

EUDAT

Towards a European Collaborative Data Infrastructure

Leif Laaksonen & Damien Lecarpentier
CSC, IT Center for Science, Finland
e-IRG workshop, 13-14 October 2011



Outline of the talk

- ❑ EUDAT concept
- ❑ EUDAT consortium
- ❑ EUDAT service approach
- ❑ Expected benefits and possibilities of a CDI



EUDAT Key facts and objectives

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.”

Building the generic data infrastructure layer

e-Infrastructures Vision

empower research communities through ubiquitous, trusted and easy access to services for data, computation, communication and collaborative work



Scientific facilities, research communities

Source: European Commission



The current data infrastructure landscape: challenges and opportunities

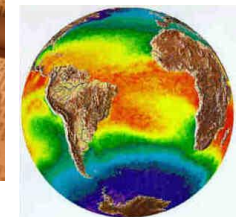
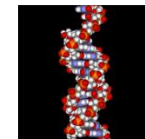
- Long history of data management in Europe: several existing data infrastructures dealing with established and growing user communities (e.g., ESO, ESA, EBI, CERN)
 - New Research Infrastructures are emerging and are also planning to build data infrastructure solutions to meet their needs (CLARIN, EPOS, ELIXIR, ESS, etc.)
 - A large number of projects providing excellent data services (EURO-VO, GENESI-DR, Geo-Seas, HELIO, IMPACT, METAFOR, PESI, SEALS, etc.)
- **However, most of these infrastructures and initiatives address primarily the needs of a specific discipline and user community**

Challenges

- Compatibility, interoperability, and cross-disciplinary research
 - how to re-use and recombine data in new scientific contexts (i.e. across disciplinary domains)
- Data growth in volume and complexity (the so-called “data tsunami”)
 - strong impact on costs threatening the sustainability of the infrastructure

Opportunities

- Potential synergies do exist: although disciplines have different ambitions, they have common basic needs and service requirements that can be matched with generic pan-European services supporting multiple communities, thus ensuring at the same time greater interoperability.



History of the EUDAT concept

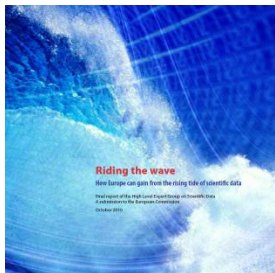
- EUDAT has its origins in the work of the **PARADE** (Partnership for Accessing Data in Europe) initiative

- **PARADE White Paper** (October 2009) defining a "Strategy for a European Data Infrastructure that should be persistent, multidisciplinary, and based on the need of user communities"



- The concept of a shared pan-European infrastructure was supported and further elaborated by a number of policy and experts bodies:

- **e-IRG and ESFRI**: e-IRG Blue Paper (September 2010) Recommends "to identify and promote common (long term) data related services across different RI"

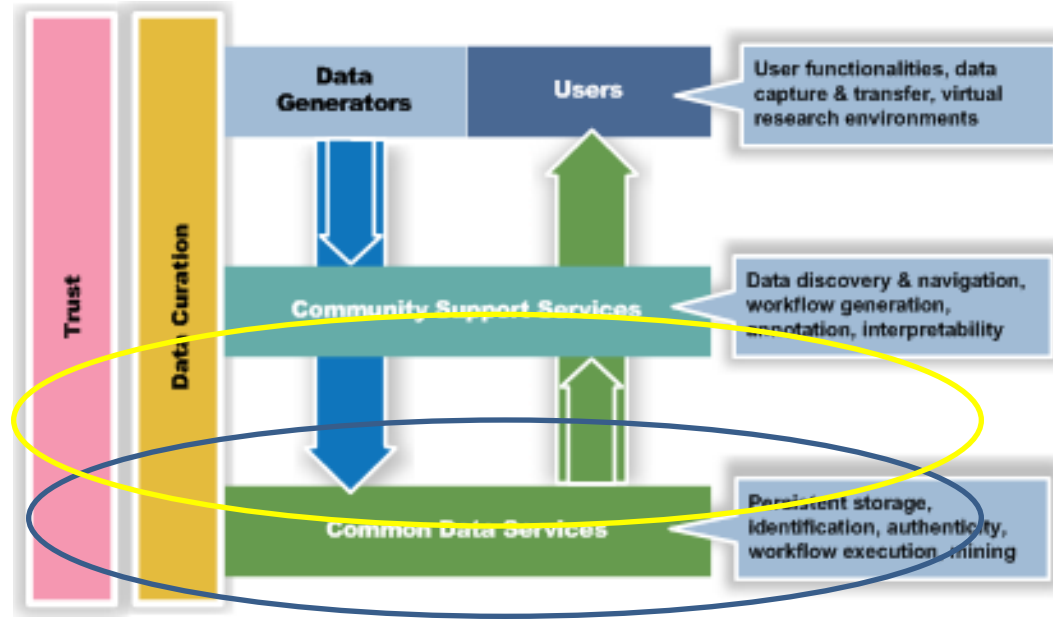


- **High Level Expert Group (HLEG)** report on Scientific Data (October 2010) Calls for a "Collaborative Data Infrastructure" for scientific data, that supports seamless access, use, reuse, and trust of data.

- **EUDAT** will materialise this vision from October 2011

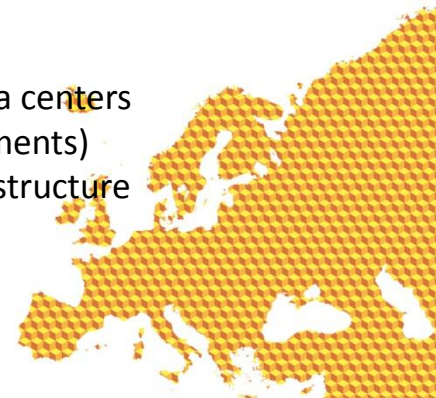


Towards a Collaborative Data Infrastructure



Source: HLEG report, p. 31

- EUDAT will focus on building a generic data infrastructure layer offering a trusted domain for long term data preservation with services to store, identify, authenticate and mine these data.
- This need be done in close collaboration with the Communities
 - Collaboration the communities involved in designing specific services and the data centers willing to provide generic solutions (core services must match communities requirements)
 - Community services can also be incorporated into the common data service infrastructure when they are of use to other communities.



The EUDAT Consortium



The EUDAT Communities

ESFRI

CLARIN
Common Language Resources and Technology Infrastructure

About CLARIN
Services
Publications
Activities

Events
Links
Glossary
Contact
Join CLARIN
Internal Web Site

Semantic data description and descriptive metadata are vital factors for determining if the data can be reused in the future. These metadata are still dependent on rapidly changing ontologies and terminologies.

John Marks
ESF 2008

Activities
Publications
Solutions
Laboratory
Consultancy
Virtual Language Observatory

January 26, 2011
New Virtual Language Observatory launch

EPOS
EUROPEAN PLATE OBSERVING SYSTEM

Research Infrastructure and E-Science for Data and Observatories on Earthquakes, Volcanoes, Surface Dynamics and Tectonics

SEARCH

Mission & Vision Objectives Architecture Partners Preparatory Phase Data Products

LIFEWATCH
e-science and technology infrastructure for biodiversity data and observatories

Home Contact About Participants Get Involved News Cases Events Press Documents

LIFEWATCH COUNTRIES
Austria Belgium Denmark Finland France Greece Hungary Italy Netherlands Norway Poland
Portugal Romania Slovak Republic Slovenia Spain Sweden Turkey United Kingdom

LIFEWATCH NEWS
2011-02-16 **LIFEWATCH RESEARCH INFRASTRUCTURE STARTS CONSTRUCTION IN 2011** - The initial country consortium establishing the LifeWatch research infrastructure agreed to finance ... [Read more](#)
2011-01-19 **LIFEWATCH CLOSING EVENT** - On this page you can download all the slides presented at the closing event of the LifeWatch preparatory project a first group of ... [Read more](#)
2011-01-17 **LIFEWATCH CONSTRUCTION KICKS OFF ON JANUARY 19TH** - On 19 January 2011, at the closing conference of the LifeWatch preparatory project a first group of ... [Read more](#)

LIFEWATCH FOCUS
LifeWatch research infrastructure starts construction in 2011
The initial country consortium establishing the LifeWatch research infrastructure agreed to finance the start-up activities for the infrastructure construction. These countries will host the Common Facilities of LifeWatch.
On 19th January 2011, representatives from organizations in Hungary, Italy, the Netherlands, Romania and Spain signed a Memorandum of Understanding to cooperate for an early start of the LifeWatch infrastructure for biodiversity and ecosystem research. The LifeWatch Stakeholders Board, representing the ten countries aiming at establishing the LifeWatch ERIC, welcomed the initiative to start early construction.

Newsletter
Subscribe to our newsletter. Send an email to newsletter@lifewatch.org

Quote
"Through our Memorandum of Cooperation GBIF and LifeWatch, based on our respective complementary mandates, now have a formal framework for co-operation and collaboration on infrastructural developments, building on GBIF's 10 years of investment to date."
Dr. Nick King
Director Global Biodiversity Information Facility (GBIF)

enes
European Network for Earth System Modelling

Welcome

ENES Townhall Meeting at EGU 2010: Here is the [announcement!](#)

For latest news on IS-ENES click [here!](#)

A major challenge for the climate research community is the development of comprehensive Earth system models capable of simulating natural climate variability and human-induced climate changes. Such models need to account for detailed processes occurring in the atmosphere, the ocean and on the continents including physical, chemical and biological processes on a variety of spatial and temporal scales. They have also to capture complex nonlinear interactions between the different components of the Earth system and assess, how these interactions can be perturbed as a result of human activities.

Accurate scientific information is required by government and industry to make appropriate decisions regarding our global environment, with direct consequences on the economy and lifestyles. It is therefore the responsibility of the scientific community to accelerate progress towards a better understanding of the processes governing the Earth system and towards the development of an improved predictive capability. An important task is to develop an advanced software and hardware environment in Europe, under which the most advanced high resolution climate models can be developed, improved, and integrated.

search...

Virtual Physiological Human
network of excellence

Home WP1 WP2 WP3 WP4 WP5 VPH-I MIP Login

Username:
Password:
Remember Me ☐
Login
[Forgot your password?](#)
[Forgot your username?](#)

Project
Home
News
Objectives
VPH 2010
Devices
VPH Links
Contact Us
Archives

Activities
European Projects
Meetings
Publications
Dissemination Material
VPH Events
Event archive

Welcome to the home page of the Virtual Physiological Human Network of Excellence (VPH NoE) and information portal for the VPH Initiative

VPH NoE
BioMed Town
VPH Initiative
VPH for the Public
VPH for Researchers
VPH for Clinicians

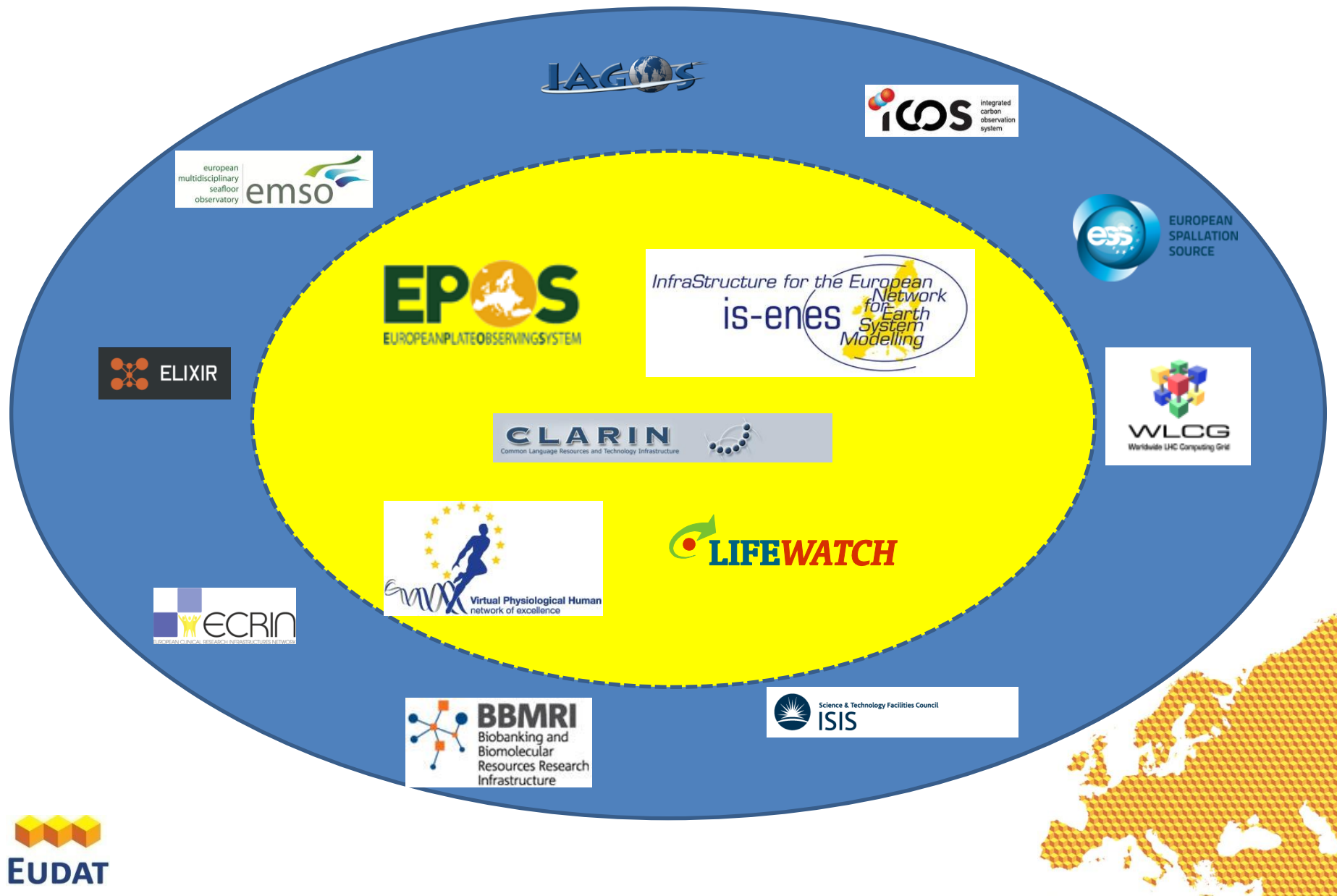
The VPH NoE is a project which aims to help support and progress European research in biomedical modelling and simulation of the human body. This will improve our ability to predict, diagnose and treat disease, and have a dramatic impact on the future of healthcare, the pharmaceutical and medical device industries.

VPH 2010
September 30th - October 1st 2010
Brussels, Belgium

HIGHLIGHTS
Building a wider VPH Community
Interface Plus special issue with VPH papers from the VPH-2010 Conference
VPH NoE and the Plasma Alliance
Example Project Call 2
Join the Public Forum of the VPH-ETC Support Action
Multi-institutional Graduate Programme for Virtual Physiological Human Scientists (VPH-MIP)
VPH Vision & Strategy Paper II
VPH NoE 5th VPH Newsletter Jan 2011 for website

LATEST VPH EVENTS
01.06.2011 - 03.06.2011 ICCS 2011 (Toulouse)
06.06.2011 - 09.06.2011 VPHS
09.06.2011 ICAPAS Workshop

The EUDAT Communities



EUDAT Services Activities – Iterative Design

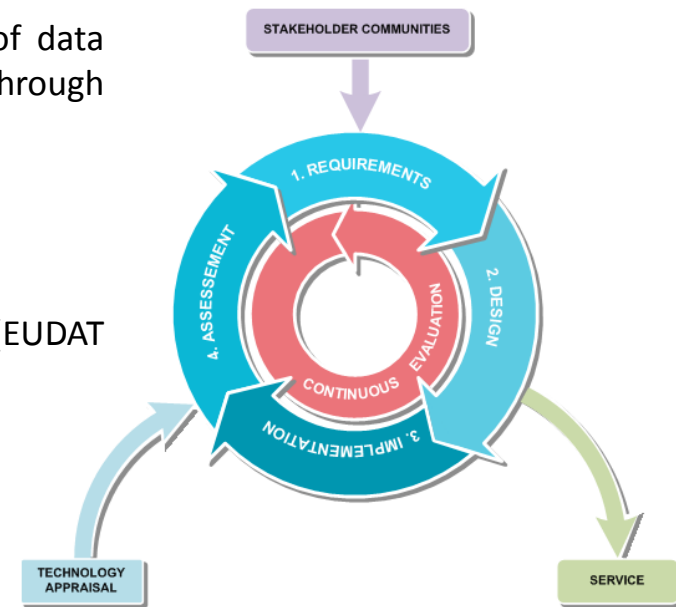
EUDAT's Services activity is concerned with identification of the types of data services needed by the European research communities, delivering them through a federated data infrastructure and supporting their users

1. Capturing Communities Requirements (WP4)

- Services to be deployed must be based on user communities needs
- Strong engagement and collaboration with user communities (EUDAT communities and beyond) to capture requirements

2. Building the services (WP5)

- User requirements must be matched with available technologies
- Need to identify:
 - available technologies and tools to develop the required services (technology appraisal)
 - gaps and market failures that should be addressed by EUDAT research activities
- Services must be designed, built and tested in a pre-production test bed environment and made available to WP4 for evaluation by their users



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

- Services must be deployed on the EUDAT infrastructure and made available to users, with interfaces for cross-site, cross-community operation
- Infrastructure must provide full life cycle data management services, ensuring the authenticity, integrity, retention and preservation of data, especially those marked for long-term archiving.



EUDAT core services

Core services are building blocks of EUDAT's Common Data Infrastructure

mainly included on bottom layer of data services

Fundamental Core Services

- Long-term preservation
- Persistent identifier service
- Data access and upload
- Workspaces
- Web execution and workflow services
- Single Sign On (federated AAI)
- Monitoring and accounting services
- Network services



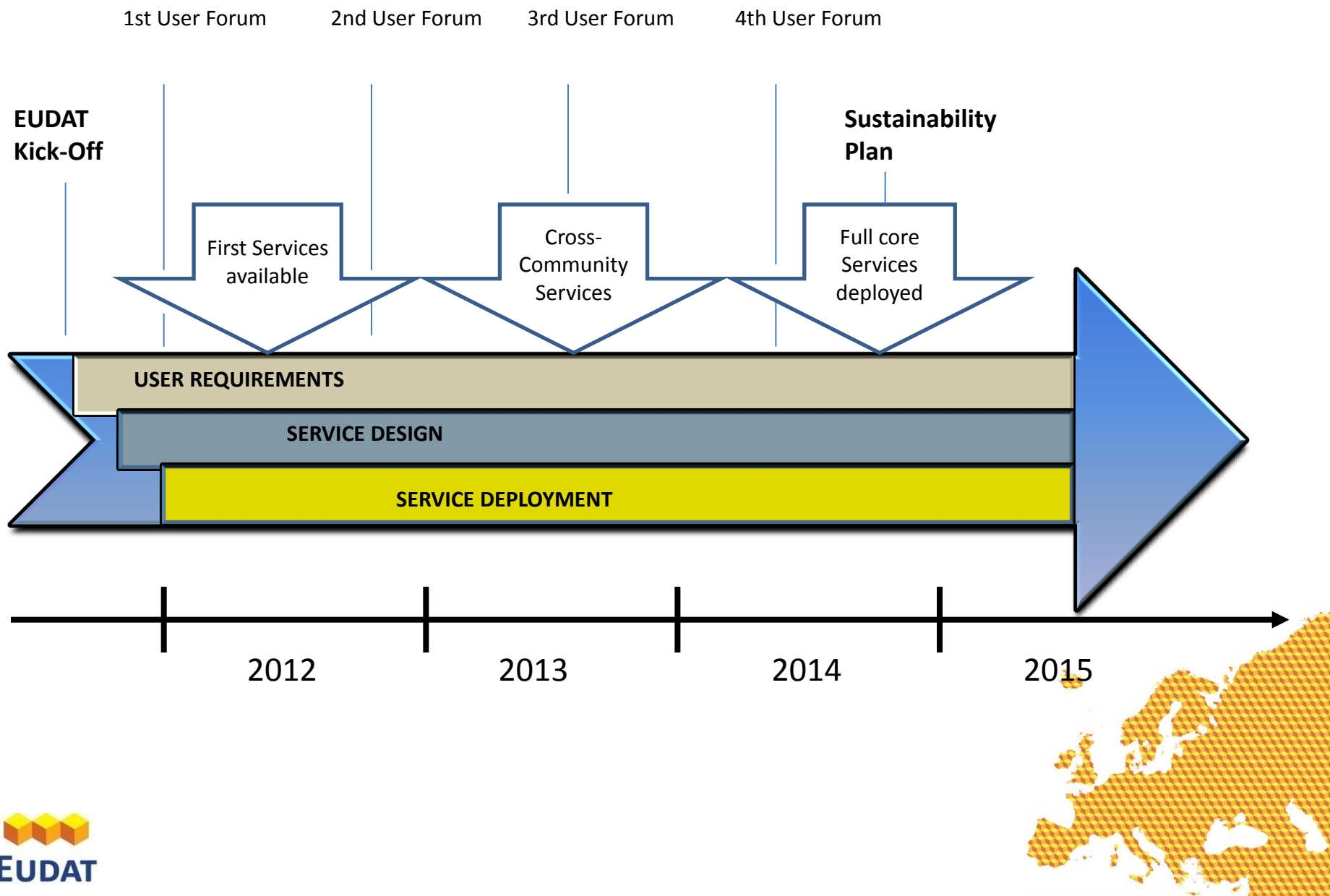
Extended Core Services (community-supported)

- Joint meta data service
- Joint data mining service

No need to match the needs of all at the same time, addressing a group of communities can be very valuable, too



EUDAT Timeline



Expected benefits of a Collaborative Data Infrastructure

▪ Enabling multi-disciplinary data intensive research and collaboration

- Development of common services supporting research communities
 - Support to existing scientific communities' infrastructures
 - Support to smaller communities through access to sophisticated services
- Inter-disciplinary collaboration and exploitation of synergies between communities
 - Communities from different disciplines working together to build services
 - Data sharing between disciplines
- Collaboration with other large-scale infrastructure
 - European e-Infrastructures: Géant, PRACE, EGI, etc.
 - Global initiatives in the US, Japan, Australia, etc.

▪ Ensuring wide access to and preservation of data in a sustainable way

- A robust generic infrastructure capable of handling the scale and complexity of data that will be generated over the next 10-20 years
 - Greater access to existing data and better management of data for the future
 - Increased security by managing multiple copies in geographically distant locations
- Put Europe in a competitive position for important data repositories of world-wide relevance

▪ Economies of scale and cost-efficiency

- Shared resources and work are less costly



How to join the EUDAT initiative?

- EUDAT has now 25 partners coming from 13 countries
- Scaling the infrastructure to other countries and partners is good (increase in complexity and richness, new solutions and practices, etc.) and is needed to build up a pan-European solution
- EUDAT project team currently defining best way to integrate new partners to the initiative
 - User forums open to all stakeholders interested in adapting their solutions or contributing to the design of the infrastructure
 - Associated membership also being defined to allow external partners to follow and to contribute to the activities of the project.



How to get in touch with EUDAT?

Kimmo Koski, CSC - IT Center for Science
EUDAT Project Coordinator
Kimmo.Koski@csc.fi

Peter Wittenburg, Max Planck Institute for Psycholinguistics at Nijmegen (MPI-PL)
EUDAT Scientific Coordinator
Peter.Wittenburg@mpi.nl

Damien Lecarpentier, CSC - IT Center for Science
EUDAT Project Manager
Damien.Lecarpentier@csc.fi

THANK YOU!

