How Identifiers Can Help you in Open Science

OSFair

17 September 2019

Horizon 2020
European Union funding
for Research & Innovation

Agenda

Introductory Presentations (40 mins)
- A PID for everything & why would you use them? (Helena Cousijn, Ivo Wijnbergen)
- Research Graphs: Getting the best out PIDs (Paolo Manghi)
- Creating a PID policy and good practices (Jessica Parland-von Essen)
- Information and training materials from the projects (Frances Madden)

Drafting an approach on how to (further) promote PIDs in your organisation (35 mins)

How to design messages for your communities (30 mins)

Action Plan: Three things you will do after this workshop (10 mins)

Are you sitting in the right place?

A PID for everything & why would you use them?

Helena Cousijn (DataCite) & Ivo Wijnbergen (ORCID)
17 September 2019

DataCite
FIND, ACCESS, AND REUSE DATA

ORCID

What is a persistent identifier?

persistent identifier

an organization
made a
promise to
keep it alive

globally unique
string

(known as PIDs to their friends)

How PIDs work (in a nutshell)

PIDs are typically backed by a **registry** that indicates what item is being identified. Different kinds of PIDs have varying degrees of descriptive metadata.

PIDs today are often expressed as **URLs**, and the registry indicates where that URL should ultimately resolve. That PID will always point to the correct item even if the item's location changes.

What kind of stuff gets a PID?

Journal articles. via Crossref (https://crossref.org)

People. via ORCID. (https://orcid.org)

Data, software, and other stuff. via DataCite. (https://datacite.org)

Research organizations. via ROR. (https://ror.org)

And others.

DOIs and ORCID IDs are persistent identifiers

DOIs (digital object identifiers) are one type of persistent identifier.

https://doi.og/10.5072/abc123 ← If you've seen this on a research paper, you've seen a persistent identifier..

An ORCID ID is also a persistent identifier, based on a 16-digit ISNI number.
https://orcid.org/0000-0001-5540-748X

Often PIDs are displayed and linked to the source by URLs

. . . but what can PIDs *do*?

PIDs Disambiguate

Robin Dasler

ORCID iD

https://orcid.org/0000-0002-4695-7874

Print view ?

Also known as

RH Dasler, RL Dasler, RL Howard, Robin Howard

Other IDs

ResearcherID: N-9035-2013

PIDs Link

This article references these other things.

References

Abd Ellah and Abouelmagd, 2016   N.H. Abd Ellah, S.A. Abouelmagd
Surface functionalization of polymeric nanoparticles for tumor drug delivery: approaches and challenges
Expert Opin. Drug Deliv., 1–14 (2016),
10.1080/17425247.2016.1213238
Google Scholar

Abouelmagd et al., 2016   S.A. Abouelmagd, F. Meng, B.-K. Kim, H. Hyun, Y. Yeo
Tannic acid-mediated surface functionalization of polymeric nanoparticles
ACS Biomater. Sci. Eng. (2016), p. 6b00497,
10.1021/acsbiomaterials.6b004
Google Scholar

Ahmed et al., 2016   S. Ahmed, S. Annu, S.S. Yudha
Biosynthesis of gold nanoparticles: a green approach
J. Photochem. Photobiol. B: Biol., 161 (2016), pp. 141-153,
10.1016/j.jphotobiol.2016.04.034
Article   Download PDF   View Record in Scopus
Google Scholar

Akhavan et al., 2011   O. Akhavan, R. Azimirad, S. Safa, E. Hasani

PIDs make research FAIR

Findable	To be Findable any Data Object should be uniquely and persistently identifiable.
Accessible	Data is Accessible in that it can be always obtained by <u>machines and humans</u>
Interoperable	Data should include qualified references to other data, and the format should use a shared vocabulary.
Reusable	To achieve this, data should comply with the above, and refer to their sources with rich metadata and provenance.

Good start, but we want more

By connecting everything, you can see the true power of PIDs

Researchers, institutions, publications, datasets, and more are already interconnected in real life, and this can be reflected and tracked through PIDs
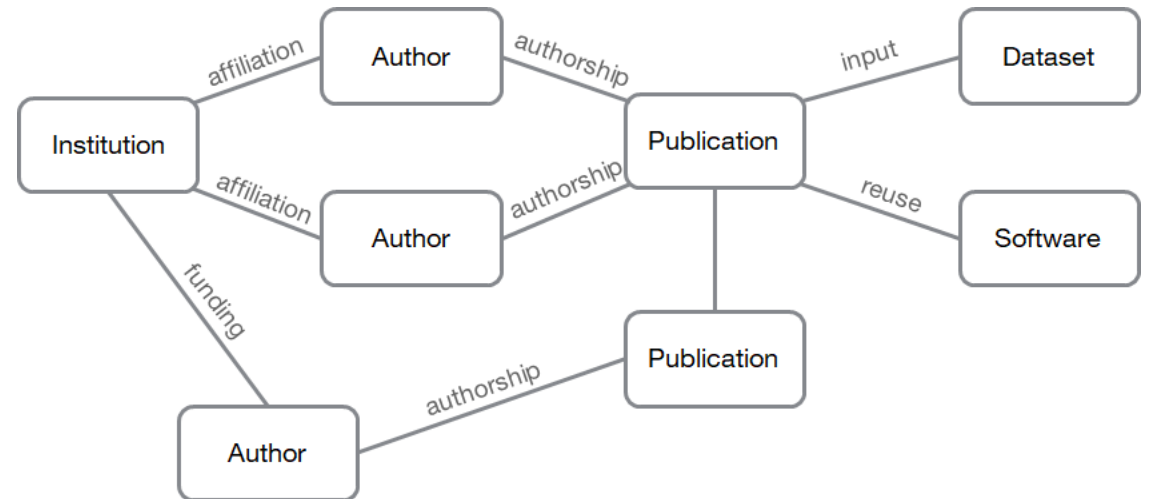
And what can you do?

Step 1: Give PIDs to your stuff

It's hard to connect things when we don't know they exist.

So get an ORCID iD for yourself → https://orcid.org

Give DOIs to your data and software → https://datacite.org, https://guides.github.com/activities/citable-code/

Put your reports and white papers into a repository that gives out PIDs → https://repositoryfinder.datacite.org or your institutional repository

Step 2: Tell your PIDs about your other PIDs

Include relevant related PIDs in the metadata for your software, dataset, and paper PIDs, even if your repository says they're optional.

In Zenodo (for example), it looks like this:

Step 3: Share these connections with the community
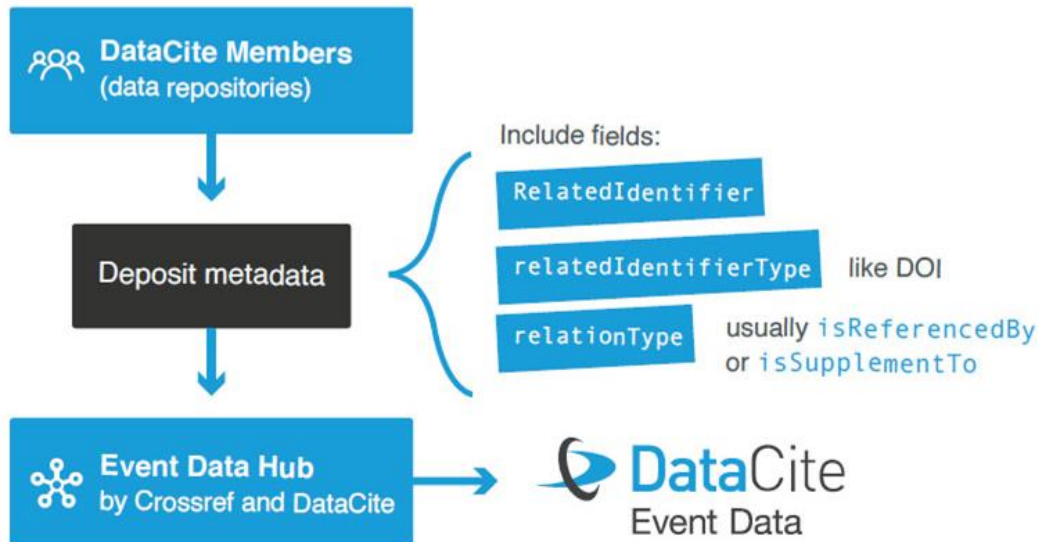
Interested in using this information? Find out more at: https://support.datacite.org/docs/eventdata-guide
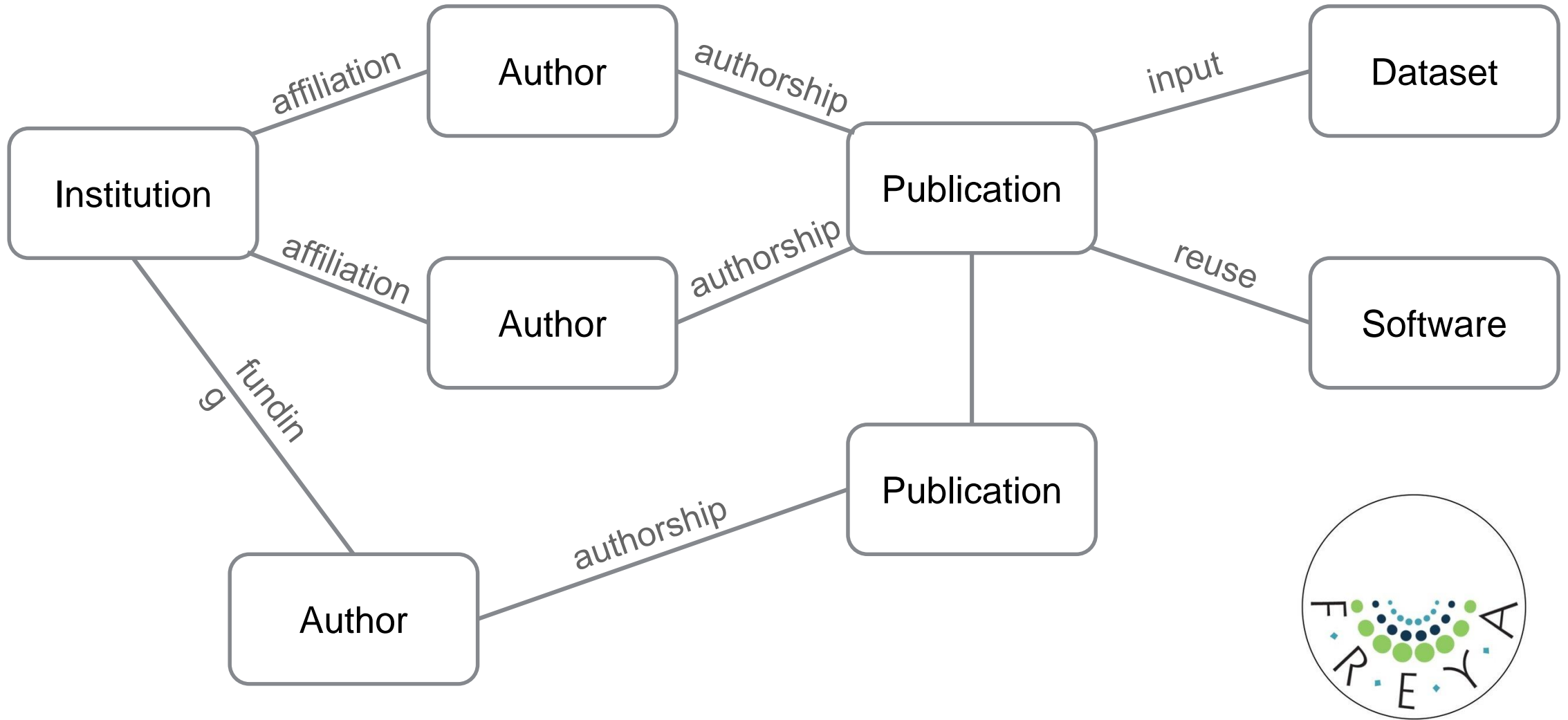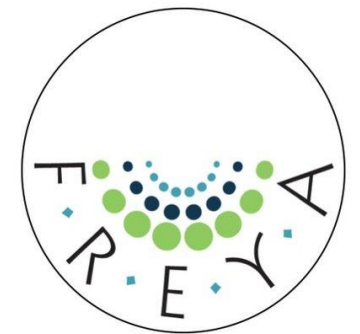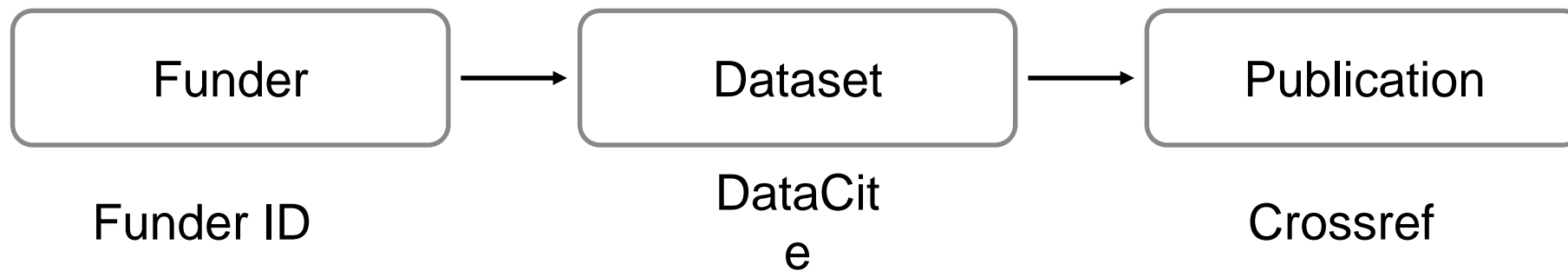
All this information feeds into a graph

Who are all the co-authors of a given researcher?

Person	→	Works	→	People
ORCID		Crossref EMBL-EBI DataCite		ORCID
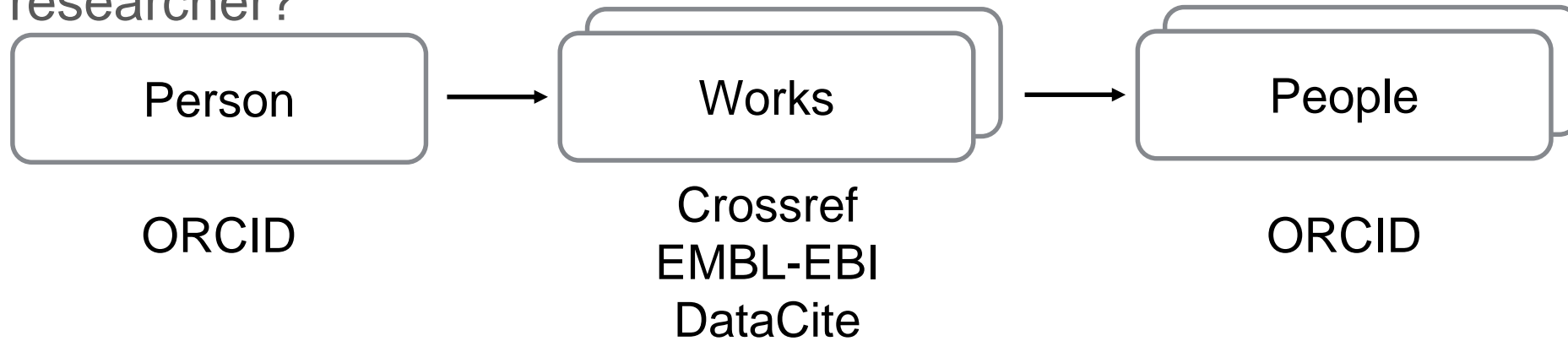
Show all datasets funded by the European Commission that have been cited by a journal article

Funder	→	Dataset	→	Publication
Funder ID		DataCite		Crossref

Which can be used to answer new questions

If you take the first steps, we'll do the rest!

Paolo Manghi

Insitute of Information Science and Technologies

National Research Council

Pisa, Italy

Research Graphs: Getting the Best out of PIDs

OpenAIRE

What's a research graph?

- **It's a graph…**

  So it must connect some objects with some links!

- **It's a research graph…**

  So objects and links must be related with research entities!

- **Which are such research entities? Do links have a meaning?**

  Depends on targeting use-cases and customers!

Some examples of research graphs

FREYA

Datasets, authors, publications, funder

With PIDs

ResearchGraph

Datasets, researchers, grants (Australian), publications

With PIDs
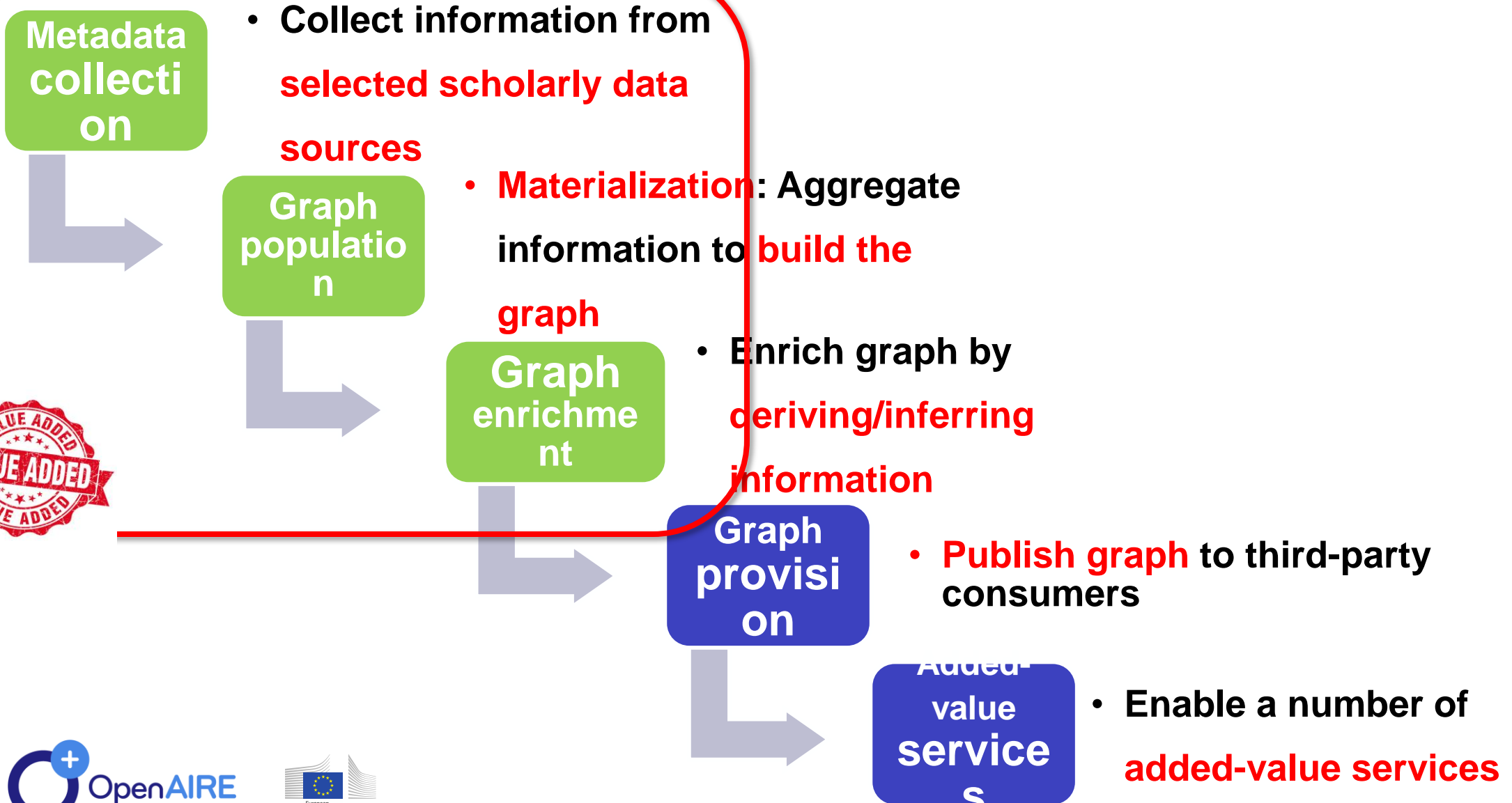
OpenCitations

Publications

With PIDs

OpenAIRE Research Graph

Publications, Datasets, software, other products, projects, funders, oganizations, data sources, research communities

With PIDs and URLs

Common use-case driven methodology

Metadata collection

- Collect information from **selected scholarly data sources**

Graph population

- **Materialization**: Aggregate information to **build the graph**

Graph enrichment

- Enrich graph by **deriving/inferring information**

Graph provision

- **Publish graph** to third-party consumers

Added-value services

- Enable a number of **added-value services**

VALUE ADDED

OpenAIRE

European Commission

Research Graph magics

- Discovery and recommendations
- Reproducing
- Scientific rewarding
- Science assessment
- Open Science Monitoring
- Research strategies planning
- ….

How can we ensure to get the best out of PIDs?

Decentralization

Exchange information with other Research Graphs
Preserve value-added information by enriching scholarly data sources

Quality

Provenance of data source PIDs

Shared understanding of quality

Openness

Licensing metadata as CC-0 as possible

Interoperability across graphs

Open Science Graphs for FAIR data RDA IG

Interoperability of research graphs

OpenAIRE Research Graph

Harvesting metadata
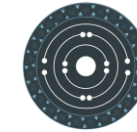
... **and more**
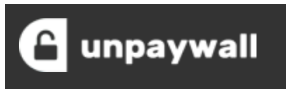
10K sources

... **and more**

... **and more**

... **and more**

... **and more**

... **and more**

Harvesting metadata records

Dataset with link to publication

Harvest

Harvest

Publication without links

Notify link to dataset to data source

Harvested content

- Records
  - 450Mi

- Links
  - 130Mi

OpenAIRE RESEARCH GRAPH

ScholeXplorer
The Data Literature Interlinking Service

OpenAIRE

European Commission

GUIDE LINES

GUIDE LINES
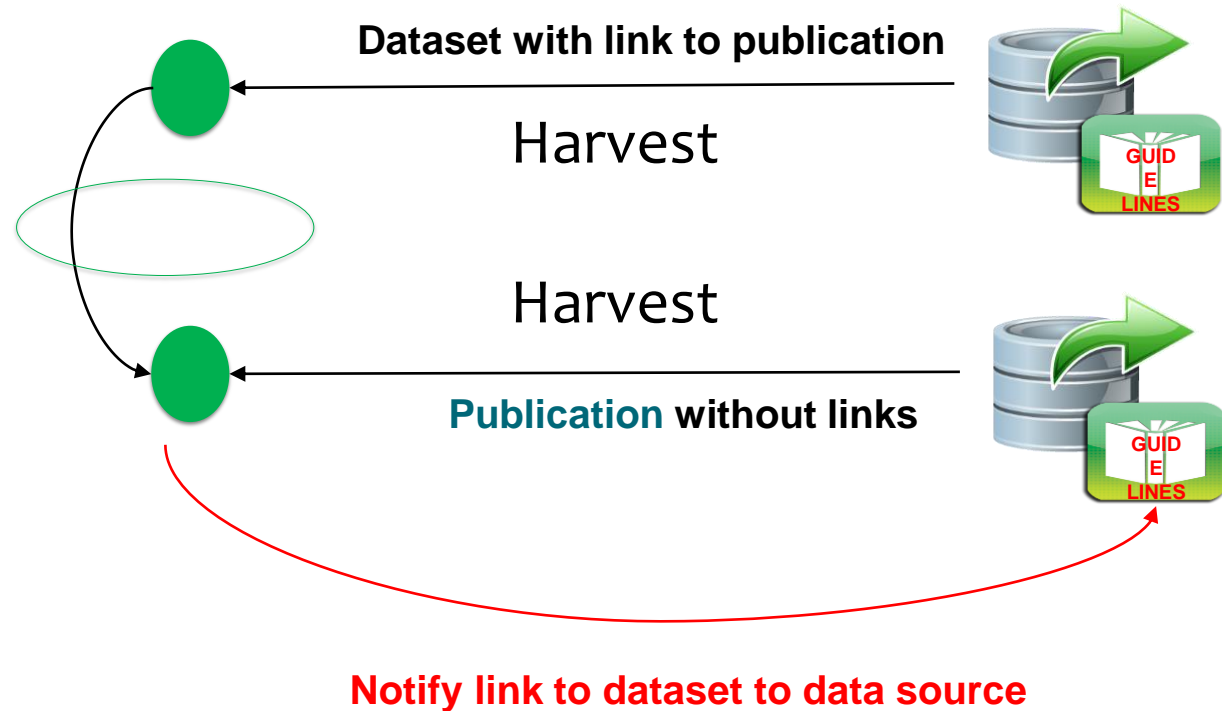
Text-mine full-text of Open Access articles

11Mi OA full-texts

Research **software with DOI and no links**

Link publication-software

Harvest

Publication without links
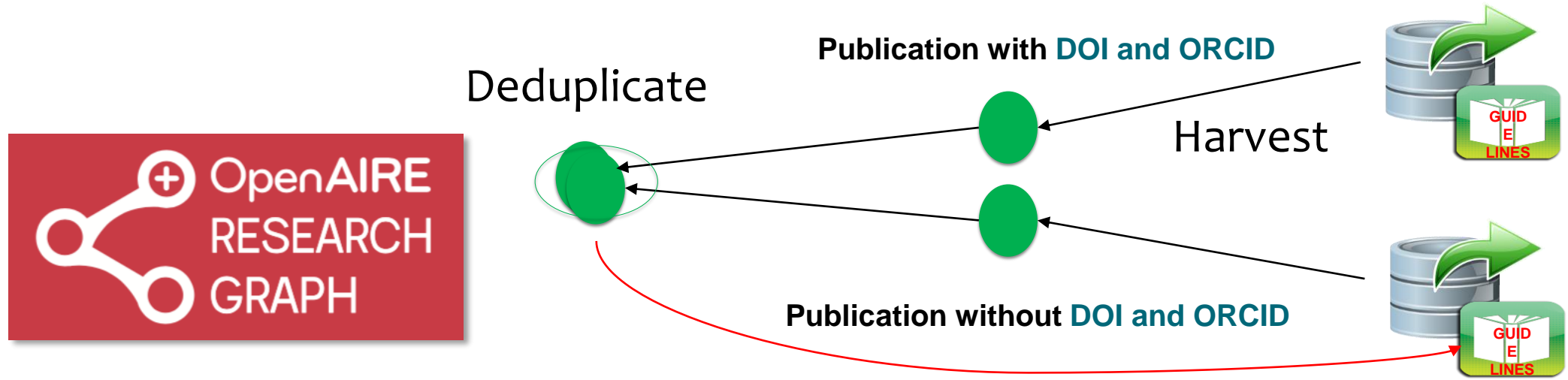
Notify link to software to data source

Links

- Text-mined links
  - 400Mi
- Text-mined values
  - 178Mi
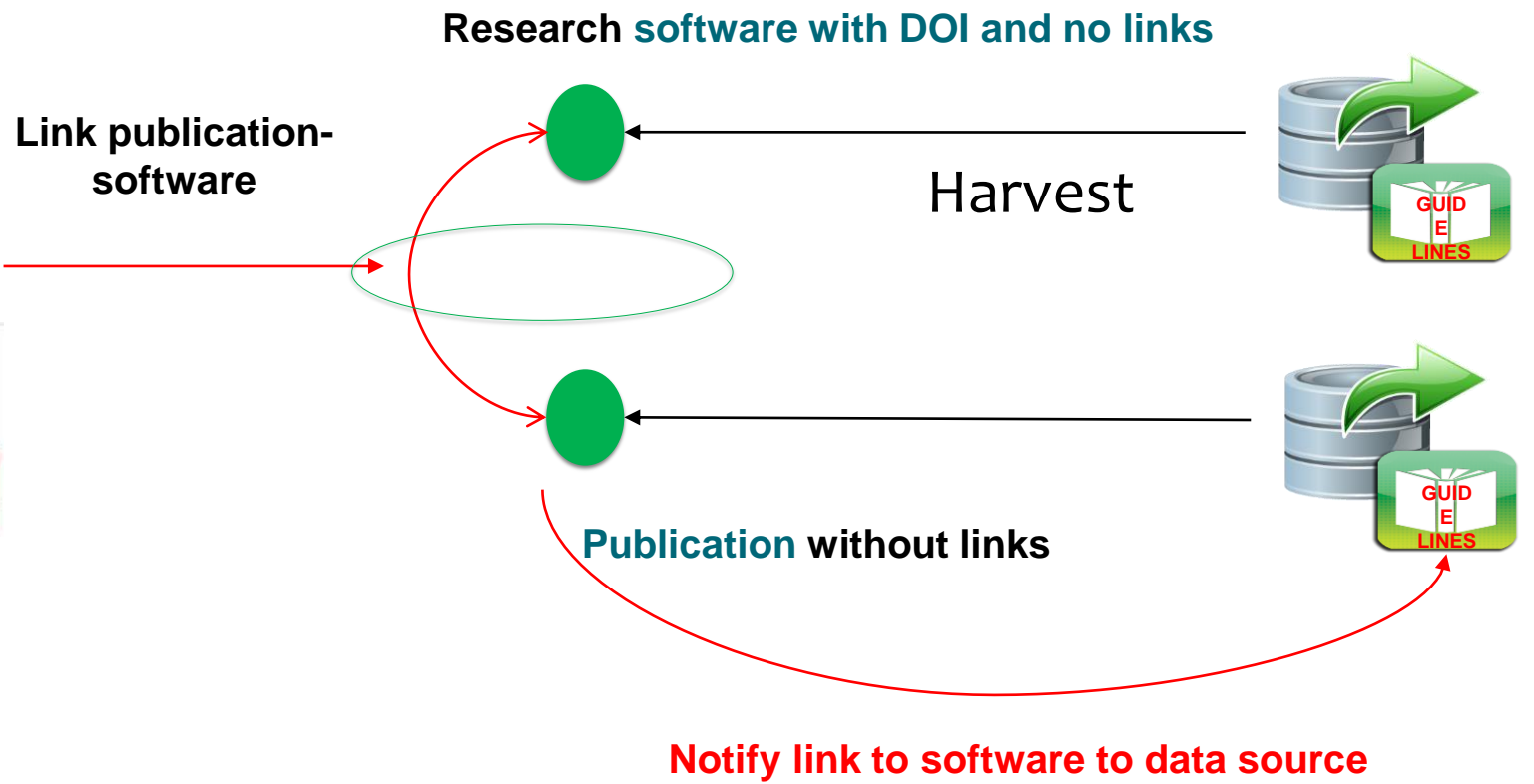
Deduplication

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

Deduplicate

Publication with DOI and ORCID

Harvest

Publication without DOI and ORCID

101mi publications, 8mi research data, 8mi other research products, 201K software
from 9,900 content providers and 28 funders linked together
for an integrated discovery of research outcomes

Notify DOI and ORCID for the record to data source

OpenAIRE

European Commission

Propagation via links

Project, countries, and communities information
from publications to other products

Publication with ORCIDs with link to datasets

8Mi

OpenAIRE RESEARCH GRAPH

Publication and dataset author names are the same

Datasets Inherits ORCIDs

Harvest

GUIDE LINES

Dataset without ORCIDs

GUIDE LINES
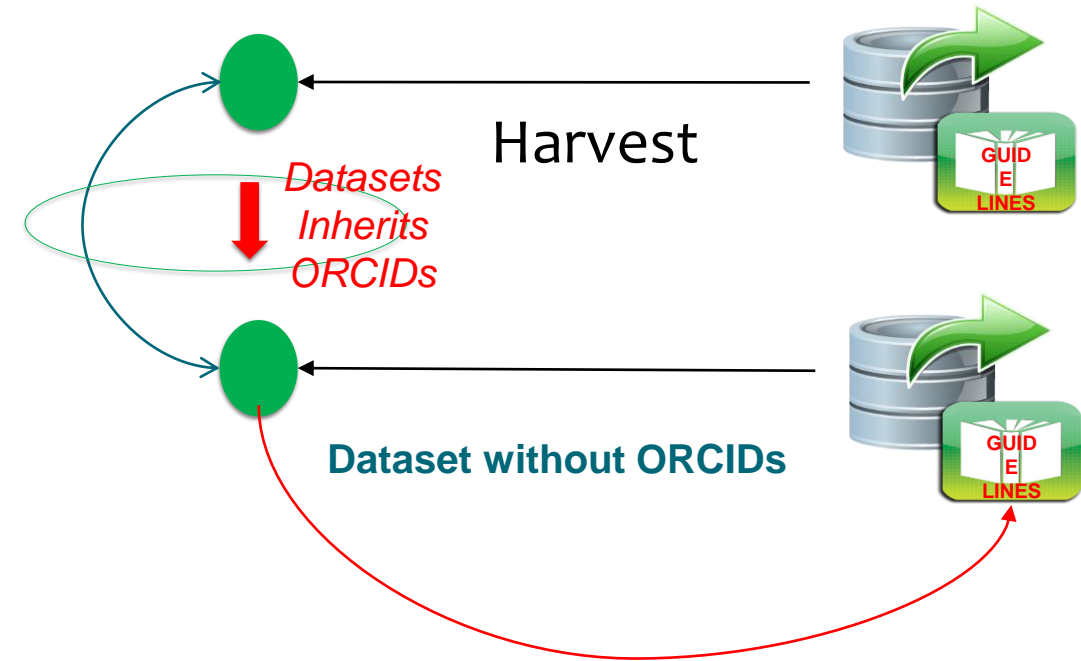
Notify ORCID IDs to data repository

OpenAIRE

European Commission

Interoperability and decentralization

Interconnecting
Research Graphs

Decentralization

CONNECT PROVIDE EXPLORE

DEVELOP MONITOR

RESEARCH **DATA ALLIANCE**
EUROPE

OpenAIRE | PROVIDE
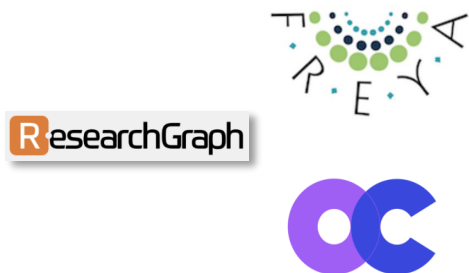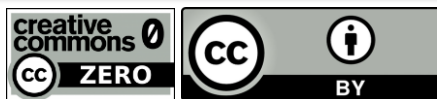
BETA Graph Open Consultation

- **September-October 2019:**

  OpenAIRE Research Graph open for consultation

  Collecting feedback via Trello (operational end of September)

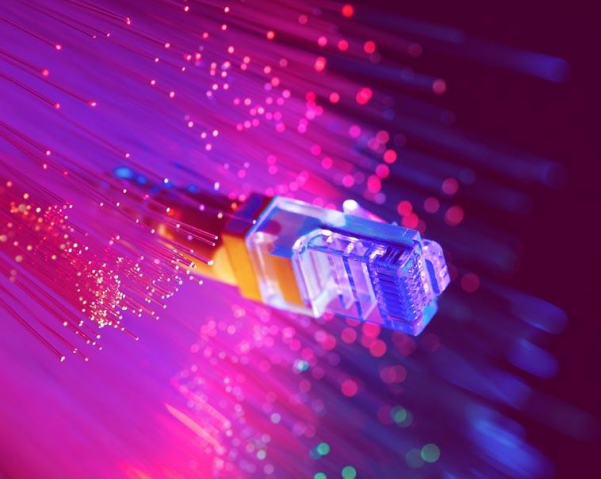- **November 2019:**

  OpenAIRE Research Graph in production

  http://beta.explore.openaire.eu

Thank you!

Paolo Manghi

paolo.manghi@isti.cnr.it

Creating a PID policy

Jessica Parland-von Essen. https://orcid.org/0000-0003-4450-3906
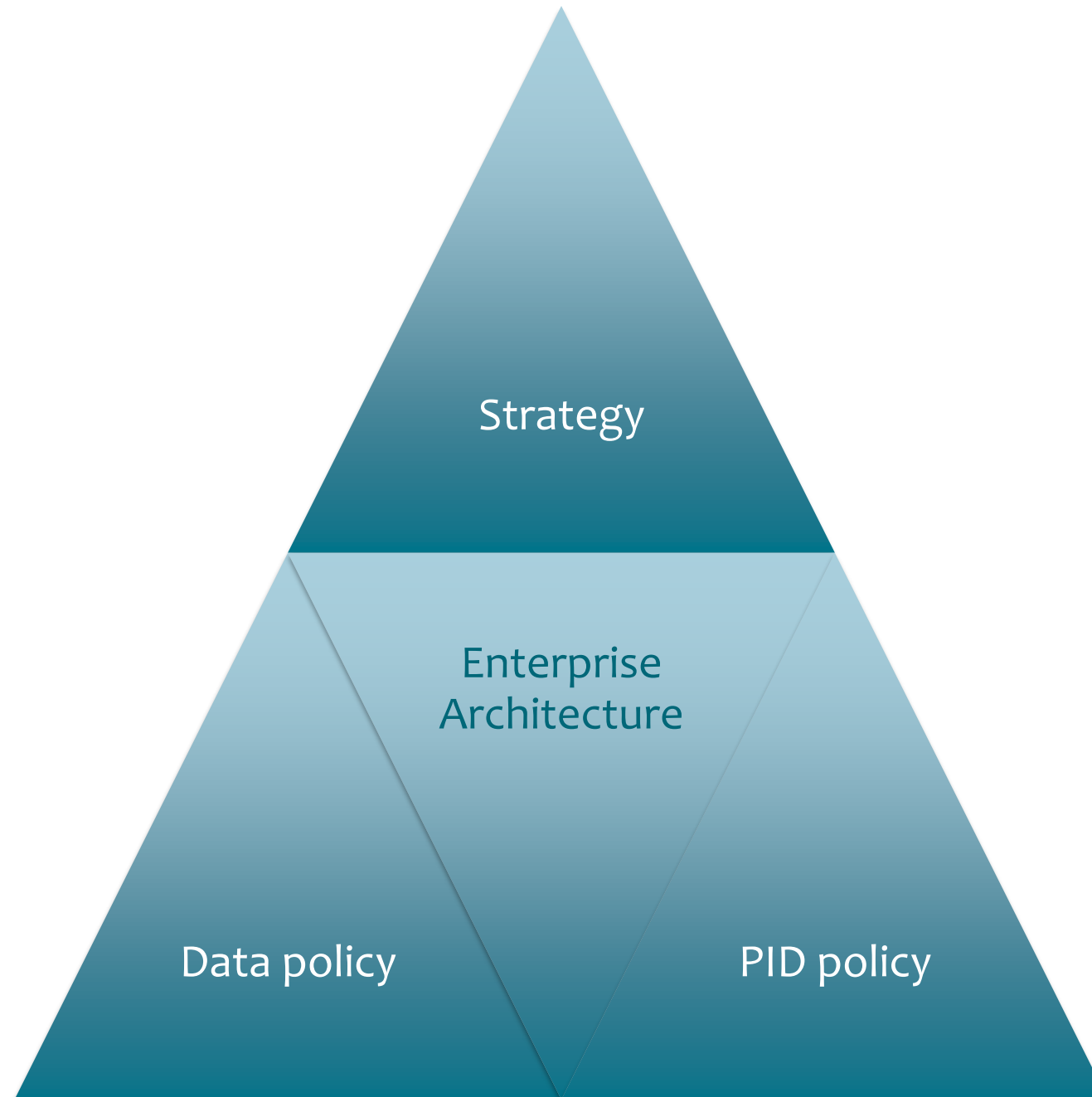
FAIRsFAIR in a nutshell
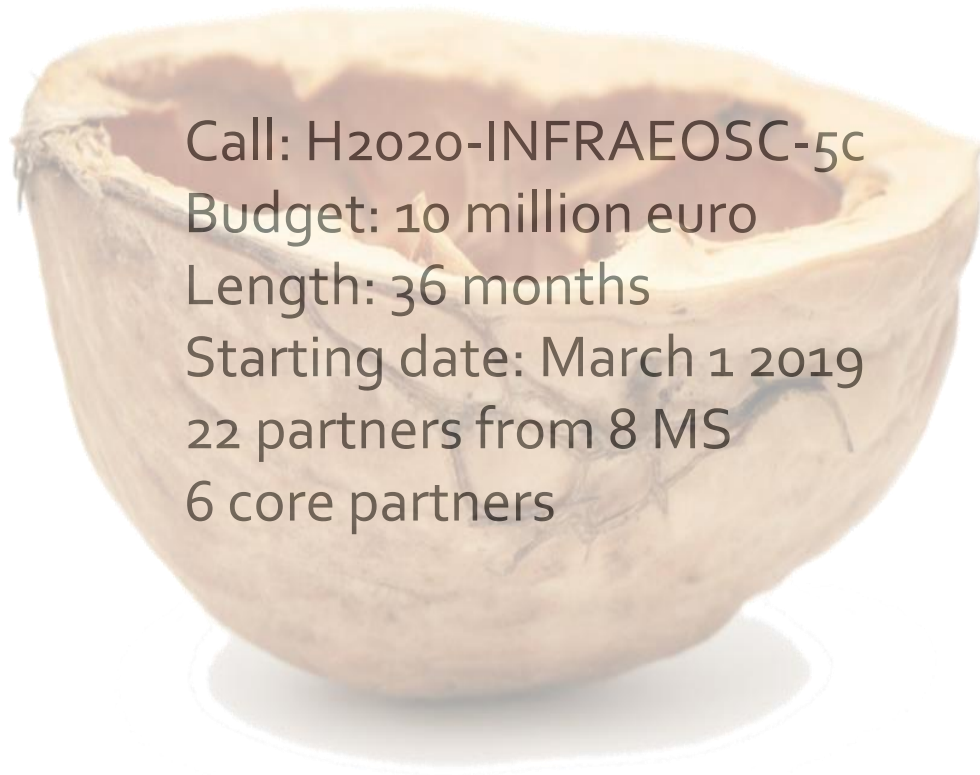
Call: H2020-INFRAEOSC-5c
Budget: 10 million euro
Length: 36 months
Starting date: March 1 2019
22 partners from 8 MS
6 core partners

- Semantic interoperability and sustainability are key features to make FAIR work

- Persistent identifiers are in the DNA of FAIR

- FAIR research data is also linked data

- Research data is often complex and dynamic

- The life cycle and deletion often not sufficiently planned and documented

- Traditional research dataset publications are often "article like", static outputs
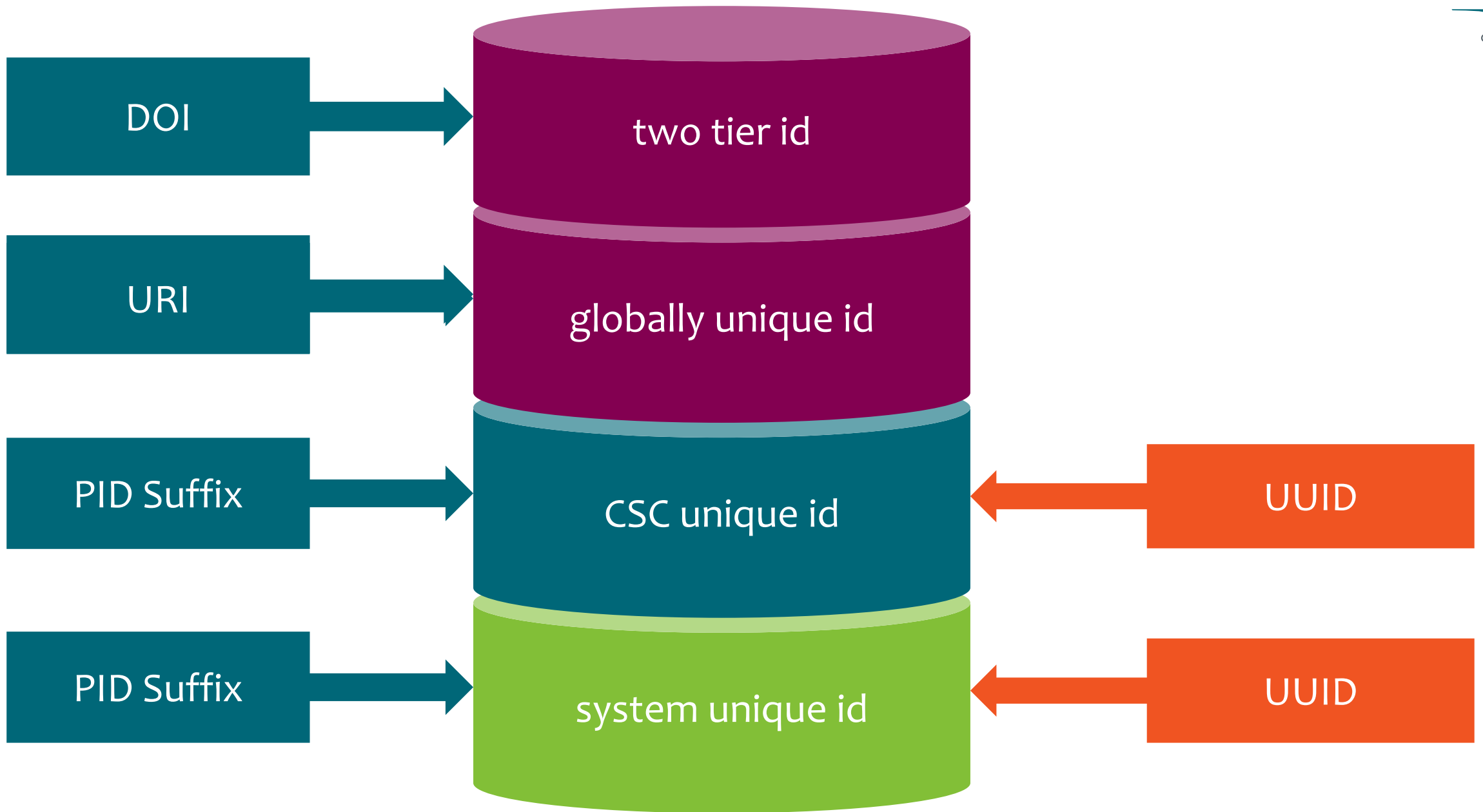
- FAIRsFAIR has a wide definition of data

Research Data Types

ACTIVE DATA
Raw, continuously updated

Documentation, validation

DYNAMIC RESEARCH DATA
Version controlled, possible to cite

Research

RESEARCH DATASET PUBLICATION
Immutable

https://doi.org/10.23978/inf.77419
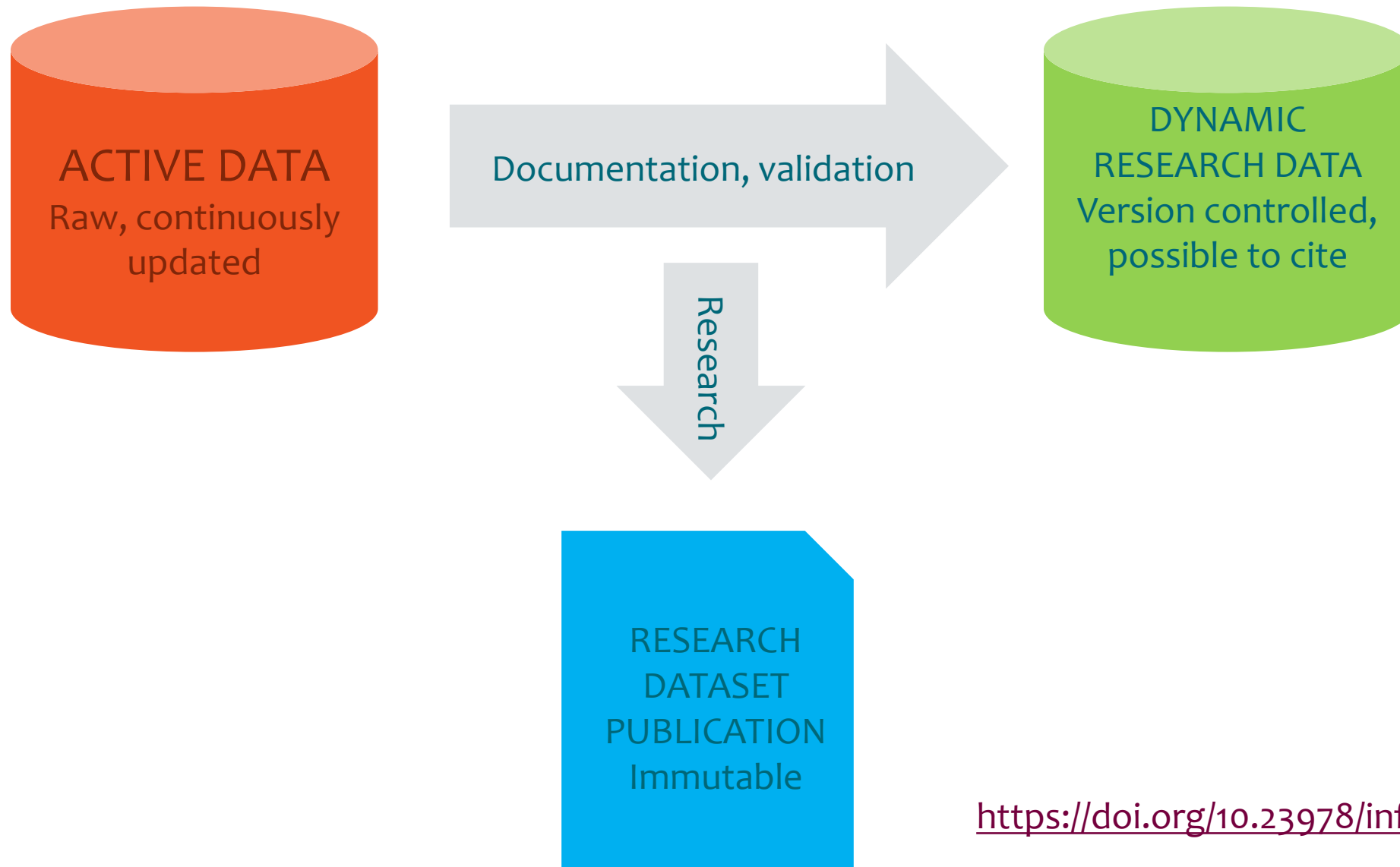
A PID is a Promise

CSC PID POLICY

1. All datasets have appropriate identifiers
2. If an object has an identifier use it
3. One object can have several identifiers
4. Identifiers are unique in their context
5. Use and management of identifiers is documented
6. No identifier is reused in its context
7. Identifiers have minimal semantic meaning and strictly defined structure
8. Identifiers comply with documented standards
9. Policies for object versioning are documented
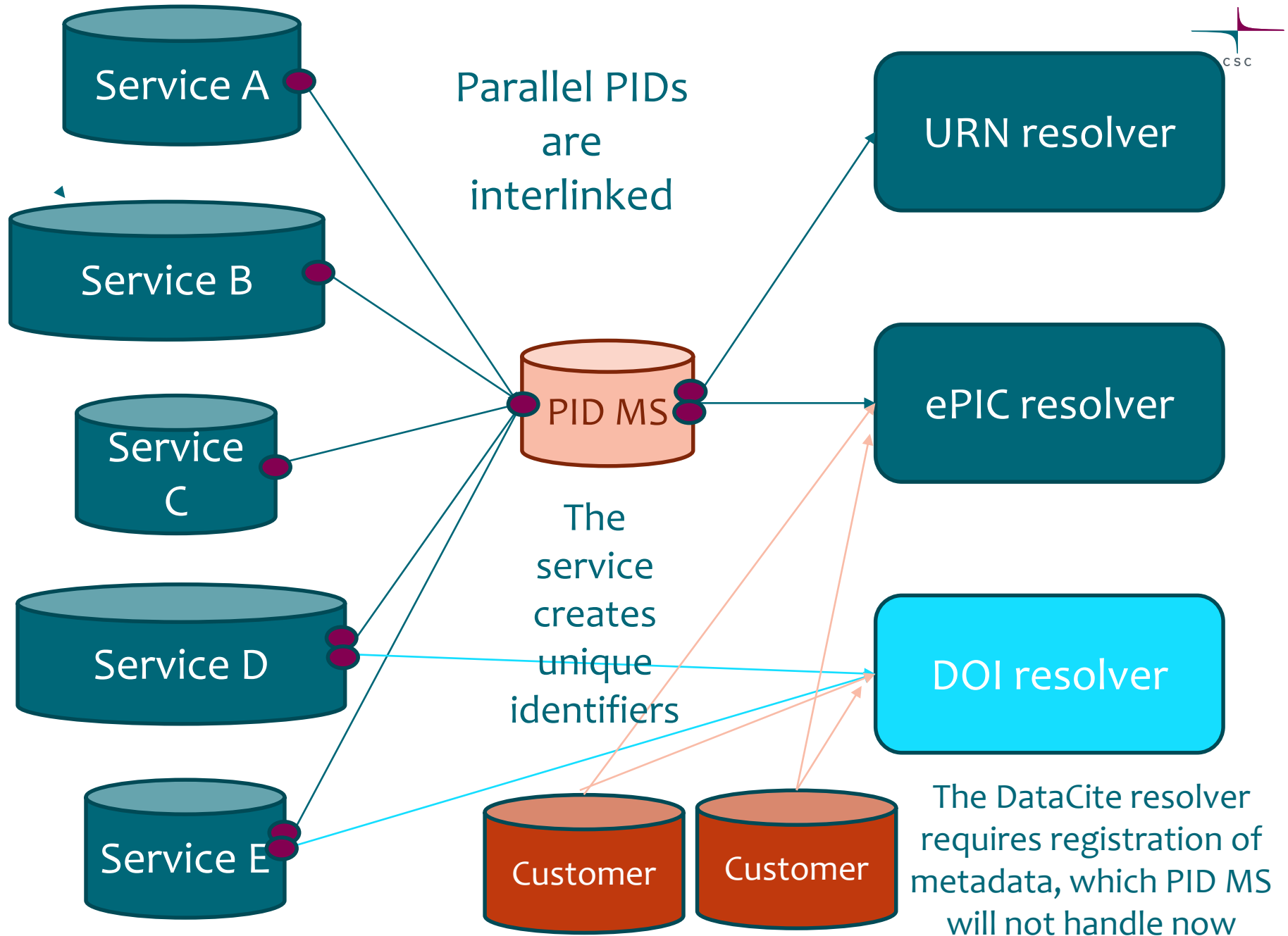10. Human readable identifiers are user friendly

CSC

2

PID MS provides PIDs that are
- standardized
- user friendly
- linked
- documented
- Centrally resolved

Service A

Service B

Service C

Service D

Service E

Parallel PIDs are interlinked

PID MS

The service creates unique identifiers

URN resolver

ePIC resolver

DOI resolver

Customer

Customer

The DataCite resolver requires registration of metadata, which PID MS will not handle now

CSC

Customer

The use of identifiers should be documented and support the needs of the research community

All research datasets that are opened or of which the metadata is published has a PID, preferably a URN or DOI

The PID directs the user to sufficient metadata

If the data is not available the landing page is a tombstone page

One dataset can have several PIDs from different systems

DataCite relation types are used to describe relations

Semantics should be used with consideration

Identifiers have a defined structure

Identifiers for human use are user friendly

Avoid creating superfluous PIDs

CSC

Jessica Parland-von Essen

Senior coordinator
parland@csc.fi

facebook.com/CSCfi

twitter.com/CSCfi

youtube.com/CSCfi

linkedin.com/company/csc---it-center-for-science
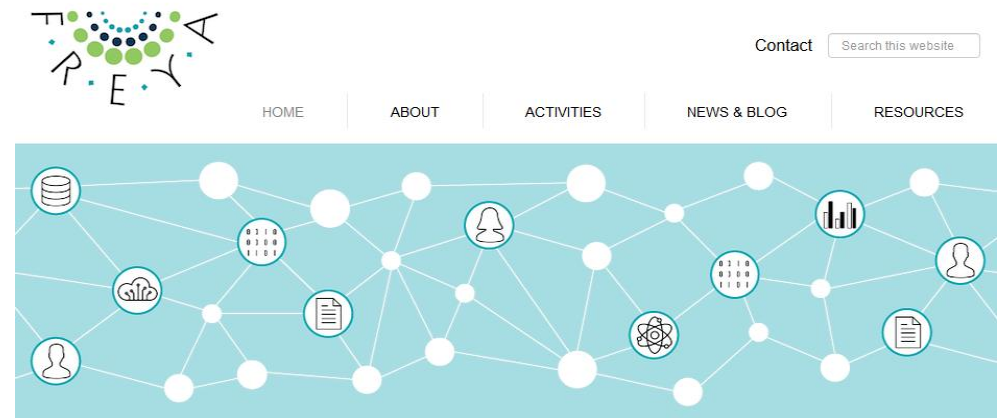
github.com/CSCfi

Where to learn more?

FREYA in a nutshell

- FREYA = persistent identifiers
  - "… iteratively extend a robust environment for Persistent Identifiers (PIDs) into a core component of European and global research e-infrastructures"

- Builds on THOR (which in turn built on ODIN)

- Started 1 December 2017

- www.project-freya.eu

Welcome to FREYA

Connected Open Identifiers for Discovery, Access and Use of Research Resources

The PID Forum

Category	Topics
General Topics that don't need a category, or don't fit into any other existing category.	3
PID Graph Discussion of the PID Graph and all related activities.	8
PID Best Practices A category to bring together information (papers, guidelines etc) and ideas on PID best practices for different communities and disciplines.	6
PID News & Blogs Share interesting PID news & blogs here	13
PID-related events Category to share any PID-related events that might be of interest to the community e.g. conferences, webinars, workshops and more!	17
PIDapalooza Discussion topics and practical announcements related to PIDapalooza	16
Knowledge Hub This category contains basic information for those new to Persistent Identifiers created by the FREYA Project. This section will continue to be updated. ■ Getting Started with PIDs ■ PIDs for Librarians and Repository Managers ■ PIDs for Funders and Policy Makers ■ PIDs for Researchers	10

Latest

⚙	📌 Welcome to the PID Forum!	0 Jun 7
	The PIDapalooza 2020 call for proposals is open! ■ PIDapalooza	0 7d
	Monitoring PID resolving ■ Questions	2 7d
	Registration open: FREYA Ambassador Webinar - 24 September 10:30am CET ■ FREYA Ambassadors Chat Room	0 10d
	ORCID for instruments ■ Questions	2 12d
	How will you use the PID Graph? ■ FREYA Ambassadors Chat Room	3 14d
	Crossref survey and annual meeting ■ PID-related events	0 19d
	Assigning PIDs to All The Things ■ User Stories	0 21d
	You're invited to Crossref LIVE19: The strategic one ■ PID-related events	0 23d

PIDForum.org

OpenAIRE

Guides for Researchers

How can identifiers improve the dissemination of your research outputs?

Connect all your research products with your person identifier

WHAT IS A
PERSON
IDENTIFIER?

HOW IT WORKS

BENEFITS

MORE
INFORMATION

What is a person identifier?

You are probably familiar with persistent identifiers like the DOI (digital object identifier) for publications and datasets. A persistent identifier or PID is a long-lasting reference to a resource - a person (you!), a place (your organisation), or a thing (your publications, data sets, software, etc). Whatever resource it refers to, the primary purpose of the PID is to provide the information required to reliably identify, verify and locate it. A PID may be connected to a set of metadata describing an item rather than to the item itself.

Support

RESOURCES

Open Science Primers

Guides

Factsheets

Use cases

Links

- https://www.pidforum.org/
- https://www.project-freya.eu/en/resources/project-output
- https://support.datacite.org/
- https://orcid.org/organizations
- https://www.fairsfair.eu/
- https://www.openaire.eu/support

How to promote PIDs in YOUR organisation

10 minutes

Name ways to promote PIDs

10 minutes

Choose the 3 most impactful

10 minutes

Report back

10 minutes

Presentations and Mentimeter

How to design messages for your communities?

What objections do you encounter

Mentimeter

Solutions!

15 minutes

Elevator pitch

1 minute per group

3 things I will do when I get back

5 minutes

Thank you!