



EUDAT

Towards a pan-European Collaborative Data Infrastructure

Damien Lecarpentier
CSC-IT Center for Science, Finland
EUDAT User Forum, Barcelona



Date: 7 March 2012



EUDAT Key facts

Project Name	EUDAT – European Data
Start date	1st October 2011
Duration	36 months
Budget	16,3 M€ (including 9,3 M€ from the EC)
EC call	Call 9 (INFRA-2011-1.2.2): Data infrastructure for e-Science (11.2010)
Participants	25 partners from 13 countries (national data centers, technology providers, research communities, and funding agencies)
Objectives	“To deliver cost-efficient and high quality Collaborative Data Infrastructure (CDI) with the capacity and capability for meeting researchers’ needs in a flexible and sustainable way, across geographical and disciplinary boundaries.”



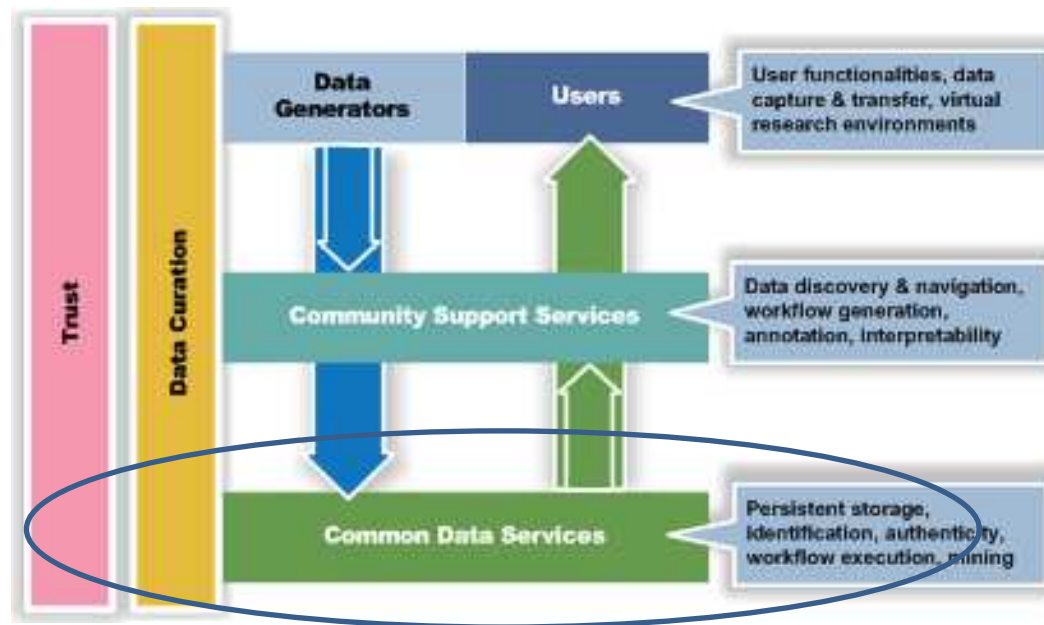
Consortium



Data centers and Communities



The CDI concept



EUDAT Core Service Areas

Community-oriented services

- Simple Data Access and upload
- Long term preservation
- Shared workspaces
- Execution and workflow (data mining, etc.)
- Joint metadata and data visibility

Enabling services (making use of existing services where possible)

- Persistent identifier service (EPIC, DataCite)
- Federated AAI service
- Network Services
- Monitoring and accounting

Core services are building blocks of EUDAT's Common Data Infrastructure
mainly included on bottom layer of data services



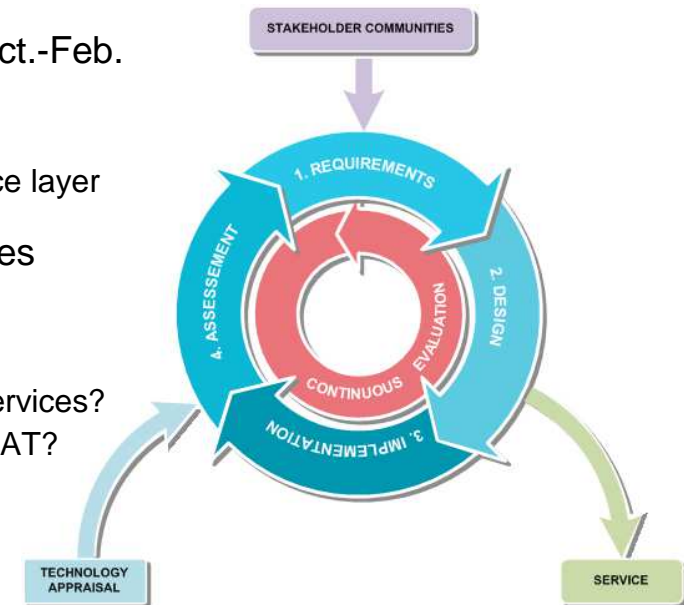
EUDAT service design activities

1. Capturing Communities Requirements (WP4)

- 1st round of interviews with the five initial communities (Oct.-Feb. 2012)
 - Understand how data is organised in each community
 - Collect first wishes and specific requirements from a common data service layer
- Next phase: refine analysis and expanding it to other communities

2. Building the corresponding services (WP5)

- Technology appraisal (ongoing)
 - What is already available at partners's sites to build the corresponding services?
 - What are the gaps and market failures that should be addressed by EUDAT?
- Next phase: Developing candidate services
 - Adapt services to match the requirements
 - Integrate with community and SP services
 - Test and evaluate with communities



3. Deploying the services and operating the federated infrastructure (WP6)

- Designing the federated infrastructure and the interfaces for cross-site operations (ongoing)
- Next phase: integrating and coordinating resource provision, operations and support



First Service Cases

■ November 2011: shortlist of 6 service/use cases identified

- **Safe replication:** Allow communities to safely replicate data to selected data centers for storage and do this in a robust, highly available way.
Task Force - Under implementation
- **Dynamic replication:** Perform (HPC) computations on replicated data. Move (part of) the safely replicated data to powerful machines and move the results back into the archives.
Task Force - Under implementation
- **Metadata:** A joint metadata domain for all data that is managed by EUDAT data centers by harvesting metadata records for all data sets in the communities. Allow to have a catalogue to demonstrate data discovery, and to have a registry which can be used for automatic operations such as data mining.
Task Force created
- **Research data store:** A function that will help researchers mediated by the participating communities to upload and store data which is not part of the officially handled data sets of the community.
- **AAI:** A solution for a working AAI scenario.
Task Force created
- **PID:** a robust, highly available and effective PID system that can be used within the communities and by EUDAT.



Why this User Forum?

- Making EUDAT's plans public and getting new input
 - Are the first requirements we identified shared by a broad range of communities
 - Are there other requirements/services we could think as being part of the common service layer?
 - Is our approach of the CDI the right approach? Are we missing something?
 - Also training needs, sustainability, etc.
- Getting other communities involved in the project
 - Discuss collaboration opportunities: NOT just exchanging contact details!
 - Plan next steps together: how would you like to be involved in EUDAT?



EUDAT USER FORUM, 7-8 MARCH 2012

TOWARDS A COLLABORATIVE DATA INFRASTRUCTURE

INVESTIGATING RESEARCH COMMUNITIES' REQUIREMENTS

Place:
Jordi Girona Street, No.31
Rectorat Building, Junes room,
08034, Barcelona

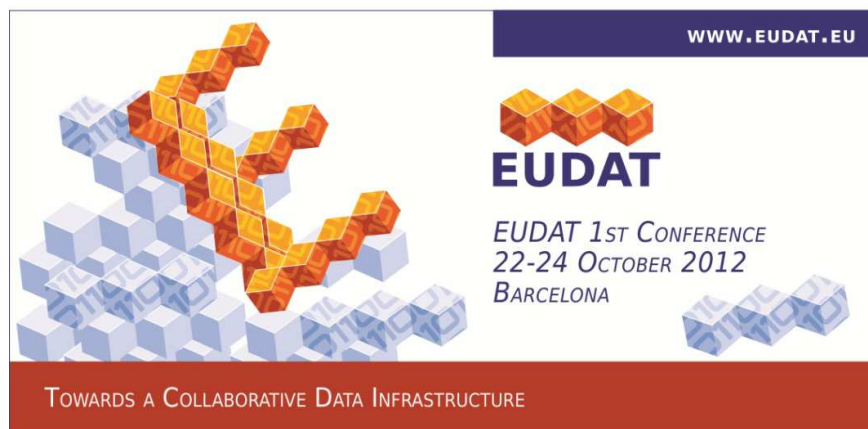
DAY 1: 7TH MARCH 2012

12:00	Registration and Snacks	
12:50	Welcome talk	Sergi Girona
12:55	Opening talk	Kimmo Koski
13:00	SESSION 1 - SETTING THE SCENE	CHAIR: KIMMO KOSKI
13:00	EUDAT - an Overview for a User Perspective	Damien Lecarpentier
13:15	CDI and EUDAT	Peter Wittenburg
13:30	Climate modeling and EUDAT	Michael Lautenschlager
13:40	Seismology and EUDAT	Alberto Micheli
13:50	Physiology and EUDAT	Stefan Zasada
14:00	Linguistics and EUDAT	Pavel Straňák
14:10	Ecology and EUDAT	Wouter Los/Michael Mirtl
14:20	Discussion	
15:00	Coffee Break	
15:30	SESSION 2 - EUDAT SERVICE CASES	CHAIR: ALBERTO MICHELINI
15:30	Replication service and its requirements	Peter Wittenburg
15:40	Staging Replicas for computations and requirements	Stefan Zasada
15:50	Researchers' Data Store and requirements	Daan Broeder
16:00	Metadata requirements	Daan Broeder
16:10	Discussion	
16:45	Short break	
17:00	Community Presentations I	
	ECRIN – European Clinical Research Infrastructure Network	Wolfgang Kuchinke
	ICOS – Integrated Carbon Observation System	Timo Vesala
	ELIXIR – Data for Life	Tommi Nyrönen
	diXa – a data infrastructure for chemical safety	Jos Kleinjans
	PanData – Open Data Infrastructure for research with Photons and Neutron	Juan Bicarregui
18:30	End	
20:30	Dinner	

DAY 2: 8TH MARCH 2012

09:00	SESSION 3 - ENABLING TECHNOLOGIES	CHAIR: MORRIS RIEDEL
09:00	Safe Replication	Jedrezej Rybicki
09:15	Data Staging	Giuseppe Fiameni
09:30	Creating a Joint Metadata Domain	Michael Lautenschlager
09:40	Distributed Authentication	Claudio Cacciari
09:50	PID Systems for Digital Objects	Ulrich Schwarzmann
10:00	The EUDAT Research Agenda	David Corney
10:10	Hosting and Service Provisioning	Johannes Reetz
10:20	Discussion	
10:40	Coffee Break	
11:00	Community Presentations II	
	INCF - Toward a collaborative research infrastructure for neuroscience	Sean Hill
	Simulation and data analysis data and data access requirements for ITER	Par Strand
	BioVeL: Biodiversity Virtual e-Laboratory	Vera Hernandez
12:00	SESSION 4 - EUDAT AND THE WAY FORWARD	CHAIR: PAVEL STRAÑÁK
12:00	Sustaining the infrastructure	Alison Kennedy
12:15	Training the new data scientist	Adam Carter
12:25	Discussion	
12:40	WRAP UP AND GENERAL DISCUSSION	D. LECARPENTIER/ P. WITTENBURG
13:00	End	

Welcome to the 1st EUDAT Conference!



22-24 October 2012, Barcelona

- International event with keynotes from Europe and US
- A forum to discuss the future of data infrastructures
- Project presentations and poster sessions
- 2nd EUDAT User Forum
- Training tutorials