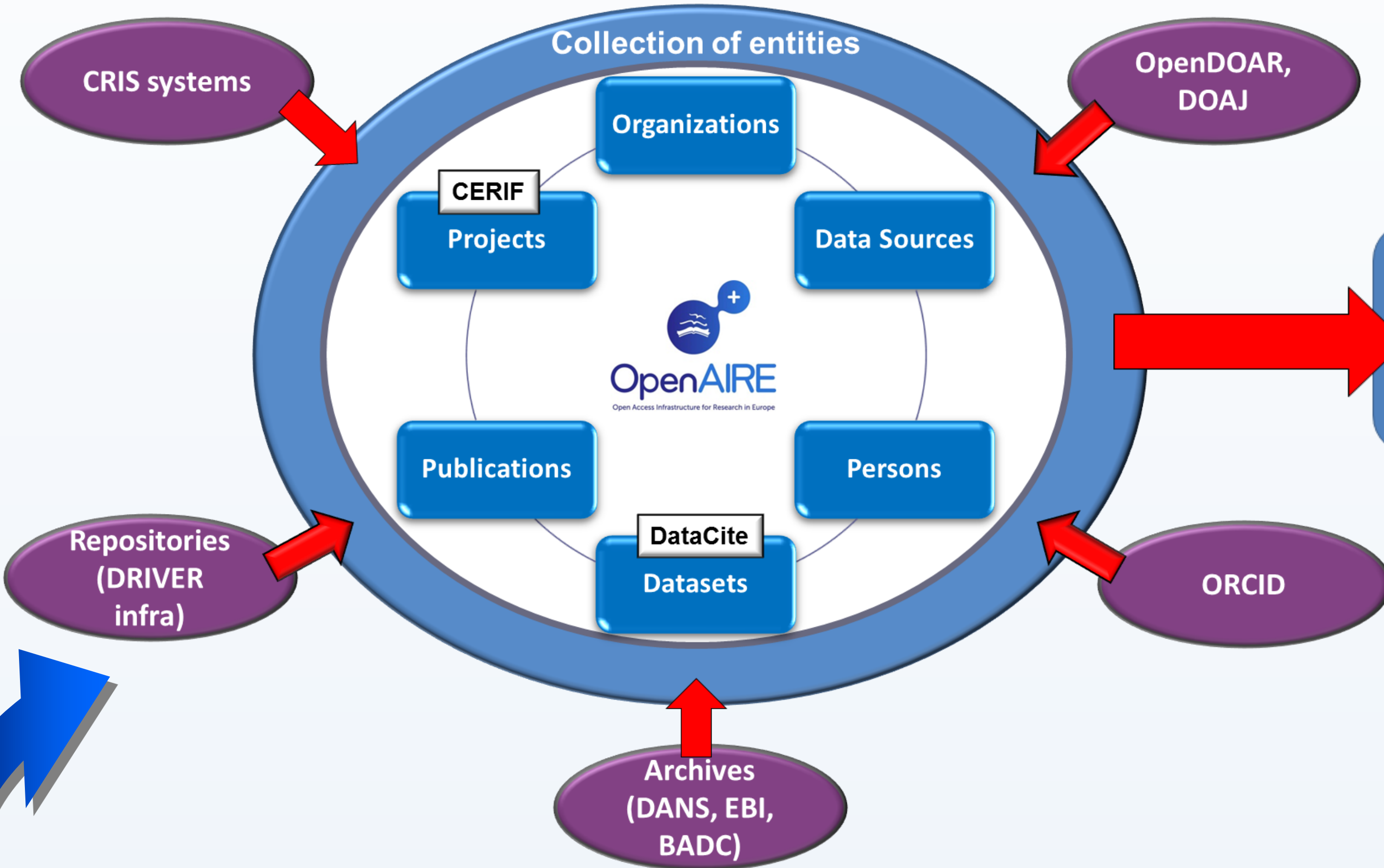


The OpenAIRE Data Infrastructure Services On Interlinking European Institutional Repositories, Dataset Archives, and CRIS Systems

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)

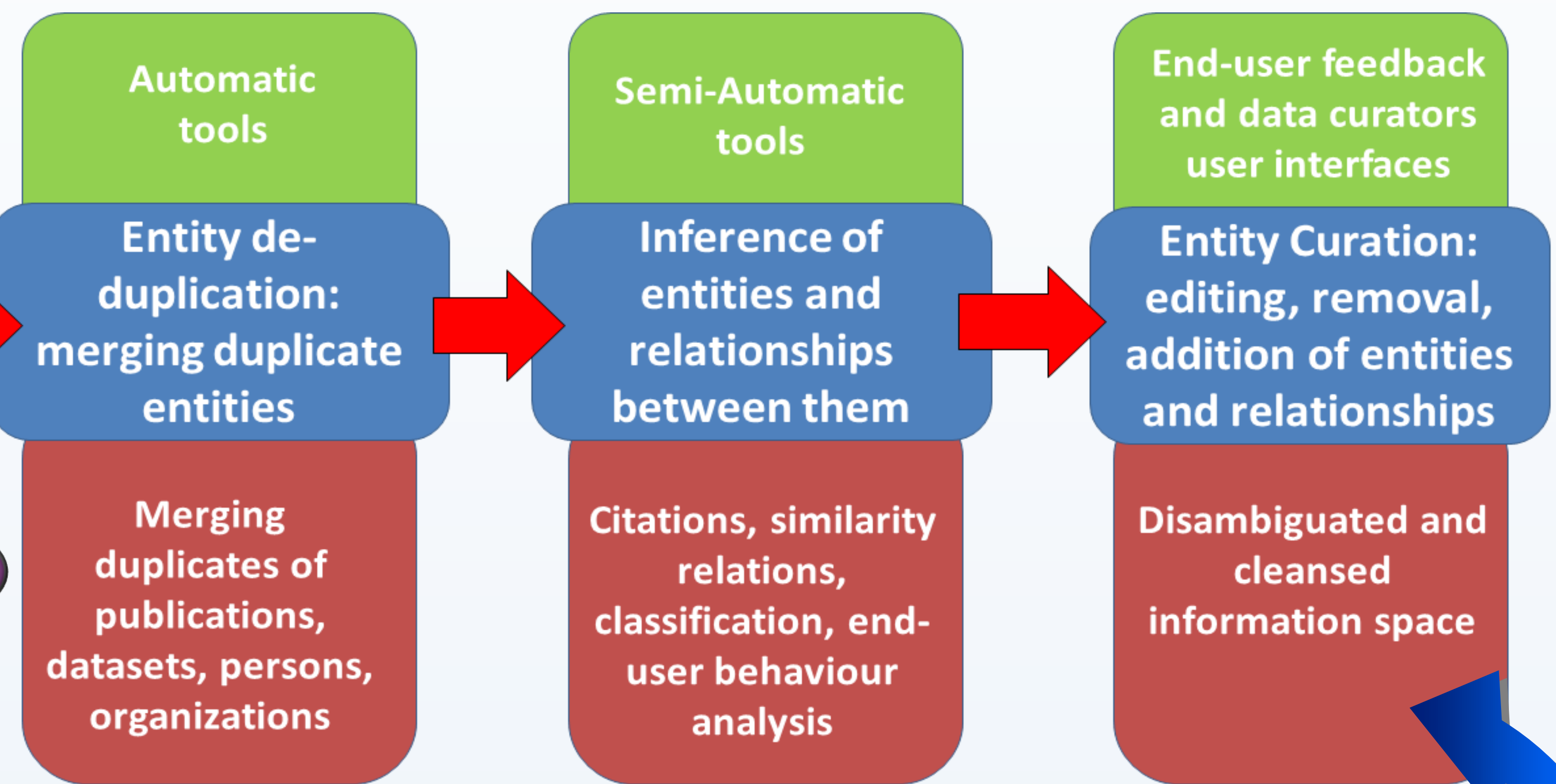
**Promoting Open Science and Open Access to publications and datasets
Realizing and maintaining the European Scholarly Communication Infrastructure**

Collecting entities from heterogeneous data sources



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 de-duplication, inference and curation



The infrastructure provides functionalities for **de-duplication**, 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

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)

Entity deposition and provision

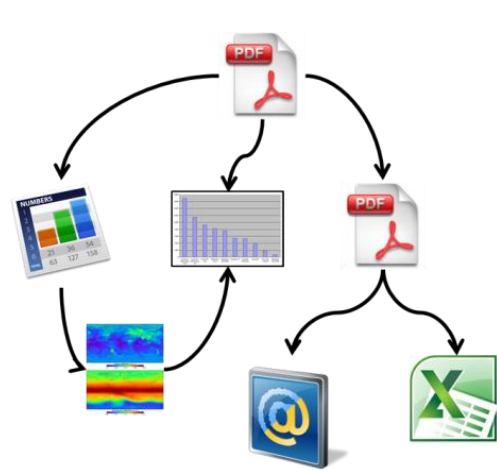
OpenAIRE Orphan Repository, enabling the deposition of publications and datasets for authors with no repository or archive of reference. The repository is powered by the INVENIO repository platform (<http://invenio-software.org>)



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



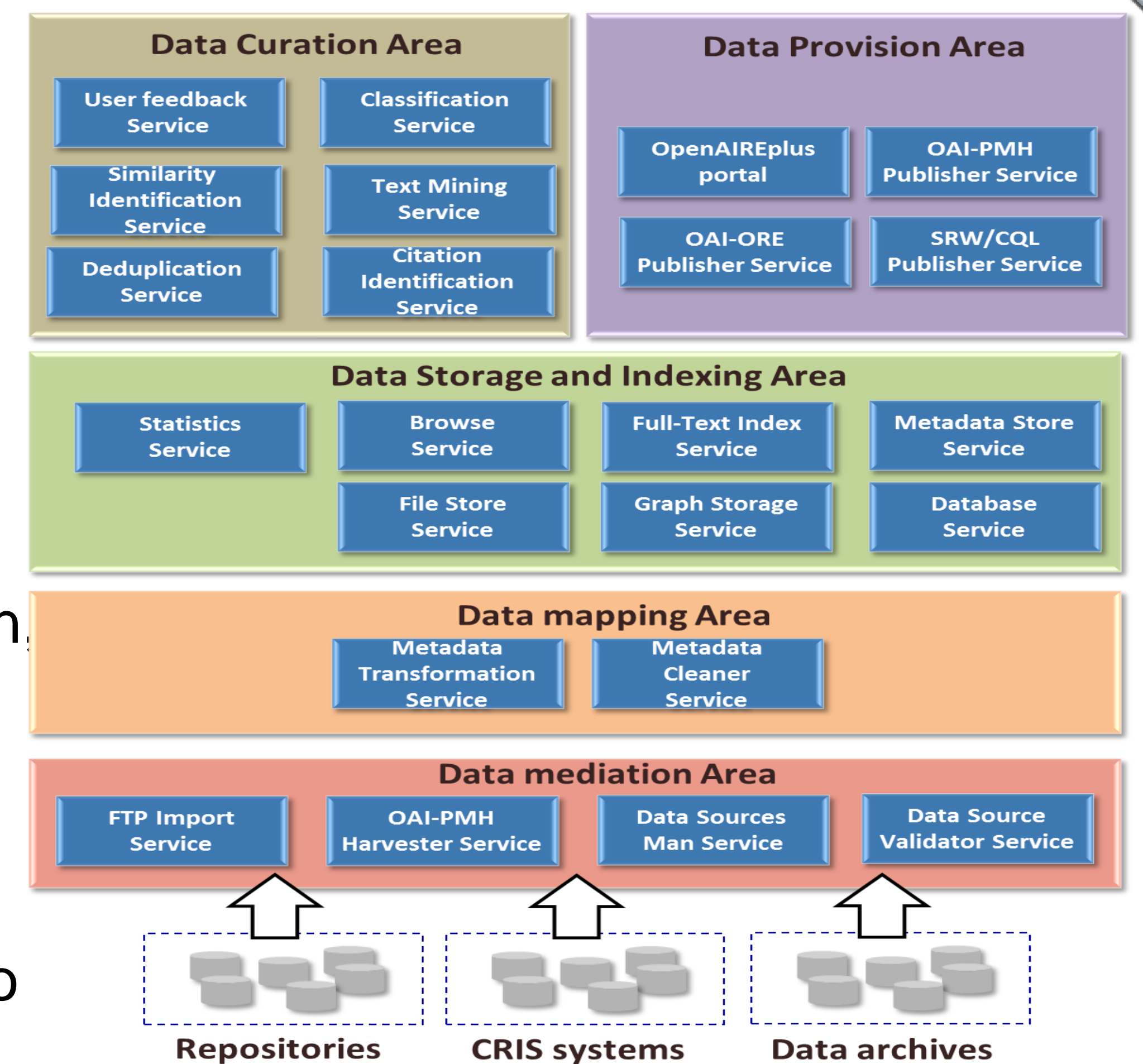
Experimenting with creation of **Enhanced Publications**: Compound Objects linking one publication to other publications or datasets



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

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>