REALIZING A SCALABLE AND HISTORY-AWARE LITERATURE BROKER SERVICE FOR OPENAIRE

Claudio Atzori  
Alessia Bardi  
**Sandro La Bruzzo**  
Paolo Manghi

claudio.atzori@isti.cnr.it
alessia.bardi@isti.cnr.it
sandro.labruzzo@isti.cnr.it
paolo.manghi@isti.cnr.it

ISTI CNR PISA
INTRODUCTION

- OpenAIRE
  - data model
  - OpenAIRE infrastructure architecture

- Literature Broker Service
  - Concept
  - Requirements
    - Subscriptions
    - Notifications
  - Data Model
  - Architecture
  - Conclusion
The point of reference for Open Access in Europe

- Establishment of an interoperable network of publication repositories
- Deposition, discovery, linking and monitoring of research products (articles, datasets, software) produced under National and EC funding
- Monitoring the compliance to EC OA mandates for publications
- Support decision makers with statistics
OPENAIRE DATA MODEL

- Datasource
- Funder
- Funding Stream
- Instance
- Project
- Result
- Organisation
- Person
- Publication
- Dataset
OPENAIRE INFORMATION SPACE: NUMBERS (JAN 2015)

- ~14M publications (de-duplicated)
- ~200K links publication-project from 5 funders
- ~9K datasets linked to publications or projects
- ~41K organizations (de-duplicated)

Collected from:
- 600+ “direct” data providers
- 5,000+ “indirect” data providers (inherited from aggregators)
- End-users …
LITERATURE BROKER SERVICE: CONCEPT

- Institutional repositories in OpenAIRE may be interested to acquire metadata information about publications that are “potentially of interest to them” that is they should/could be part of their collection.
  - They can enrich the records they already have with extra metadata information (Enrichment)
    - properties or relationships relative to publication objects in their collection that do not appear in their local metadata
  - They can enrich their collection with records they were unaware of (Addition)
    - publication objects in OpenAIRE that do not appear in their collection but may be pertinent to it
LITERATURE BROKER SERVICE: FUNCTIONAL REQUIREMENTS
SUBSCRIPTIONS: ENRICHMENT

- Publication collected from the repository by OpenAIRE
  - Enriched by OpenAIRE inference algorithms
    - relationships to projects and datasets, citation lists, document classification properties
  - Merged with richer publication objects
    - DOI of a publication, Open Access version of the publication
Exploits relationships
publication → author → organization → repository
collection relationship
Inferred relationship
Publication is relevant to the repository

SUBSCRIPTIONS ADDITION: AFFILIATION REPOSITORY
Exploits de-duplication and author’s submission frequency in repositories
Exploits relationships
publication $\rightarrow$ project $\rightarrow$ organization $\rightarrow$ repository
high chances to yield false positive notifications
LITERATURE BROKER SERVICE: NOTIFICATIONS

- **Mail postcards:**
  - notified by email at given interval of times

- **Programmatic access:**
  - Pull mode
  - Push Mode

- **Web interface**
LITERATURE BROKER SERVICE: DATA MODEL

Repository
  +repositoryID
  +repositoryName
  +repoManagerEmail

Notification
  +timestamp
  +sent
  +openaireID
  +subscriptionType
  +trust
  +title
  +authors
  +publicationDate
  +DCRecord

Subscription configuration
  +trustThreshold
  +timestamp
  +criteria
  +type
  +notification_scheduling
  +notification_granularity
  +notification_mode

notifiedRepository

repositoryOfInterest

Potential Notification
  +openaireID
  +subscriptionType
  +trust
  +title
  +authors
  +publicationDate
  +DCRecord
LITERATURE BROKER SERVICE: ARCHITECTURE
Addition
- repositories miss an average of 20% of their publications
- 400 institutional repositories with an average of 13,000 publications
- Worst case: the estimate of missing publications is around 1M

Enrichment
- Deduplication: 1,4M
- Inference: 3 M

Notifications, for a total of around 27M
CONCLUSION

- First beta available June 2016
  - Portuguese repositories

- Next steps
  - Make core of Broker Literature Service general purpose
  - notification broker services, literature subscriptions/notifications descriptions, subscription/notification APIs should be agreed on, conform to existing standards (SWORD)
THANK YOU