



EUDAT

A European Collaborative Data Infrastructure

Daan Broeder

The Language Archive – MPI for Psycholinguistics

CLARIN, DASISH, EUDAT



OpenAire Interoperability Workshop
Braga, Feb. 8, 2013

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



First EUDAT Communities

ESFRI

CLARIN
Common Language Resources and Technology Infrastructure

January 20, 2011
New Virtual Language Observatory launch

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

About CLARIN

Events

Links

Glossary

Contact

Join CLARIN

Internal Web Site

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
19 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... [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 organisations 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.es

Quote
"Through our Memorandum of Understanding, 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

News

IS-ENES

The Rationale

The Aims

The MOU

Related Projects

Partners

Meetings

Sitemap

Contact

Imprint

search...

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.



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

Overview

VPH Links

Contact Us

Activities

Example Projects

Resources

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

VPH Toolkit

VPH Training

Integrative Research

Dissemination & Networking

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

Interface Focus special issue with best papers from the VPH2010 Conference

VPH NoE and the Pictos Alliance

Example Project Call 31

Join the Public Forum of the VPH-FET Support Action

Multi-Institutional Graduate Programme for Virtual Physiological Human Scientists (VPH-MIP)

VPH Vision & Strategy Paper II

VPH MAIL
5th VPH Newsletter (Jan 2011) now available

LATEST VPH EVENTS

01.06.2011 - 03.06.2011 ICCS 2011 (Toulouse)

06.06.2011 - 08.06.2011 VPHS

08.06.2011 3DAPAS Workshop

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

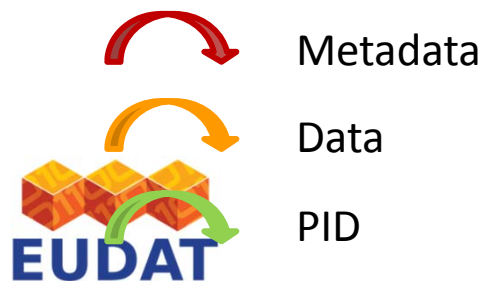
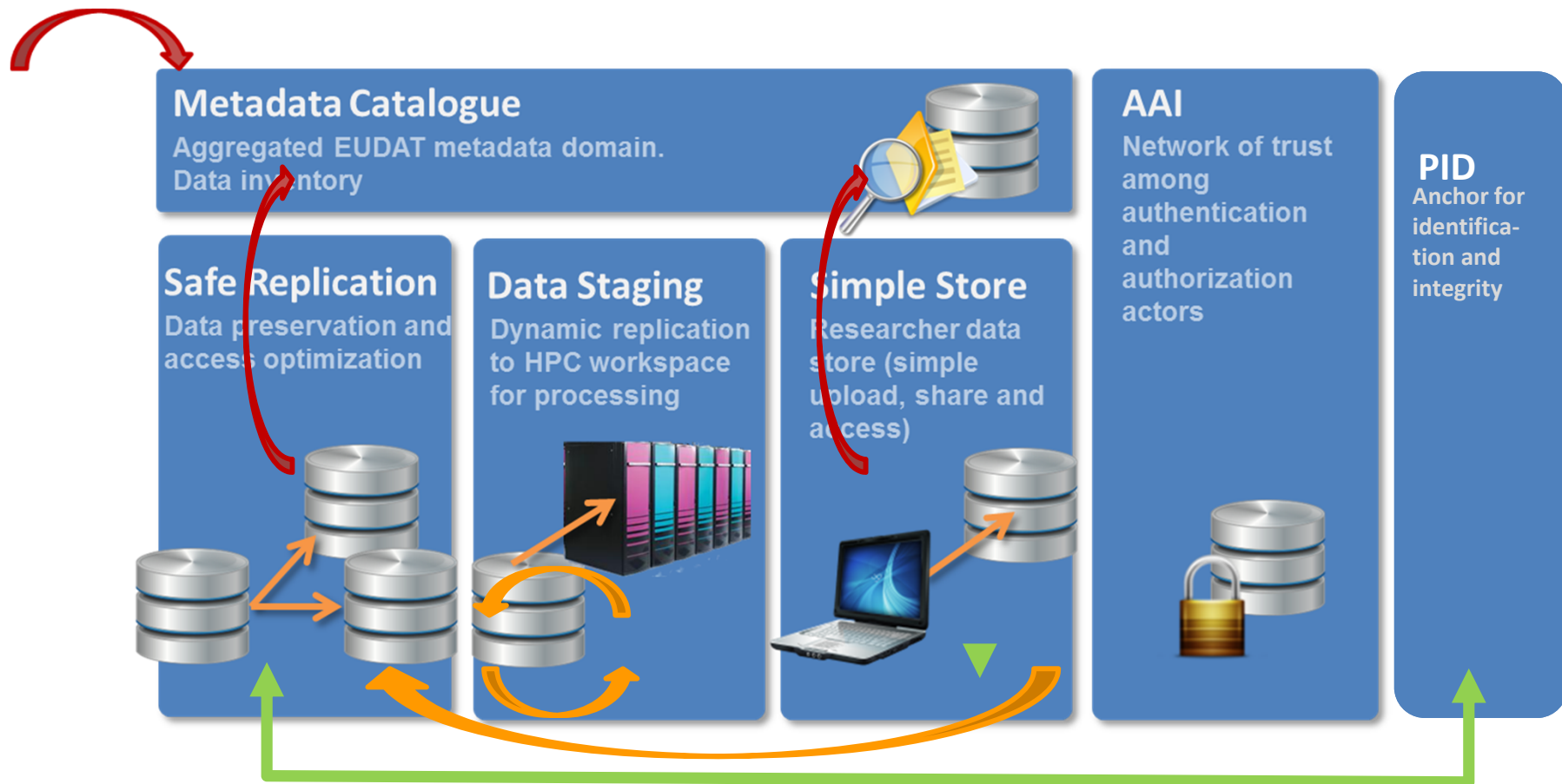


First Service Cases

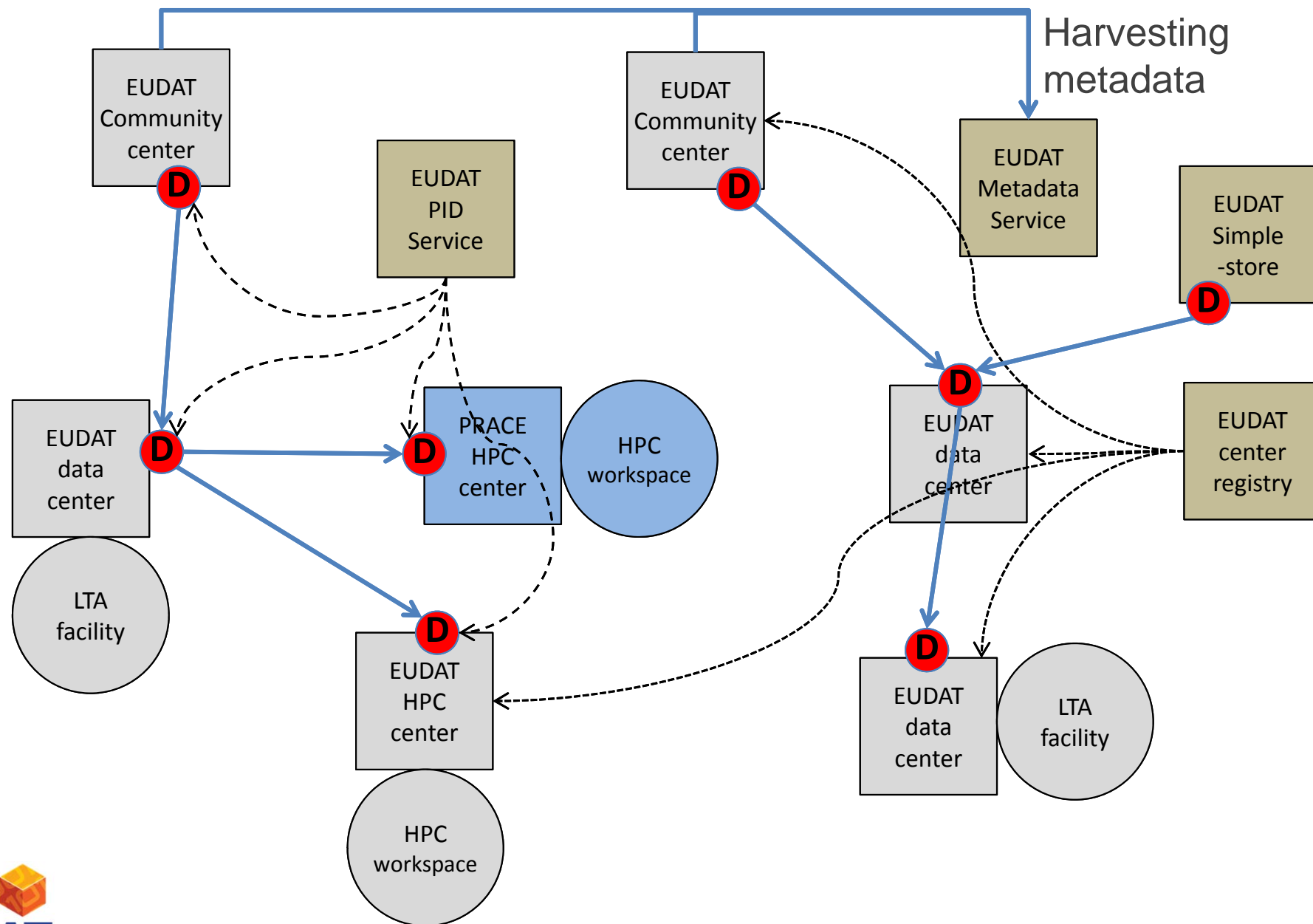
▪ List of 6 service/use cases identified

- **Safe replication:** Allow communities to safely replicate data to selected data centers for storage and do this in a robust, reliable and highly available way.
- **Dynamic replication:** Perform (HPC) computations on the replicated data. Move (part of) the safely replicated data to a workspace close to powerful machines and move the results back into the archives.
- **Metadata:** A joint metadata domain for all data that is stored by EUDAT data centers by harvesting metadata records for all data objects from the communities. Allow to have a catalogue to demonstrate what EUDAT stores, and to have a registry which can be used for automatic operations such as data mining.
- **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 system in a federation scenario.
- **PID:** a robust, highly available and effective PID system that can be used within the communities and by EUDAT.

EUDAT – Architecture



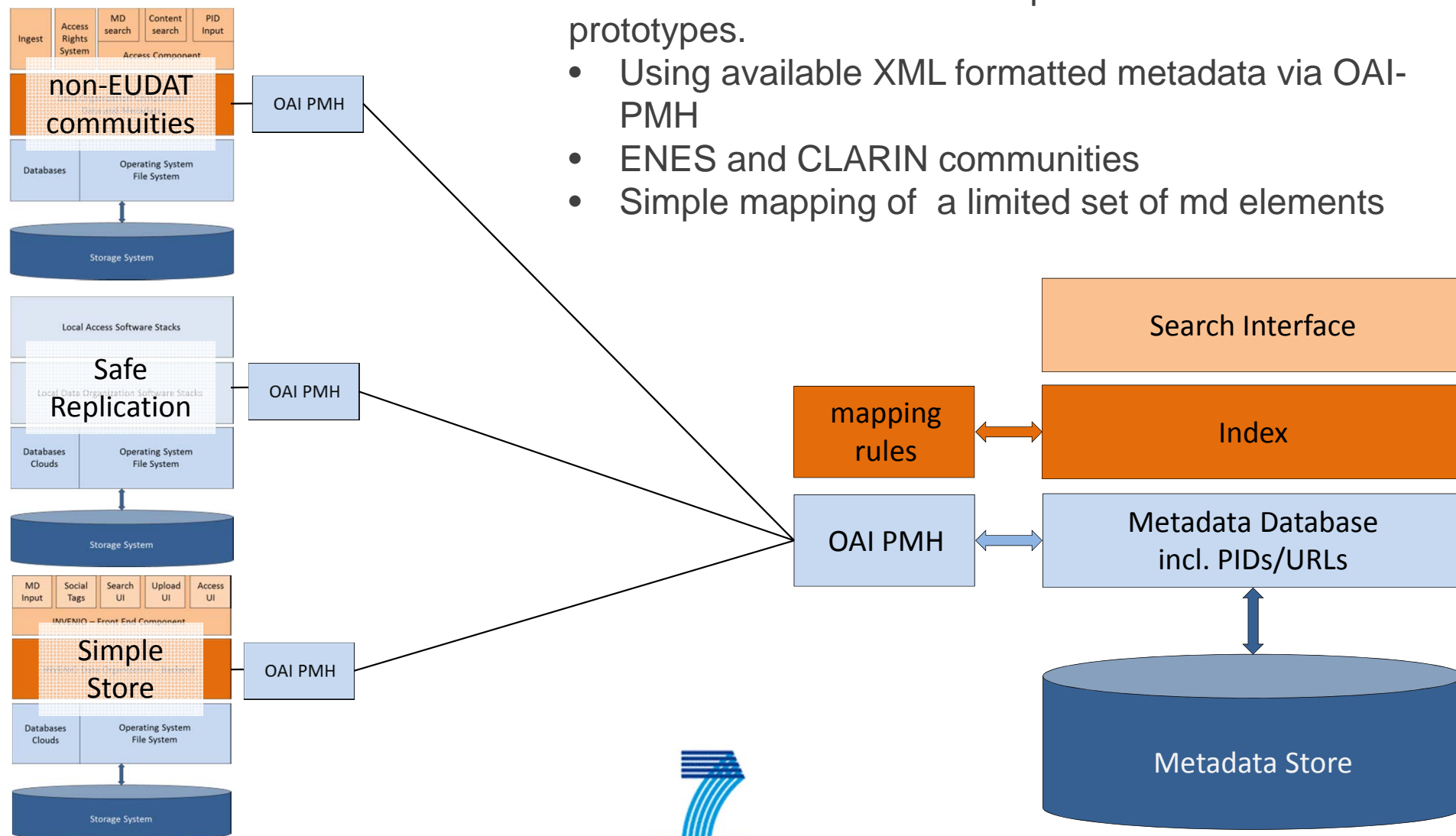
EUDAT Data & Community Centers



EUDAT – Architecture Joint Metadata Catalogue

Joint metadata will be developed as a series of prototypes.

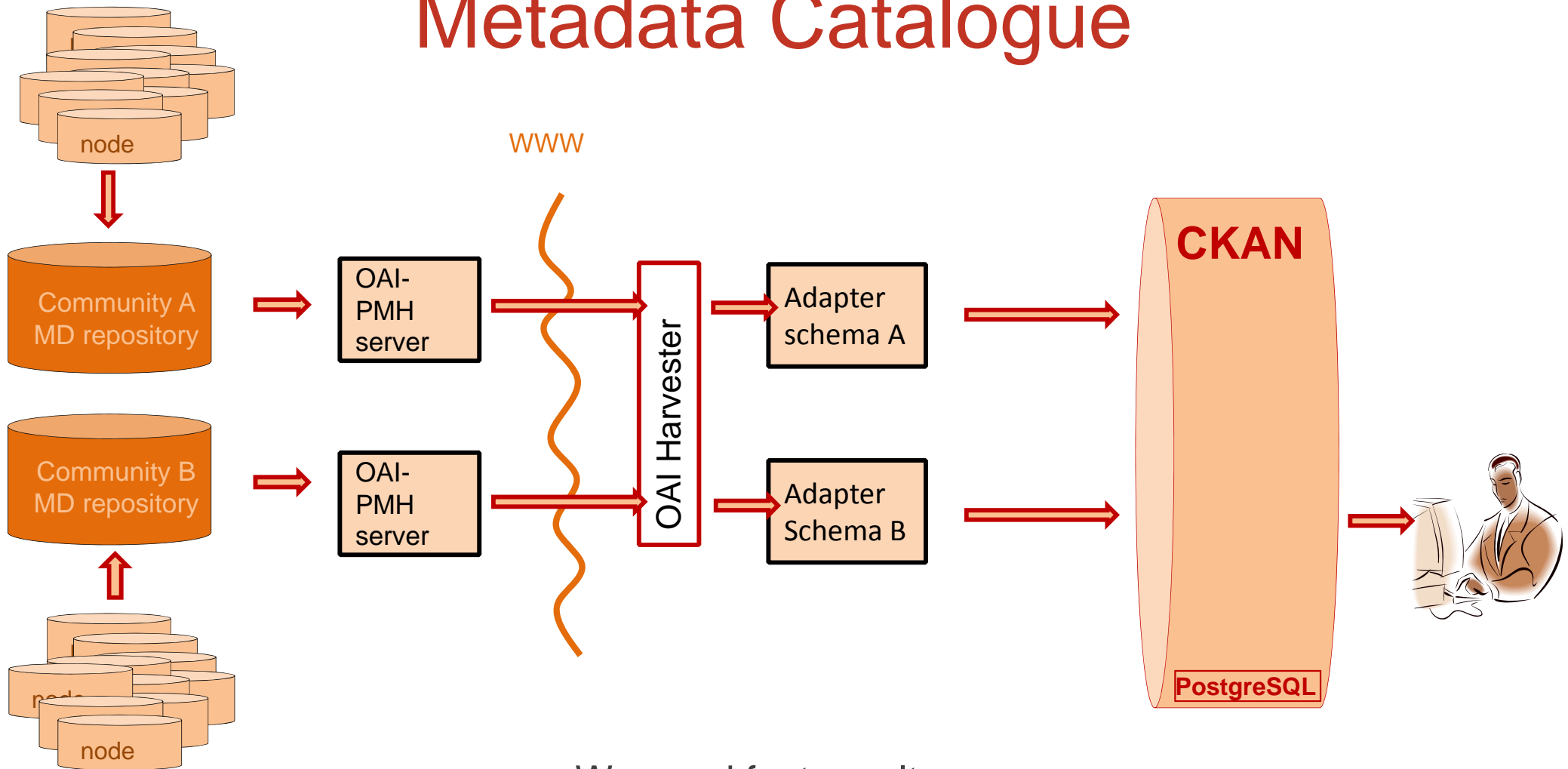
- Using available XML formatted metadata via OAI-PMH
- ENES and CLARIN communities
- Simple mapping of a limited set of md elements



User perspective joint metadata catalogue

- From the user's perspective the following points are considered important:
 - What metadata terminology can be used in the catalogue interface
 - What are the browsable/searchable dimensions
 - How is the data presented e.g. what visualization options are available
 - What added value can there be?
 - Cross community data
 - Commenting facility (also for metadata quality improvement)
 - Direct visibility of data uploaded in the EasyStore
 - Good visualization of different community data types
 - Linking to original metadata records and resource
 - Harmonized data citation facility
 - Export as LoD
- *Consider also its use for the community own purposes*
 - *Starting communities, that do not yet have a proper catalogue*

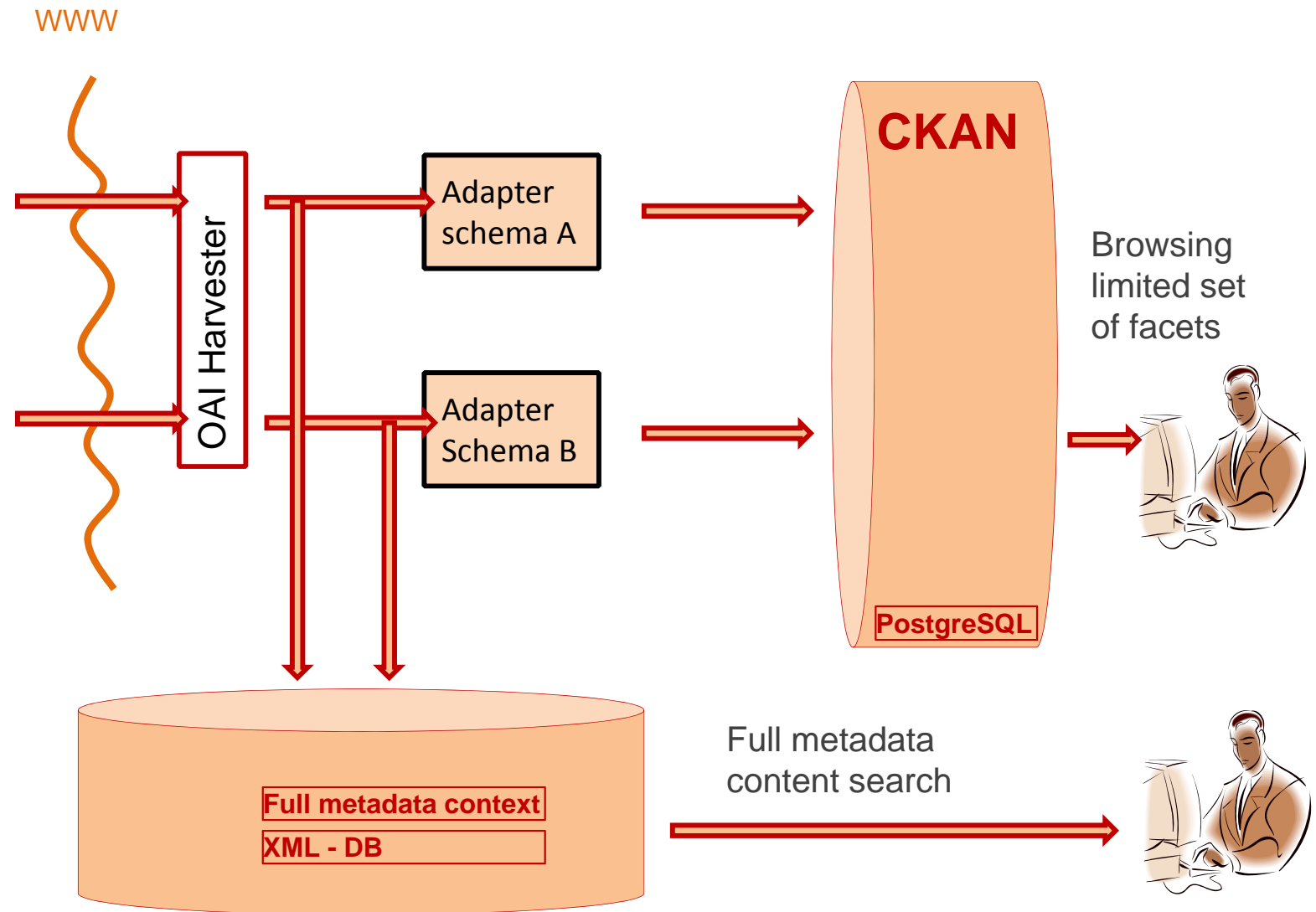
Metadata Catalogue



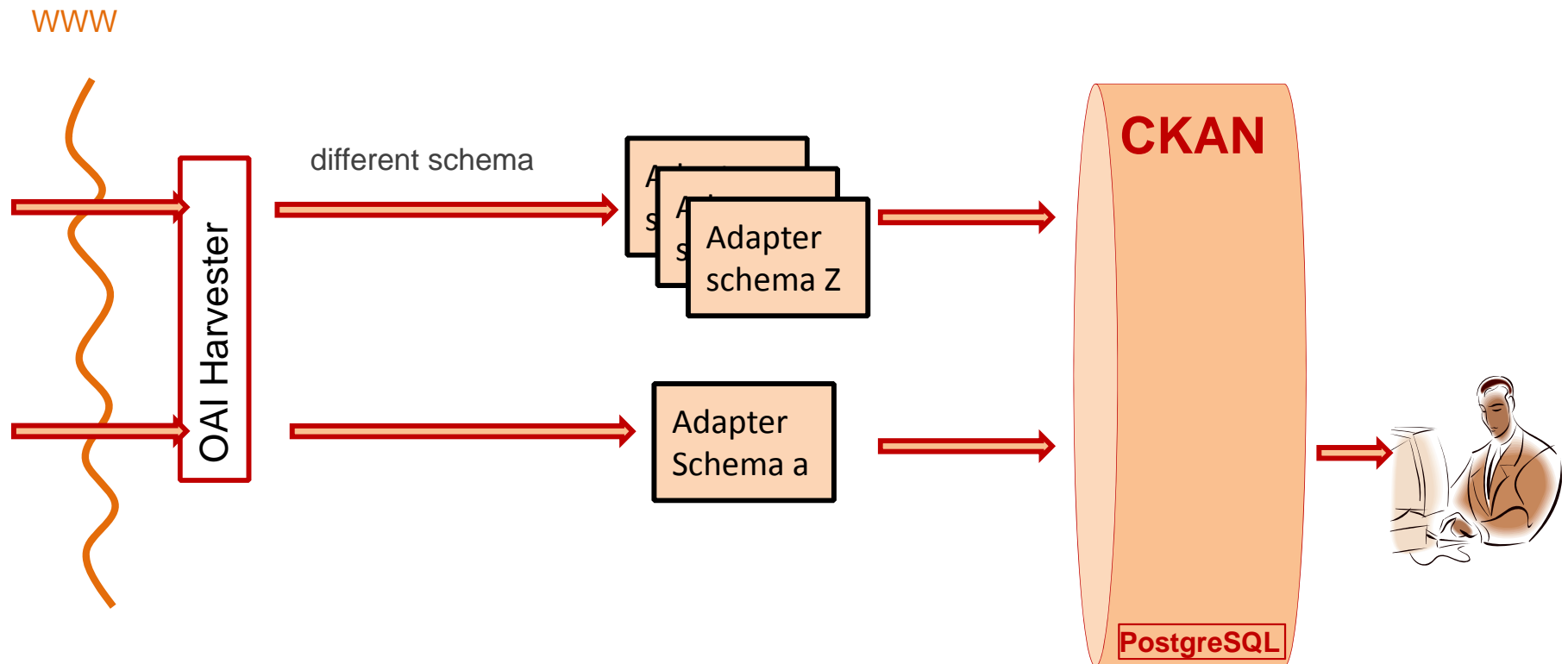
We need fast results

- Based on community practices
- No new schema developments
- Using metadata that is already out there
- Using existing and proven technology if possible

More Functionality



CLARIN particulars



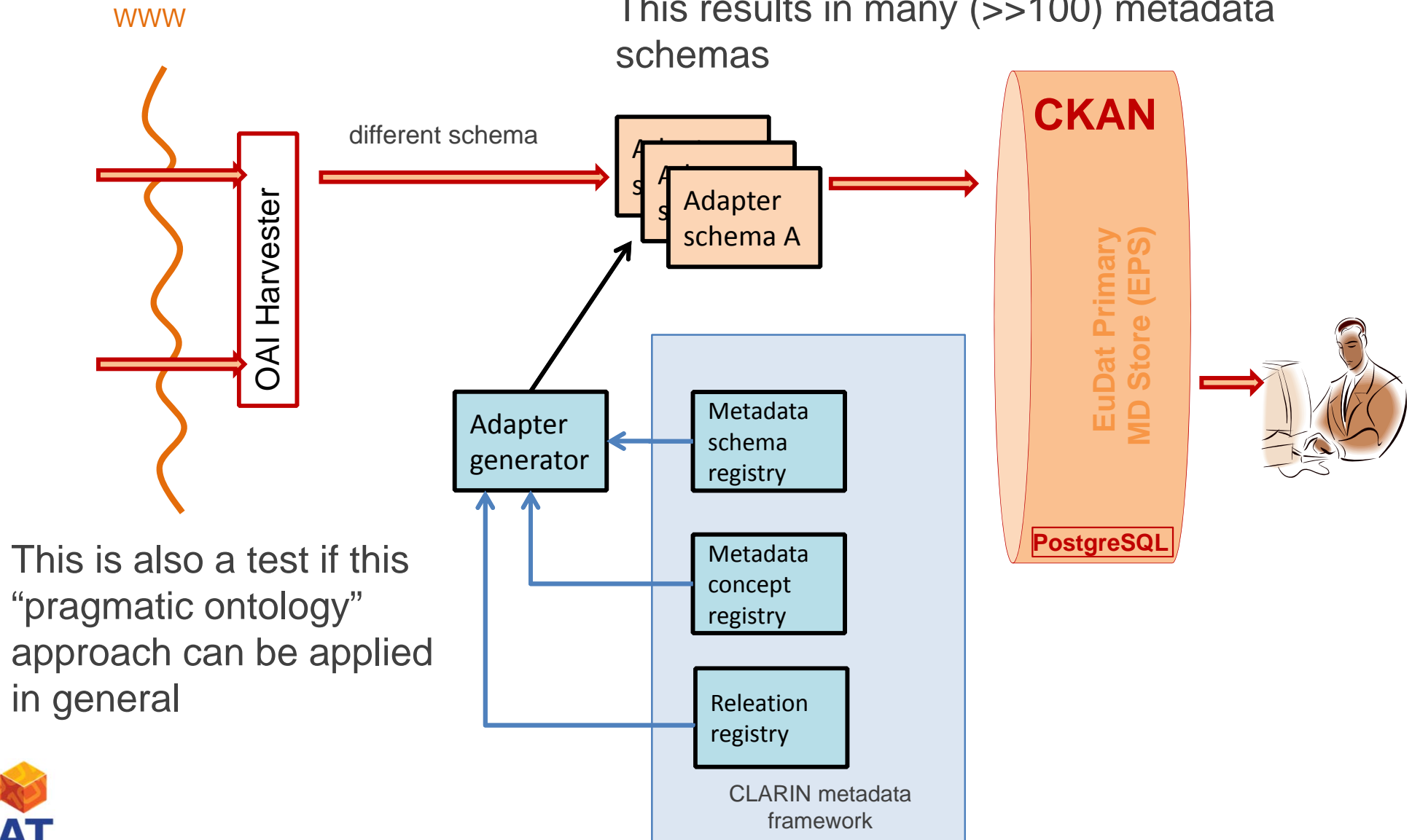
CLARIN uses CMDI as a metadata infrastructure
Users can specify metadata schema as they see
fit to describe a particular resource type.
This results in many (>>100) metadata schemas

CLARIN particulars

CLARIN uses CMDI as a metadata infrastructure

Users can specify metadata schema as they see fit to describe a particular resource type.

This results in many (>>100) metadata schemas



EUDAT & Interoperability

In this context we use a 'narrow' definition from the IEEE Glossary that defines interoperability as:

the ability of two or more systems or components to exchange information and to use the information that has been exchanged

Interoperability consequences within EUDAT between data and community centers with maybe some consequences for the community practices

- Use of iRods as a data replication software component.
 - Used within that a set of policies (general) governing the exchange of data between the centers.
- Use of FIM and/or X509 certificates to access EUDAT services
- Use of a specific PID system; with some added information items
 - EPIC, DataCite both based on Handle System technology
- Use of OAI-PMH harvesting of metadata:
 - possibly developing an EUDAT standardized use of a metadata concept and relation registry
- *Other technology choices for data-infrastructure components as CKAN as a metadata catalogue or INVENIO as CMS for the SimpleStore are not so intrusive since they are central facilities that can be replaced*



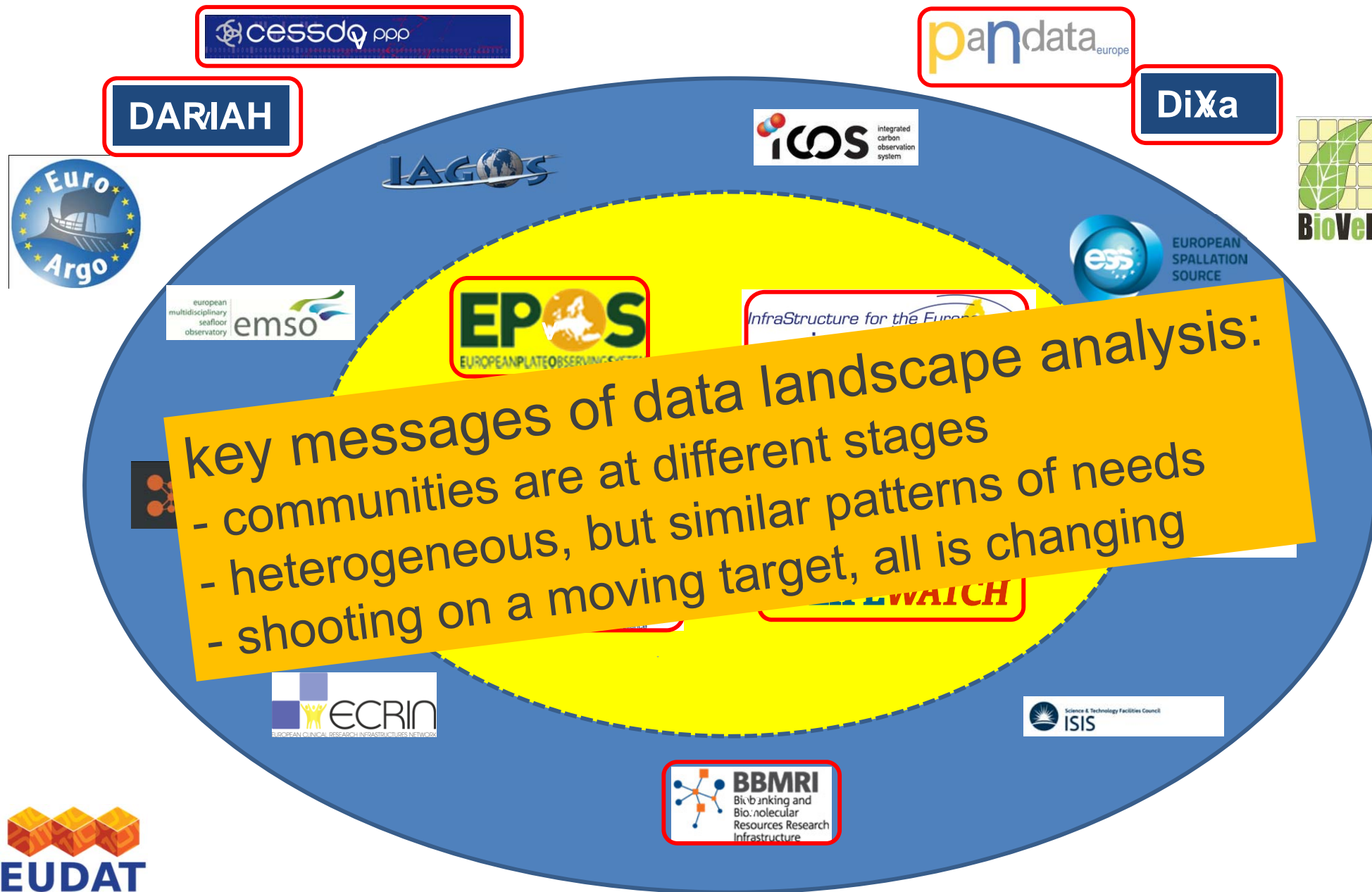
Interoperability of ideas

- There are many communities active with data management
- Developing ideas, tools & infrastructure or both
- To make ideas come together or at least get informed about differences. We need to talk!
- RDA – Research Data Alliance
 - Supported by EC, US, Aus
 - EUDAT, OpenAire via the iCordi project
 - Bottom up, community driven, (including data-scientists)

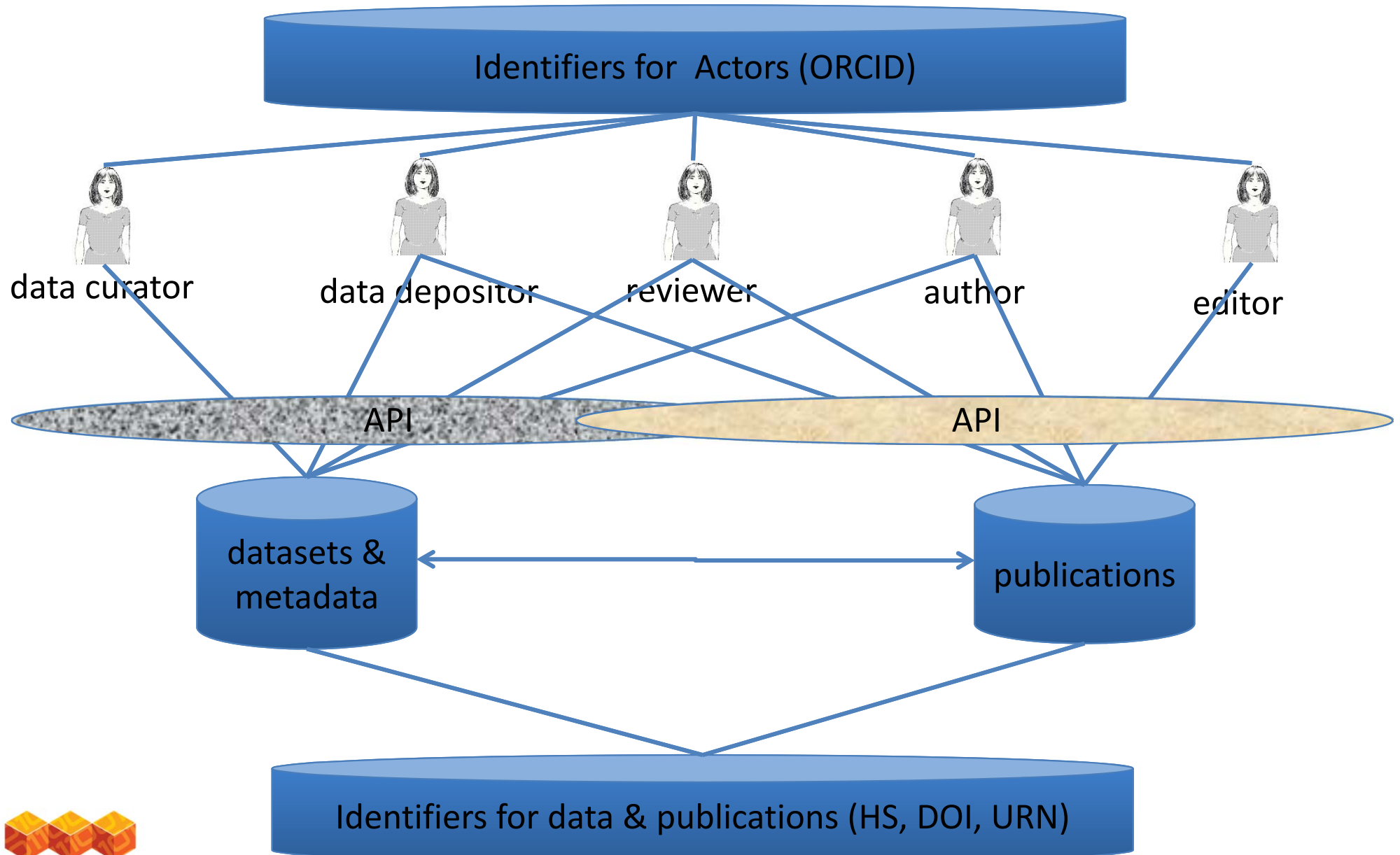


Thank you for your attention

EUDAT Services for Communities!?



Interlinking data and publications

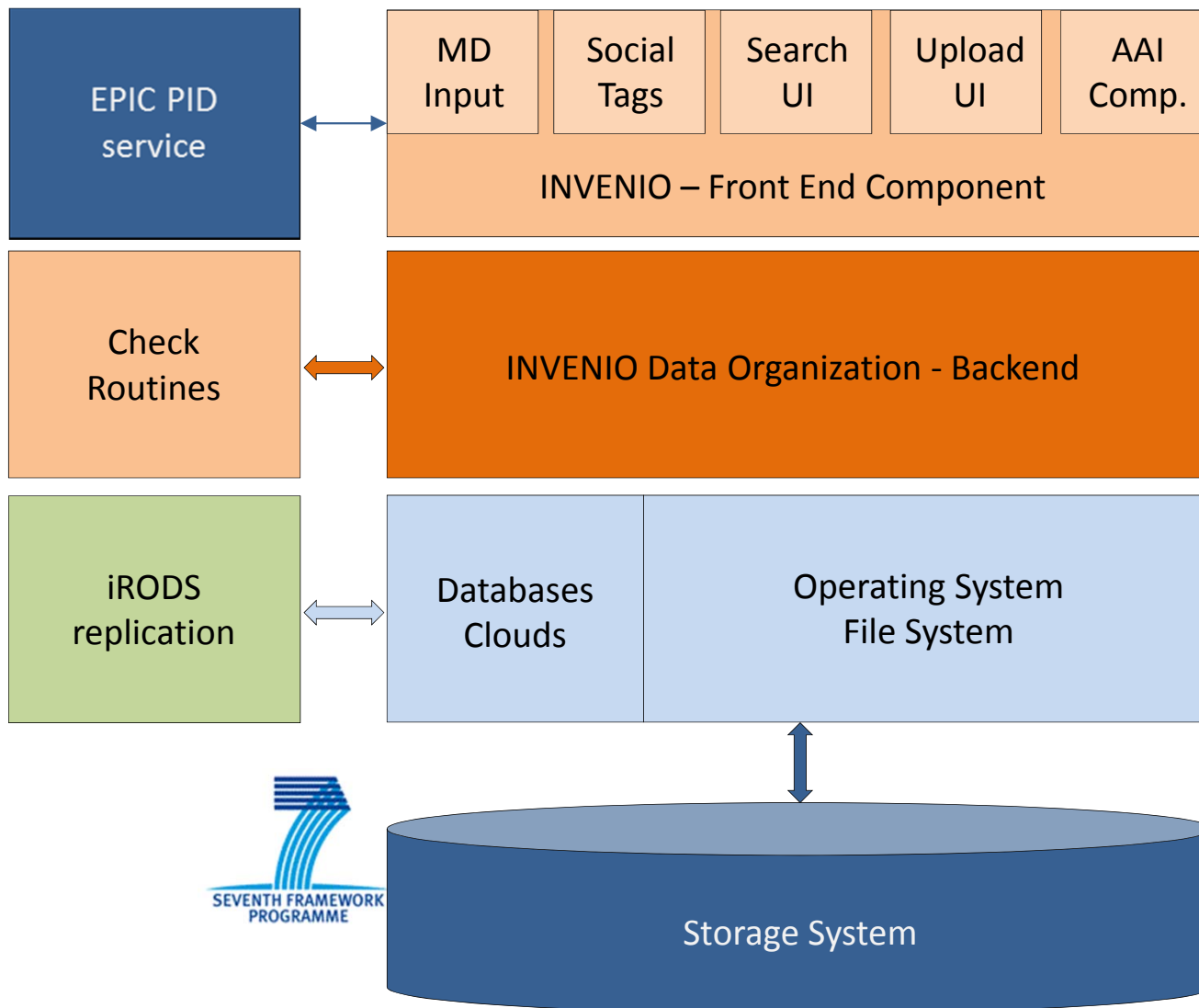


EUDAT – Architecture Simple Store (SS)

The architecture of the Simple Store service is driven by the Invenio backend machinery and an adapted and optimized front end that has all functionality as required by EUDAT. The backend system makes use of underlying standard components, i.e. all metadata, social tag data etc. is stored in a SQL database with references to files stored in a file system or cloud dependent on the kind of component the data center is using.

Important for EUDAT is that we

- will not interfere with the setup at the hosting data center
- will not be dependent on Invenio but be able to export all metadata and data to new technology if necessary
- will associate PIDs with all objects once uploaded
- will carry out checks on the uploaded objects to facilitate later curation and access (check routines can come from libraries such as JHOVE or are self-defined)
- will be able to make use of the SR service to replicate all stored data (microservices to support this need to be developed)
- AAI is supporting many users from communities registered with IdPs and beyond.





Interoperability

From Wikipedia, the free encyclopedia

Interoperability is the ability of diverse systems and organizations to work together (inter-operate). The term is often used in a technical [systems engineering](#) sense, or alternatively in a broad sense, taking into account social, political, and organizational factors that impact system to system [performance](#).

While **interoperability** was initially defined for IT systems or services and only allows for information to be exchanged (see definition below), a more generic definition could be this one:

Interoperability is a property of a product or system, whose [interfaces](#) are completely understood, to work with other products or systems, present or future, without any restricted access or implementation.

This generalized definition can then be used on any system, not only information technology system. It defines several criteria that can be used to discriminate between systems that are "really" interoperable and systems that are sold as such but are not because they don't respect one of the aforementioned criteria, namely:

- non-disclosure of one or several interfaces
- implementation or access restriction built in the product/system/service

The [IEEE](#) Glossary defines interoperability as:

the ability of two or more systems or components to [exchange information](#) and to use the information that has been exchanged.^[1]

James A. O'Brien and George M. Marakas define interoperability as:^[2]

Being able to accomplish end-user applications using different types of computer systems, operating systems, and application software, interconnected by different types of local and wide area networks.