OpenAIRE Research Graph & the broker service

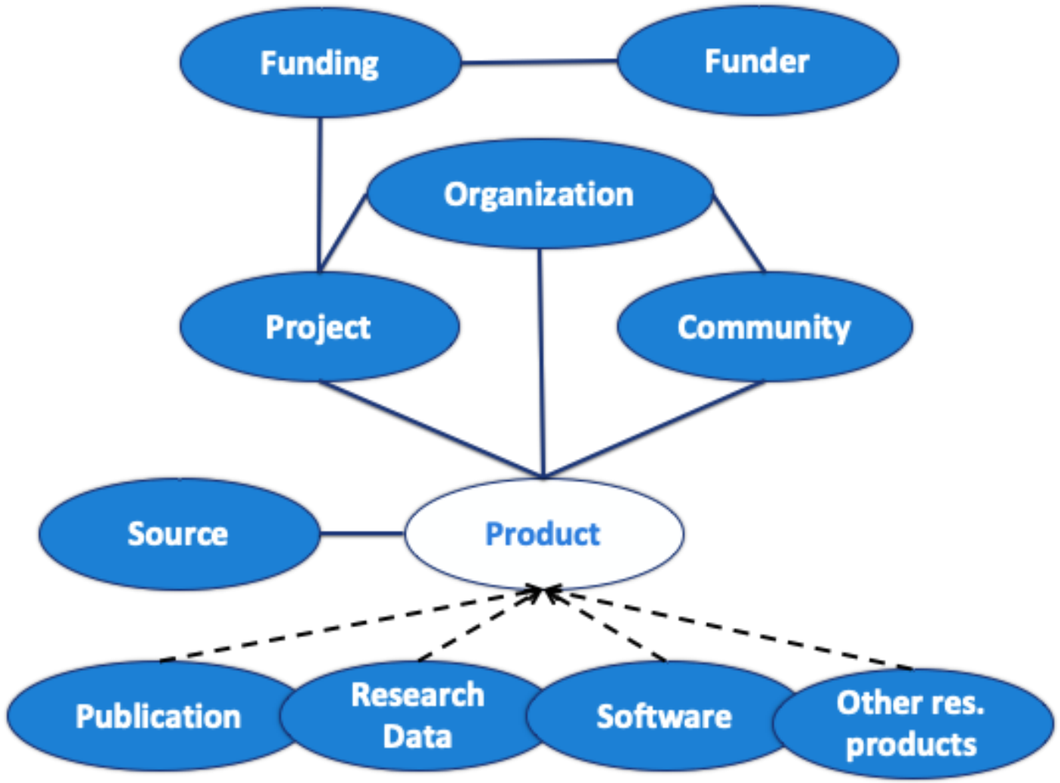
Claudio Atzori (CNR-ISTI)
OpenAIRE week 2020

OpenAIRE

European Commission

@openaire_eu

An open metadata research graph of interlinked scientific products, with access rights information, linked to funding information and research communities

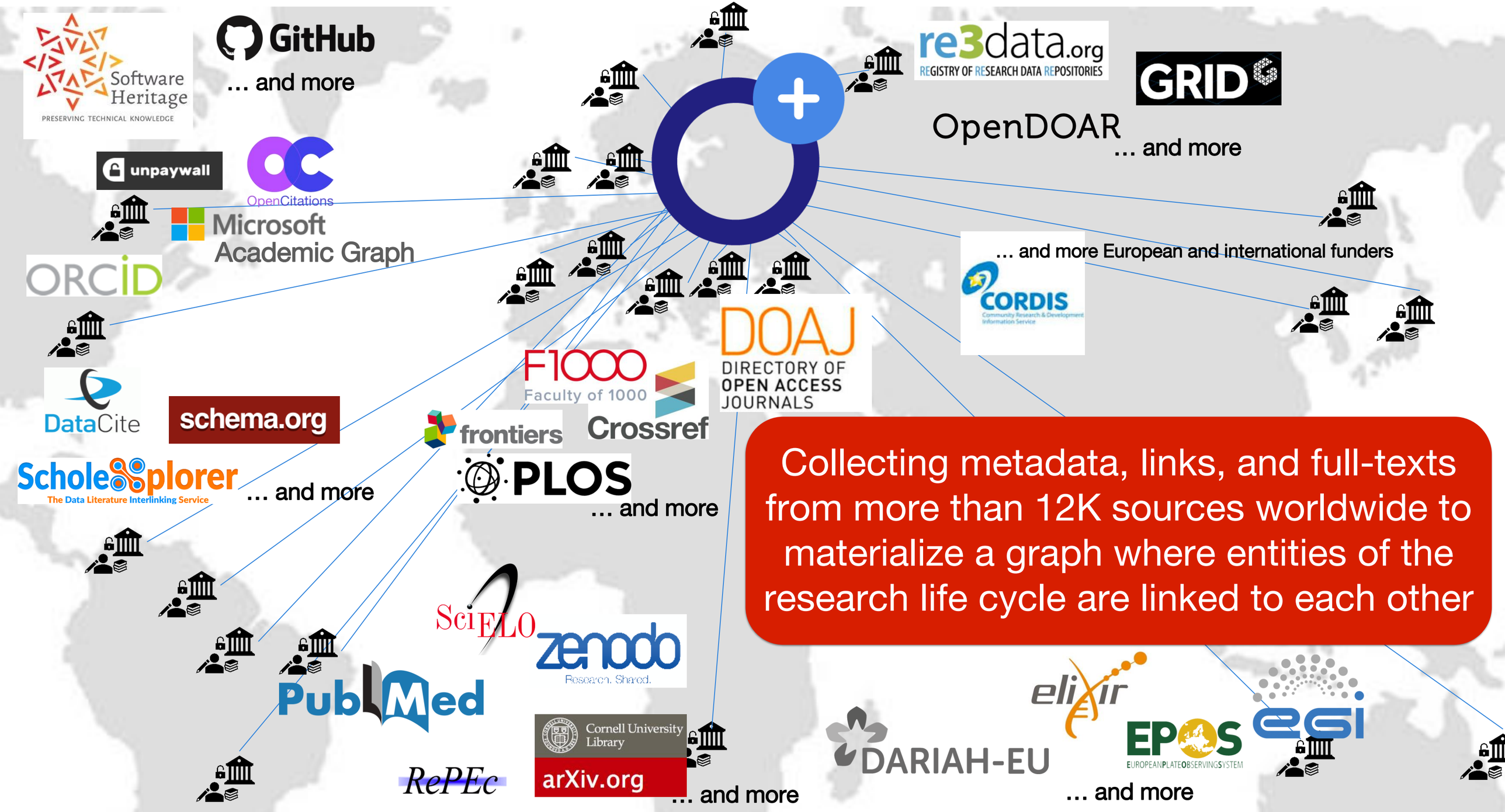


Discover more at graph.openaire.eu

Graph: model for the representation of information

OpenAIRE uses it to represent **objects** in the scholarly communication domain and the **relationships** that exist among them.

Edges of the graph are annotated with a label that specifies the **semantics** of the relationships between two objects, each represented as a node in the graph.


Software Heritage
PRESERVING TECHNICAL KNOWLEDGE
GitHub
... and more
re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES
GRID
OpenDOAR
... and more
unpaywall
OpenCitations
Microsoft Academic Graph
ORCID
... and more European and international funders
CORDIS
Community Research & Development Information Service
DOAJ
DIRECTORY OF OPEN ACCESS JOURNALS
F1000
Faculty of 1000
Crossref
DataCite
schema.org
frontiers
PLOS
ScholeXplorer
The Data Literature Interlinking Service
... and more
... and more
SciELO
zenodo
Research. Shared.
PubMed
Cornell University Library
arXiv.org
RePEc
... and more
elixir
DARIAH-EU
EPOS
EUROPEANPLATEOBSERVINGSYSTEM
egi
... and more
Collecting metadata, links, and full-texts from more than 12K sources worldwide to materialize a graph where entities of the research life cycle are linked to each other

Registries:
data sources offering authoritative lists of entities

Registries of researchers

ORCID

Registries of organizations

GRID

CORDIS Community Research & Development Information Service

and other funders

Registries of projects

FCT Fundação para a Ciência e a Tecnologia

CORDIS Community Research & Development Information Service

ACADEMY OF FINLAND

NWO

Science Foundation Ireland Sfi For what's next

Australian Government
Australian Research Council

and other funders

Registries of data sources

re3data.org REGISTRY OF RESEARCH DATA REPOSITORIES

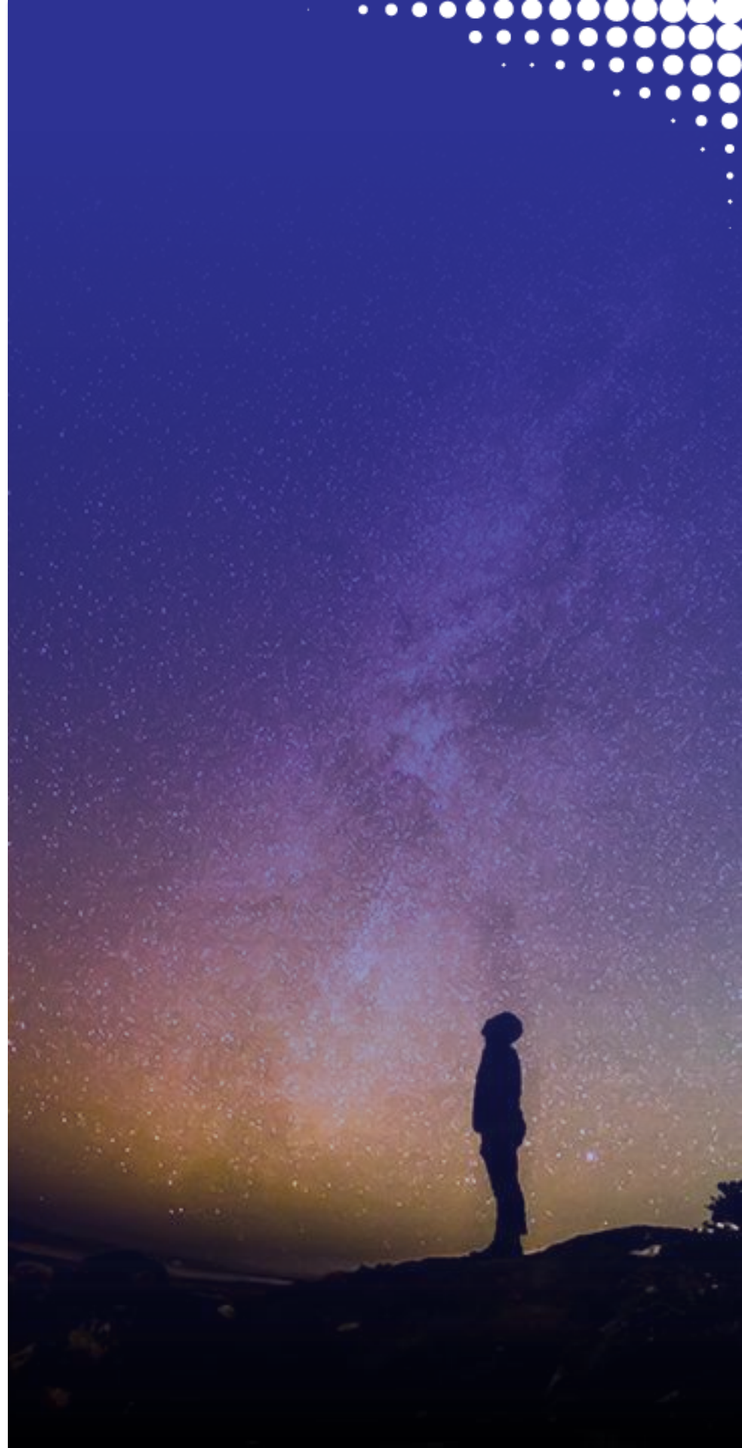
OpenDOAR

DOAJ DIRECTORY OF OPEN ACCESS JOURNALS

\+ journal lists from publishers

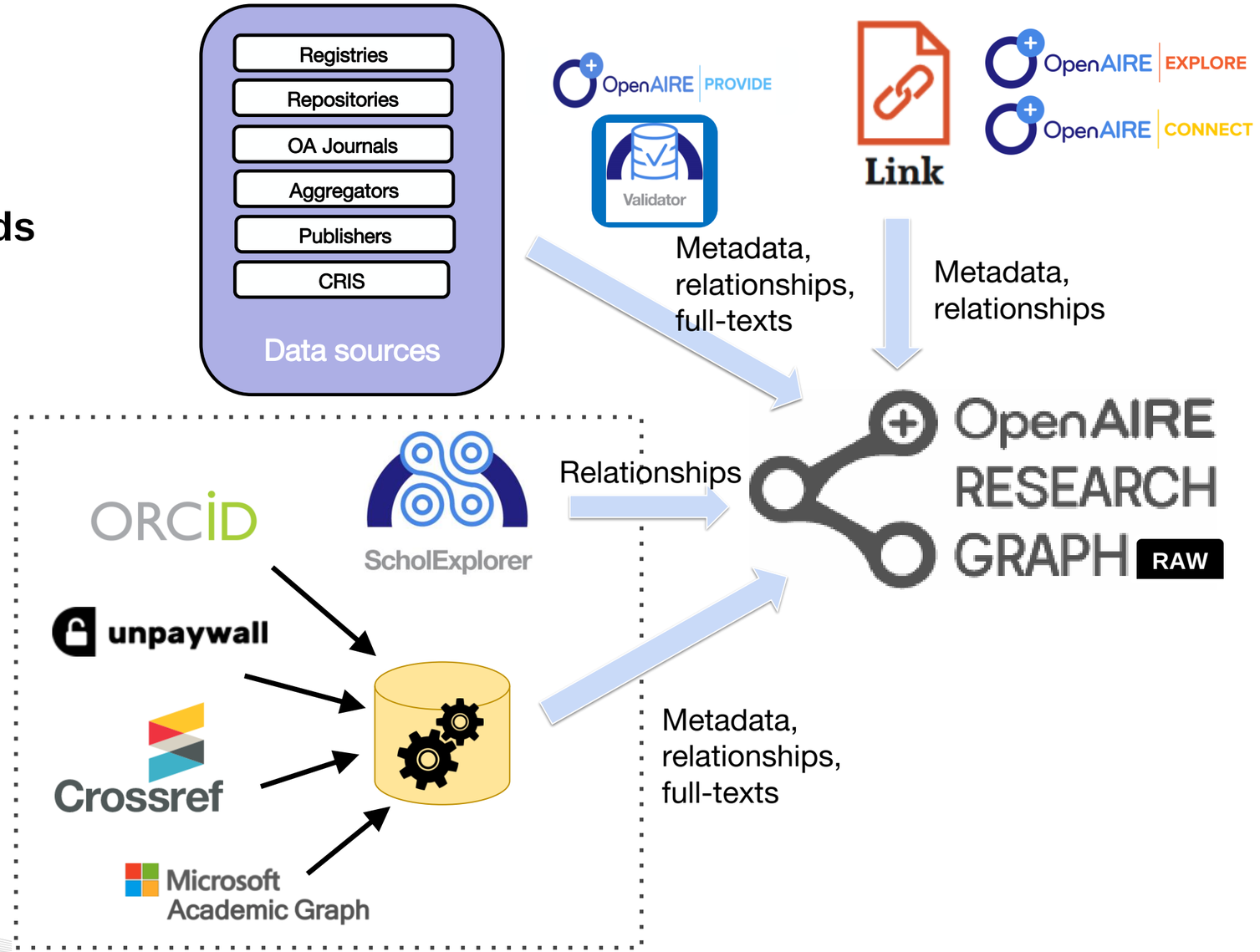
OpenAIRE

European Commission

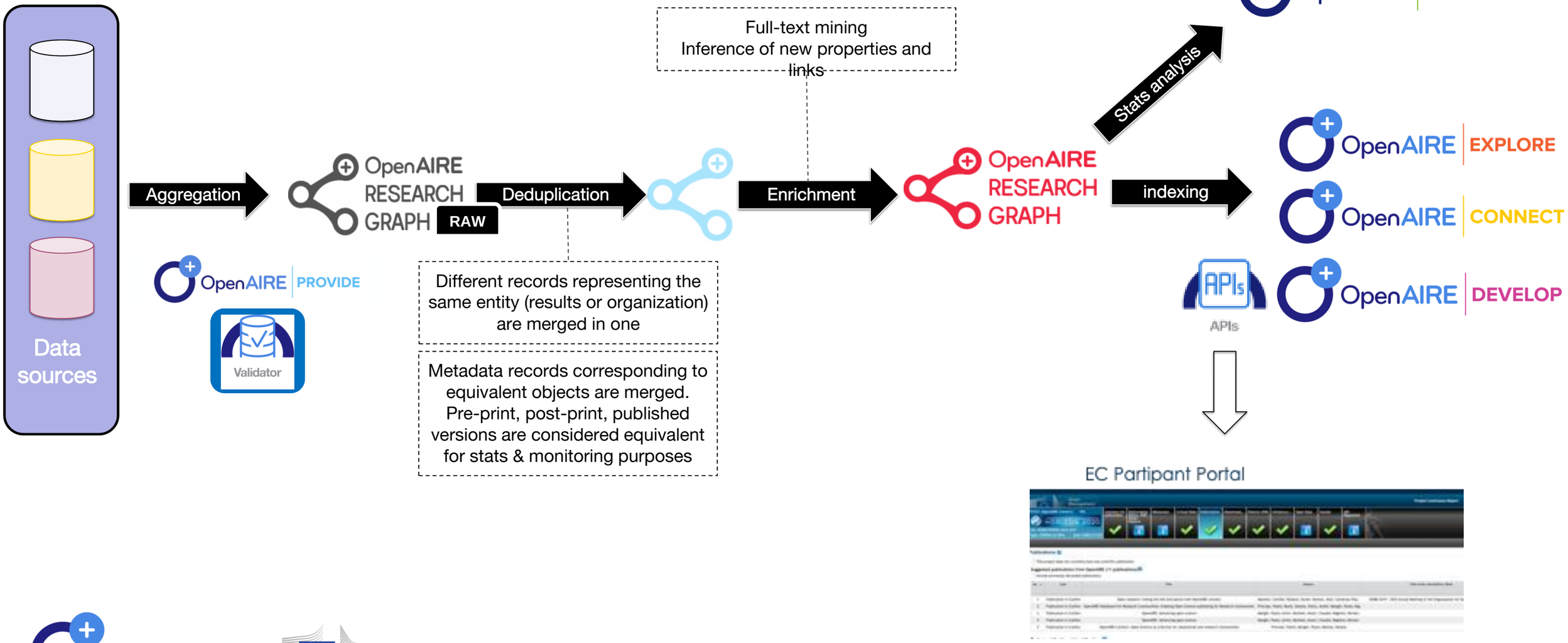


Aggregated data sources

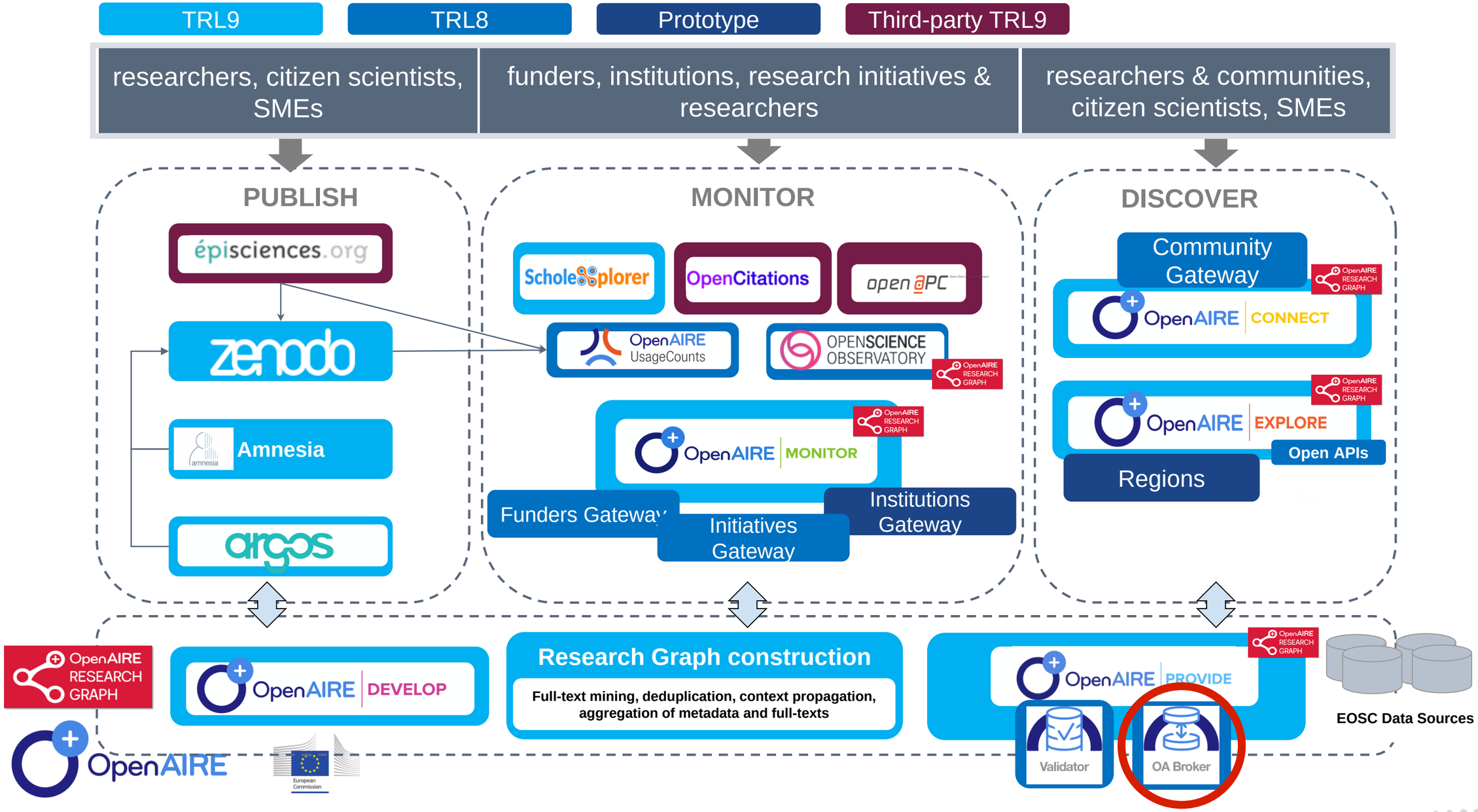
450Mi harvested records
500Mi bi-lateral links
13Mi+ full-texts



OpenAIRE Research Graph the supply chain



OpenAIRE Services



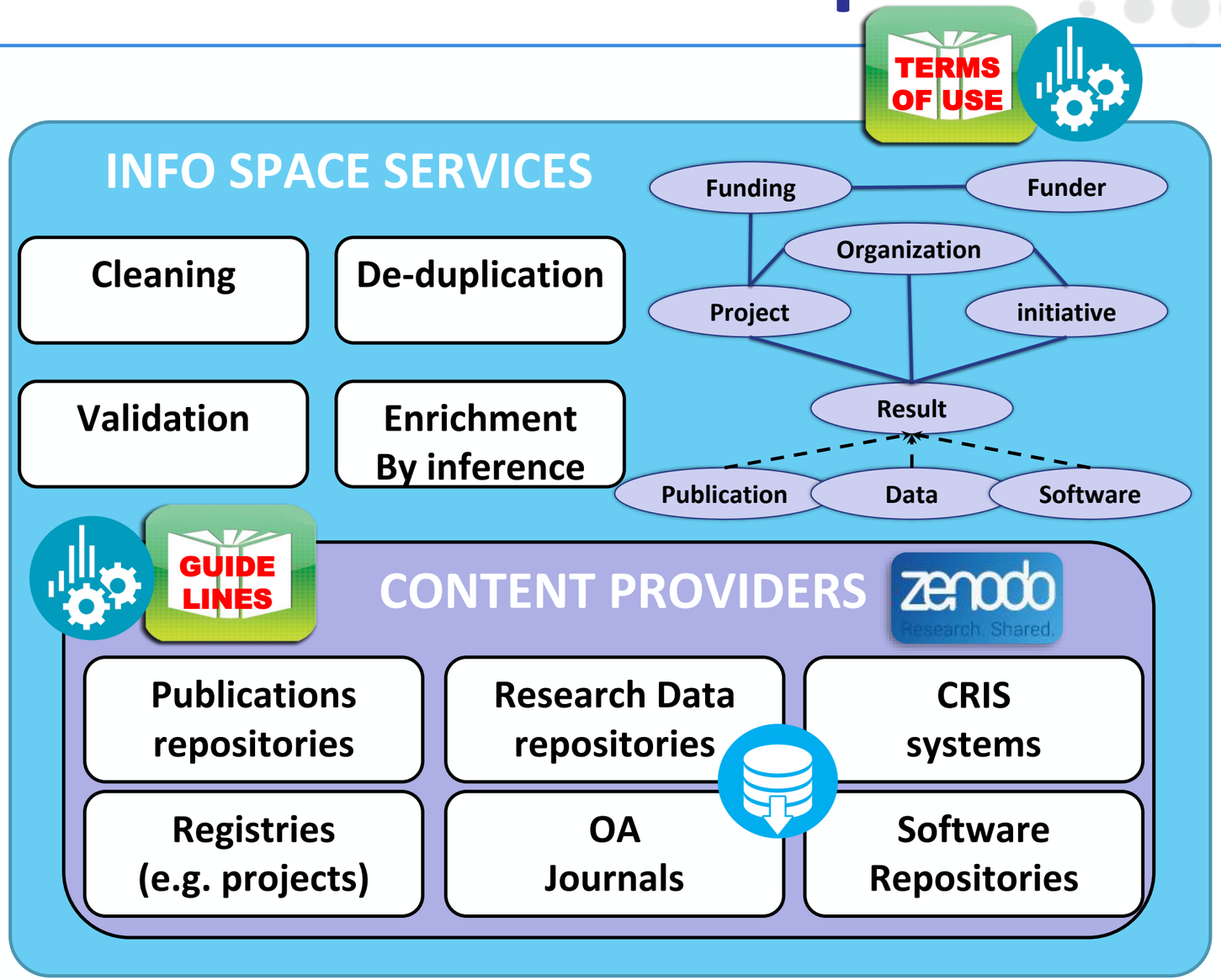
OpenAIRE Broker Service - Concept

Repositories in OpenAIRE may be interested to acquire metadata information about **publications** that are

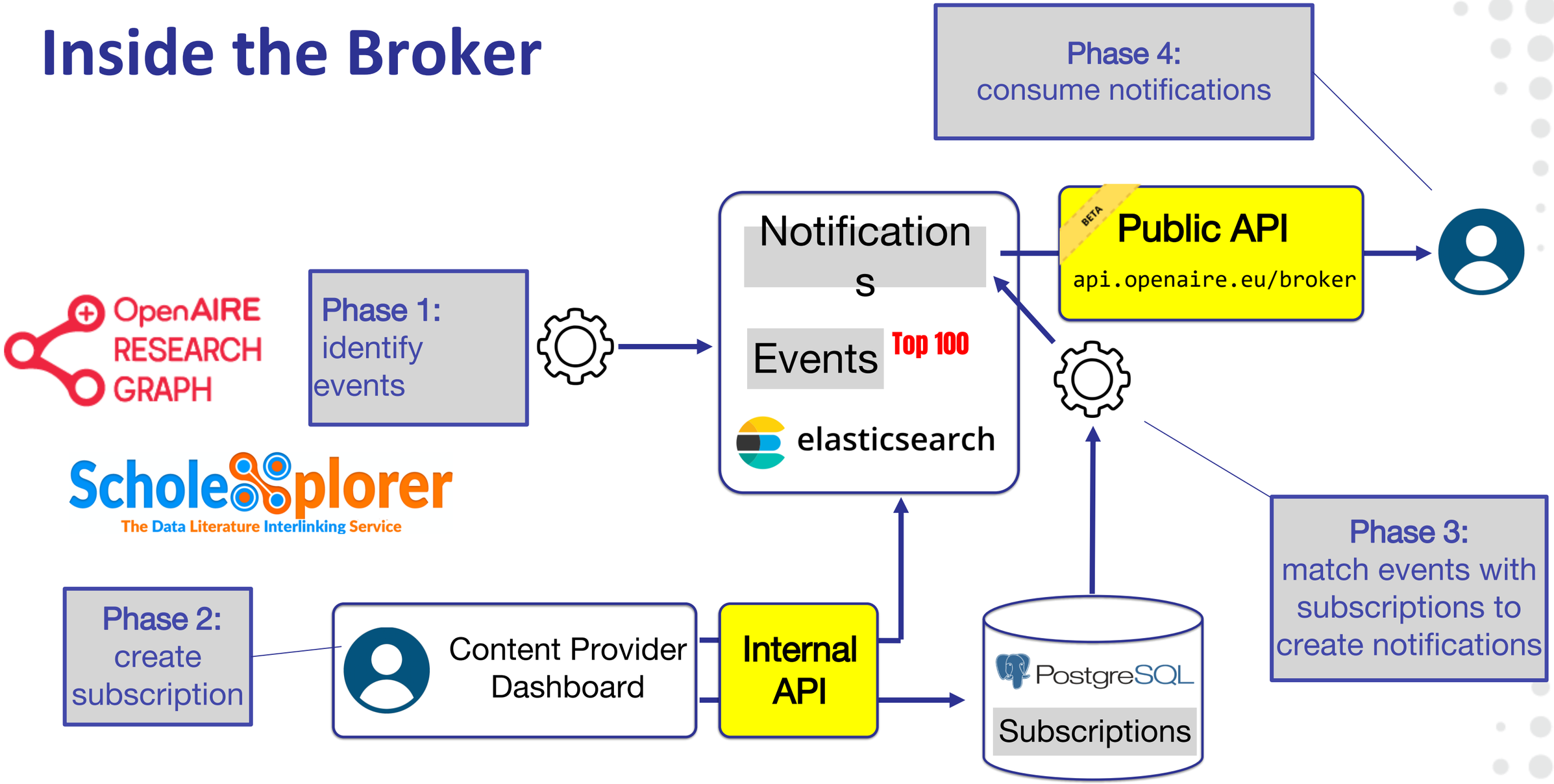
“potentially of interest to them”

i.e. be part of their collection:

- ***enrich the records*** *with* ***extra metadata*** *information*



Inside the Broker



OpenAIRE

European Commission

OpenAIRE | PROVIDE

Calculation of Broker events

Analysis of the groups of duplicated records

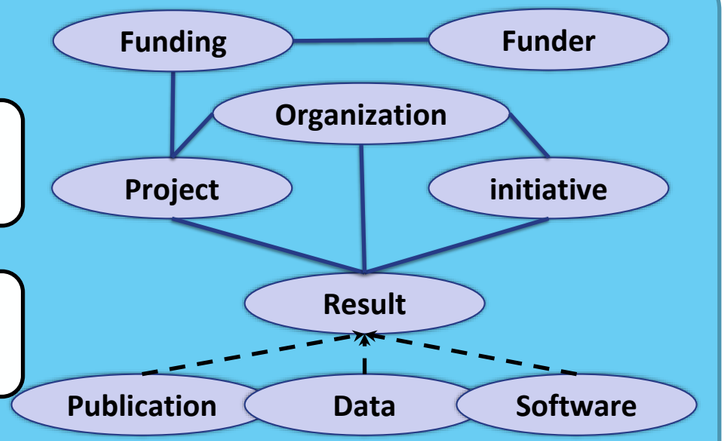
INFO SPACE SERVICES

Cleaning

De-duplication

Validation

Enrichment
By inference



ScholeXplorer
The Data Literature Interlinking Service

OpenAIRE RESEARCH GRAPH

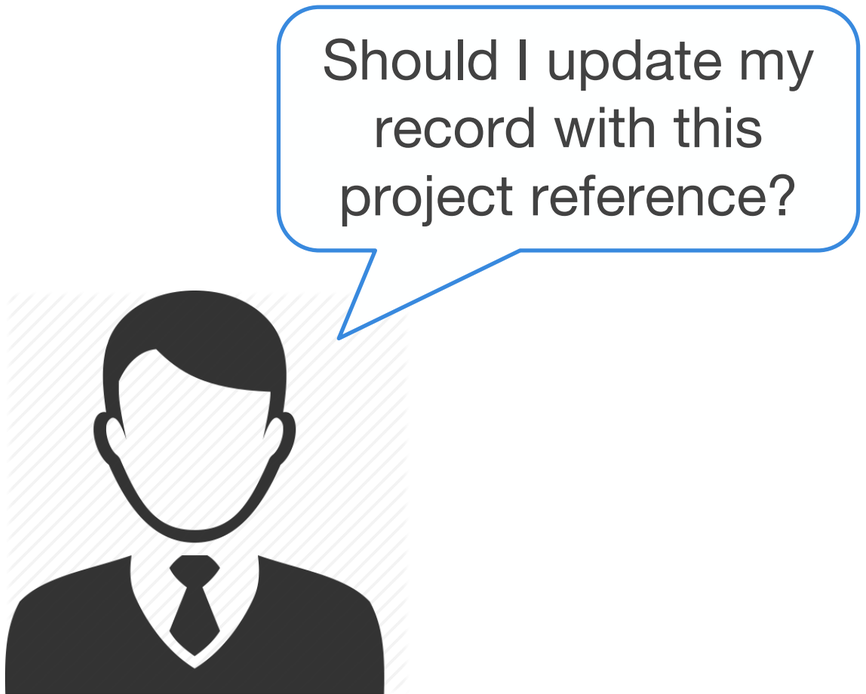
OpenAIRE

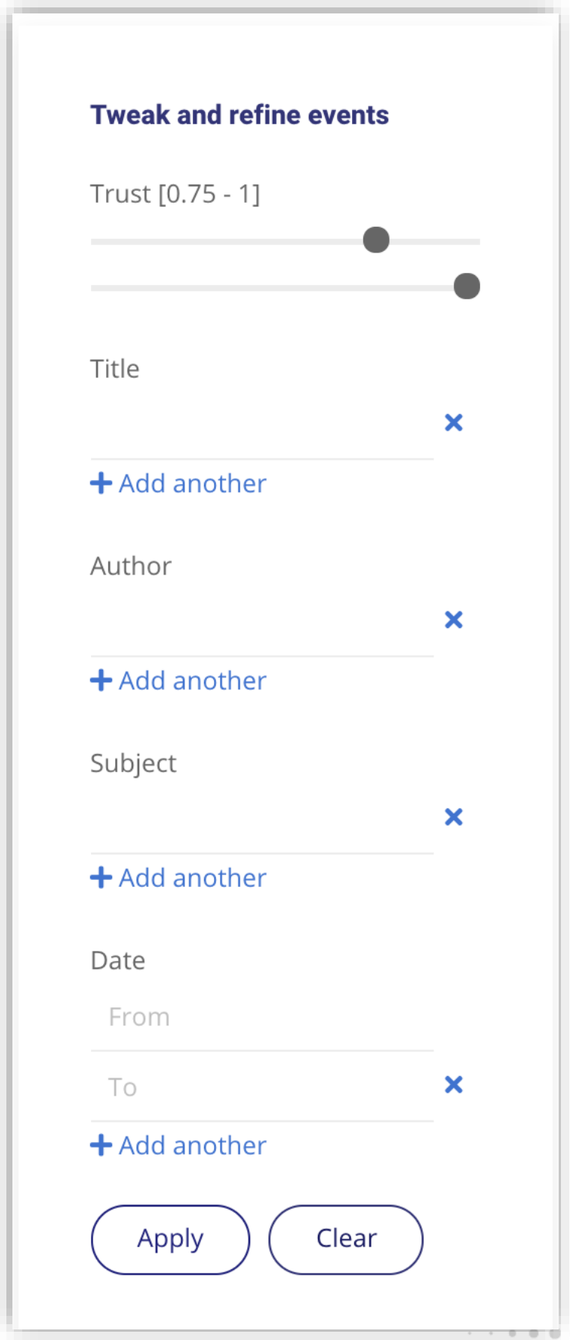
European Commission

Event trust

Events are produced by ***algorithms*** that takes decisions using ***non-authoritative information;***

Trust as a manner to expose uncertainty





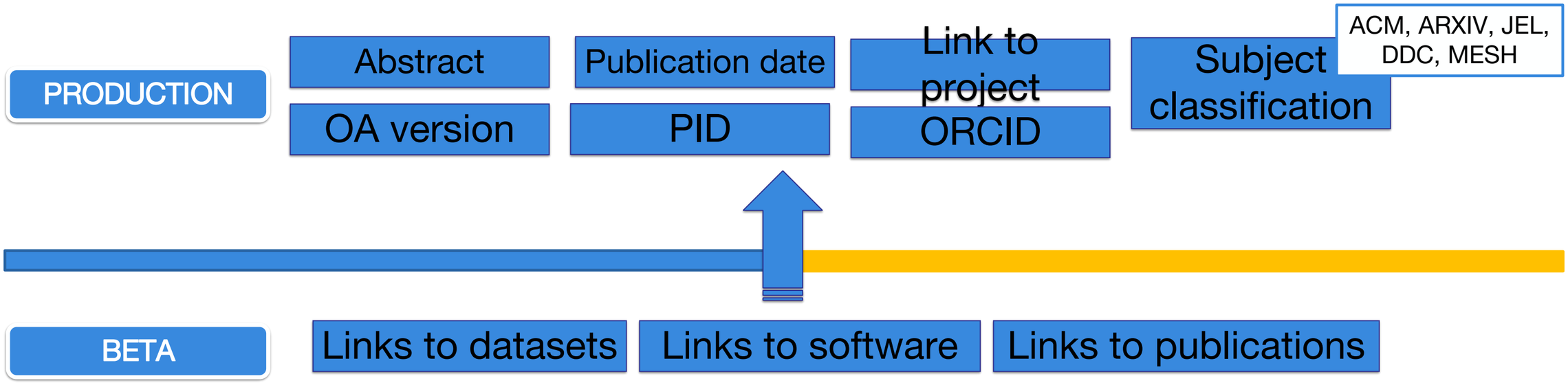
Broker Enrichment Events

ENRICH / MORE

Macro-category that groups events about field values that **differ** from those available in the repository.

ENRICH / MISSING

Macro-category that groups events about field values that are **not present** in the metadata of the repository.



OpenAIRE

European Commission

Events produced

Topic	#Events
ENRICH/MORE/OPENACCESS_VERSION	17054736
ENRICH/MORE/PID	4167583
ENRICH/MORE/SUBJECT/MESHEUROPEPMC	1429089
ENRICH/MISSING/PID	1292499
ENRICH/MISSING/PROJECT	857956
ENRICH/MORE/SUBJECT/ARXIV	1025148
ENRICH/MISSING/ABSTRACT	806453
ENRICH/MORE/SUBJECT/JEL	648996
ENRICH/MORE/SUBJECT/DCC	320023
ENRICH/MORE/SUBJECT/ACM	608756
ENRICH/MISSING/OPENACCESS_VERSION	245912
ENRICH/MISSING/SUBJECT/ARXIV	165230
ENRICH/MISSING/SUBJECT/MESHEUROPEPMC	122031
ENRICH/MISSING/SUBJECT/JEL	84043
ENRICH/MISSING/SUBJECT/DDC	70748
ENRICH/MISSING/PUBLICATION_DATE	37931
ENRICH/MISSING/SUBJECT/ACM	61210
ENRICH/MISSING/AUTHOR/ORCID	**69120**

Produced by OpenAIRE (for ~700 repositories) in production

Topic	#Events
ENRICH/MISSING/DATASET/IS_REFERENCED_BY	4542704
ENRICH/MISSING/DATASET/IS_RELATED_TO	52870
ENRICH/MISSING/DATASET/IS_SUPPLEMENTED_BY	152
ENRICH/MISSING/DATASET/IS_SUPPLEMENT_TO	325
ENRICH/MISSING/DATASET/REFERENCES	24672
ENRICH/MISSING/PUBLICATION/IS_REFERENCED_BY	31026
ENRICH/MISSING/PUBLICATION/IS_RELATED_TO	288375
ENRICH/MISSING/PUBLICATION/IS_SUPPLEMENTED_BY	3653
ENRICH/MISSING/PUBLICATION/IS_SUPPLEMENT_TO	55921
ENRICH/MISSING/PUBLICATION/REFERENCES	164084
ENRICH/MISSING/SOFTWARE	17380

Produced by ScholeXplorer (for ~60 data repositories) in beta

Thank you!

Claudio Atzori (CNR-ISTI)

claudio.atzori@isti.cnr.it