EUDAT

Towards a pan-European Collaborative Data Infrastructure

Martin Hellmich
Slides adapted from Damien Lecarpentier
DCH-RP workshop, Manchester, 10 April 2013
Research Infrastructure trends:

- Internationalisation
- Diversification
- Increasingly relying on ICT
- Data deluge is a common challenge

European Ris:

- Around 500
- €100 billion investment
Data trends

- Where to store it?
- How to find it?
- How to make the most of it?
- How to ensure interoperability?
Collaborative Data Infrastructure - A framework for the future?
European Data

EUDAT

- Start date: 1st October 2011
- Duration: 36 Months
- Budget: 16.3 M€ (9.3M€ EC)
- EC Call: INFRA-2011-1.2.2
- Consortium: 25 partners from 13 countries
  • National data centers, technology providers, research

- Objectives:
  • Cost-efficient and high-quality CDI
  • Meetings users' needs in flexible and sustainable way
  • Across geographical and disciplinary boundaries

http://www.eudat.eu
Data Centers and Communities
Five research communities on Board

- **EPOS**: European Plate Observatory System
- **CLARIN**: Common Language Resources and Technology Infrastructure
- **ENES**: Service for Climate Modelling in Europe
- **LifeWatch**: Biodiversity Data and Observatories
- **VPH**: The Virtual Physiological Human

All share common challenges:
- Reference models and architectures
- Persistent data identifiers
- Metadata management
- Distributed data sources
- Data interoperability
Communities ↔ Data Centers
Building Blocks of the CDI

EUDAT Portal
Integrated APIs and harmonized access to EUDAT facilities

Metadata Catalogue
Aggregated EUDAT metadata domain.
Data inventory

Data Staging
Dynamic replication to HPC workspace for processing

Safe Replication
Data curation and access optimization

Simple Store
Researcher data store (simple upload, share and access)

AAI
Network of trust among authentication and authorization actors
Infrastructure – first pilots
Principles – where we want to be (1)

• **1**: Data deposited with the EUDAT CDI will be preserved in perpetuity
  – Costs will (probably) have to fall on data producers (or their funders) at time of data deposit

• **2**: Access to data in the EUDAT CDI is free at the point of use
  – Charging users for access or use creates a barrier to use that runs counter to the principles of open access
Principles – where we want to be (2)

• **3**: Data are best curated in their own communities
  – Data producers are central to discussions about its long term preservation

• **4**: EUDAT will operate as a federation of community-facing repositories and “back-office” hosting providers
  – EUDAT must operate as a partnership model with a distributed infrastructure hosting common services
Principles – where we want to be (3)

• 5: EUDAT services and infrastructure must be a suitable target for “Trustworthy Digital Repository” TDR outsourcing (cf. datasealofapproval.org)
  – opens the door for existing TDRs to join the EUDAT federation

• 6: EUDAT will not assert ownership of any data that it holds
Work plan

• Moving the services to a production environment
• Integrating new partners to EUDAT (in particular research communities)
  – Working groups, pilots, observers and associate partners
• Collaborating with other initiatives
  – European e-Infrastructures: EGI, PRACE, DANTE, HELIX NEBULA, SCIDIPS-ES, TERENA, etc.
  – Global initiatives: RDA, CODATA, etc
• Defining EUDAT’s path to sustainability
  – Cost and funding models
  – Governance
Your Questions

• How to submit feedback? EUDAT User Forums, service contact lists
  http://eudat.eu/2nd-eudat-user-forum -> Fact sheets

• We welcome Use Cases from DCH-RP! We want to know how you want to be involved

• D7.2.1: Managing data curation and long-term preservation in a federated environment (not avail. yet)

• Contact: Damien Lecarpentier
damien.lecarpentier@csc.fi
Contact Us

Project Coordinator: Kimmo Koski
kimmo.koski@csc.fi

Scientific Coordinator: Peter Wittenburg
peter.wittenburg@mpi.nl

Project Manager: Damien Lecarpentier
damien.lecarpentier@csc.fi

Dissemination Manager: Nagham Salman
nagham.salman@bsc.es

Industry Task Force: David Manset
dmanset@maatg.fr

http://www.eudat.eu

eudat-info@postit.csc.fi

EUDAT
Hierarchy of data needs

Value creation, though openness and sharing...

... relies on more basic data needs being met first

Combining data

Sharing data

Improving data usability and reusability

Keeping data safe and accessible

Storing and archiving data

Combining data in imaginative new ways to solve problems

Sharing data across communities

Metadata catalogue, persistent identifiers

Data security, data standards, data curation

Data archive, data storage facilities

Value creation, though openness and sharing...

... relies on more basic data needs being met first