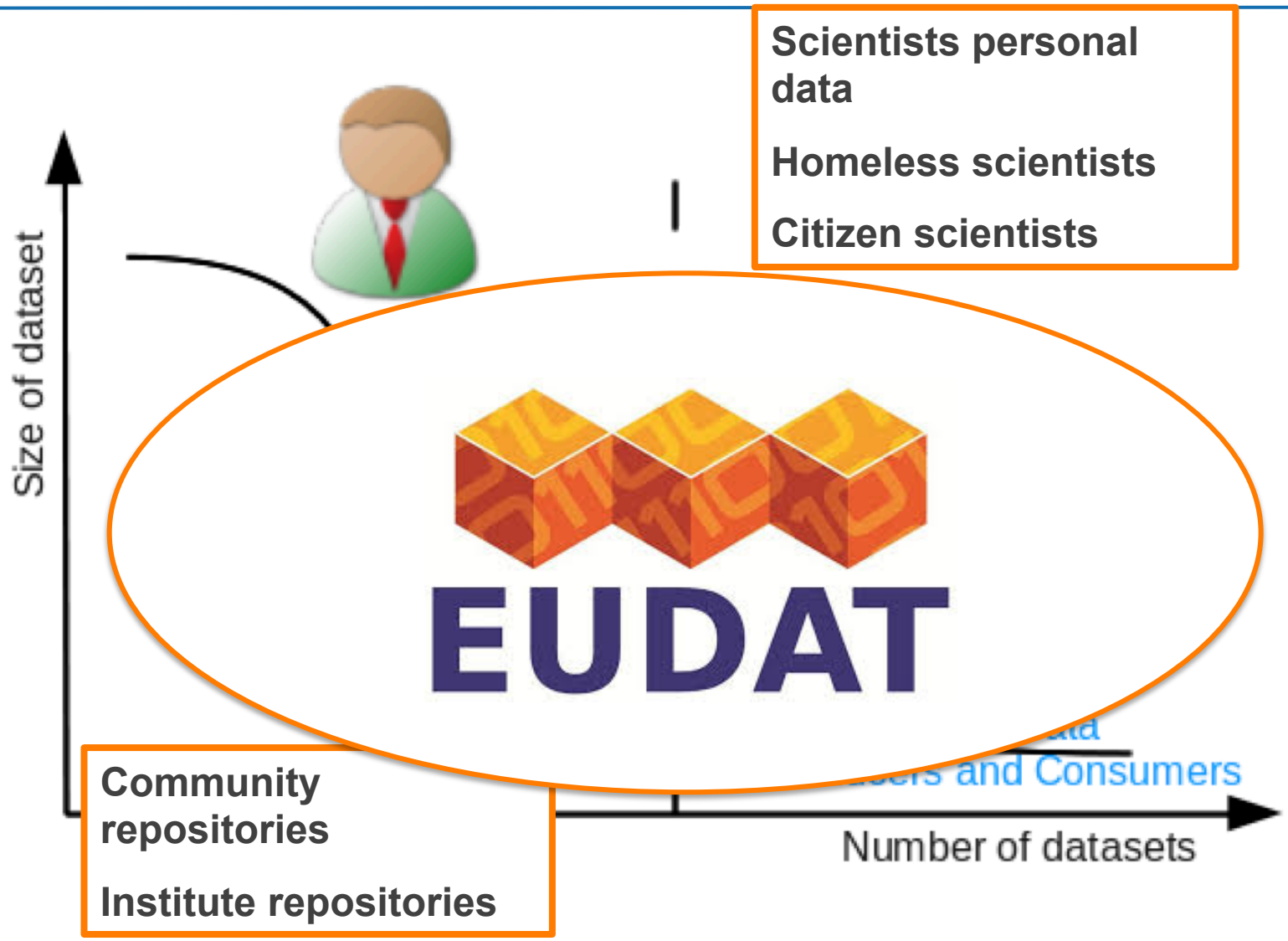
Recommendation 9 - e-infrastructure

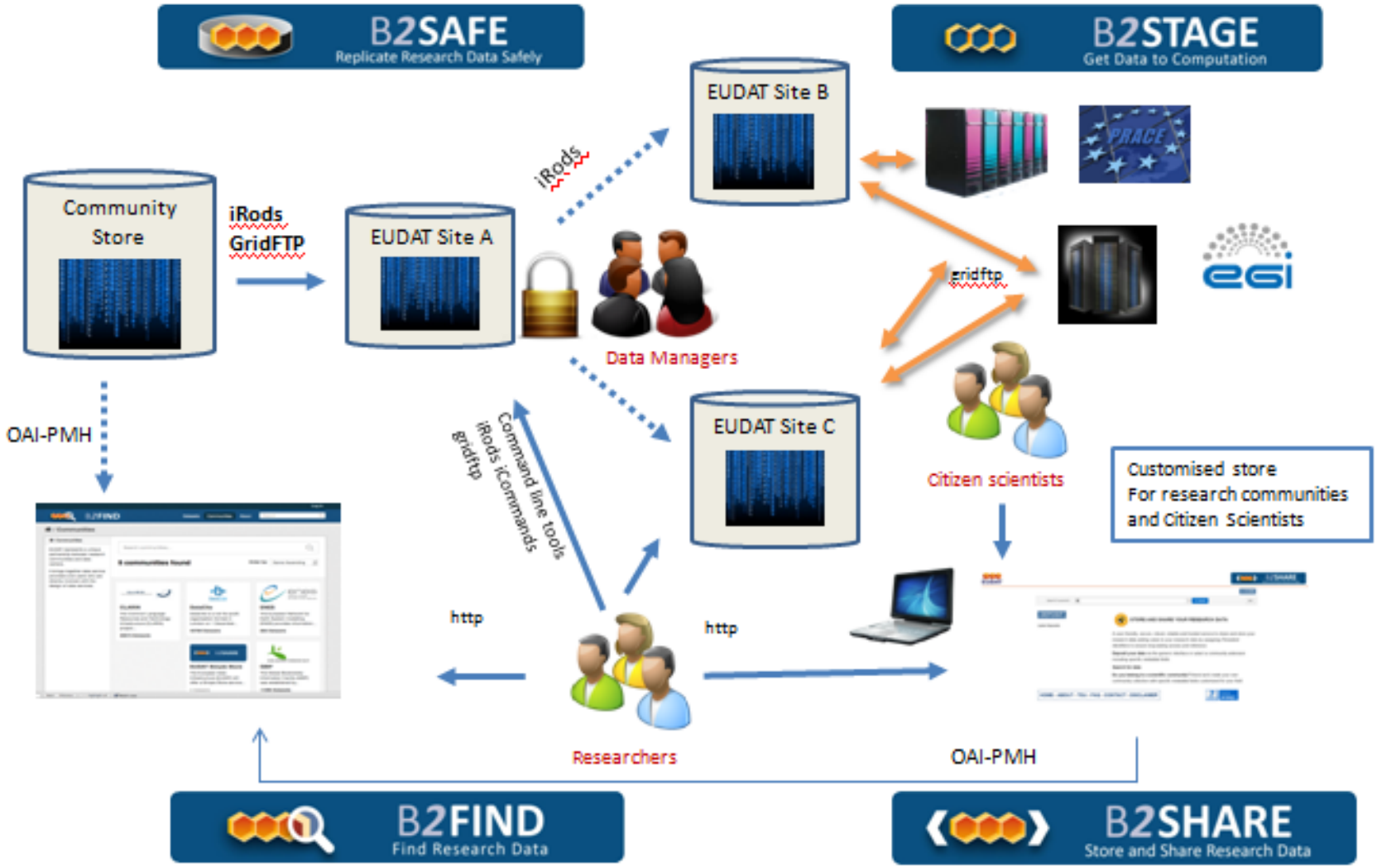
Global research infrastructure initiatives should recognize the utility of the **integrated use of advanced e-infrastructures**, services for accessing and processing, and curating data, as well as remote participation (interaction) and access to scientific experiments.

Recommendation 10 - Data exchange

Global scientific data infrastructure providers and users should recognise the utility of **data exchange and interoperability** of data **across disciplines and national boundaries** as a means to broadening the scientific reach of individual data sets.

Data Domain

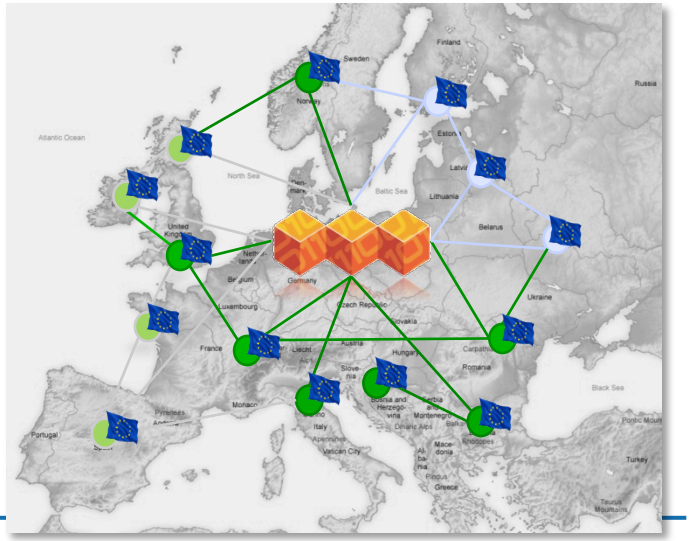

Scientists personal data
Homeless scientists
Citizen scientists
Size of dataset
EUDAT
and Consumers
Number of datasets
Community repositories
Institute repositories


B2SAFE
Replicate Research Data Safely
B2STAGE
Get Data to Computation
EUDAT Site B
Community Store
iRods
GridFTP
EUDAT Site A
iRods
gridftp
Data Managers
OAI-PMH
EUDAT Site C
Command line tools
iRods iCommands
gridftp
Citizen scientists
Customised store
For research communities
and Citizen Scientists
http
http
Researchers
OAI-PMH
B2FIND
Find Research Data
B2SHARE
Store and Share Research Data

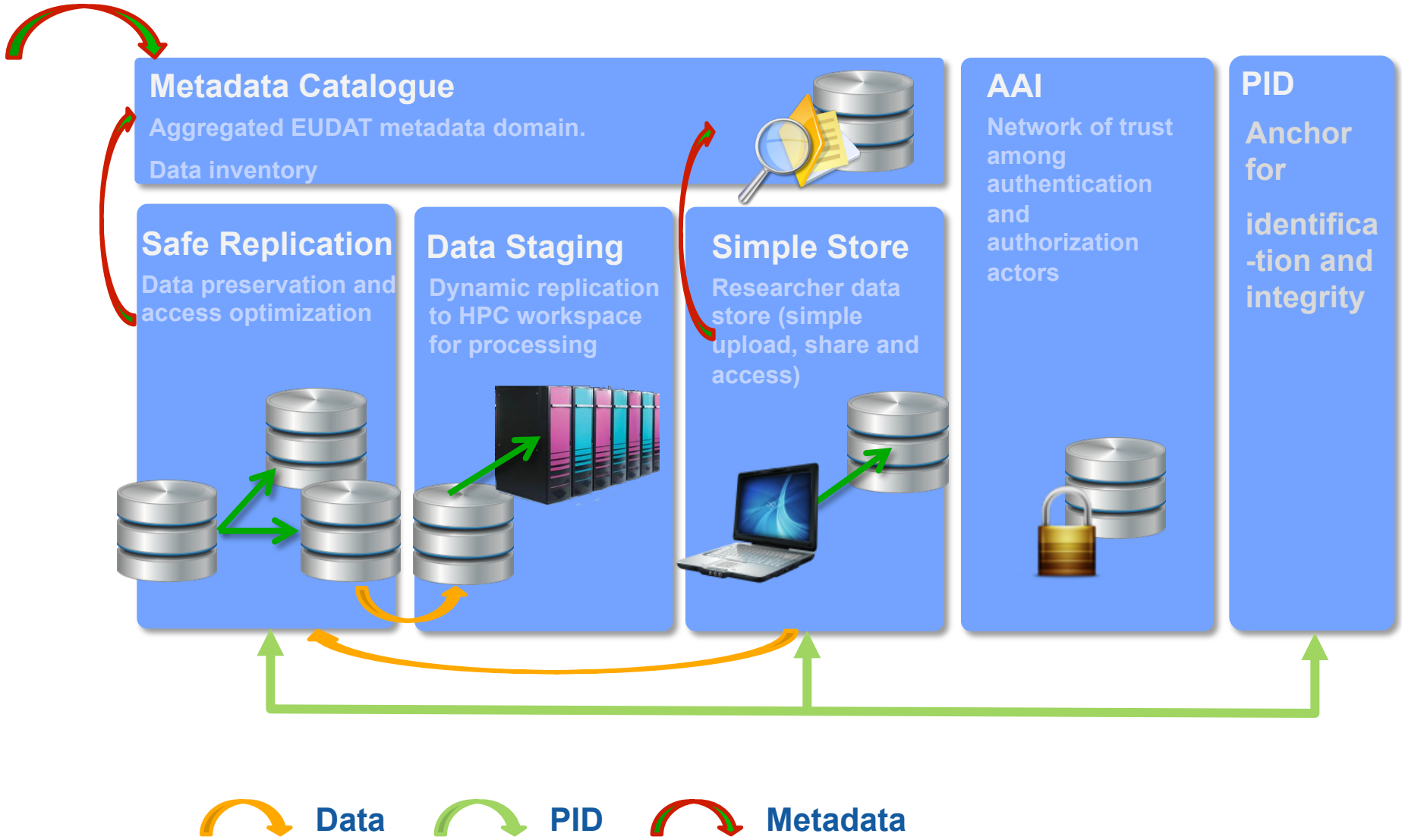
EUDAT is …

- a pan-European initiative building a sustainable cross-disciplinary and cross-national data infrastructure providing a set of shared services for accessing and preserving research data

- supporting multiple research communities working closely with them to deliver these technical services as part of the EUDAT Collaborative Data Infrastructure (CDI)



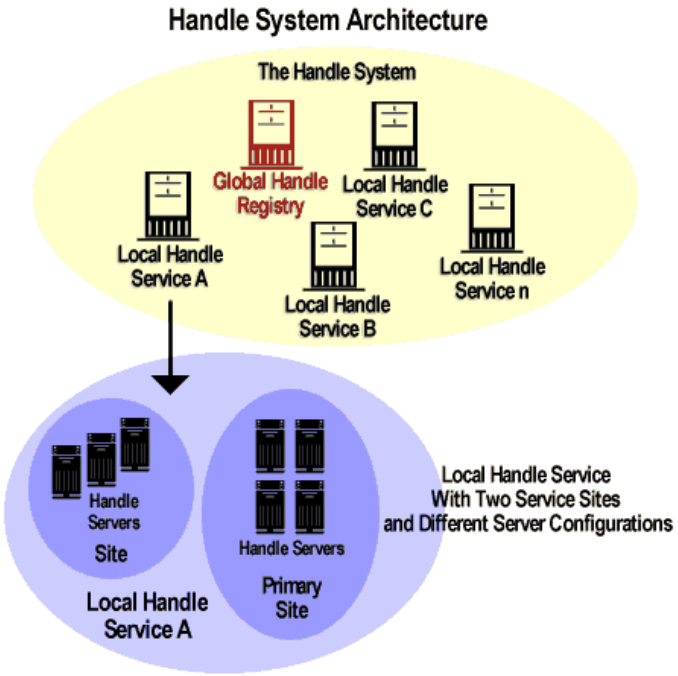
Services

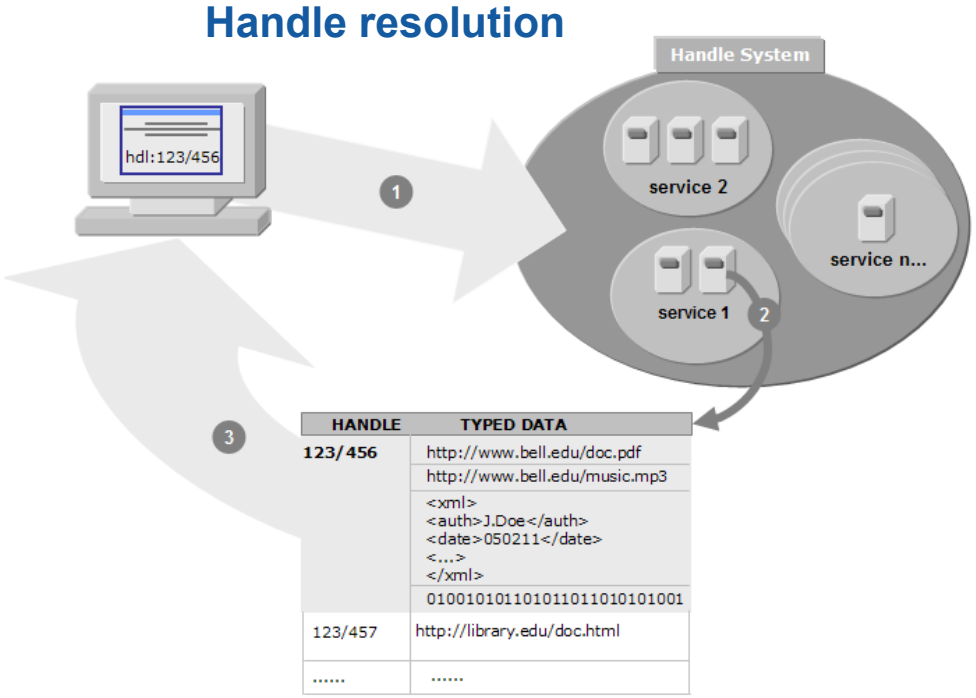

Metadata Catalogue
Aggregated EUDAT metadata domain.
Data inventory
Safe Replication
Data preservation and access optimization
Data Staging
Dynamic replication to HPC workspace for processing
Simple Store
Researcher data store (simple upload, share and access)
AAI
Network of trust among authentication and authorization actors
PID
Anchor for identifica-tion and integrity
Data
PID
Metadata

Persistent Identifiers (PID)

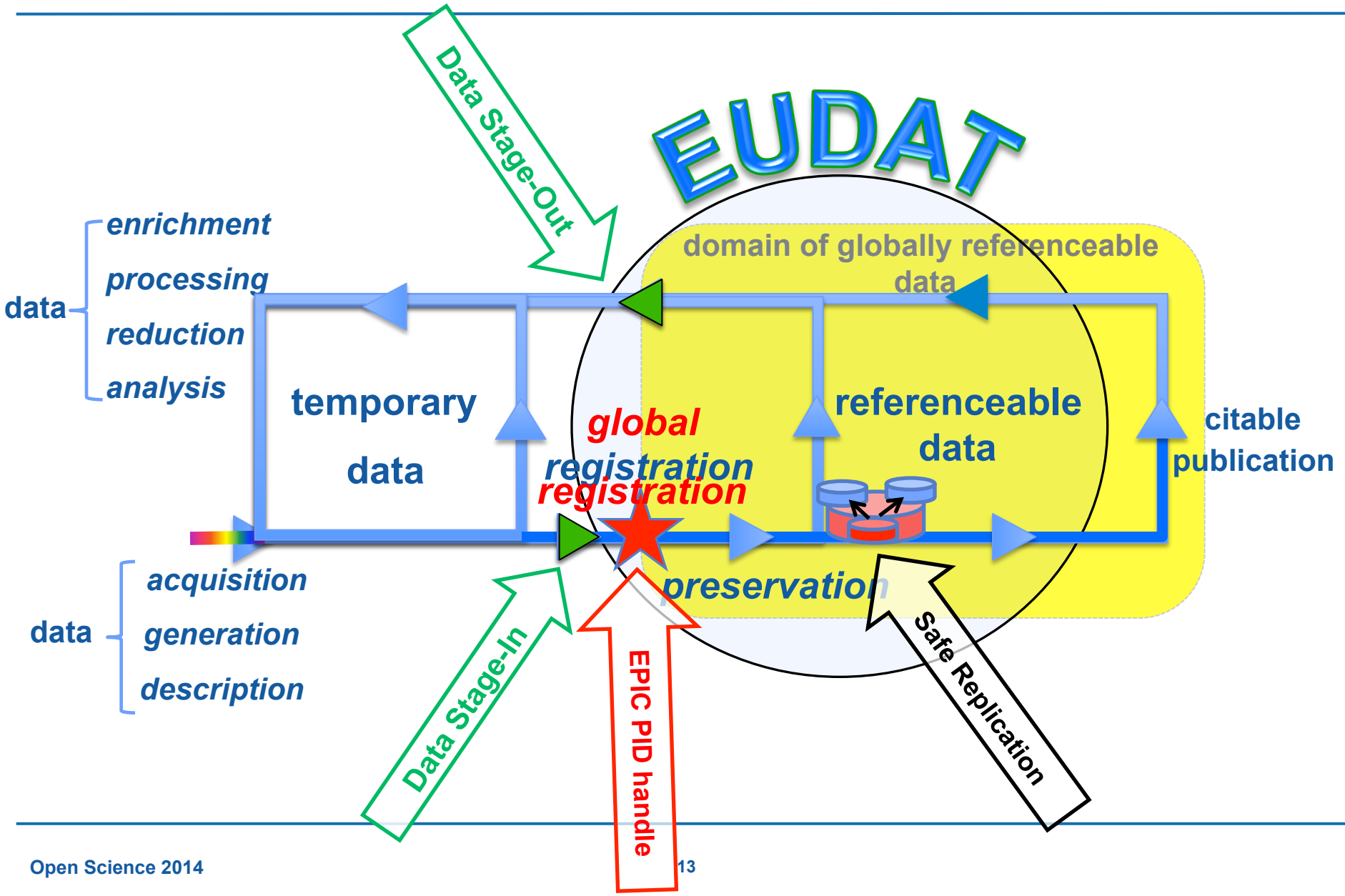
- EUDAT relies on the **EPIC service** to associate persistent identifier to digital objects

- EPIC is an identifier system using the **Handle infrastructure**. Its focus is the registration of data in an early state of the scientific process, where lots of data is generated and has to become referable to collaborate with other scientific groups or communities, but it is still unclear, which small part of the data should be available for a long time period.





EUDAT domain



EUDAT-EPOS success story

- The whole archive of the INGV's data center in Rome is replicated to CINECA, the EUDAT's data center in Bologna (years:1990-2013 ~25 TB)
- Granularity is at single file level (the collection of sesimic waveforms related to 1 sensor for 1 day)
- Metadata are replicated too, as files (dataless).
- The replicas are stored, synchronized and accessed by a single user, the INGV community manager.

Initiatives (HPC – HPDA)

HPC

- Open Access
- Peer Review
- ISCRA, PRACE, LISA

HPDA

2014

Repository

- Access
- Reference
- Curation
- Replica
- Preservation

2013

2014

Many thanks for your attention!

CNR - PISA

Open Science 2020: Harmonizing current practices with Horizon 2020 guidelines

How to support universities and research centres to implement Horizon 2020 OA guidelines

Paola Gargiulo, **Information and Knowledge Management Dept, CINECA**
p.gargiulo@cineca.it

Pisa, April 8th 2014

Outline

- OpenAIRE initiative and support infrastructure
- NOAD's task and role of CINECA
- Next steps

OpenAIRE https://www.openaire.eu/it/support/helpdesk

NOAD- National Open Access Desk- Tasks 2009- 2013/14

- Support and Assistance

- help EC-funded project (FP7, ERC)) Italian coordinators to comply with EC OA Pilot Project (2008-2013), point them to appropriate repositories when available or to ZENODO, help them with OA, EC OA Policies, copyright issues etc.
- assist repository managers (literature and data), CRIS managers and journal publishers to make their repositories and journals compliant with the OpenAIRE Guidelines
- answer questions/tickets received via OpenAIRE helpdesk
- provide assistance with every author claims/deposits his/her publication in OpenAIRE

- Outreach and Dissemination

- work for national funding project information to be added to the OpenAIRE portal (namespace and grant numbers) to increase visibility of research outputs funded nationally

  close communication with the EC National Reference Point to keep track of policy issues (national policies harmonized with EC) Infrastructural developments (promoting OpenAIRE's guidelines and services)
- identify events to present OpenAIRE and make presentations/poster sessions;
- disseminate information about upcoming OpenAIRE workshops in your country and invite relevant people to attend them.

NOAD activities and Horizon 2020

The Eu expects the NOAD to continue support and dissemination activities to comply with Horizon 2020 OA requirements and also add two new areas of action:

RESEARCH DATA

Support projects with data management plans by pointing to appropriate resources

Promote policies at national or funder level for policies and research data repositories.

Promote the EC's open data pilot and the harmonisation of policies.

Contact Open Data Pilot project coordinators and research data repository managers to promote the OpenAIRE Guidelines for Data Managers

Identify any links between the OA publication in OpenAIRE and underlying datasets.

OpenAIRE Gold OA WORK PACKAGE

Promote the Gold OA fund to institutions and to publishers in Italy

Raise awareness of the existence of funds among projects coordinated by Italian researchers

Provide input for negotations with publishers for reduced APC fees

Next steps (1)

- Merge of SURplus and U-Gov solutions (23 open-access repositories and 56 CRIS), 2013
- Roadmap to merge repositories and CRIS (> 80 installations) 2014-2015
- Release of a cloud service Research-Link
  - ✓ check copyright policy and authors' rights
  - ✓ automatic enriching of bibliographic publications metadata
  - ✓ through semantic annotation and automated tagging and research information processing to compare statistical information among universities and expose the academic research information as Linked Data

Next steps (2)

- move forward to reach 100% of OpenAIRE compliant and populated OA repositories integrated with CRIS
  - ✓ Horizon 2020 OA guidelines, the Italian Law on OA
  - ✓ Italian institutional mandates, funders mandates,
- continue to develop added value - services to expose and increase visibility of Italian research output
- work closely with all the stakeholders to
  - ✓ build an effective national OA infrastructure for research results (publications and data)
  - ✓ OA research data: a national and istitutional action required!
  - ✓ work on roadmap to implement the research data infrastructure together with universities and research centres

Thanks !

p.gargiulo@cineca.it