



EUDAT

Common Data Services for Science

Peter Wittenburg, EUDAT Scientific Coordinator
Kimmo Koski, EUDAT Project Coordinator

ERF Meeting, Brussels, 7 May 2014

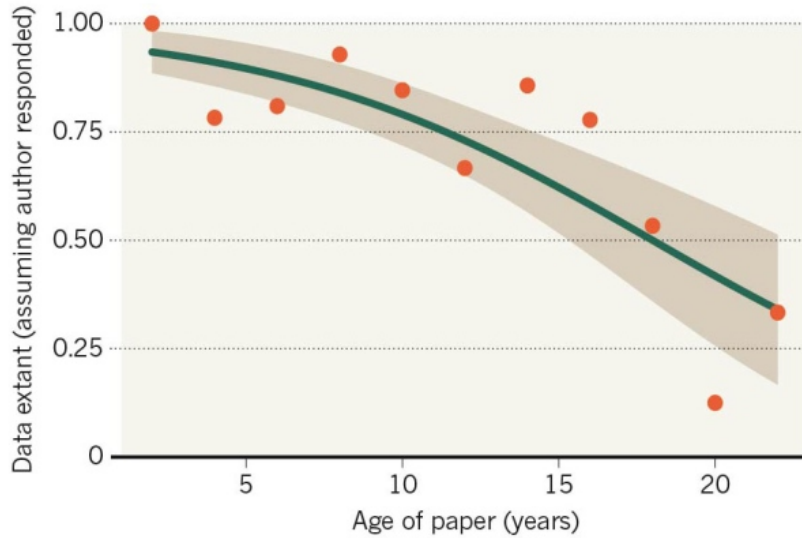


Data Loss is real and significant ...



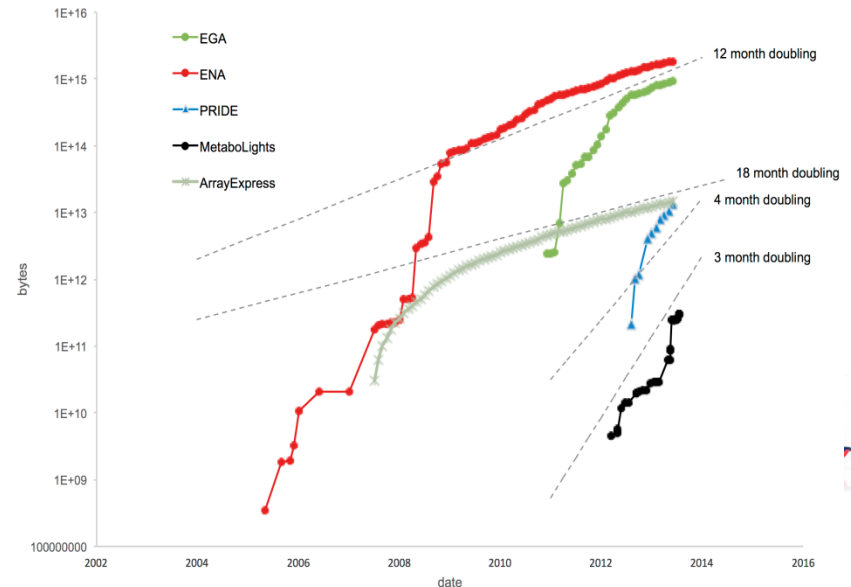
MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.



Nature news, 19 December 2013

...and so is Data growth

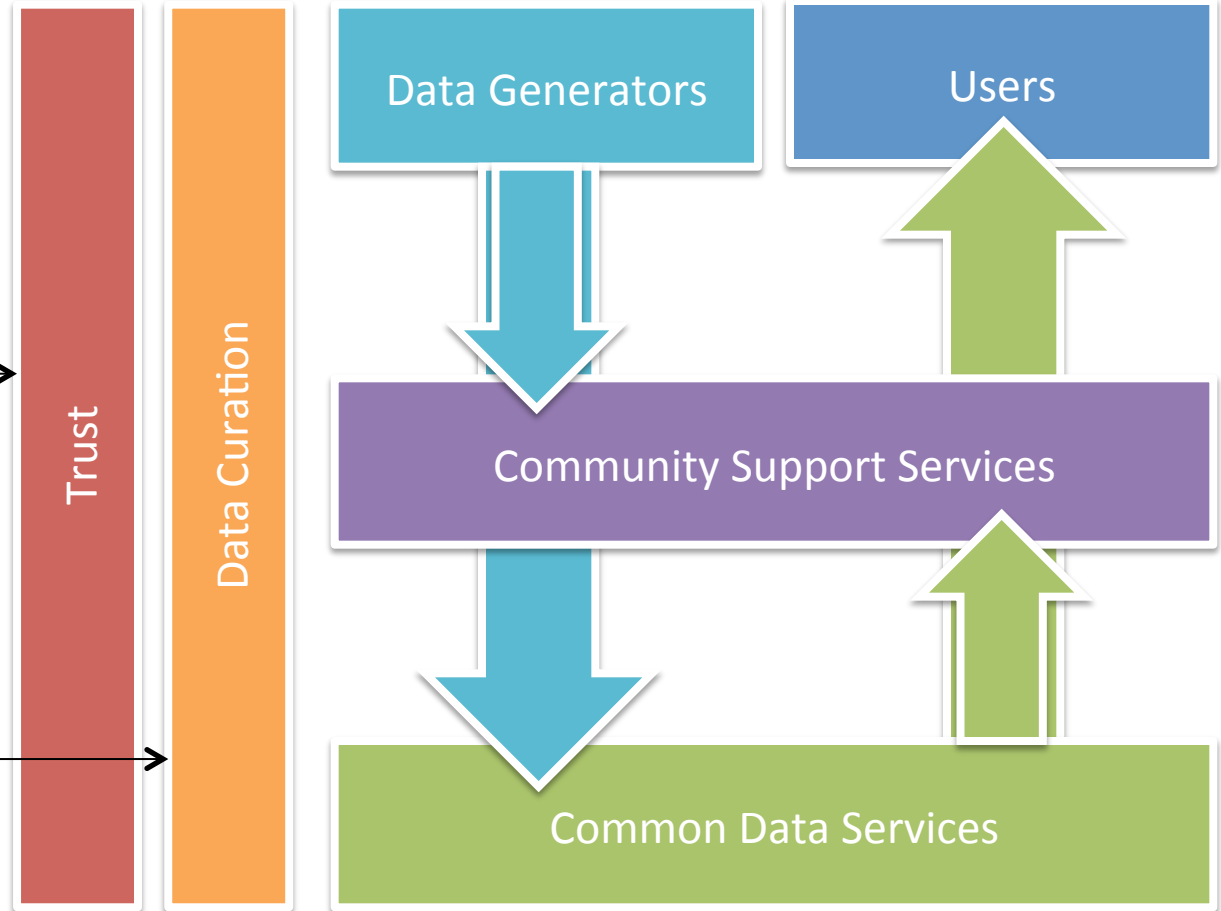




Collaborative Data Infrastructure

organizational and cultural closeness are primary for TRUST

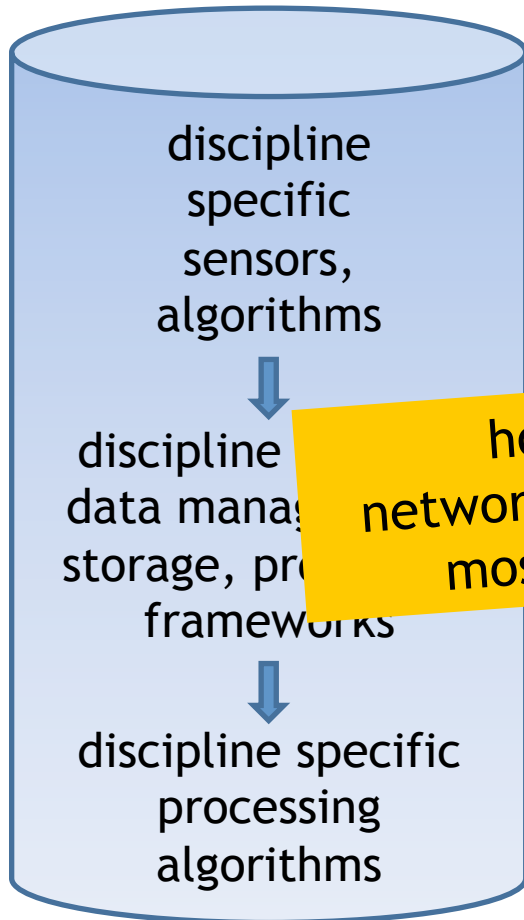
curation is a task of all parties



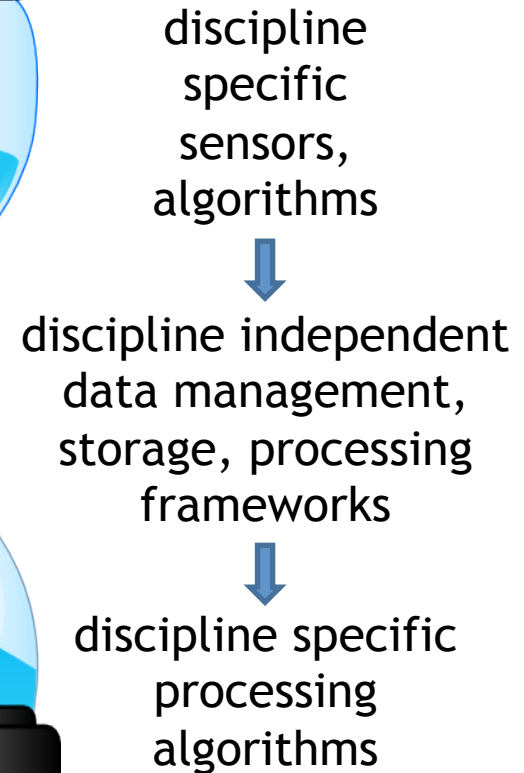
Widely observed trend

data staid in institutions

data can move around



how far can we go?
network capacity and security
most restricting factors



all at different locations

EUDAT - Federation of Centers

25 European partners

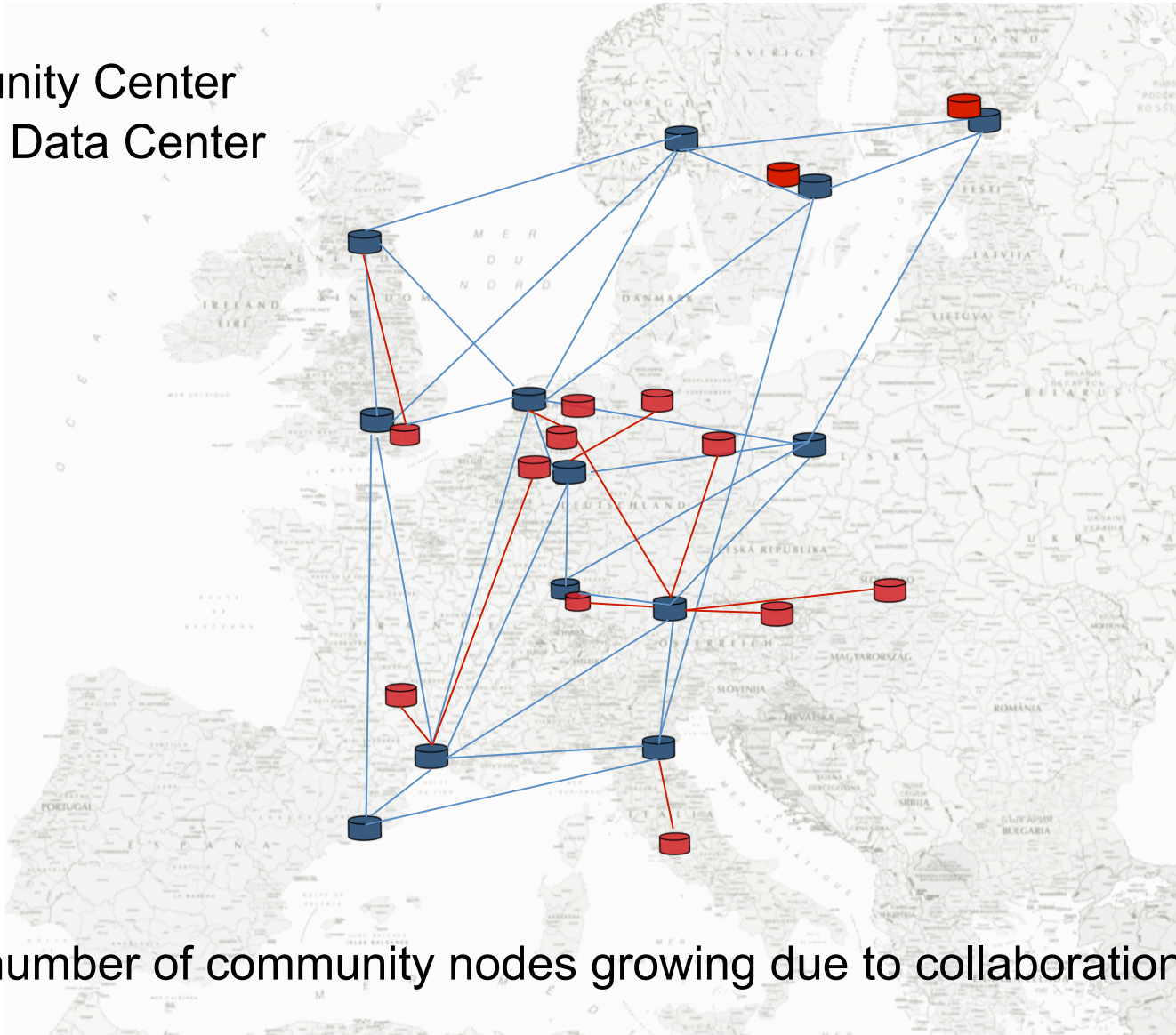
AND: communities were put into the driving seat results:

- we have services ready but lack an overall architecture 😊
- we rejected all technology not mature enough such as AAI



EUDAT Nodes

- Community Center
- EUDAT Data Center



number of community nodes growing due to collaboration projects

- being offered
- in progress
- to come

EUDAT Service Overview

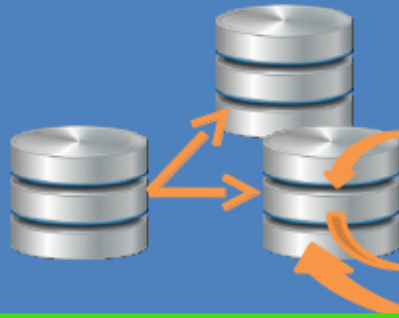
B2FIND

Aggregated EUDAT metadata domain.
Data inventory



B2SAFE

Data curation and access optimization



B2STAGE

Dynamic replication to HPC workspace for processing



B2SHARE

Researcher data store (simple upload, share and access)



AAI

Network of trust among authentication and authorization actors



PID

Identity
Integrity
Authenticity
Locations



EUDAT Box

dropbox-like service
easy sharing
local synching



B2DROP

Semantic Anno

checking, referencing and annotating



B2NOTE

Generic Workflow

automating data



installed Working Groups

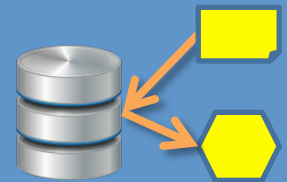
Dynamic Data

immediate handling



Type Registry

data intelligence



RDA Outcome

DRIHM - B2SHARE for Citizen Data

Community Domain

Search 37 records for

Search

Home

Login >>>

Deposit >>>

STORE AND SHARE YOUR RESEARCH DATA



A user-friendly, secure, robust, reliable and trusted service to share and store your research data *adding value to your research data by assigning Persistent Identifiers to ensure long-lasting access and reference*

Deposit your data via the generic interface or select a community extension including specific metadata fields

Share your data with others in a safe and trusted environment

Do you belong to a scientific community? Brand and create your own community collection with specific metadata fields customized for your field

Latest Deposits

2014-01-29

FIM4R report

by Tommi Nyönen, FIM4R community
Report from the federated identity management for research collaborations conference held in Helsinki.

2014-01-23

Syntactically annotated Czech legal texts

by Barbora Hladka
Two legal texts syntactically manually annotated according to the Prague dependency treebank framework. Dependency trees

2014-01-23

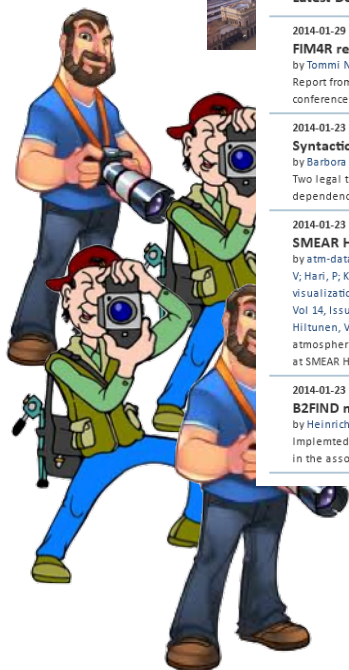
SMEAR Hyytiälä -2009; 23-01-2014

by atm-data@helsinki.fi Junninen, H; Lauri, A; Keronen, P; Aalto, P; Hiltunen, V; Hari, P; Kulmala, M. Smart-SMEAR: on-line data exploration and visualization tool for SMEAR stations. BOREAL ENVIRONMENT RESEARCH (BER) Vol 14, Issue 4, pp.447-457; Junninen, H; Lauri, A; Keronen, P; Aalto, P; Hiltunen, V; Hari, P; Kulmala, M.[...] atmospheric, flux, soil, tree physiological and water quality measurements at SMEAR Hyytiälä research station of the U

2014-01-23

B2FIND mapping files

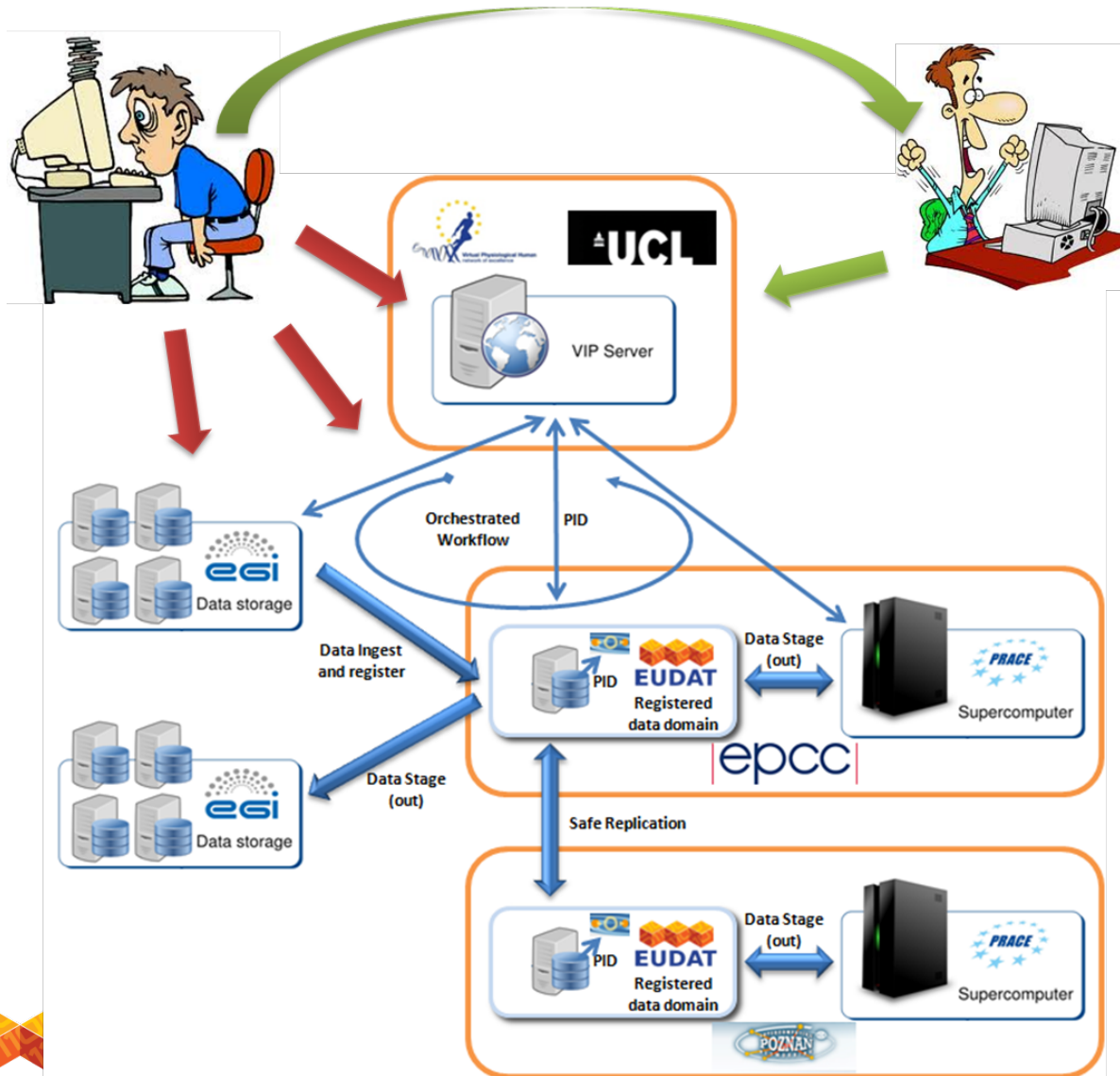
by Heinrich Widmann
Implemented in xpath2 syntax for each community and each metadata format in the associated xml files - .xml the mapping ru



Citizens



VPH B2SAFE Replication + B2STAGE Processing Data



Replication at physical layer is simple (File, Cloud-Objects, etc)

Replication including logical information is complex (all different)

Staging to HPC complex as well

Collaboration Projects

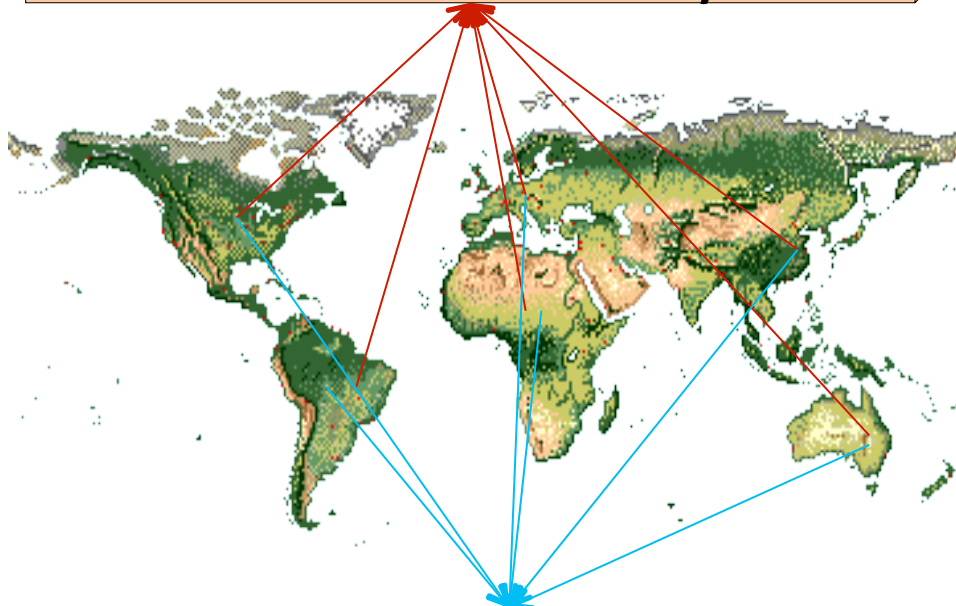
Over 30 applications from a vast range of scientific communities covering, but not limited to, life sciences, atmospheric sciences, marine biology, physics, chemistry, nano-science and materials science, medical science, biodiversity and climate modelling

Service	Collaboration
B2FIND	17
B2SAFE	16
B2SHARE	14
B2STAGE	9
BNOTE	4
PID	3
AAI	1
WF	1

Community / Focus
Atmospheric science
Biobanks
Biodiversity
Biology
Cross-disciplinary
Libraries
Climate & Climate Modelling
Toxico genomics
Environmental Science
GeoSpatial
Life science
Marine biology
Medical science
Molecular dynamics
Neuro informatics
Physics, chemistry, nano-science and materials science
Seismology
Solid Earth GeoScience
Vulcanology

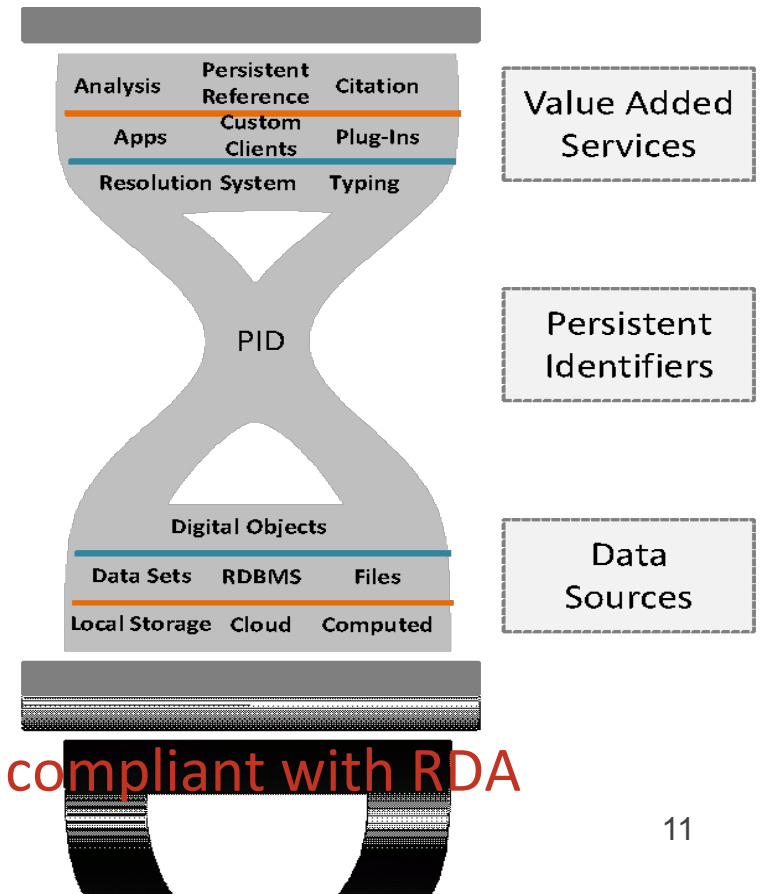
DONA – world wide PID System

Digital Object Numbering Authority
Senior Experts from all continents
Stewards of the Handle System



Worldwide Registration Authorities
Datacite, EPIC, CNRI, etc.

DONA is a Swiss foundation under the umbrella of ITU.



Sustainability from User Perspective

- CSC, SARA, BSC, CINECA, FZJ, RZG etc. will have a long lifetime as far as we can see – they are anchors in science landscape
- CLARIN, EPOS, ENES, etc. have their natural “entry nodes” which i often based on some history
- Proper agreements between community and entry node (SLA, etc)
- EUDAT Data centers also have their agreements (SLA, OLA, joint business model, etc) – some agreements driven by requirements are in place
- moving ahead to EUDAT 2 – but there will definitively be no EUDAT3
EUDAT must have a sustainable funding model based on services and contracts

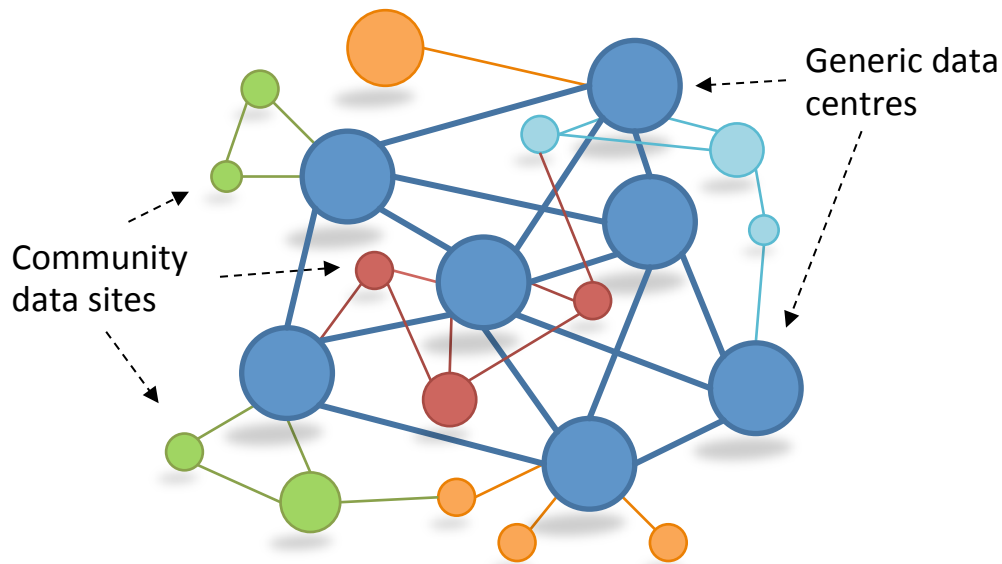


EUDAT is ready to serve science!

<http://www.eudat.eu>

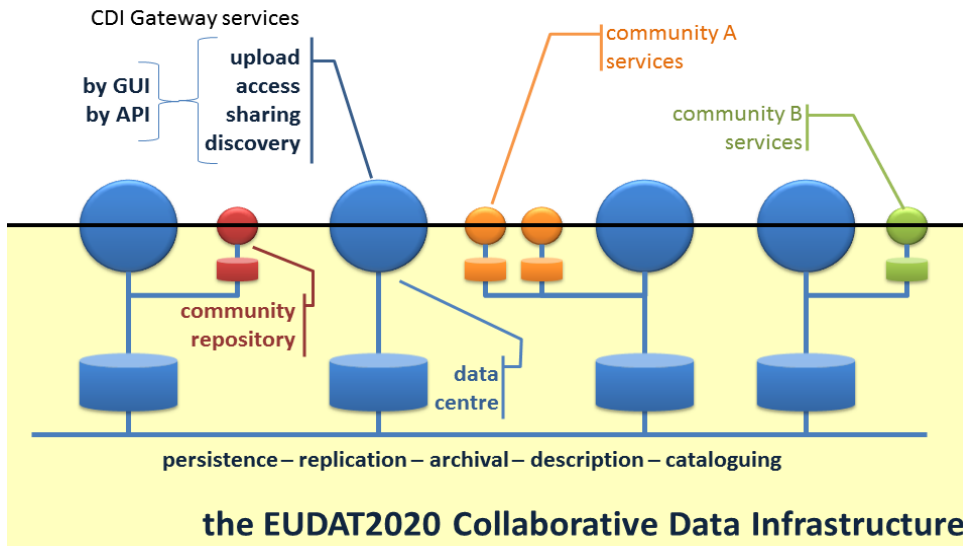
Thanks for your attention.

A Network of Trusted Centers



- Independent Centers working within a Common Framework to develop Services
- Generic Data Centers with long experience of working together
- Community Data Sites with connection to at least one Generic Center
- EUDAT is about providing solutions in a federated environment

Using Vs Joining



- **using EUDAT services:** finding and accessing data, for instance, or storing smaller data sets by interacting with one of the CDI public front-end services

VS

- **joining the CDI :** implies a tighter integration with at least one of the EUDAT centre.