EUDAT

Towards a pan-European Collaborative Data Infrastructure

Norbert Meyer

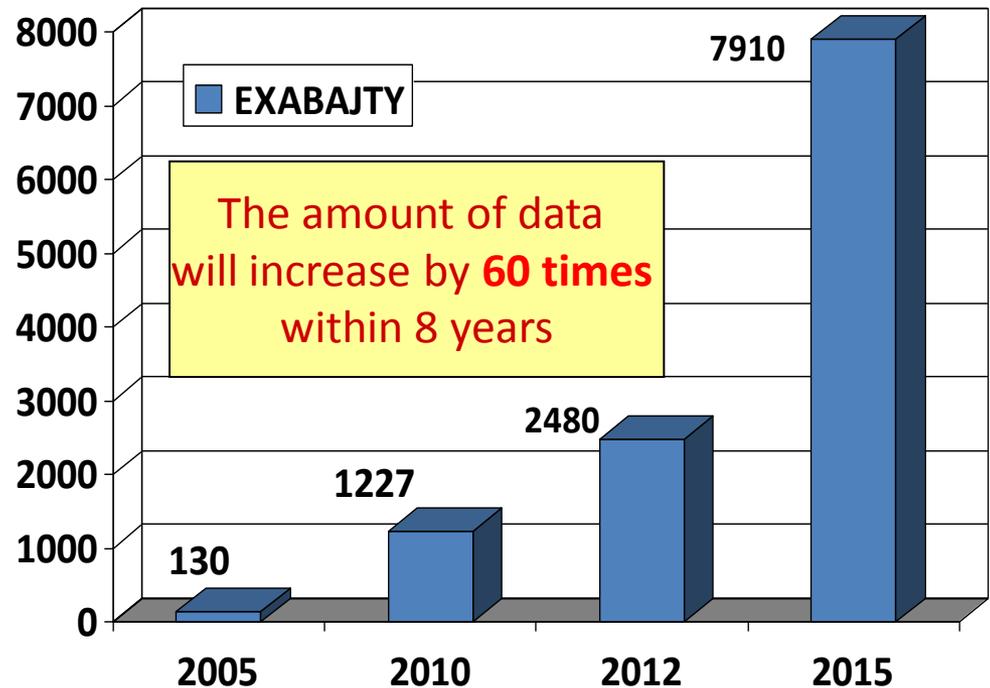Poznan Supercomputing and Networking Center

Poland

ICT 2013, Vilnius, November 7, 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 annualy**

The amount of data will increase by **60 times** within 8 years

Year	EXABAJTY
2005	130
2010	1227
2012	2480
2015	7910

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

EUDAT

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? -
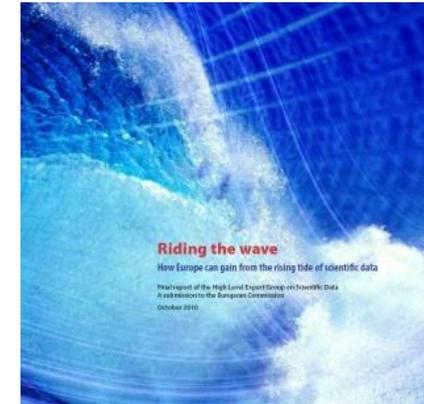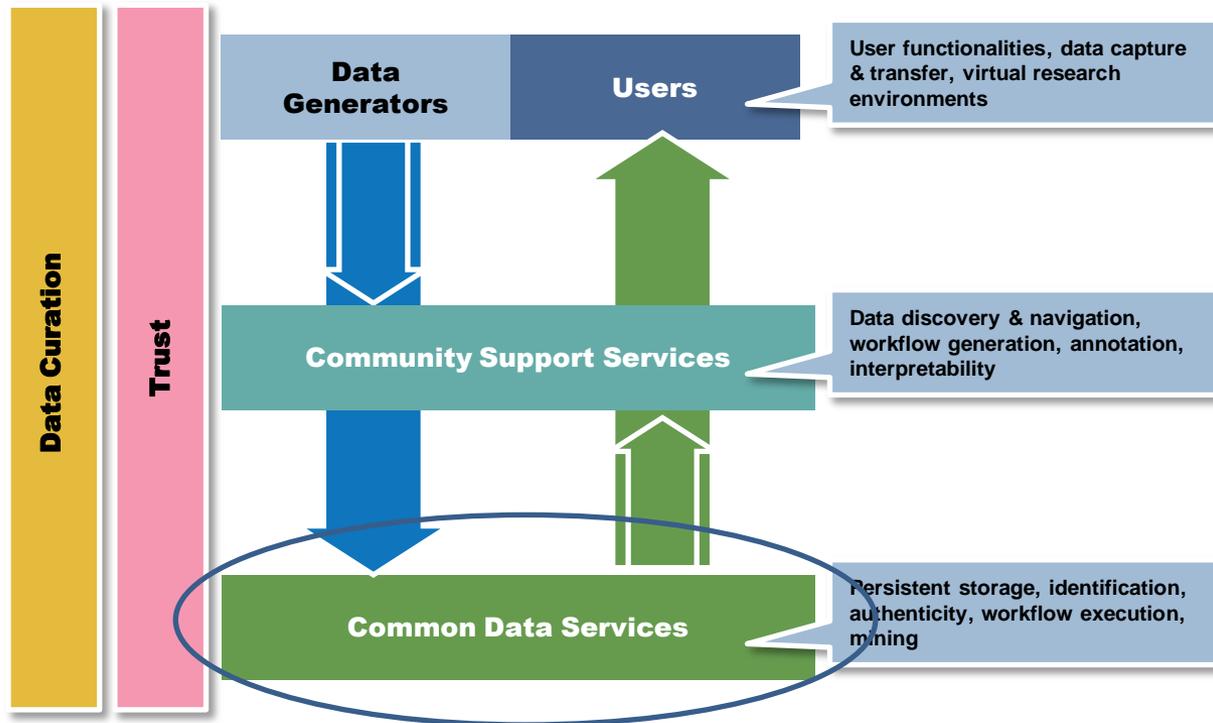
Selected Services

Metadata Catalogue
Aggregated EUDAT metadata domain.
Data inventory

AAI
Network of trust among authentication and authorization actors

PID
Identity
Integrity
Authenticity
Locations

Data Staging
Dynamic replication to HPC workspace for processing

Safe Replication
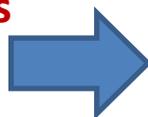Data curation and access optimization

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

New services to come

EUDAT Box
dropbox-like service
easy sharing
local synching

Semantic Anno
checking & referencing

Dynamic Data
immediate handling

EUDAT

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

EUDAT

Three priorities towards H2020

- **Stabilise EUDAT data services**
  - Increased performance, new functionalities, AAI, etc.
  - Develop tools and policies to facilitate usage: data management plans, licensing, etc.
- **Sustainability**
- **Interoperability**
  - E-Infrastructures
  - National initiatives
  - RDA

Consortium