Open Access e-Infrastructure for Research in Europe

Donatella Castelli
CNR-ISTI
Pisa (Italy)
Outline

- e-Infrastructure: some preliminaries
- OpenAIRE e-Infrastructure (and related products): past, present and future
Infrastructure is a fundamentally **relational concept**, becoming real infrastructure in relation to organised practices

Infrastructure **is sunk into and inside** of other structures, social arrangements, and technologies

*S. Leigh Star & K. Ruhleder (1996) «Steps towards an ecology of infrastructure: Design and access for large information space»*
Platform operated by a **responsible entity** offering an open set of **basic enabling facilities** (including access to resources) to a **Community of Practice**. By sharing these facilities the members of the Community of Practice realise **economies of scale**.
Leveraging on existing platforms
An ecosystem of e-Infrastructures

**Interaction**
- Each e-Infrastructure can outsource required services to other e-Infrastructures
- The same e-infrastructure can play both provider and consumer roles

**Competition**
- The most effective and sustainable e-Infrastructures will survive
e-Infrastructure dimensions

- Responsible organization
- Software system
- Data resources
- Governing policies
- Other factors (e.g. logical and physical architecture)
Our focus

Open Access e-Infrastructure for Research

DRIVER

DRIVER II

OpenAIRE

OpenAIREPlus
Simplify retrieval and access of OA scientific documents

Basic components:
- search, browse
- repository registration & validation
- create profile and communities
- user interface

5.160.000 documents in 331 repositories from 45 countries

CNR, NKUA, UniBi, ICM

24x7 since 2007

Service sites hosted at CNR, content storage and curation in local repositories

DRIVER Guidelines 2.0
OpenAIRE Pilot

Support the OA mandate for peer-reviewed publications across Europe and monitor its impact

2009-2012

COAR General Assembly, 21-22 May, 2012, Uppsala
OpenAIRE in numbers

- 26,030 FP7 publications from 5000 projects
- ~10,300 Open Access
- 94 OpenAIRE compliant OA Repositories & Journals
Support the OA mandate for research outputs, monitor its impact and facilitate the exploitation of these outputs
OpenAIREPlus services

- Publication repositories
  - Institutional & Thematic Open Access Journals
- Data repositories
- Deposit/Ingest Publications & data
- Visualize - Manage Enhanced Publications
- Curate & collaborate
- Search & Browse
- Linked Content Statistics
- Research impact
  - Citations, usage statistics
- Get support (NOADs)
- Linked Content

5,200,000 OA publications
331 validated repositories

Orphan

Metadata on data sets

Usage data

Text Mine

Link

Classify

De-duplicate

Cite

APIs

ResearchID (ORCID, ..)
OpenDOAR
...

EC funding

National funding

CRIS Systems
For researchers

- Linked data through "Enhanced publications"
  - The new generation publications
  - Representation, end user creation and visualization

- Repository for "homeless" authors
  - Publications & data

- End-user data (metadata) curation
  - Send "content feedbacks" to OpenAIRE data curators
Alerts & notifications

Delivery of information to users and services

- Subscribe to topics of interest
- Receive notifications through various means
For research administrators

*Research and funding evaluation*

- **Statistics on research output**
  - Volume, trends for
    - project/institution/funder/country/scientific area/...
  - Co-authorship graphs

- **Statistics on research impact**
  - Citation analysis
  - Usage statistics aggregation
  - Research metrics and impact applied to
    - project/organization/country/scientific area/...
Text mining of various sources to help them in OA compliance

Complete lists of project publications
- Use in dissemination – dynamic lists retrieved in XML, HTML format
- Use in EC reporting
Backend services for automated object interlinking

- Authority file management
  - Identification and semi-automated de-duplication of entities: articles, data, organizations, authors, ...

- Automated content classification
  - Text/data mining techniques for publication or data clustering

- Advanced citation and similarity analysis
  - Citation discovery in publications & data
  - Author identification tools

- Usage pattern analysis tools
  - Content linking from implicit user actions
For data providers

Guidelines and tools

- Tools for data source validation and registration
- Guidelines for publication providers
- Guidelines for data providers - accessing (metadata of) datasets from providers while minimizing effort to comply
  - Metadata (synergy with DataCite, EuroCRIS, 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
Other D-Net enabled e-infra

- Under construction:
  - Poland
  - Spain
  - Argentina

- Operational:
  - European Film Gateway (EFG)
  - Heritage of the People’s Europe (HOPE)
www.driver-repository.eu/
www.openaire.eu/
www.europeanfilmgateway.eu/
www.peoplesheritage.eu/