



# EUDAT

## Towards a pan-European Collaborative Data Infrastructure

Norbert Meyer

Poznan Supercomputing and Networking Center

Poland



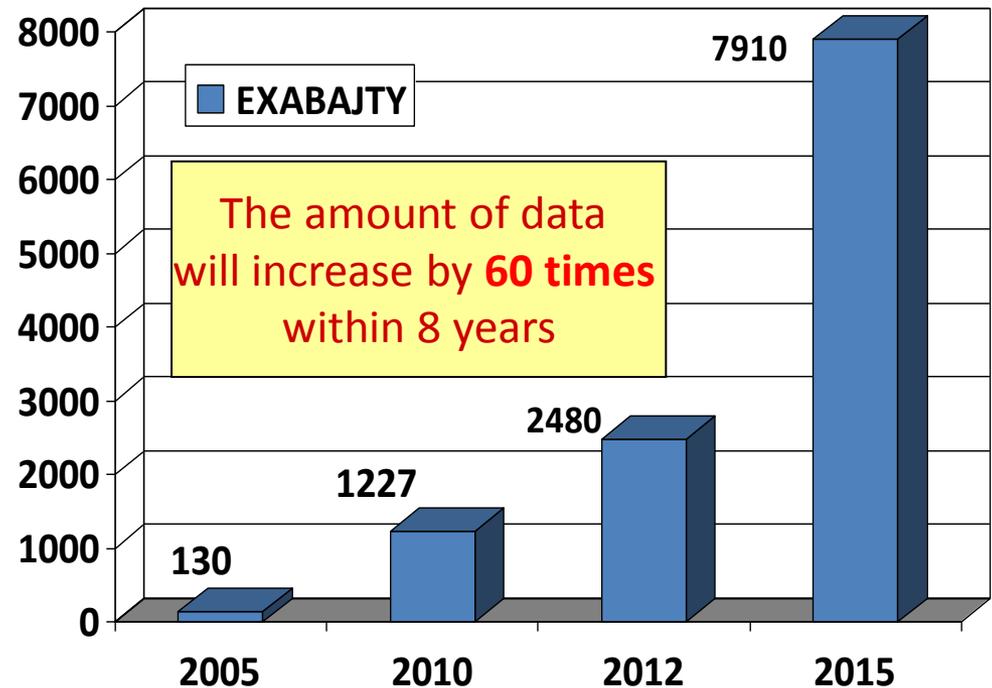
TPDL conference, September 24, 2013



# The Digital Universe

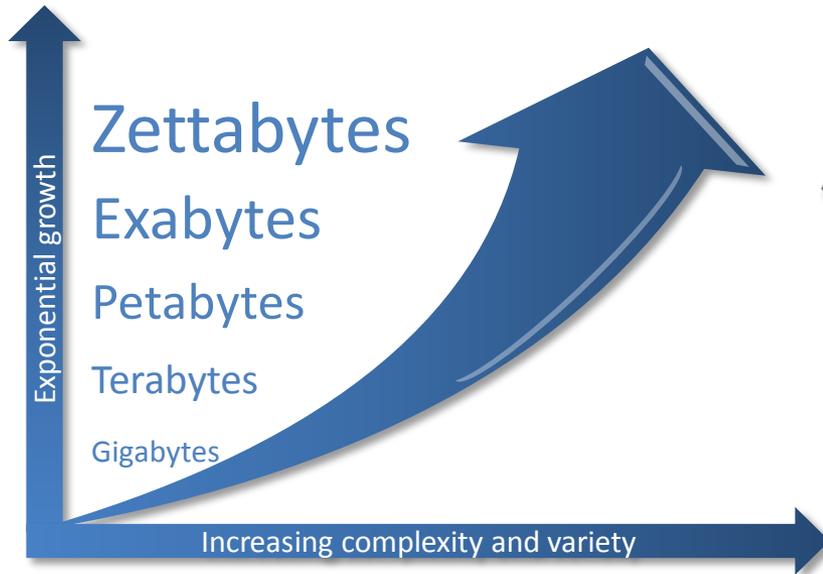
- 1,987 ZBytes generated since 01.01.2011
- 1,987 Zeta bytes = 1.987.000.000.000.000.000.000 bytes ....
- 2012 - 2,5 ZB (doubled within 12 months)
- 60+ % data lost due to missing hardware capacity .....

Digital info bytes created, moved, copied, sent annually



\*) source : „The 2011 IDC DIGITAL UNIVERSE STUDY sponsored by ECM2”

# Data trends



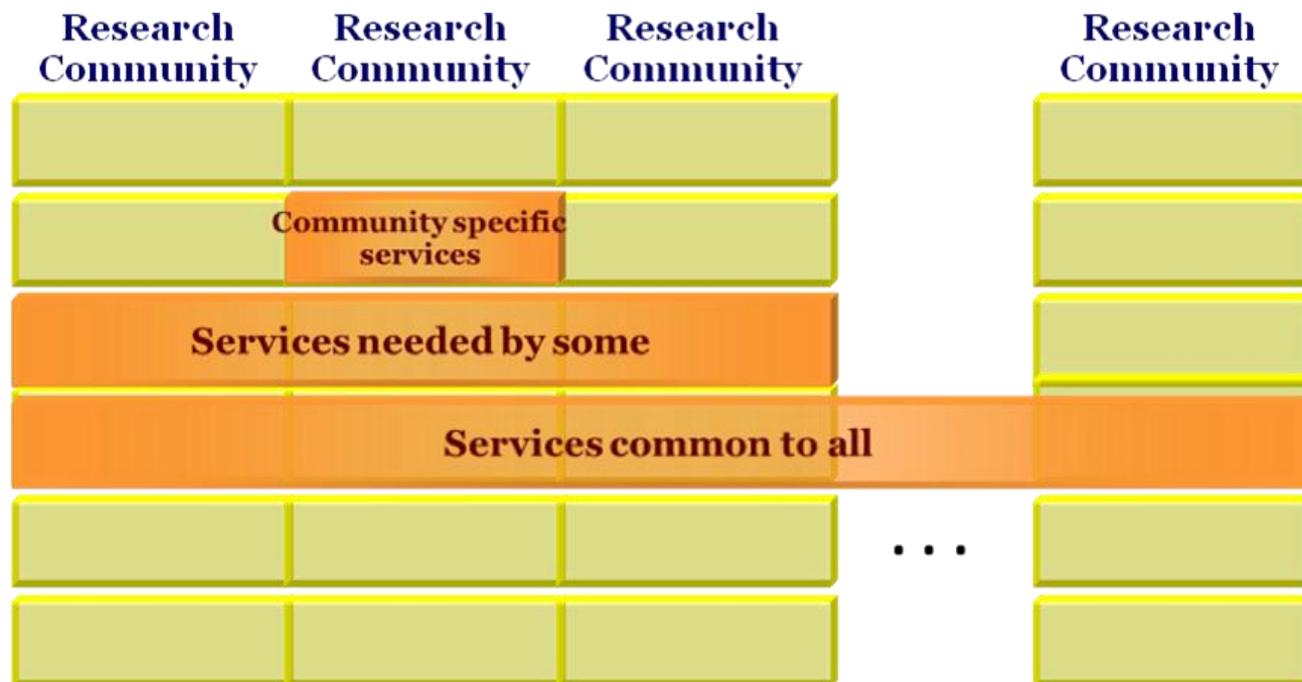
- Where to store it?
- How to find it?
- How to make the most of it?



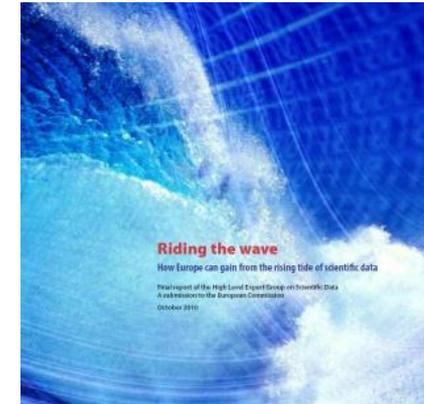
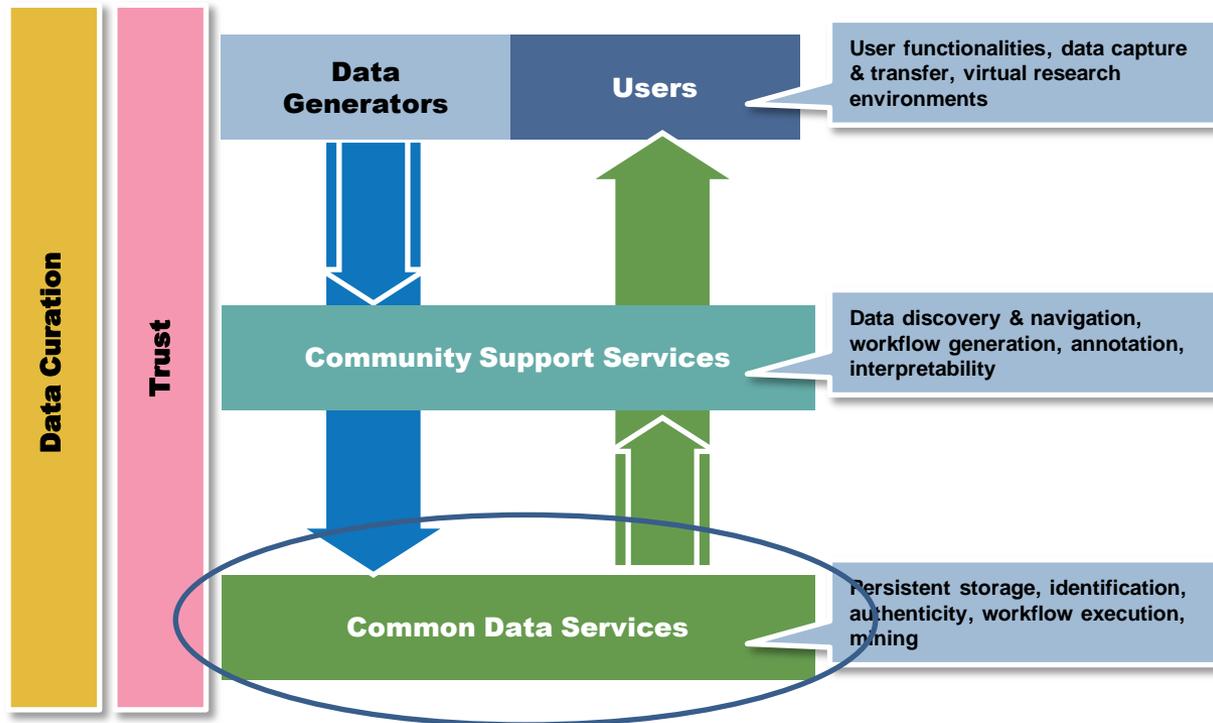
- How to ensure interoperability?
- Sustainability of services ?

# The EUDAT Case

If there are hundreds of Research Infrastructures, how many different data management systems can we sustain?



# Collaborative Data Infrastructure -A framework for the future? -



# European Data



Researchers, citizens,  
industry and society...

- 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
  - Meeting users' needs in a 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



# Building Blocks of the CDI



## EUDAT Portal

Integrated APIs and harmonized access to EUDAT facilities

## Metadata Catalogue

Aggregated EUDAT metadata domain.  
Data inventory



## AAI

Network of trust among authentication and authorization actors

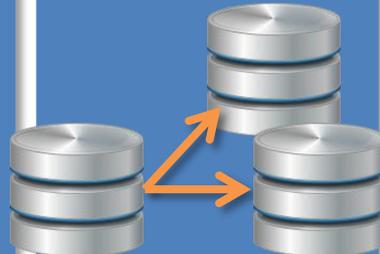
## Data Staging

Dynamic replication to HPC workspace for processing



## Safe Replication

Data curation and access optimization



## Simple Store

Researcher data store (simple upload, share and access)



# Expected benefits of the CDI

## ▪ Cost-efficiency through shared resources and economies of scale

- Better exploitation of synergies between communities and service providers
- Support to existing scientific communities' infrastructures and smaller communities

## ▪ Trans-disciplinarity

- Inter-disciplinary collaboration
  - Communities from different disciplines working together to build services
  - Data sharing between disciplines – re-use and re-purposing
  - Each discipline can solve only part of a problem

## ▪ Cross-border services

- Data nowadays distributed across states, countries, continents, research groups are international

## ▪ Sustainability

- Ensuring wide access to and preservation of data
  - 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

# Blue Paper

Invited ESFRI cluster projects

BioMedBridges, DASISH, ENVRI,  
CRISP (**pilot projects**)

also DC-NET, PaNdata, ITER



e-IRG “Blue Paper” on  
Data Management

---

FINAL VERSION

30 October 2012



**Norbert Meyer**

[meyer@man.poznan.pl](mailto:meyer@man.poznan.pl)