

The OpenAIRE Data Infrastructure Services On Interlinking European Institutional Repositories, Dataset Archives, and CRIS Systems

Paolo Manghi (ISTI-CNR), Lukasz Bolikowski (ICM, University of Warsaw), Natalia Manola (University of Athens), Jochen Schirrwagen (University of Bielefeld), Najla Rettberg (University of Gottingen), Tim Smith (CERN)

Promoting Open Science and Open Access to publications and datasets Realizing and maintaining the European Scholarly Communication Infrastructure

Collecting entities from heterogeneous data sources

Collection of entities OpenDOAR, **CRIS** systems DOAJ **Organizations** CERIF **Data Sources Projects** OpenAIRE **Publications Persons** DataCite Repositories (DRIVER ORCID **Datasets** infra) Archives (DANS, EBI, BADC)

The OpenAIRE data infrastructure collects and interlinks content from repositories (i.e., OA and non-OA publications), dataset archives, and CRIS systems (i.e., EC and National funding schemes). Moreover, it supports advanced metrics to measure impacts of Open Access mandates and funding over research.

Entity deposition and provision

OpenAIRE Orphan Repository, enabling the deposition of publications and datasets for authors with no repository or archive of reference. The repository is powered by the INVENIO repository platform (http://invenio-software.org)

End-user Claims, enabling registered users to ingest publication or dataset metadata by providing the relative DOIs doi>®

> Experimenting with creation of Enhanced Publications: Compound Objects linking one publication to other publications or datasets

OpenAIRE portal and standard APIs, providing access to information space to endusers and third-party service providers

Entity de-duplication, inference and curation

Automatic tools

Entity de-Inference of entities and duplication: merging duplicate relationships between them entities

Merging Citations, similarity duplicates of publications, classification, enddatasets, persons, organizations

End-user feedback Semi-Automatic and data curators tools user interfaces

> **Entity Curation:** editing, removal, addition of entities and relationships

Disambiguated and cleansed information space

The infrastructure provides functionalities for deduplication, i.e., identification and merging of objects representing the same entity, inference, i.e., analysis of collected metadata and digital objects (e.g., PDFs) to infer new objects or relationships, and curation, i.e., addition, removal or update of collected objects by a set of experts or registered end-users

relations,

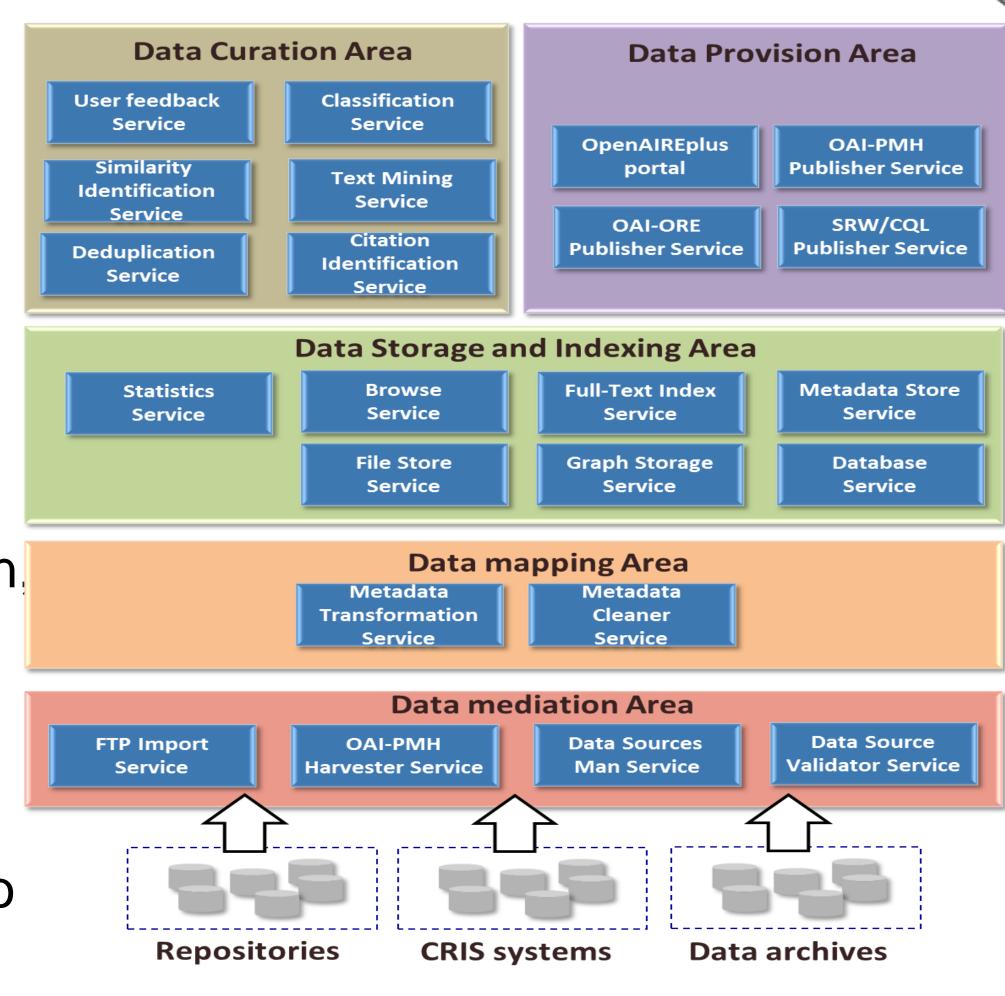
user behaviour

analysis

The design of the data model and the architecture of the infrastructure is supported by scientific partners following diverse paradigms - DANS (Social Sciences), EBI (Life Sciences), BADC (atmospheric sciences)

D-NET Software Toolkit: Data infrastructure technology

The OpenAIRE infrastructure is powered by the **D-NET Software** Toolkit. D-NET services can be customized to form workflows for entity collection processing and provision that satisfy a specific data aggregation scenario



http://www.d-net.research-infrastructures.eu