

Data Sources and Persistent Identifiers in the Open Science Research Graph of OpenAIRE

Alessia Bardi

<https://orcid.org/0000-0002-1112-1292>

Institute of Information Science and Technologies - CNR

Jochen Schirrwagen

<https://orcid.org/0000-0002-0458-1004>

Bielefeld University



IDCC | Dublin | 19. February 2020



Agenda

Data sources in OpenAIRE - a heterogeneous landscape

Why, what and how of OpenAIRE Guidelines

Data processing pipeline

Role of PIDs in OpenAIRE

Value added services and the OpenAIRE Research Graph

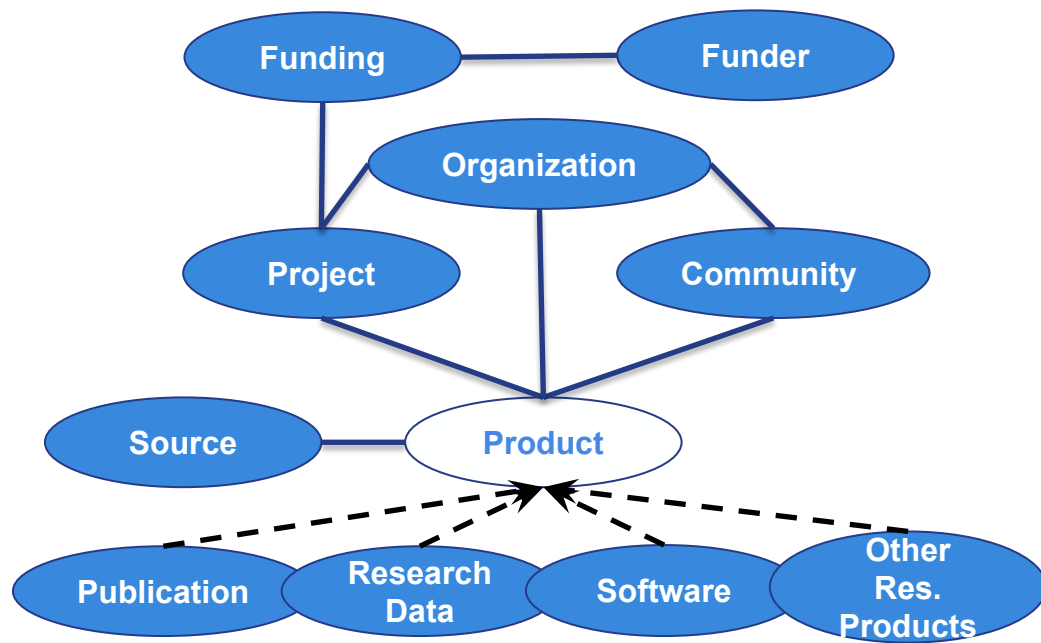
Materializing the Open Science Graph

CONNECT PROVIDE EXPLORE



Scientific product catalogue

MONITOR DEVELOP



Harvesting / Transformation

Deduplication

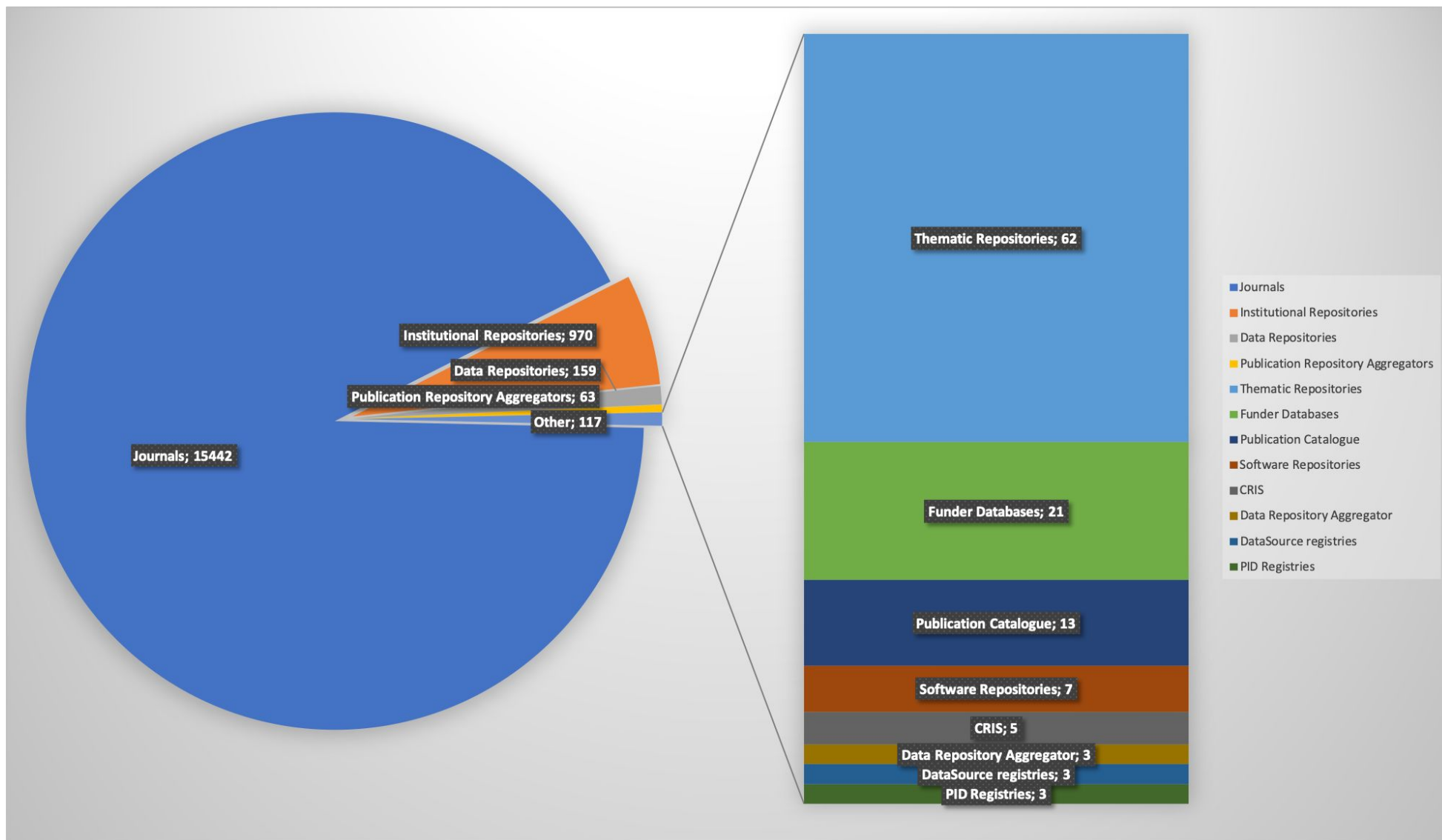
Mining

End-user feedback

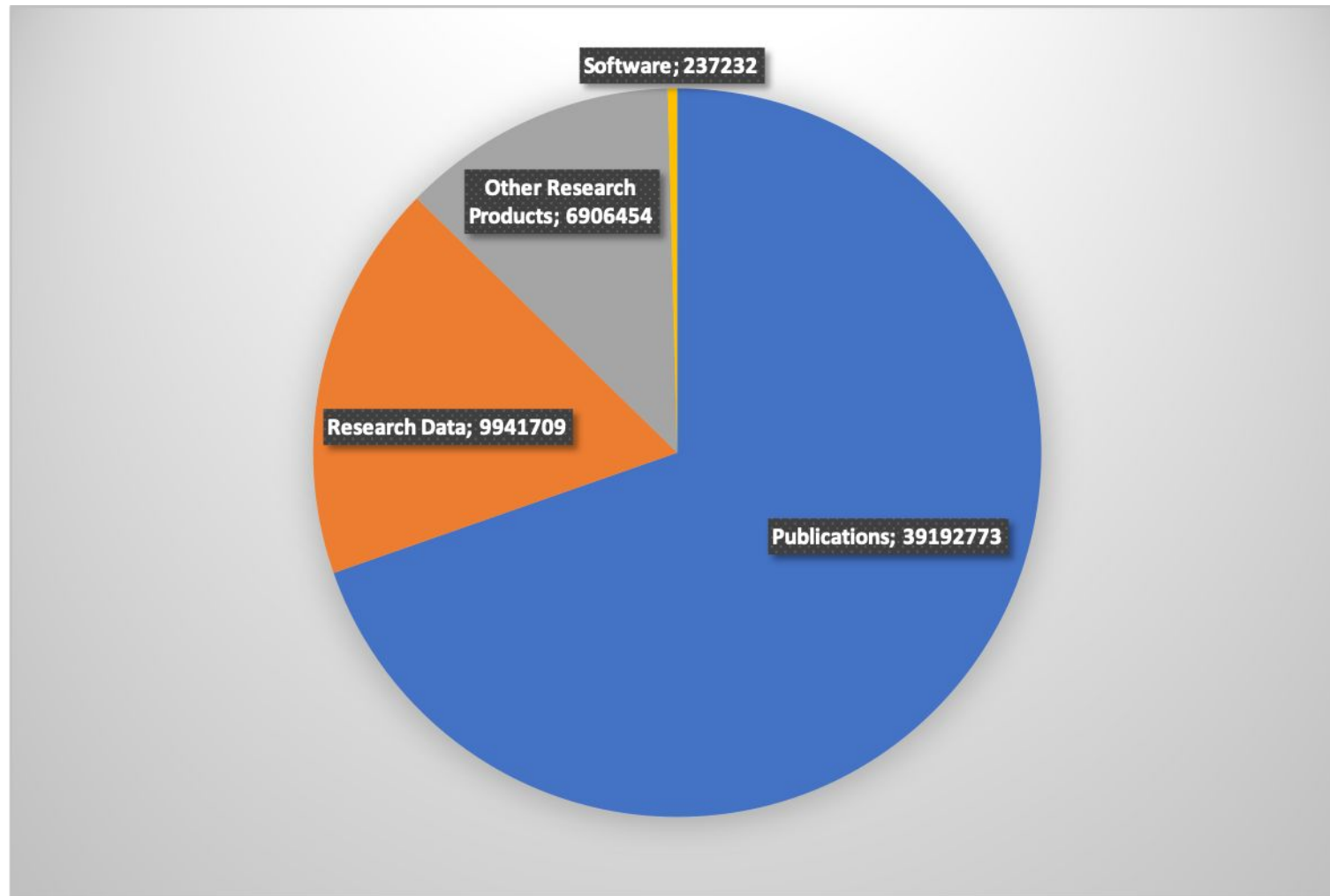


Data Sources and Content in OpenAIRE

DataSource Types and Quantity



Product Types and Quantity



OpenAIRE Guidelines

FROM Guidelines for Data Providers TO Guidelines for Open Science Content Providers

RESEARCH LITERATURE

Institutional, thematic, data &
catch-all Repositories
Aggregators
E-journals

RESEARCH SOFTWARE

Software Repositories
Catch-all Repositories



RESEARCH DATA

Data, catch-all, institutional &
thematic Repositories
Aggregators

OTHER RESEARCH PRODUCTS

Data & catch-all Repositories

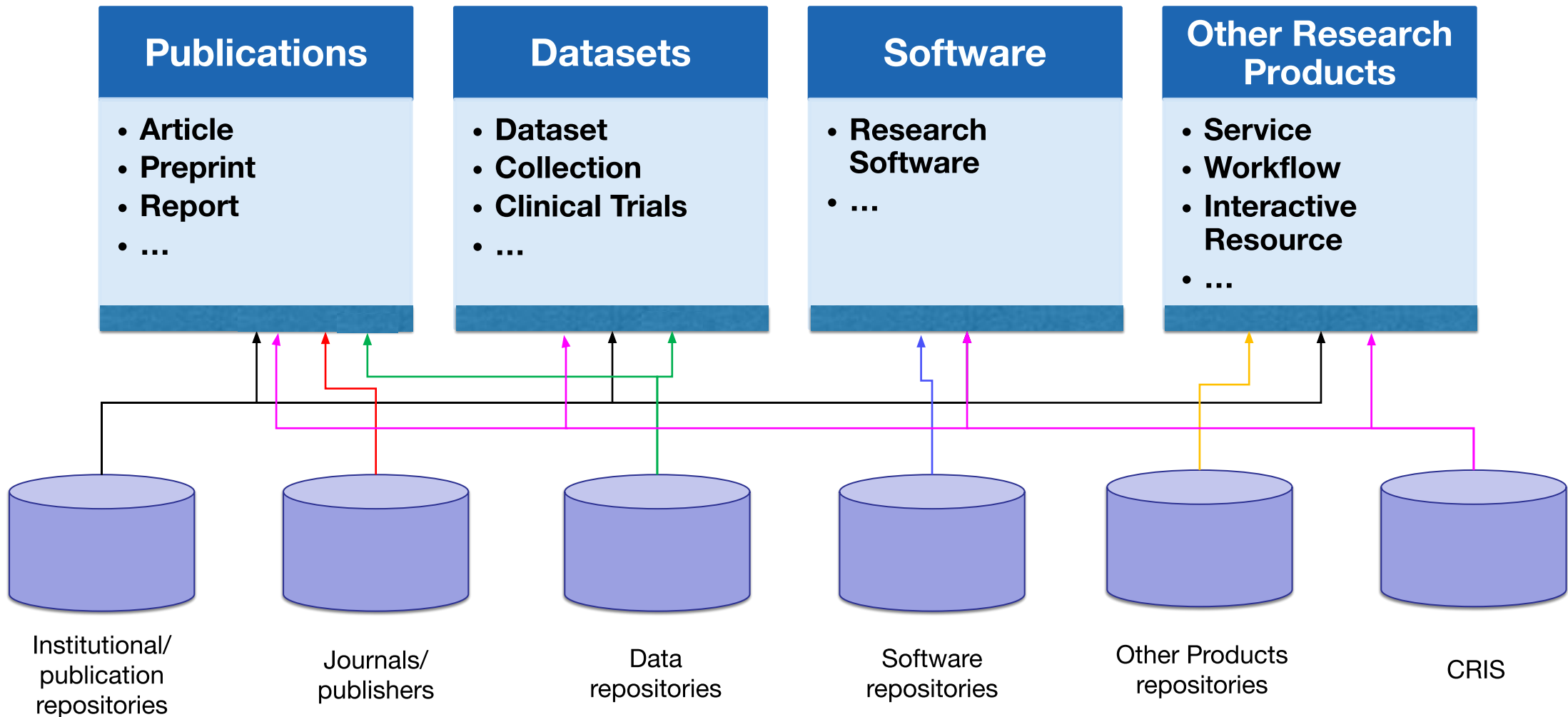
RESEARCH INFORMATION

CRIS

OpenAIRE Guidelines Objectives

- to support FAIR Data Principles
- to describe different kinds of research products by specific application profiles
- by re-using existing standards (Dublin Core, DataCite)
- by extending vocabularies when necessary
 - e.g. for PID types
- to facilitate value added services
 - monitoring, reporting, usage metrics, broker (enrichment)

Aggregation of Research Results by OpenAIRE



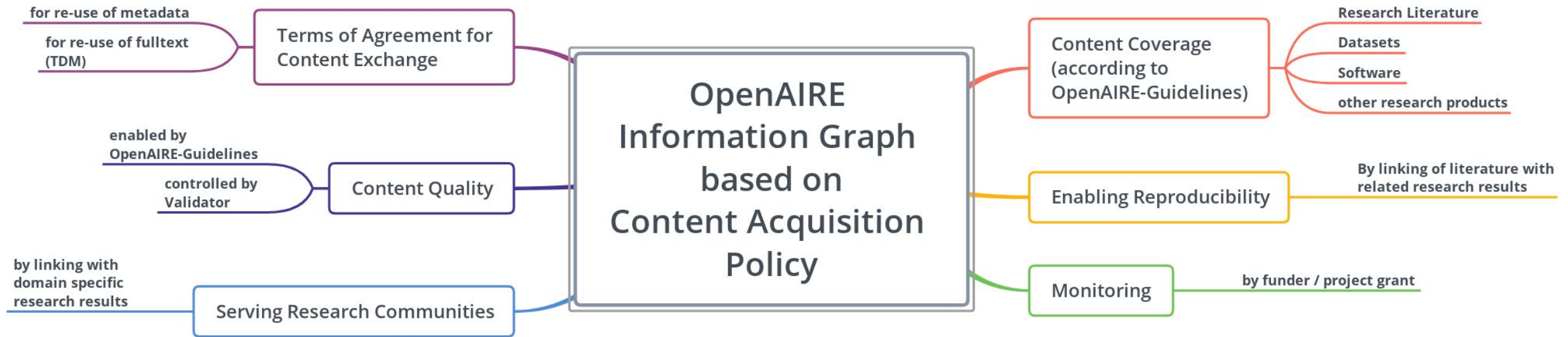
“

OpenAIRE accepts the metadata records of all scientific products whose structure respect the model and semantics as expressed by the OpenAIRE guidelines. This means that both Open Access and non-Open Access material will be included and links to other products will be resolved where this is possible (i.e. the provided PIDs have a resolver)

*as stated in the
Content
Acquisition Policy*

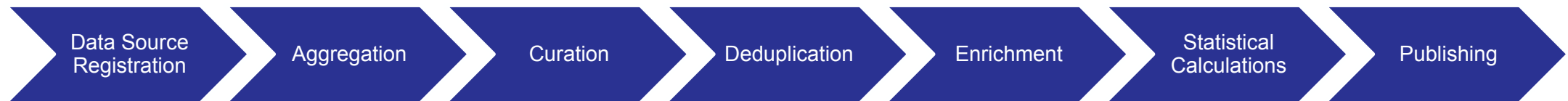
”

Objectives of OpenAIRE's Aggregation Policy



Content Acquisition Policy released 05-Oct-2018, <https://doi.org/10.5281/zenodo.1446408>

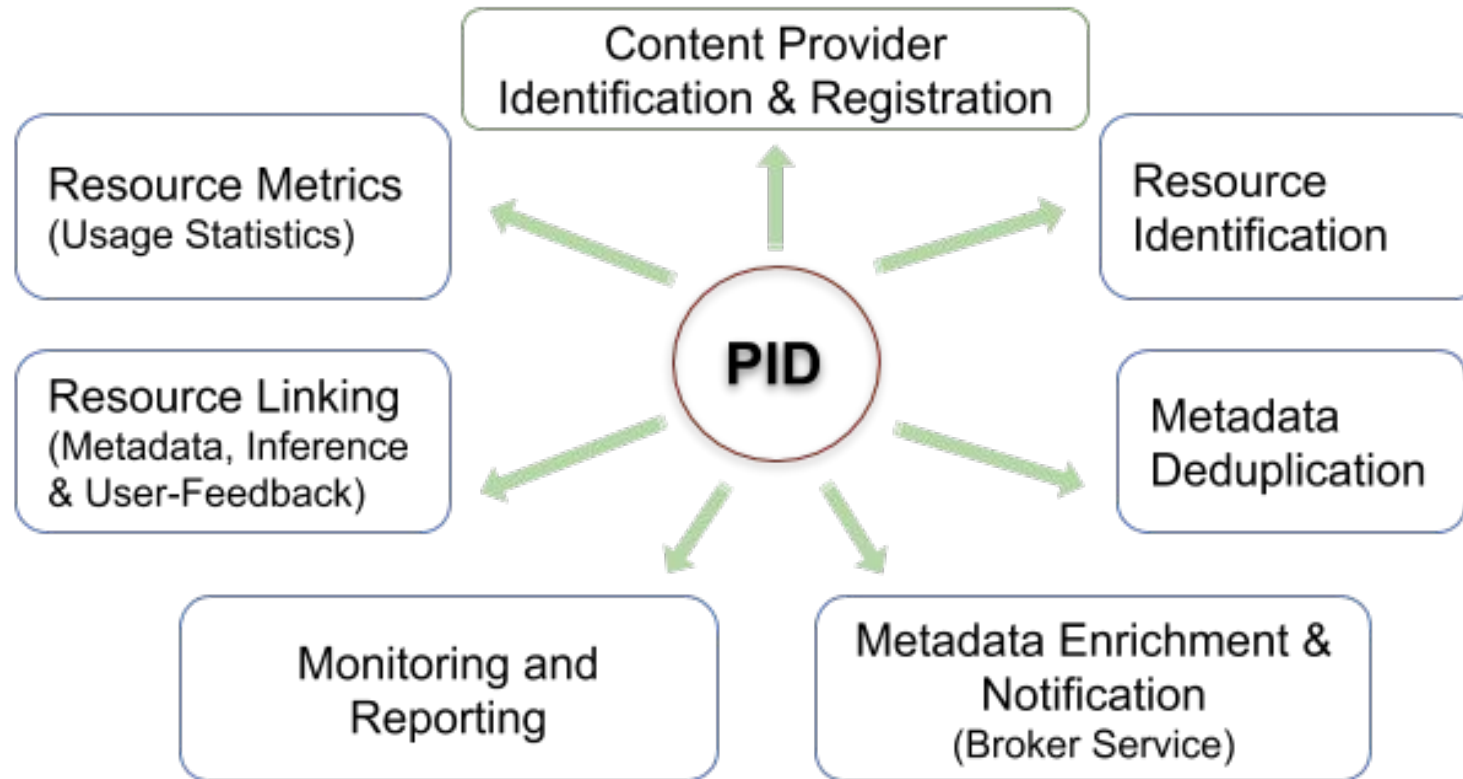
Data Processing Pipeline



Metadata Quality Characteristics

Timeliness	Metadata should reflect current state incl. recent changes
Completeness	All relevant statements are leveraged
Accuracy	Information is veritable, correct, non-contradictory; PIDs are valid and resolvable
Legibility	Metadata descriptions are comprehensible
Consolidation	No duplicates; contextualization, i.e. linking with other related records (citations, versions, ...)
Wastelessness	No test records, no records out of scope, ...
Format conformance	Compliance with format standards, utilization of vocabularies / thesauri

Role of PIDs in OpenAIRE



PIDs in Value-Added Services

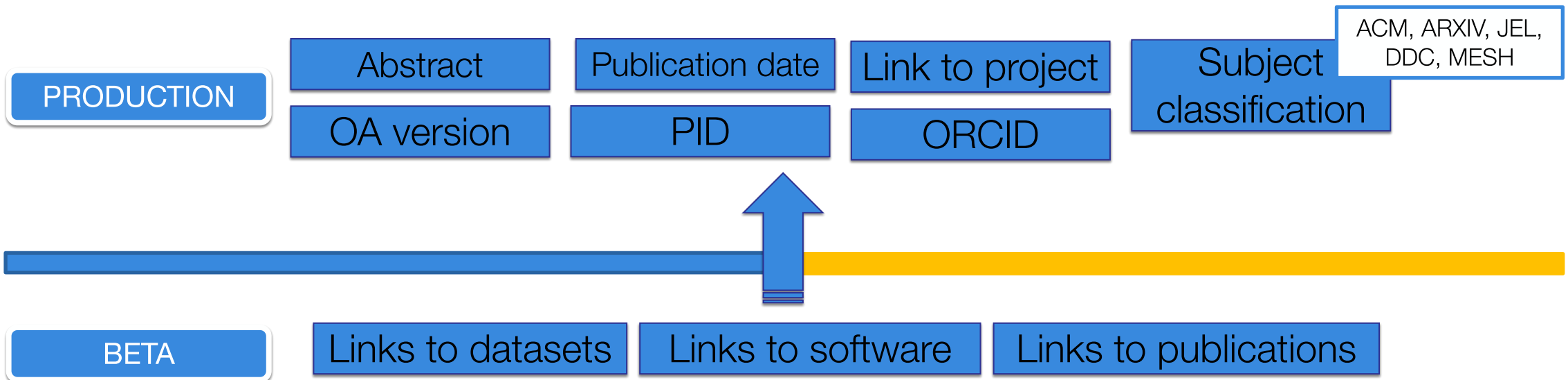
OpenAIRE Broker Service

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 Link-Service

Publication

Teaching Research Data Management for Students

ARTICLE ENGLISH OPEN

Wiljes, Cord ; <https://orcid.org/0000-0003-2528-5391>; Cimiano, Philipp; (2019)

Publisher: Ubiquity Press
Journal: Data Science Journal (issn: 1683-1470, eissn: 1683-1470)
[Publisher copyright policies & self-archiving](#)

Related identifiers: [doi: 10.5334/dsj-2019-038](#)

Subject: Good Scientific Practice | teaching; research data management; RDM | Science (General) | Reproducibility | Q1-390 | research data management | Educatioun | teaching |

LINK THIS PUBLICATION TO...

Projects Research results Communities

Download from

[Data Science Journal via Data Science Journal \(Article, 2019\)](#)

Dataset

LINK TO RESEARCH RESULTS ↕

Start searching Research Results and add them to the Basket to Link

Research Data Management Course: Survey Data

[OPENAIRE \(2\)](#) [CROSSREF \(10,271,522\)](#) [DATACITE \(6,040,308\)](#) [ORCID \(0\)](#)

Entity Type (4)

2 results, page 1 of 1

dataset

Research Data Management Course: Survey Data

Authors Wiljes, Cord
Publisher Bielefeld University
Published in 2018

LINKS BASKET

SOURCE

[Teaching Research Data Management for Student...](#)

LINK TO (1)

Entities to link with the sources

[Upload a DOI's CSV file](#)

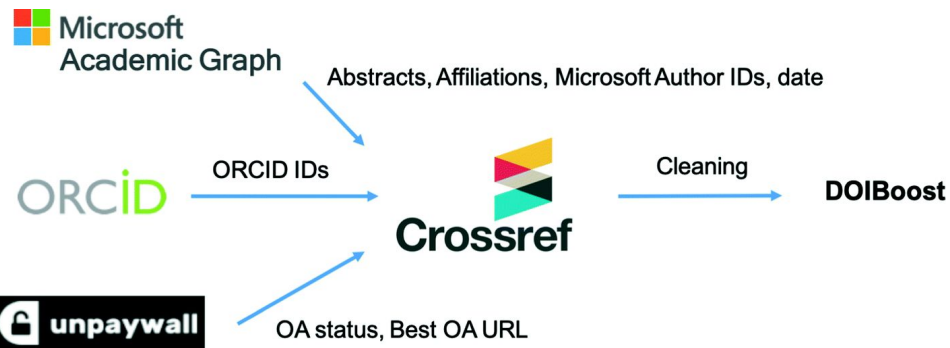
dataset

Research Data Management Course: Survey Data

Authors Wiljes, Cord

And A Few More Examples ...

DOIBoost



source: La Bruzzo S., Manghi P., Mannocci A. (2019). OpenAIRE's DOIBoost - Boosting CrossRef for Research. In: Digital Libraries: Supporting Open Science. IRCDL 2019, Communications in Computer and Information Science, vol.988, Springer International Publishing.

Scholar Explorer

The Data Literature Interlinking Service

- accepts publications-data or data-data links
- builds a de-duplicated graph
- provides access (search / query) to the graph

open @PC

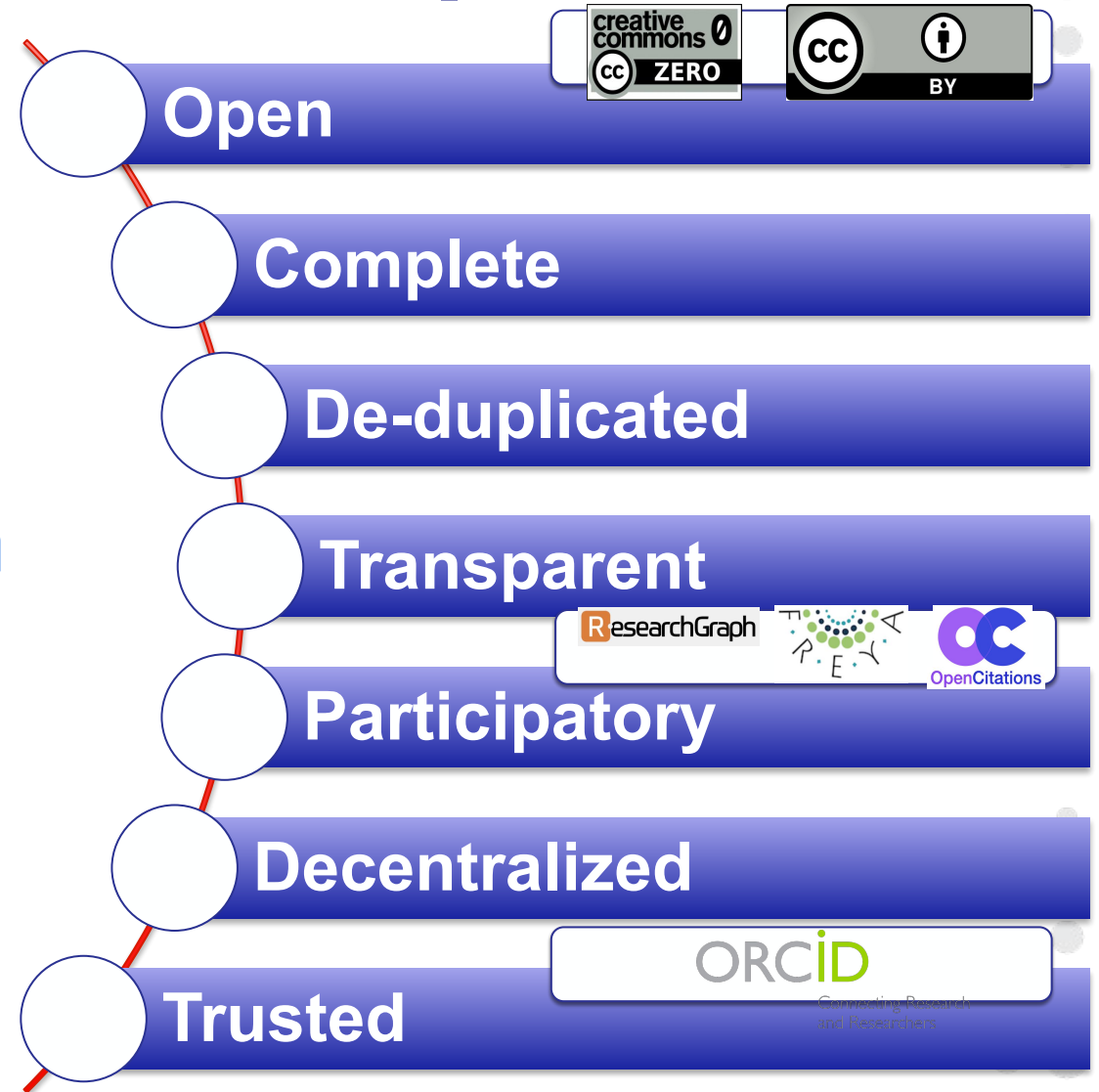
Open Data on Article Charges

- datasets on fees for OA-articles
- OA-articles minted with DOI
- match with OpenAIRE Research Graph
- Cost information per publications per project per funder

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**



Thank you!

Alessia Bardi

Andreas Czerniak

Aenne Loehden

Paolo Manghi

Mike Mertens

Najla Rettberg

Jochen Schirrwagen

