



**Paolo Manghi**

Institute of Information Science and Technologies

National Research Council

Pisa, Italy

# OpenAIRE Research Graph: interlinking research entities



# Materializing the Open Science Graph

CONNECT PROVIDE EXPLORE

DEVELOP MONITOR



End-user feedback Deduplication



Harvesting



Mining

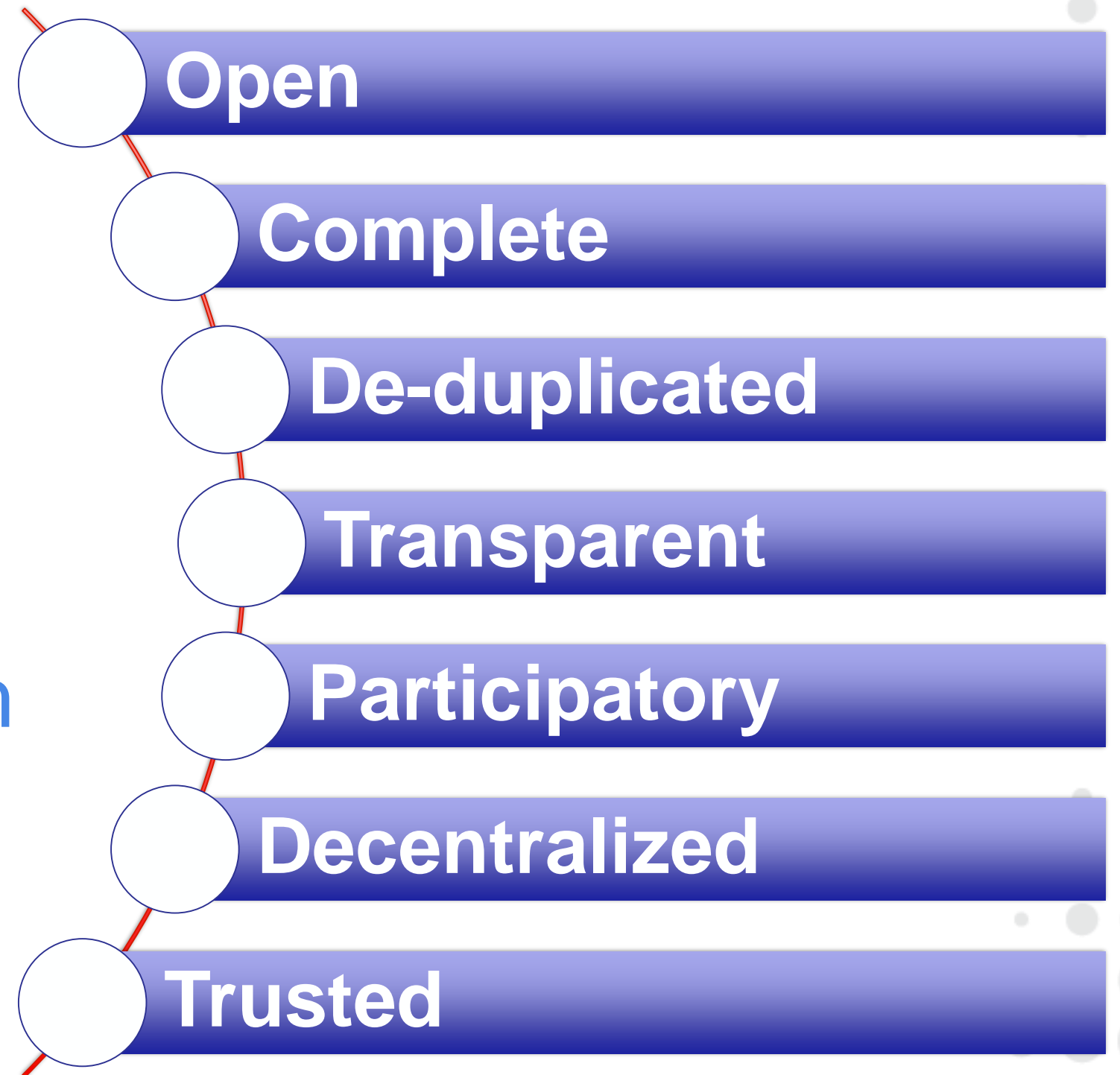




# The OpenAIRE research graph



Providing an **open metadata research graph** of interlinked **scientific products**, with **Open Access information**, linked to **funding information** and **research communities**



# Open metadata



Repositories  
OA Publishers  
Open Archives

...

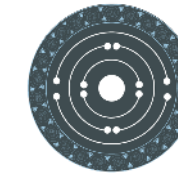


Springer  
Microsoft

# Complete: community-trusted sources



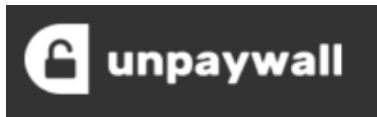
... and more



PARTHENOS  
Pooling Activities, Resources and Tools  
for Heritage, research, Networking,  
Optimization and Synergies



... and more



OpenDOAR



... and more



... and more



# De-duplicated

**Metadata records**  
corresponding to equivalent  
objects are **merged**

Scientific products  
Organizations



More information about the de-duplication framework used by OpenAIRE can be found searching on Zenodo for :

- [“De-duplicating the OpenAIRE Scholarly Communication Big Graph”](#) (poster)
- [“GDup: De-Duplication of Scholarly Communication Big Graphs”](#)

# Participatory

- Rely on quality scholarly communication sources of different kinds
- Include **solutions and content** from any interested and **known** content provider in scholarly communication

Institutional repositories  
Aggregators  
Data archives  
Software repositories  
Research infrastructure sources  
Funder grant databases  
Authors & Orgs entity registries  
Publishers & journals





# Transparent

- Metadata in the graph includes **provenance** when harvested and **reliability indicators** when obtained from mining





# Decentralized

- **Preservation** and **ownership** beyond OpenAIRE
  - **Exchanged** with other graph initiatives
  - **Broker Service**: Redistributed via subscription and notification to contributing data sources  
([provide.openaire.eu](https://provide.openaire.eu))
- Openly accessible via APIs  
([develop.openaire.eu](https://develop.openaire.eu))

 ResearchGraph



 OpenCitations



# Trusted (November 2019)

- Authors in the loop to enrich their ORCID record
- Validation of end-user "claims"

ORCID

Connecting Research  
and Researchers





Content



# Production: Open Access CAPs

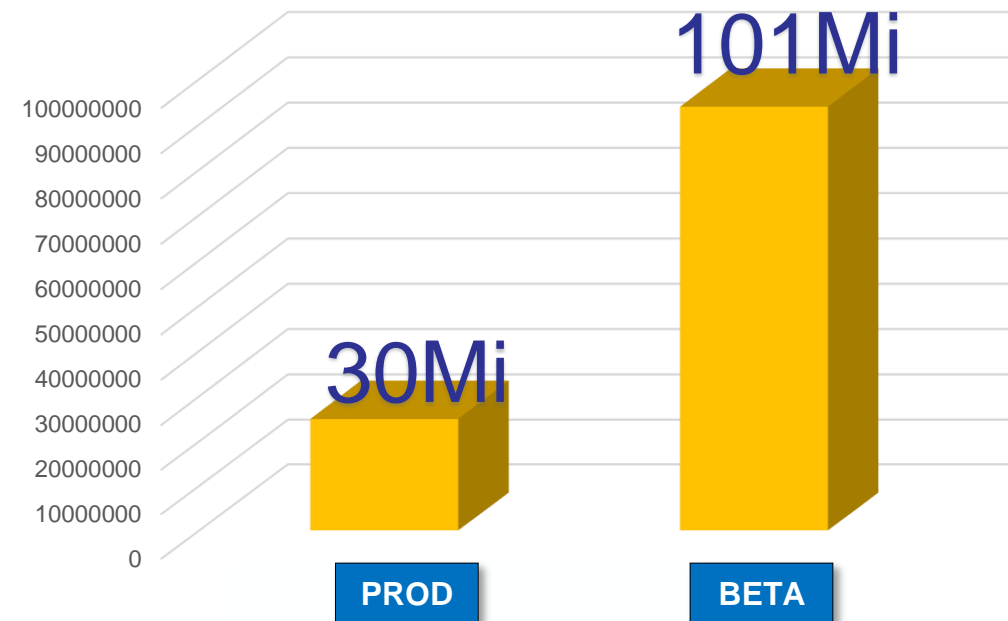
## BETA: Open Science CAPs

### Harvested content

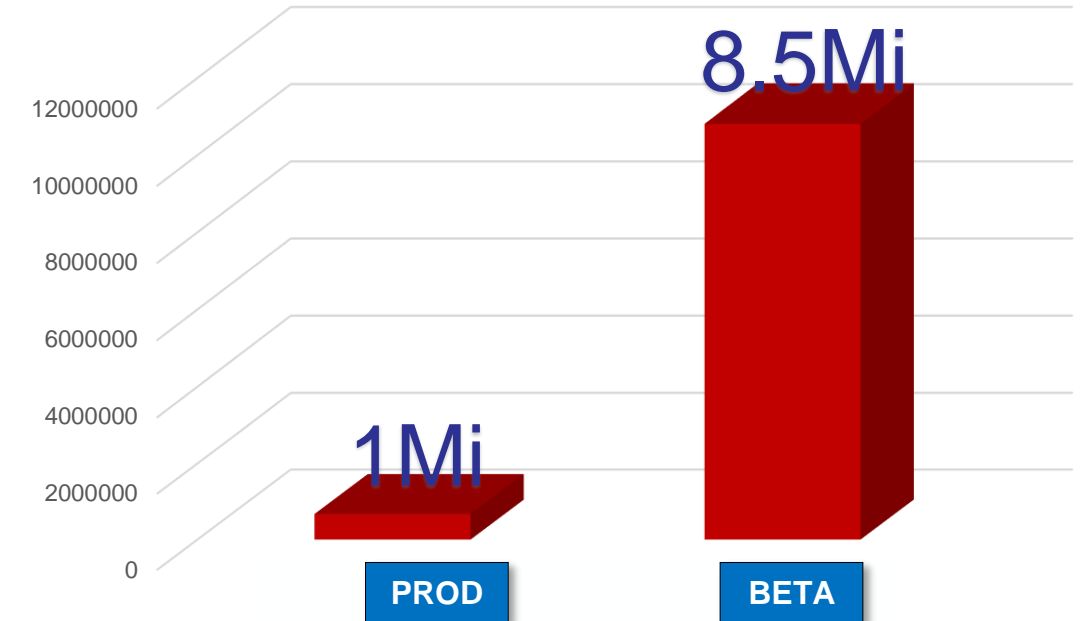
- Data sources
  - 10K+
- Records
  - 450Mi
- Publication full-texts
  - 11Mi
- Links (also text-mined)
  - 480Mi



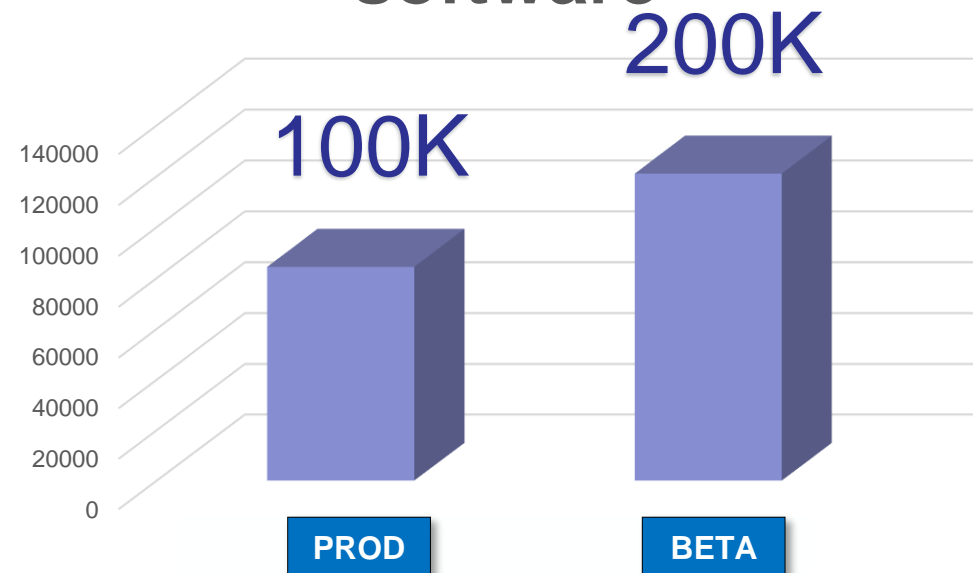
### literature



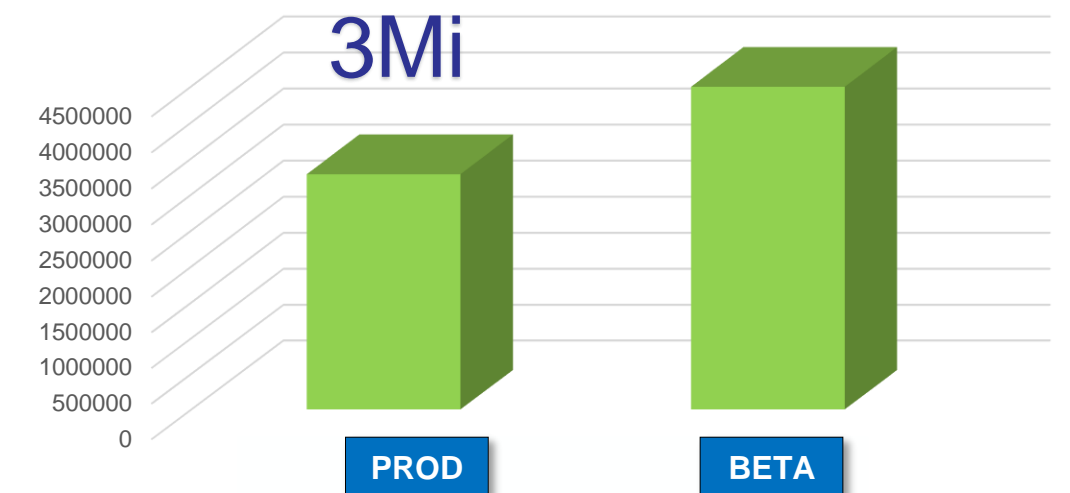
### research data



### software



### other 7.5Mi

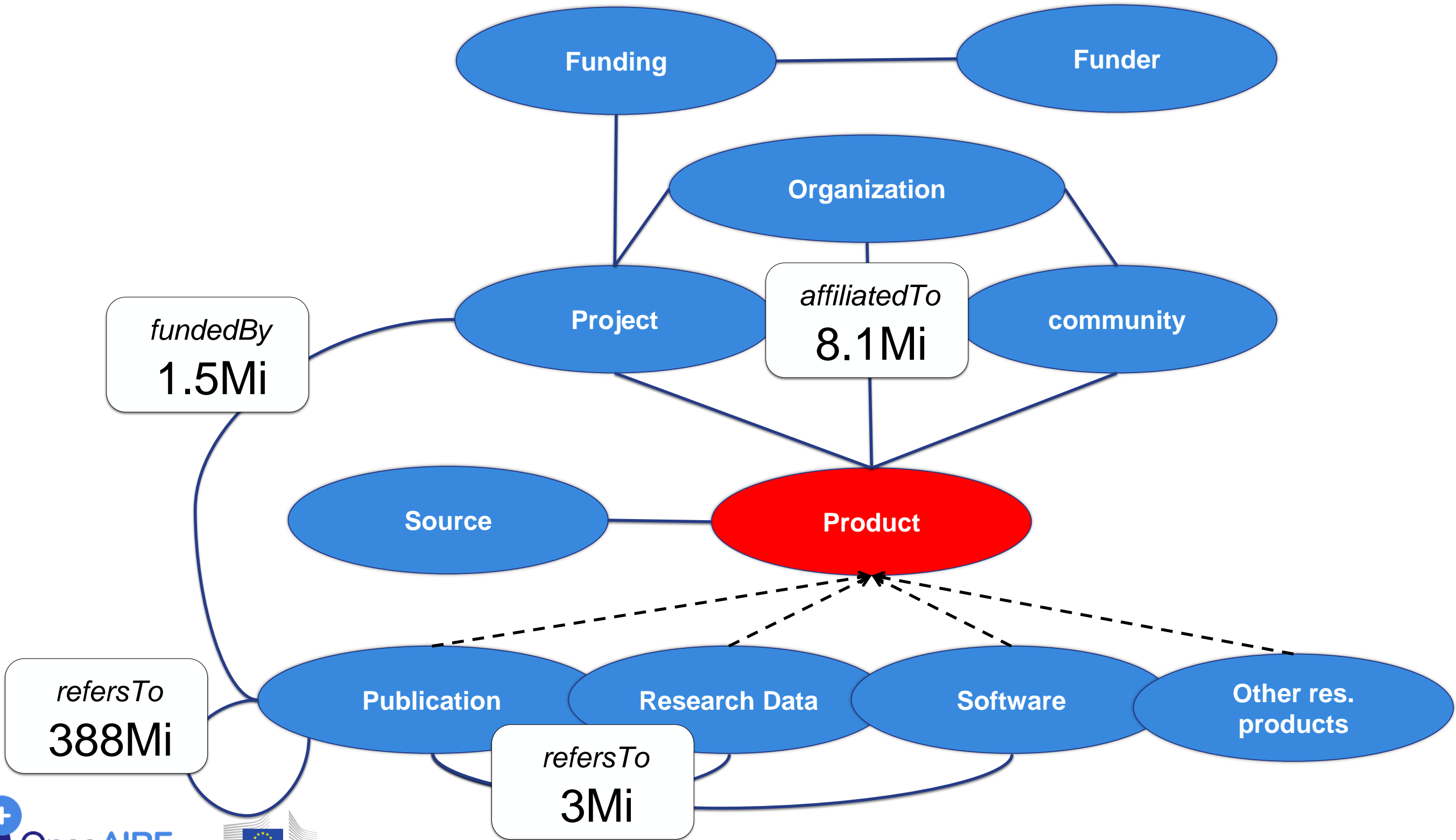


# Open Access articles sources

- **Repositories and publishers**
  - **Download from URLs in harvested metadata: 7.2Mi**
  - **Machine-learning on OA URLs from large aggregators (DOAJ, CrossRef/Unpaywall): 3.8 Mi**
- **Publishers metadata/PDFs via CORE-UK**
  - **ResourceSync**
  - **Springer Open Access, etc.: 750K**



# Mining results: inferred links





# Mining results: inferred properties

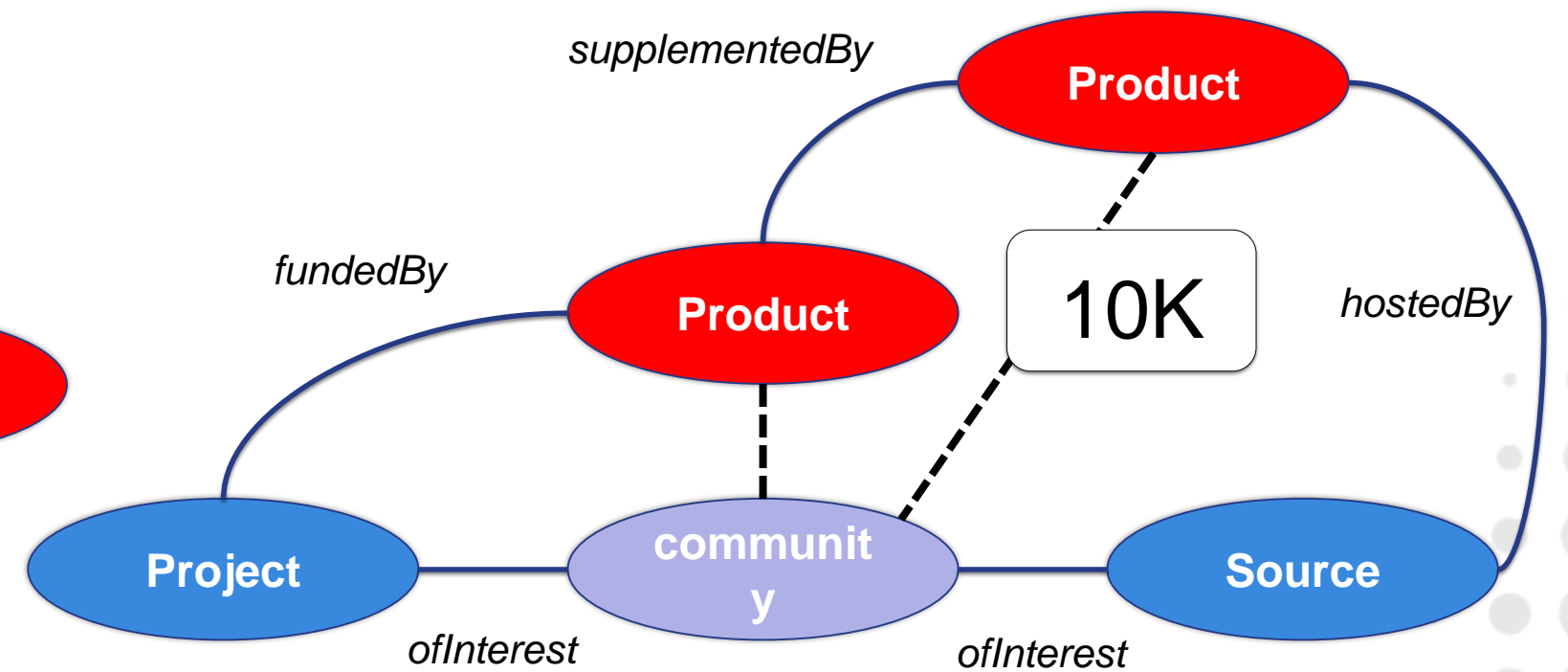
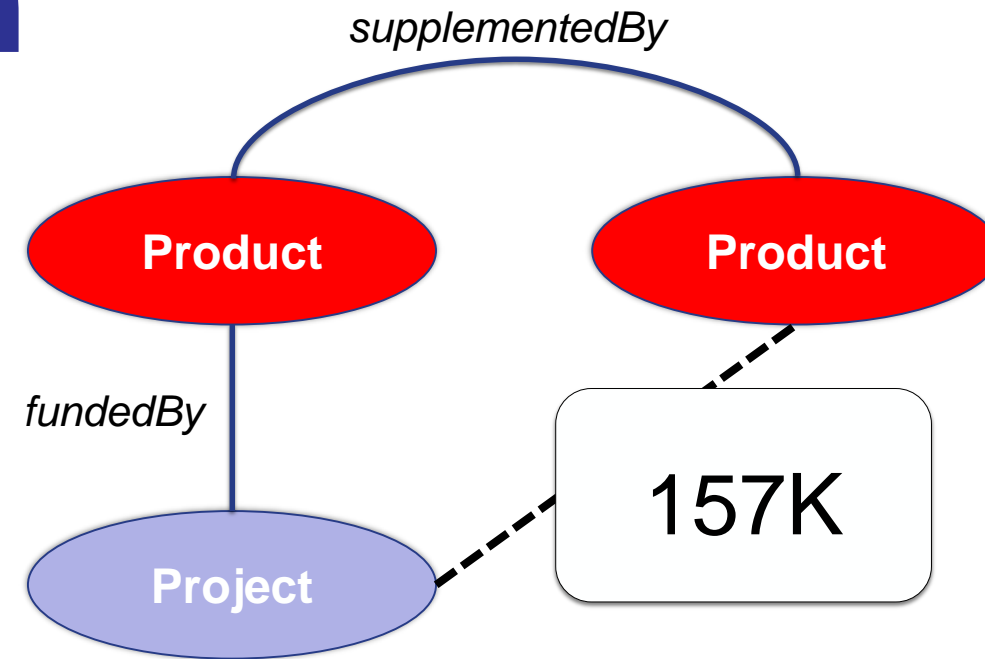
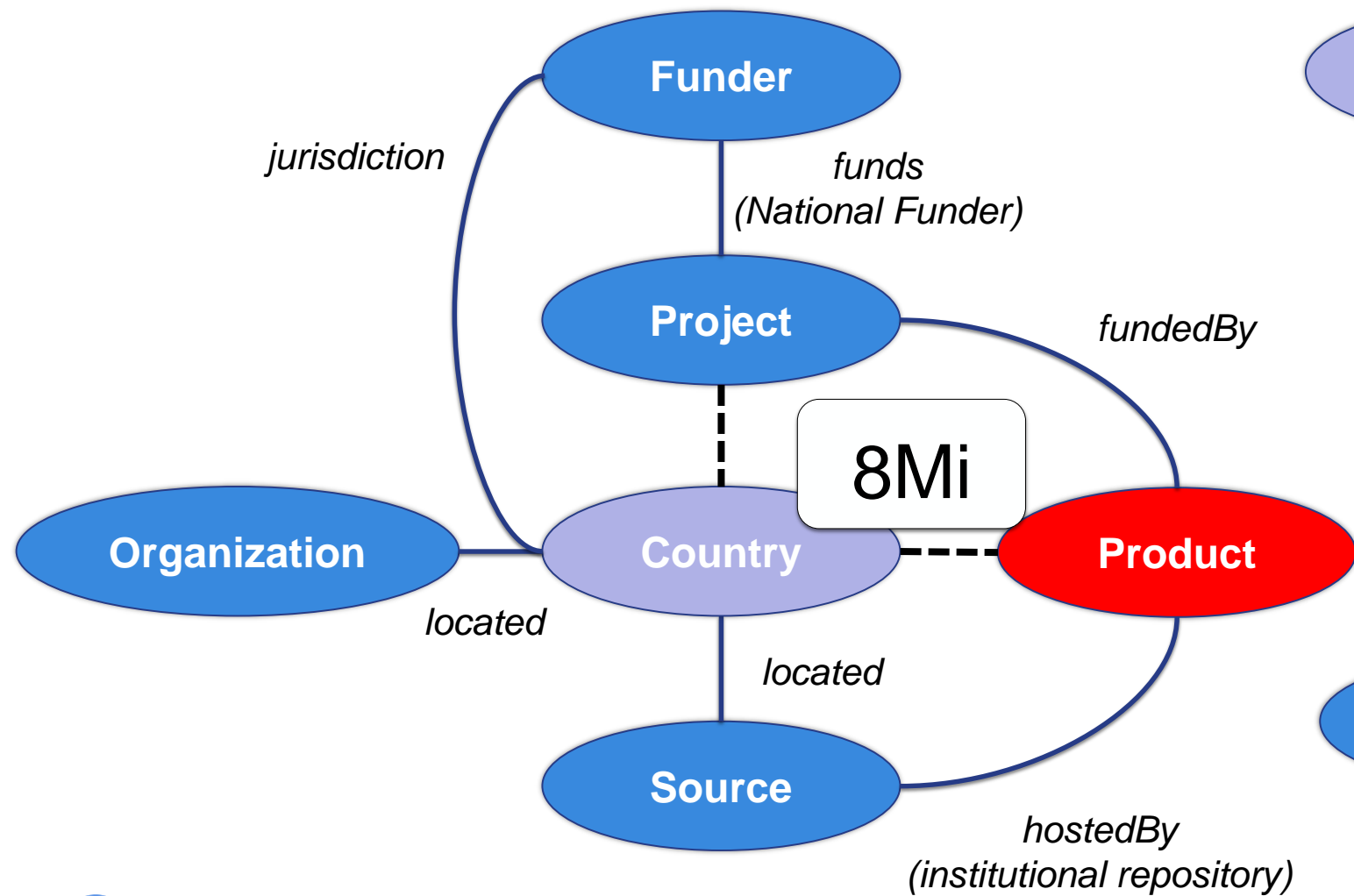
- **Document classification: 3.86M of pubs with at least one class assigned:**
  - **arxiv: 2.35Mi, meshEuroPmc: 3.64Mi, acm: 832k**
- **Document properties**
  - **New abstracts: 1.3Mi**
- **Document references**
  - **168.44M bibliographic references for 5.33M pubs**
- **Document external links**
  - **PDB reference extraction: 320k references (68k of unique pubs)**

# Bulk tagging

- **Associating products to communities:**
  - via Zenodo communities, projects, data sources, keywords

2.2Mi

# Context Propagation





# Strategic for Open Science



Making the research graph  
an **EOSC** resource

Open, Trusted, Complete, De-duplicated,  
Participatory, Transparent, Decentralized

## Actors

Institutions, research organizations, funders,  
content providers, researchers, SMEs, etc.



## Added value services

e.g. discovery services, monitoring services



# BETA Graph Open Consultation

- **September-October 2019:**
  - **OpenAIRE Research Graph open for consultation**
  - **Collecting feedback via Trello (operational end of September)**
- **November 2019:**
  - **OpenAIRE Research Graph in production**

<http://beta.explore.openaire.eu>



# Thank you!

**Paolo Manghi**  
**[paolo.manghi@isti.cnr.it](mailto:paolo.manghi@isti.cnr.it)**