OpenAIRE

Open Access Infrastructure for Research in Europe

The OpenAIRE*plus* Project
e-IRG Workshop, 12th – 13th of October, 2011 - Poznan, Poland

Paolo Manghi

Istituto di Scienza e Tecnologie dell'Informazione
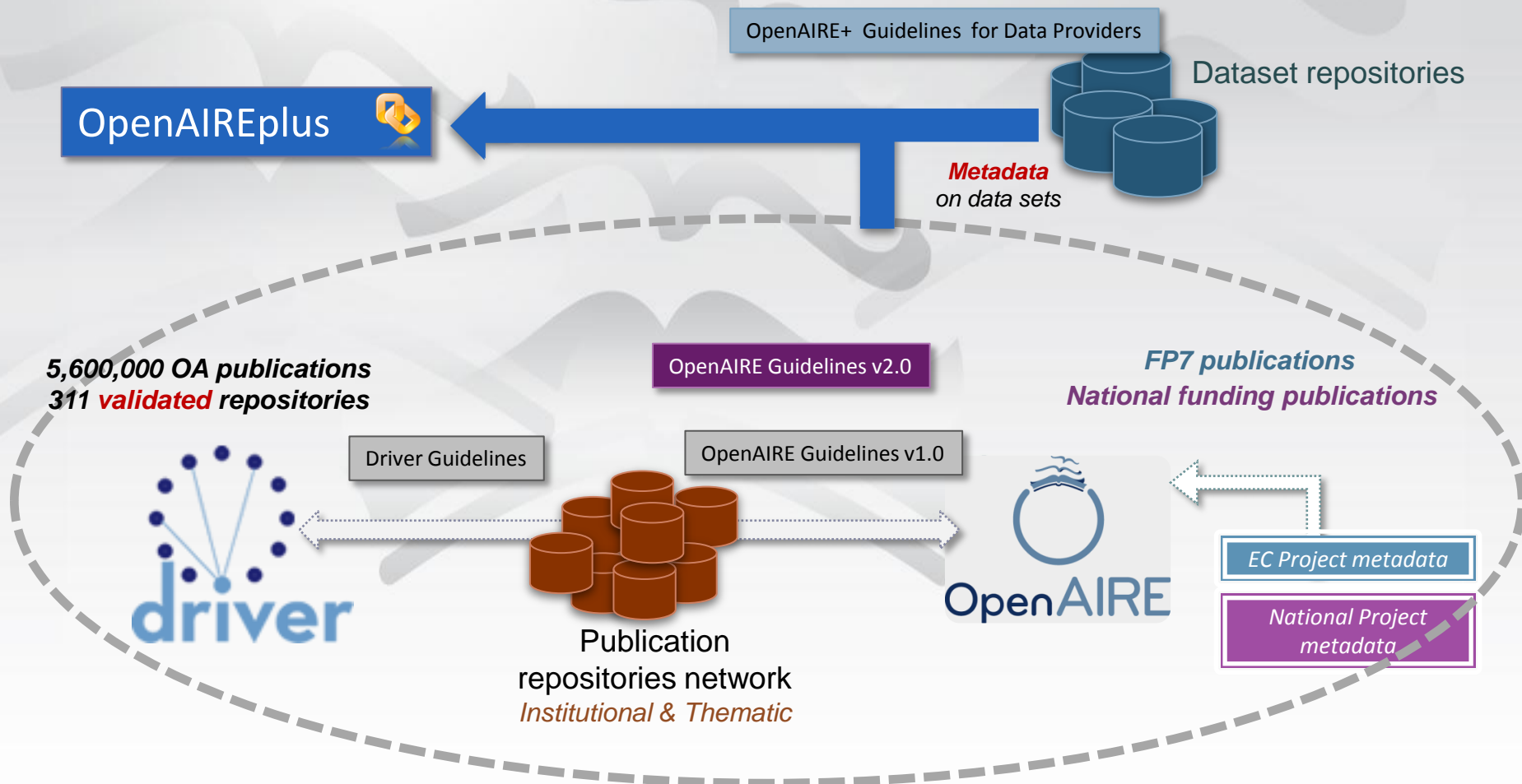*Consiglio Nazionale delle Ricerche, Pisa, Italy*

OpenAIREplus

- Develops an **Open Access**, participatory infrastructure for scientific information that includes **publications**, **datasets**, **funding** and **interlinking** between them

  - Cooperates with scientific communities to investigate and experiment with the concept of linking publications and associated datasets

- **Disseminates** OA for publications, datasets and interlinking retaining a **European footprint**

  - Broad network of people and repositories in all 27 EU member states plus Norway, Switzerland, Croatia, Turkey and Iceland

- Capitalizes on the **OpenAIRE infrastructure,** built for connecting FP7 and ERC funded articles with relative projects (measuring the impact of EC SC39)

  - Expands OpenAIRE base of publications including all OA publications indexed by the **DRIVER infrastructure**

  - Expands **OpenAIRE networks of repository** reaching out, thematic repositories and dataset repositories

- **Opens its content** through the OpenAIREplus portal (search, browse, stats) and to third-party applications or service providers through standard protocols

Past, present and OpenAIREplus

OpenAIRE+ Guidelines for Data Providers

Dataset repositories

OpenAIREplus

Metadata
on data sets

5,600,000 OA publications
311 *validated* repositories

FP7 publications
National funding publications

OpenAIRE Guidelines v2.0

Driver Guidelines

OpenAIRE Guidelines v1.0

EC Project metadata

National Project metadata

Publication
repositories network
Institutional & Thematic

High-level objects
Networking activities

- Bringing together representatives from **all 27 European member states** plus Norway, Switzerland, Croatia, Iceland and Turkey

- Aligning national and European **scholarly publication and data infrastructural initiatives** on topics related to Open Access

- Studying **IPR issues** and their specific applications on publications and research data

- Identifying **feasibility/financing OA models** that will allow the envisioned expansion of Open Access to all European funded rogrammes

High-level objectives
Joint research activities

- Design of **OpenAIREplus data model**

  - Representation of interlinked research datasets, publications and funding entities

- **Design of content management services** for the OpenAIREplus infrastructure

  - Functionalities for collecting publications, datasets, and projects and for curating, discovering and visualizing links between them

- Delivering the **OpenAIREplus Orphan repository**

  - By extending the OpenAIRE *Orphan Repository*

  - Deposition and Web publishing for publication/datasets to authors who do not have a repository of reference

High-level objectives
Service activities

- Delivering a **24x7 production quality infrastructure**

- Continuous **population and curation** of the OpenAIREplus European Research Information Space

  - Collection of content from EC CORDA database, OpenDOAR, publication repositories, OpenAIREplus Orphan Repository, and dataset repositories

- Support to **software life-cycle** for the services designed in JRA

- **Helpdesk operation**

  - Supporting researchers over the 30 partner European Countries towards meeting the demands of the EC Open Access mandate

  - Drafting and promoting **OpenAIREplus guidelines**: dataset exports (metadata and protocols) and publication-dataset linking process and exports

Networking activities
National Open Access Desks

OpenAIRE*plus* Networking Organization			

Operational Support (eIFL)	Strategic Coordination (UGOE)				Helpdesk Management (UMINHO)

Region 1 North (DTU)	Region 2 South (UMINHO)	Region 3 East (eIFL)	Region 4 West (UGENT)
Denmark	Cyprus	Bulgaria	Austria
Finland	Greece	Croatia	Belgium
Iceland	Italy	Czech Republic	France
Norway	Malta	Estonia	Germany
Sweden	Portugal	Hungary	Ireland
	Spain	Latvia	Luxemburg
	Turkey	Lithuania	Netherlands
		Poland	Switzerland
		Romania	UK
		Slovakia	
		Slovenia	

University of Goettingen
Technical University of Denmark
University of Minho
Stichting eIFL.net
University of Ghent

Human Network

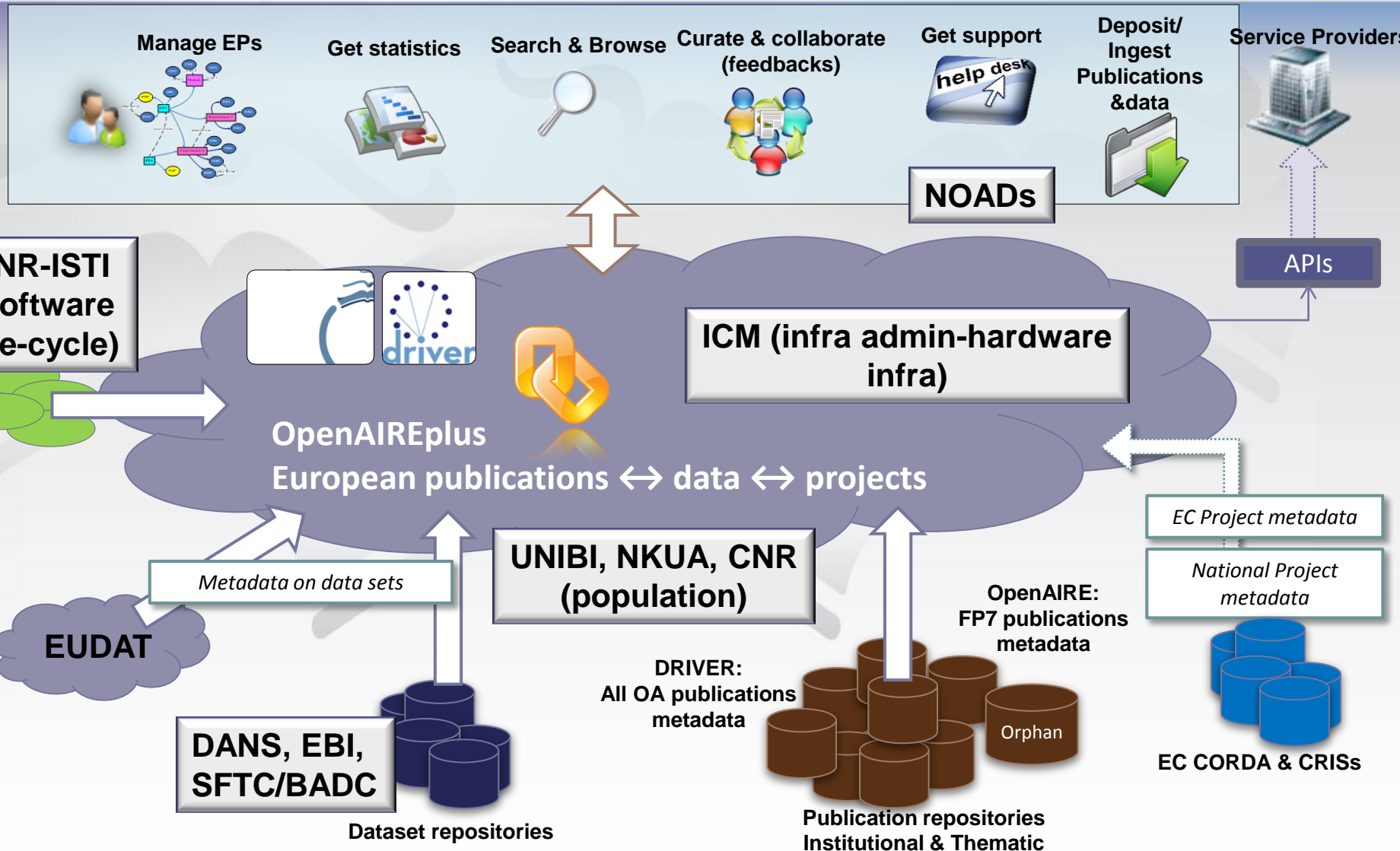# Networking activities
Main objectives

- **Top-down strategy**: dissemination, liaison activities, community building (NOADs)

  - Engaging all stakeholders in the OA dissemination, linking and usage of publications and related datasets

- **Bottom-up strategy**: aligning with national, European and international initiatives

  - e.g., DataCite, ESFRI-infrastructure projects (such as CLARIN, DARIAH) EuroCRIS (CERIF), EUDAT, Europeana …

  - Organizations of four national and international workshops

- **Subject-specific pilots** for enhanced publications

  > **EMBL-EBI, DANS, BADC, CNR-ISTI, NKUA, ICM**

  - Linking datasets and publications in Life Sciences and Social Sciences

  - Resource exchange between generic infrastructures (OpenAIRE) and subject specific infrastructures (UK PubMedCentral and DANS archives

- Study on **licensing of publications and research data**

- Study on **sustainability and business models for (Open Access) infrastructures**

Service activities
Users

- **Registered end-users** (e.g., EC personnel, project coordinators, researchers, authors)
  - Search, browse and access statistics over the information space
  - Deposit files and metadata of publications and datasets into the Orphan Repository
  - Ingest (claim) into the information space metadata relative to publications and datasets
  - Search, visualize and create enhanced publications out of the information space content
  - Send "content feedbacks" to OpenAIREPlus data curators

- **Content provider** managers (e.g. datasets and publications repository managers)
  - Registration and validation (OpenAIREPlus guidelines) of publication and dataset repositories

- **Data curators** (administrative tasks)
  - Collect and aggregate publications, project data and dataset metadata
  - Refine the aggregated data: authority file control, citation identification, end-users feedbacks

- **Third-party application developers**
  - Bulk-fetch content from the (curated) information space

Service activities
The OpenAIREplus's Guidelines for Content Providers

- For publication providers - OpenAIRE's Guidelines for repository managers

  - Metadata: DC + project encoding + license encoding

  - Protocols: OAI access protocols (and more)

- For data providers: accessing (metadata of) datasets from providers while minimizing effort to comply

  - *Metadata* (synergy with EBI-EMBL, DANS, STFC/BADC): indications on **minimal metadata** about datasets (e.g., identifiers, date of creations, title, URLs) and best-practices for interlinking datasets and publications

  - *Access protocols:* no requirements for adopting precise protocols (e.g., OAI, FTP) or ID/URL frameworks (e.g., OpenURL, DOI) to comply
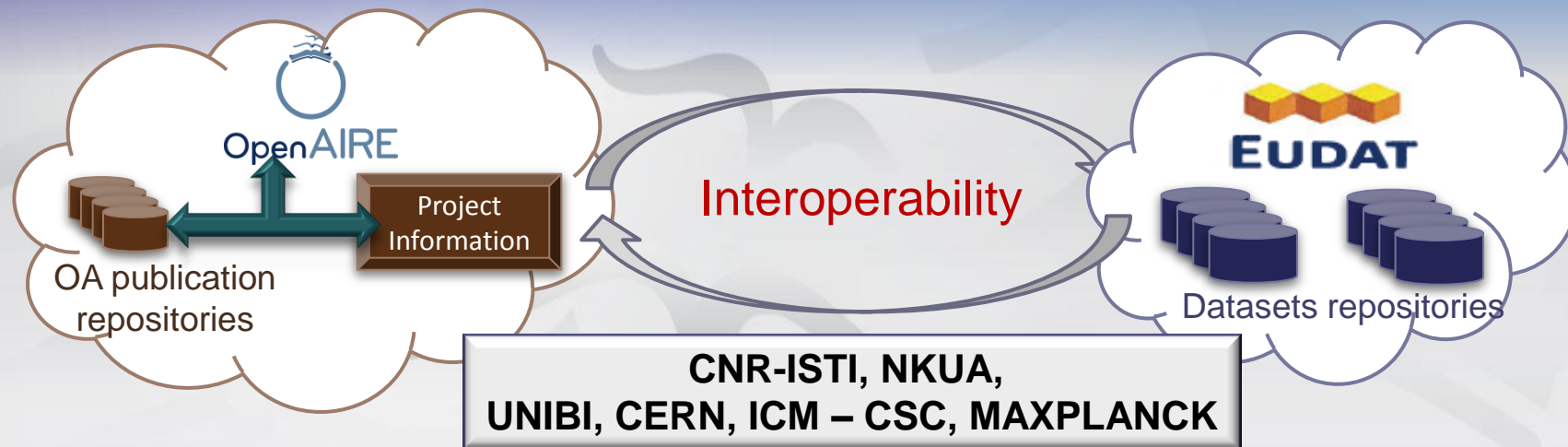
validated 1.1

validated driver

DTU, CERN, UNIBI, UNIMINHO

Service activities
Data Access and Interoperability Task Force (DAITF)

OpenAIRE

OA publication repositories

Project Information

Interoperability

EUDAT

Datasets repositories

**CNR-ISTI, NKUA,
UNIBI, CERN, ICM – CSC, MAXPLANCK**

- Compare requirements from the publications world with those from the datasets world and computing-centres world

- High-level strategic meetings on co-operation to design clear action points on

  – Standards and guidelines for data infra interoperability

  – Community outreach activities: workgroups

Service activities
Infrastructure enabling technology

- **D-NET Software toolkit**

    – **Service-oriented data infrastructure enabling technology**, support for developers implementing data infrastructures for data sources aggregation

    – Openness (to new functionality and data formats), autonomicity, sharing and reuse (economy of scale), standard-based

    – Adoption: DRIVER infra, OpenAIRE infra, EFG infra, HOPE infra, Spain-Recolecta

    – By: CNR-ISTI (Italy), ICM (Poland), NKUA (Greece), UNIBI (Germany)

- **INVENIO Repository**

    – **Customizable repository platform**: workflows and data models

    – Adoption: CERN digital library and 30+ institutions world-wide

    – By: CERN (Switzerland), collaboration from DESY, EPFL, FNAL, SLAC

Research Joint activities
Main objectives

- Data model definition

    - Entities: projects, publications, datasets, relationships between such entities and enhanced publications

- D-NET services for content management/processing

    - Store (according to data model), populate, search, browse, calculate statistics, export through standard APIs, and curate quality and richness of data

    - Data-export components on the dataset providers side

    - OpenAIREPlus Orphan Repository (Invenio repository platform

NKUA, CNR-ISTI, UNIBI, ICM, CERN

Research Joint activities
Main objectives

- D-NET services for content analysis and text mining

    - Relationships extraction (e.g. citations, similarity, funding), similarity search, content classification, and enhanced authority file management

- D-NET services for end-users and service providers access

    - UIs required to search, browse, access stats, deposit, curate data by sending feedbacks, and construct enhanced publications

    - User patterns-driven semantics extraction

    - Access to content via OpenSearch, SRW/CQL or harvesting via OAI-PMH, OAI-ORE

NKUA, CNR-ISTI, UNIBI, ICM, CERN

Partners and roles: factsheet

- Coordination
  - University of Athens - GR
  - Goettingen University Library - DE
  - CNR-ISTI - IT

- Technical production & operation
  - 5 partners with expertise in technologies for Digital Libraries and Data Infrastructures

- General
  - Starting date: Dec 1, 2011
  - Duration: 30 months
  - Total budget: 5.2 Mi

- Communities
  - EBI – biology
  - DANS – social sciences
  - STFC/BADC – climate

- Networking Organization
  - 5 libraries, active in OA movement

- National Open Access Desks
  - All member states
  - Norway, Switzerland, Turkey, Iceland

Questions?

For more:

OpenAIRE infrastructure: http://www.openaire.eu

DRIVER infrastructure: http://www.driver-community.eu

D-NET Software toolkit: http://www.d-net.research-infrastructures.eu
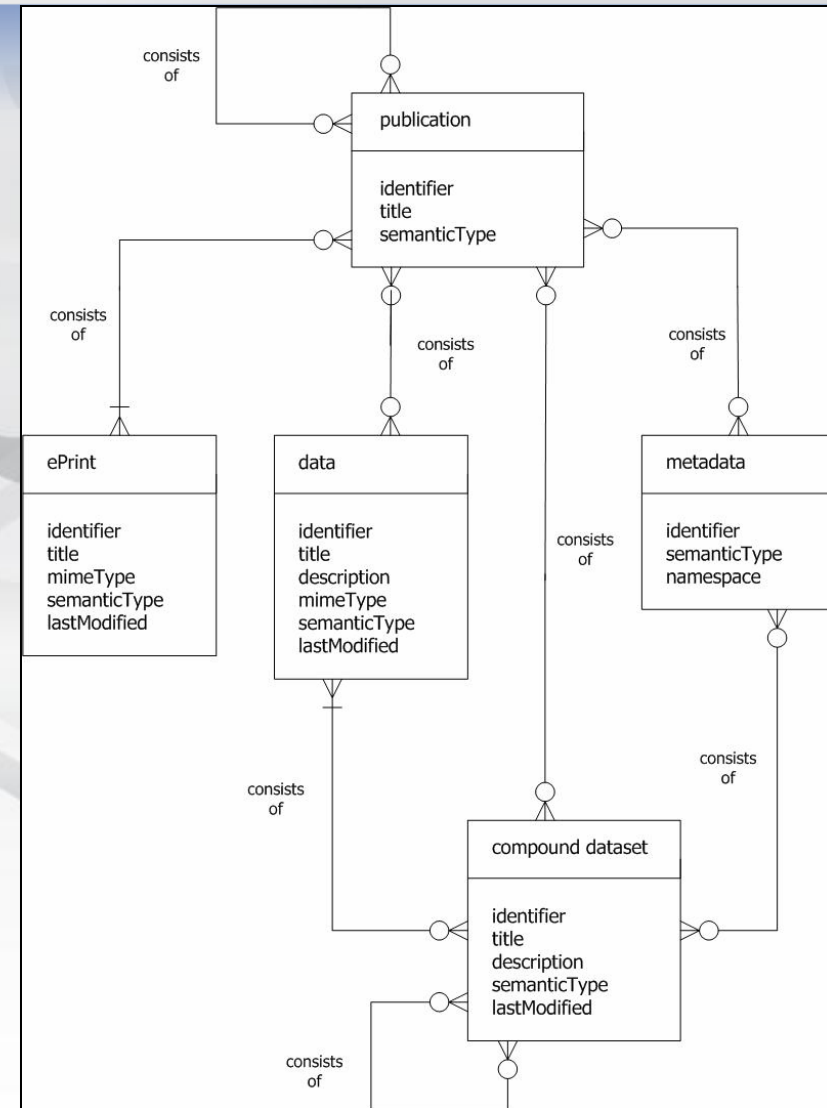
INVENIO Repository: http://invenio-software.org

Enhanced Publications (EPs)

Compound information objects: information objects that represent the aggregation of distinct information objects through meaningful relationships

Example of SURF-EPs: textual publications enhanced with links to datasets (e.g., e-learning material, queries, db entries)

OpenAIREplus refines EP data model and provides EP services (repository) for:

- ○ Management: creation and curation

- ○ Visualization, browsing, querying

- ○ Import: OAI-PMH/ORE harvesting of EPs from external providers, others

- ○ Export: OAI-PMH/ORE publishing of EPs, Linked Data representation, others

High-level objectives
Networking activities

- **Consolidation of the network** of European Open Access institutional repositories (OpenAIRE NOADs)

  - Expansion to dataset providers and OAJs from domain specific scientific areas

  - Reaching out researchers, research managers, policy makers

  - For publications and datasets and their linking

- **Liaisons** towards exchanging resources with infrastructures and research content systems

  - e.g., D4Science, UKPubMed, DANS, other repositories, data archives, CRISs, DataCite

- Study on "**licensing of publications and research data**" in the European research area and beyond

- OpenAIREplus infrastructure sustainability and exploitation plan

High-level objectives
Service activities

- Open Access infrastructure to the scientific production of the European Research Area

  – 24x7 service portal, where scientists can deposit, link, search, get stats over publications and datasets

- Services for cleansing and enrichment of harvested and linked metadata

  – Semi-automatic and manual identification of projects-publications-datasets links and de-duplication

- Services for exposing scientific information resources

  – To third-party applications, providers of value-added services, communities, sections of society

this is a demo slide presentation to show you all the layouts

High-level objectives
Joint research activities

- Integrating data archives and interlinking publications and datasets (enhanced publications):

  - Working with scientific communities

  - Feedback to technical and networking infrastructures

- Expert feasibility study on OA publishing models on the European level for both publications and datasets

High-level objectives
Service activities

- Helpdesk – National Open Access Desks

  - Open Access policies and IPR issues for publications and datasets

  - Deposition: when – how - where

  - Institutional, thematic, data repository issues

  - Open Access sustainable business models

- NOADs repository network

  - Today: 311 validated Open Access repositories

  - Guidelines

- Liaison and international outreach

  - Four workshops on Open Access related issues

  - Liaise with COAR, LIBER, SPARC Europe, …

this is a demo slide presentation to show you all the layouts