FREYA mid-way webinar

9 May 2019

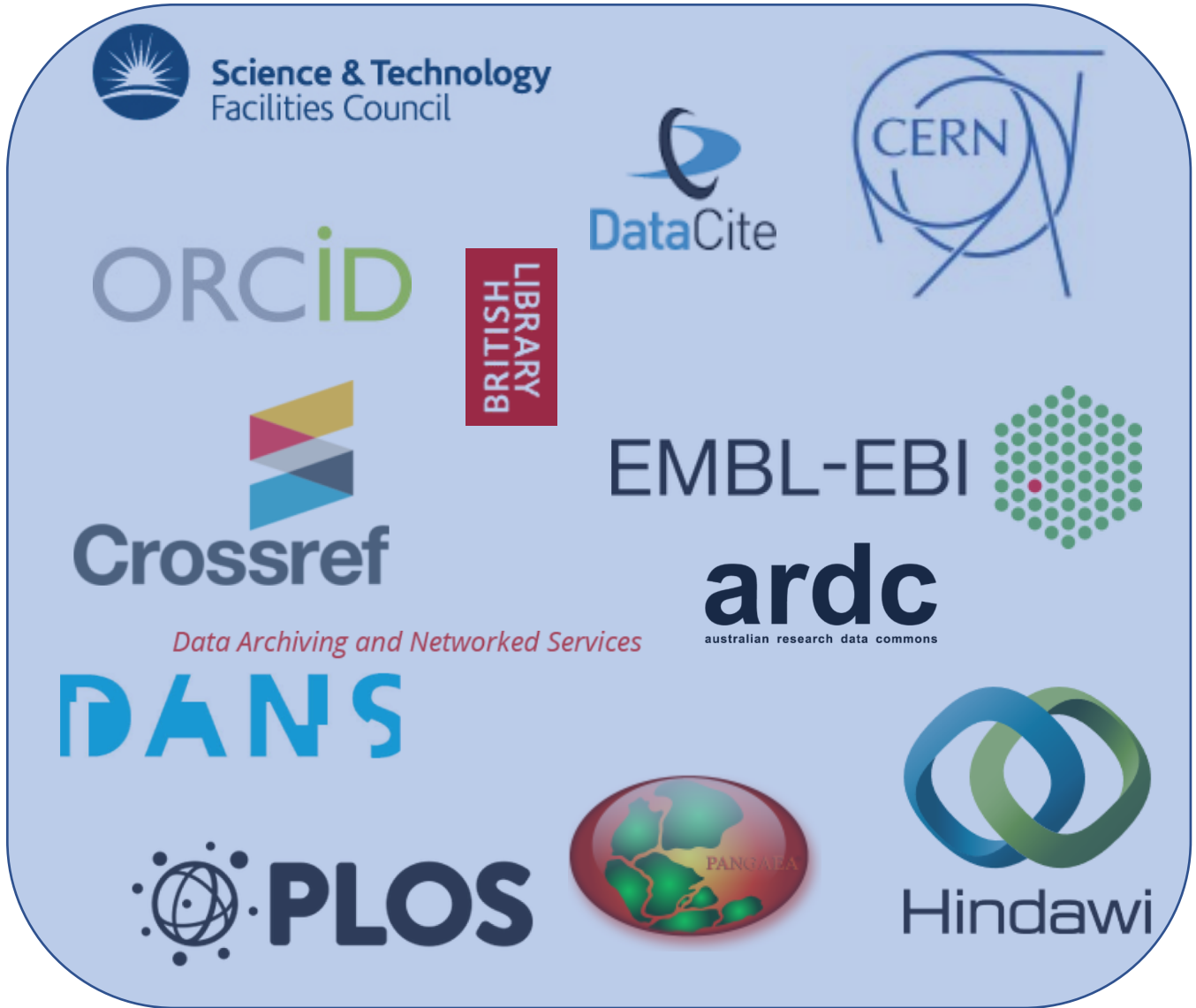
Introduction to FREYA

Simon Lambert
UKRI-STFC
FREYA Project Coordinator

FREYA: infrastructure for persistent identifiers

- Improved discovery and access of research data
- Develop new types of PIDs
- Demonstrate disciplinary PID systems
- Integration with EOSC

Take a look at www.pidforum.org !



Why this webinar?

- FREYA results ready to present
- Engagement with the community as we move into the second half of the project
- Growing interest in “FAIR” data

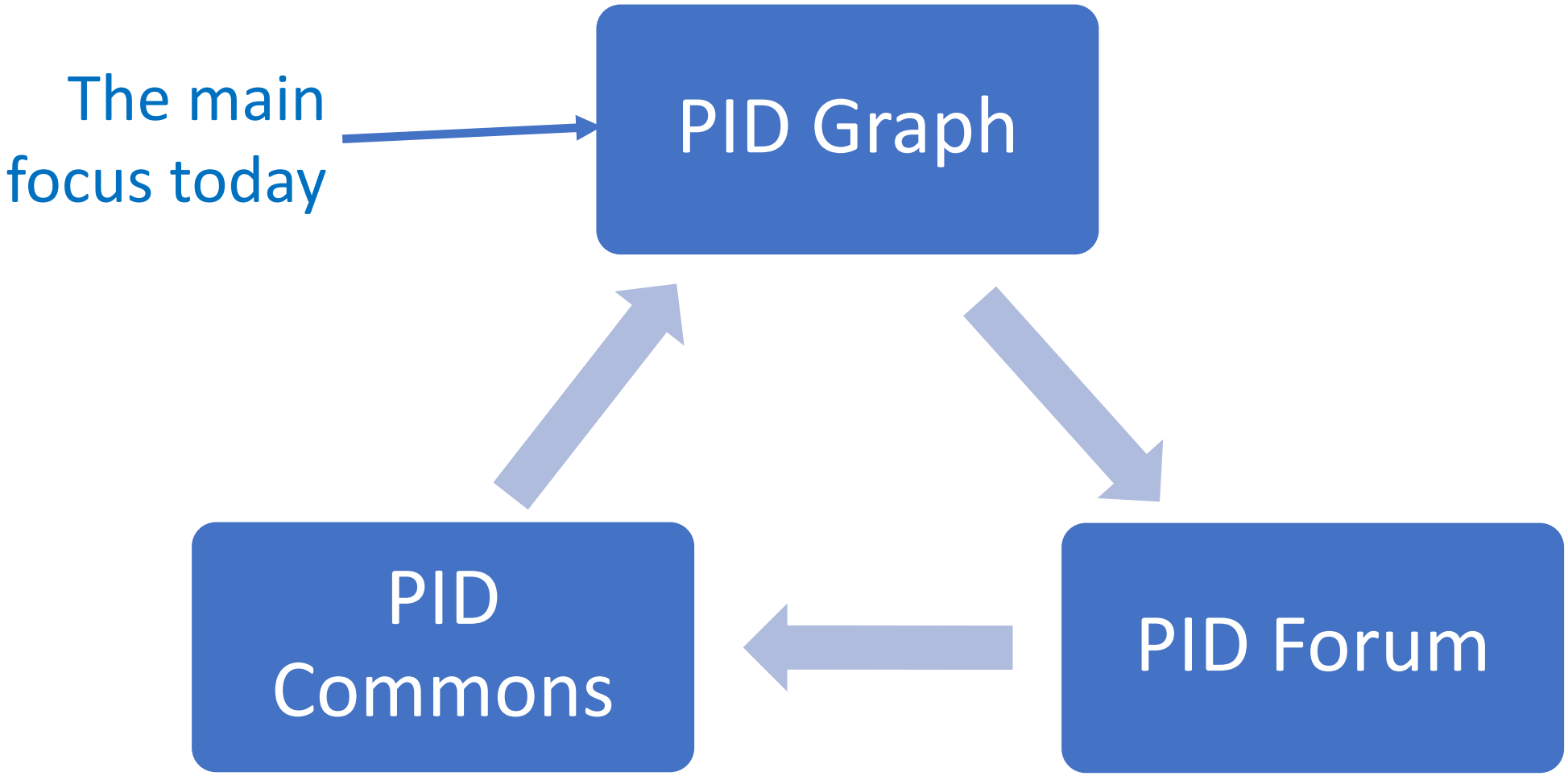
To be Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. (meta)data are registered or indexed in a searchable resource.
F4. metadata specify the data identifier.

TO BE ACCESSIBLE:

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2 metadata are accessible, even when the data are no longer available.

The three pillars of FREYA



Introduction to the PID Graph

Midway Webinar
9 May 2019
Robin Dasler, DataCite

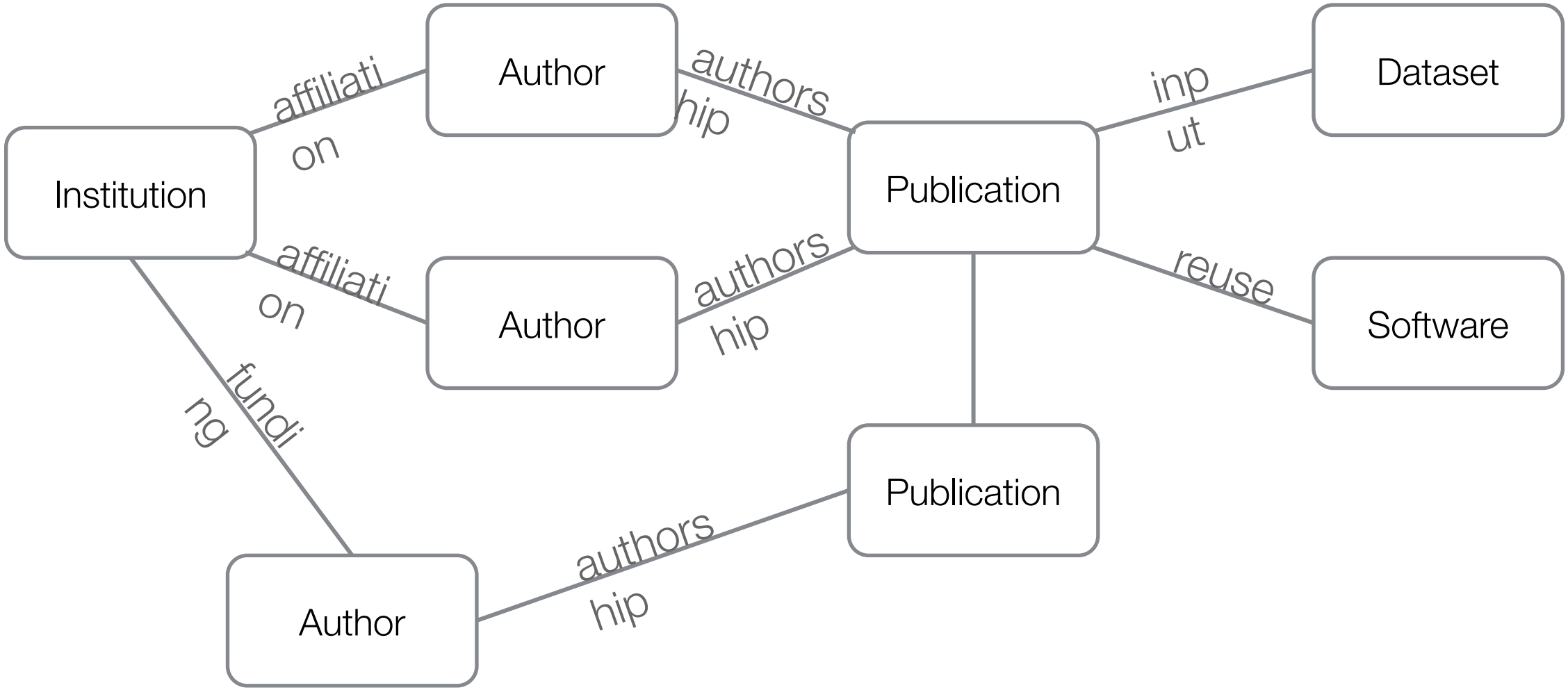
What is the PID Graph?

Research is already a graph

Researchers, institutions, publications, datasets, and more are interconnected.

Entities and the relationships between them form a conceptual graph of the connected research landscape.

It could look like this



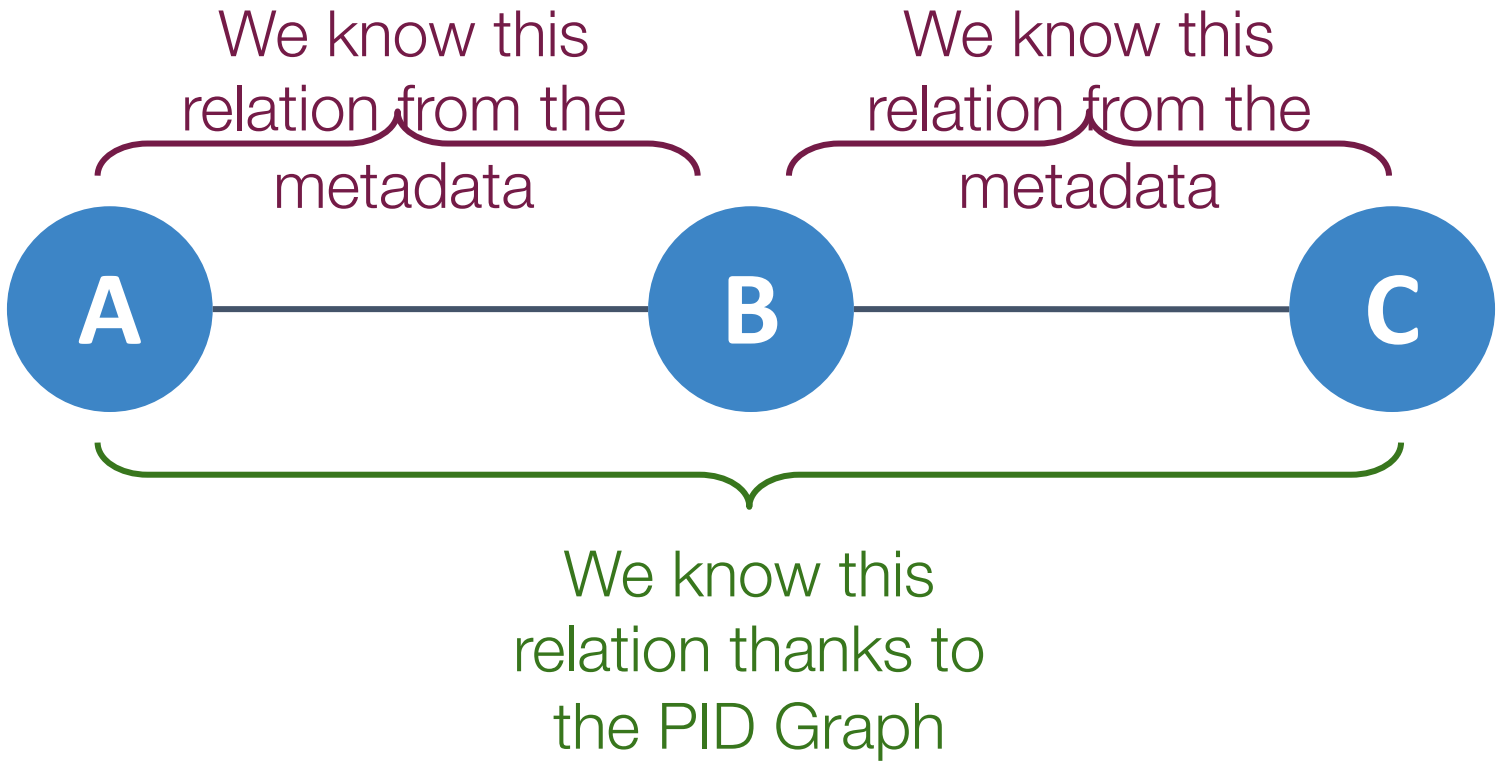
PIDs are the backbone of connected research

Having unique persistent identifiers for researchers and their outputs is crucial to connecting pieces of the research landscape together.

PIDs already have the potential to enable the connected research graph, but we're not yet taking full advantage of their connecting powers.

Enter the PID Graph

We can link PIDs together via relations in their metadata to enable the discovery of connections at least two “hops” away.



The PID Graph concept

FREYA partners will implement services that enable their local PID Graph.

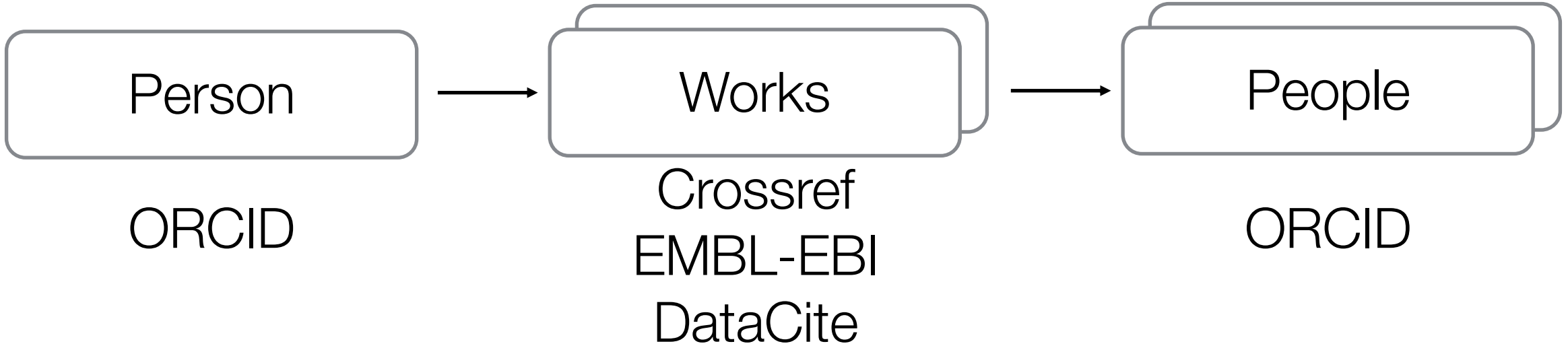
Looking up a single PID in these services should return the graph for that PID.

Infrastructural partners and the magic of related identifiers bring these local graphs together.

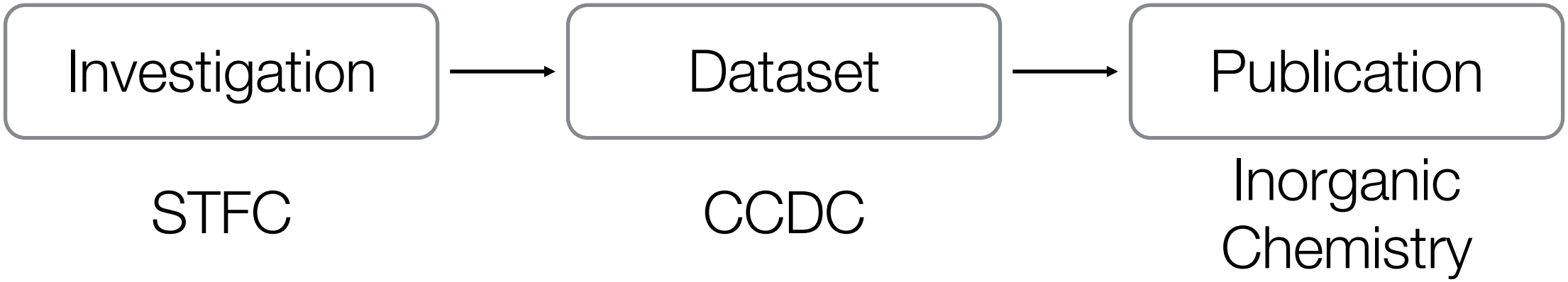
Users can tap into this PID Graph for use in their own applications.

What questions could the PID Graph answer?

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

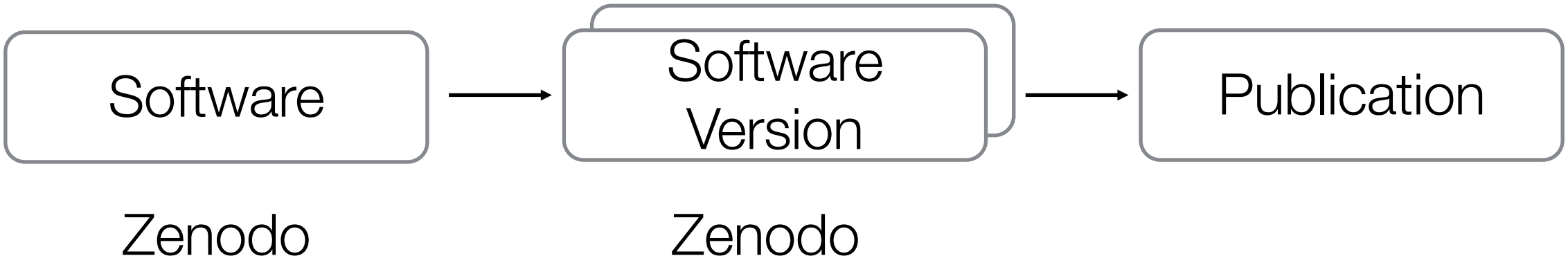


Which publications used STFC investigations for the underlying data?



Which publications cite papers based on PANGAEA data?

Which publications cite any version of a piece of software?



Enabling the PID Graph at DataCite

Event Data - Backend service

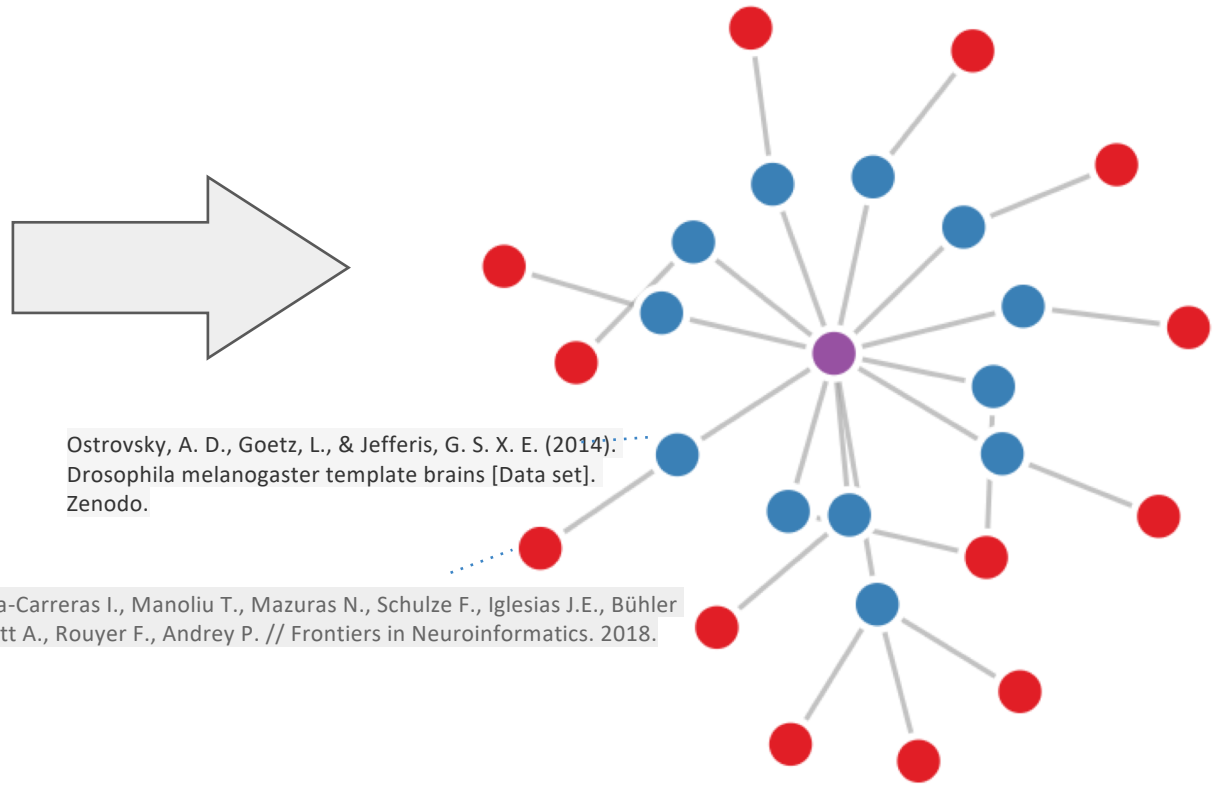
We have developed a backend service for the PID Graph by extending the Crossref-DataCite Event Data service.

Event Data was already a collection of relational “events” between Crossref and DataCite DOIs as well as other related identifiers in the DataCite metadata.

We tweaked our API so you can now retrieve the kinds of “two hop” relations the PID Graph enables.

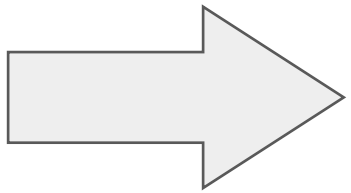
Which datasets, funded by the EC, have been cited by a journal article?

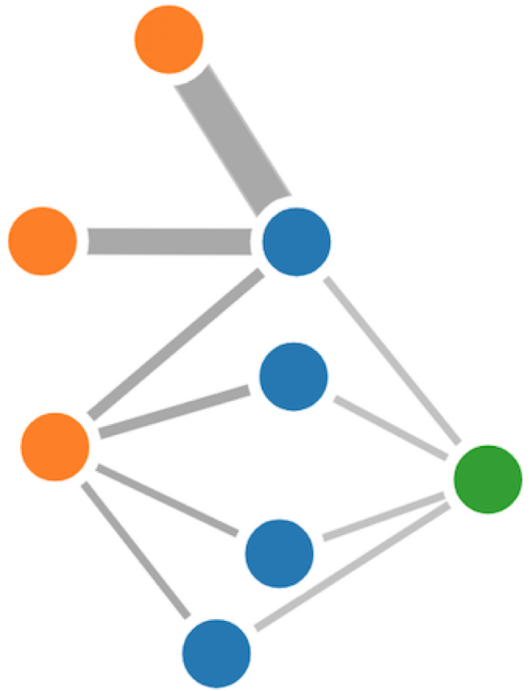

```
{
  + data: […],
  - meta: {
      total: 14,
      total-pages: 1,
      page: 1,
    - sources: [
      - {
            id: "crossref",
            title: "Crossref to DataCite",
            count: 14
        }
      ],
    - prefixes: [
      - {
            id: "10.13039",
            title: "10.13039",
            count: 14
        },
      - {
            id: "10.5281",
            title: "10.5281",
            count: 14
        },
      - {
            id: "10.5194",
            title: "10.5194",
            count: 5
        },
      - {
            id: "10.1007",
            title: "10.1007",
            count: 2
        },
```

How many citations and repository usage counts do a researchers' datasets have?


```
{
  + data: […],
  - meta: {
      total: 8,
      total-pages: 1,
      page: 1,
    - sources: [
        - {
            id: "datacite-orcid-auto-update",
            title: "datacite-orcid-auto-update",
            count: 4
          },
        - {
            id: "datacite-usage",
            title: "DataCite Usage Stats",
            count: 4
          }
      ],
    - prefixes: [
        - {
            id: "10.7272",
            title: "10.7272",
            count: 8
          }
      ],
    - citation-types: [
        - {
            id: "Dataset-Person",
            title: "Dataset-Person",
            count: 4,
          - year-months: [
              - {
                  id: "2019-03",
                  title: "March 2019",
                  sum: 4
                }
            ]
```



Dataset
Article
Researcher

Where do we go from here?

Consuming the PID Graph with GraphQL

DataCite is currently using Event Data as our PID Graph backend service, which is part of our REST API.

DataCite has started to implement GraphQL, because it is easier to consume.

GraphQL could also make federation and standardization simple. Just implement a GraphQL API (or a wrapper for your REST API), and your service can easily slot into the PID Graph.

GraphQL Endpoint https://api.datacite.org/graphql Method POST Edit HTTP Headers

GraphiQL Prettify History Docs

```
{
  publications(query: "creators.name:dasler") {
    id
    titles {
      title
    }
    descriptions {
      description
    }
    creators {
      id
      name
      familyName
    }
    fundingReferences {
      funderIdentifier
      funderName
      awardTitle
      awardNumber
    }
  }
}
```


```
{
  "data": {
    "publications": [
      {
        "id": "https://doi.org/10.5281/zenodo.1064000",
        "titles": [
          {
            "title": "Pid Service Adoption"
          }
        ],
        "descriptions": [
          {
            "description": "This presentation describes how the uptake of persistent identifiers can be measured and gives an overview of the main results of the ORCID adoption study."
          }
        ],
        "creators": [
          {
            "id": null,
            "name": "Dasler, Robin",
            "familyName": "Dasler"
          }
```

Working with other initiatives

To expand and improve the PID Graph, FREYA needs to work with other initiatives.

We have started these collaborative activities via RDA, such as the recent Research Data Graph BoF that involved representatives from FREYA, OpenAIRE, Scholix, and DDRI. The BoF will be taken forward as an Interest Group at future RDA meetings.

PID Graph Demonstrators

Tina Dohna, Christine Ferguson and Vasily Bunakov

Landscape – *Disconnected Data Silos*

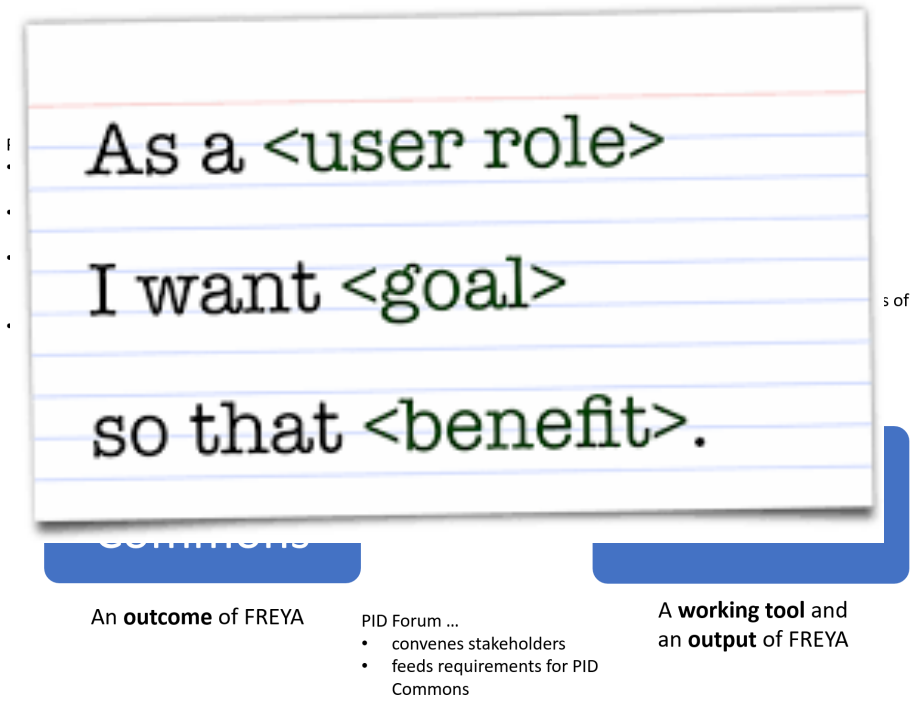
Source: https://odutrephotography.photoshelter.com/image/I0000eqxG8fTWs5Q

FREYA – Access to *Graphed PID Connectivity*

Source: https://commons.wikimedia.org/wiki/File:Saint-Martin-sur-Ouanne--vue_ouest_de_Boissel-silos_CAPROGA.JPG#filelinks

PID Graph as the driving vision

- Harmonising PID services and providers
- New PID types for entities
  - e.g. Instrument PIDs
- Support for research workflows for disciplinary partners
  - Bottom-up approach
  - FREYA User stories
  - Graph - Federated network among disciplinary partners



https://www.pidforum.org/

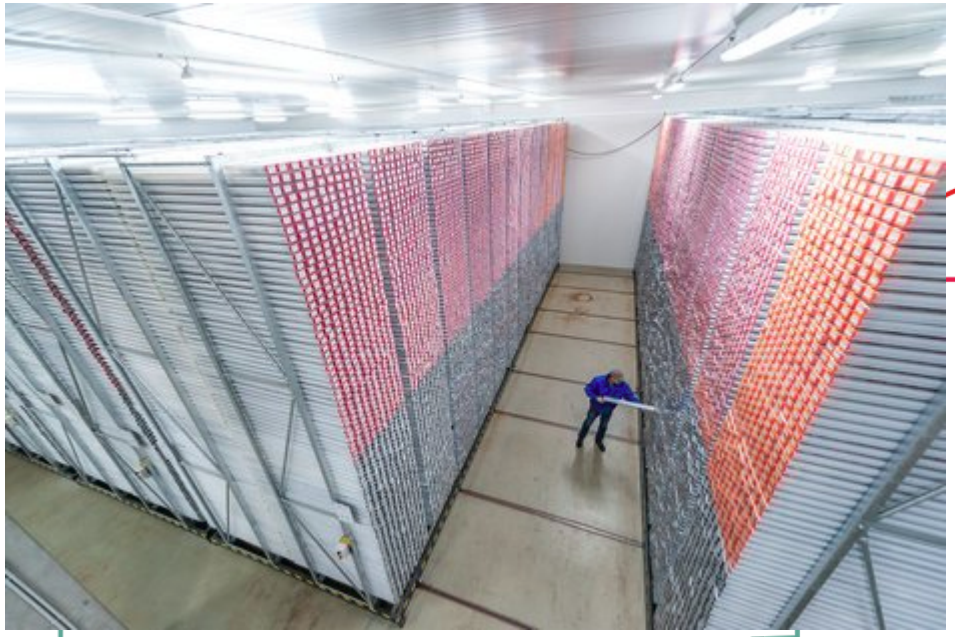
Starting from user stories

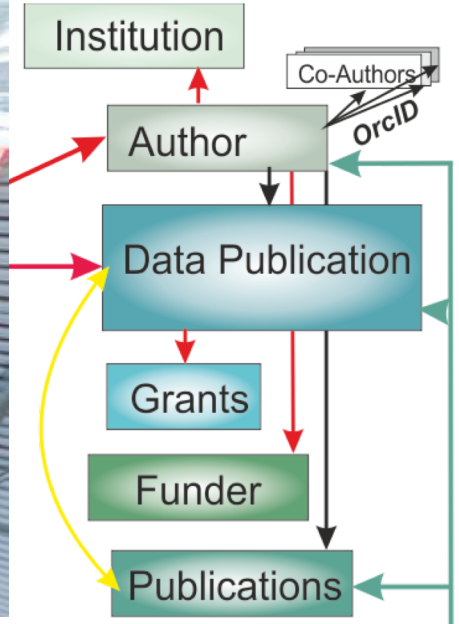
Example: Research Campaigns

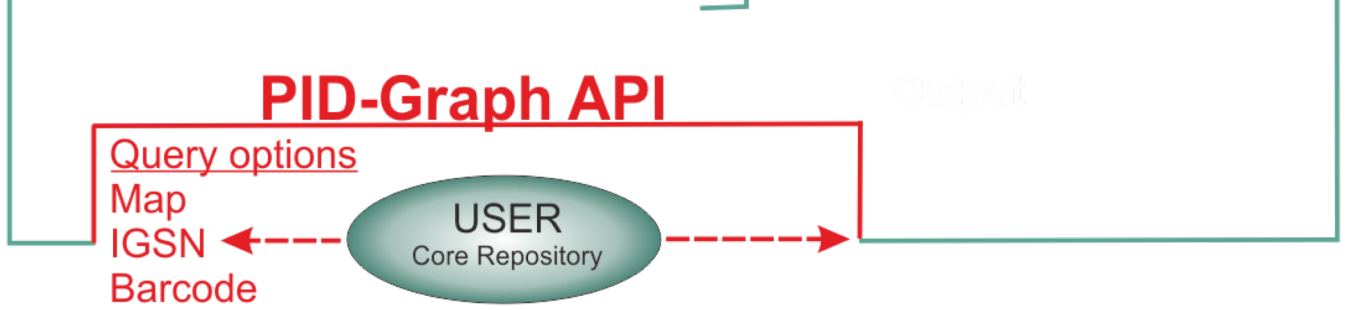
'As a funding agency, I would like to trace the outcome of my financial contribution to a marine research cruise (Cruise ID) by tracking the data generated (data-PID) and articles (publication-PID), and the use of physical samples taken (IGSN).'

Image: MARUM©

Demonstrator #1
Sample Use Case (IGSN)

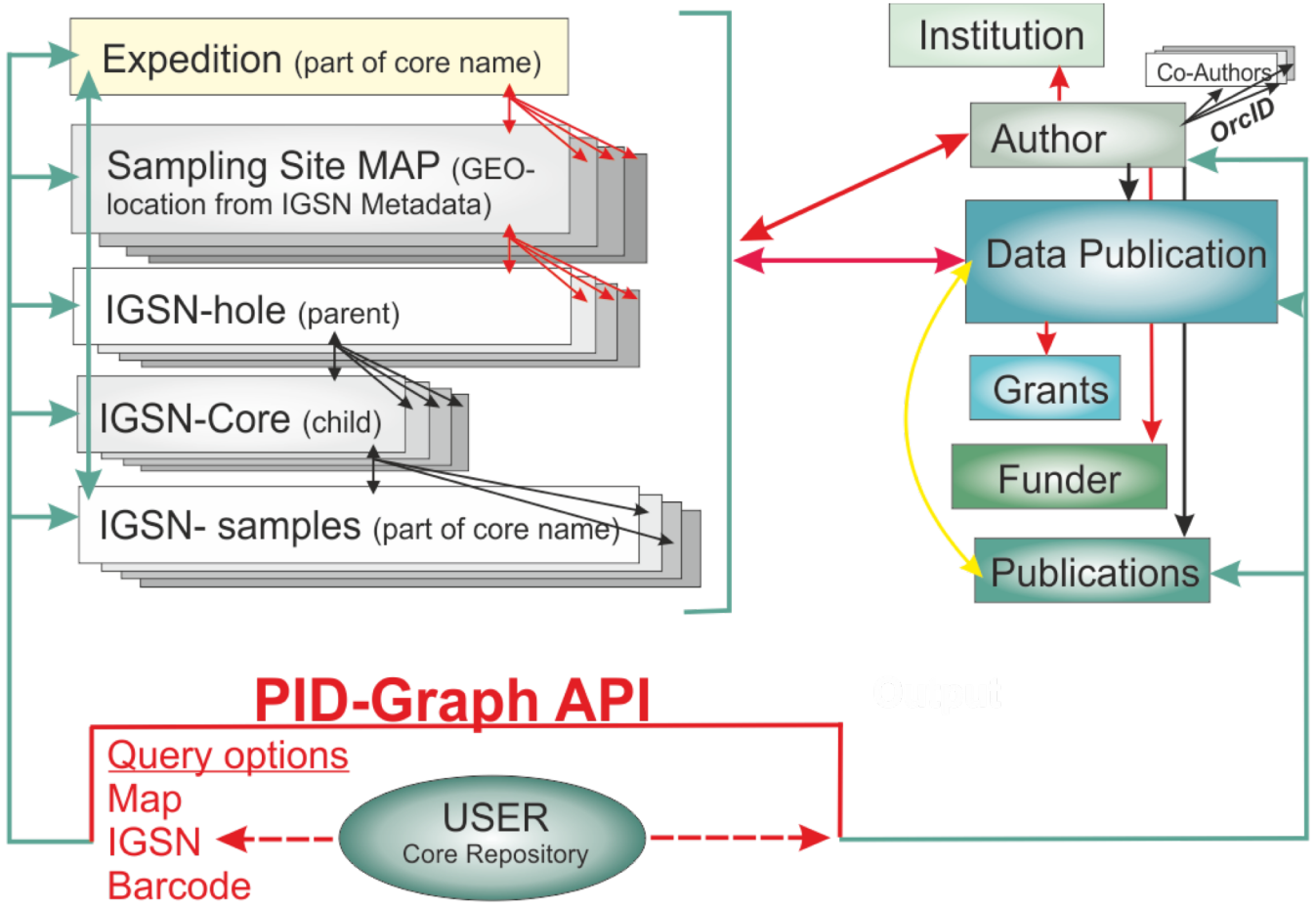






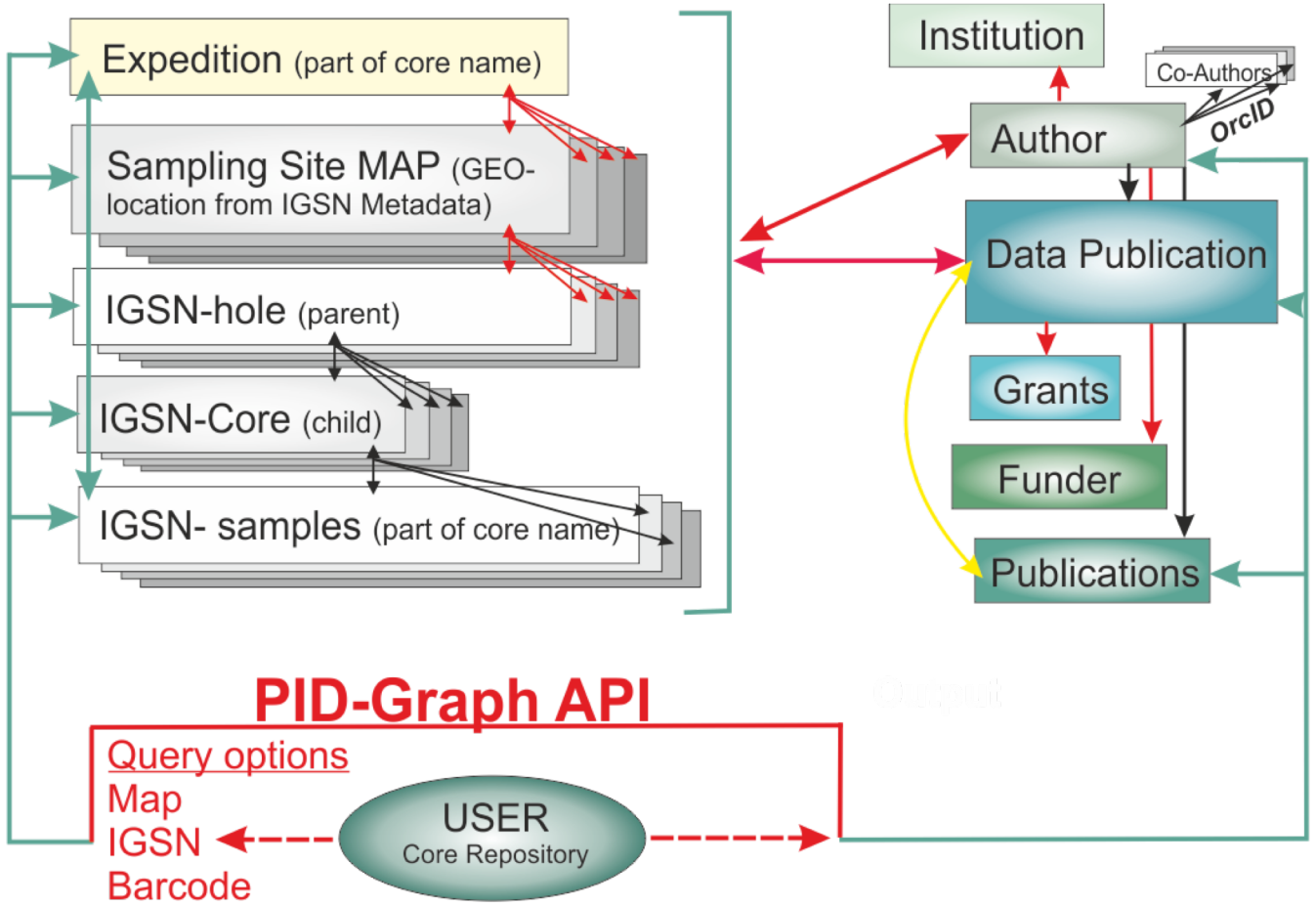
1) A function allowing automated PID-graph explorations

Sample Use Case (IGSN)



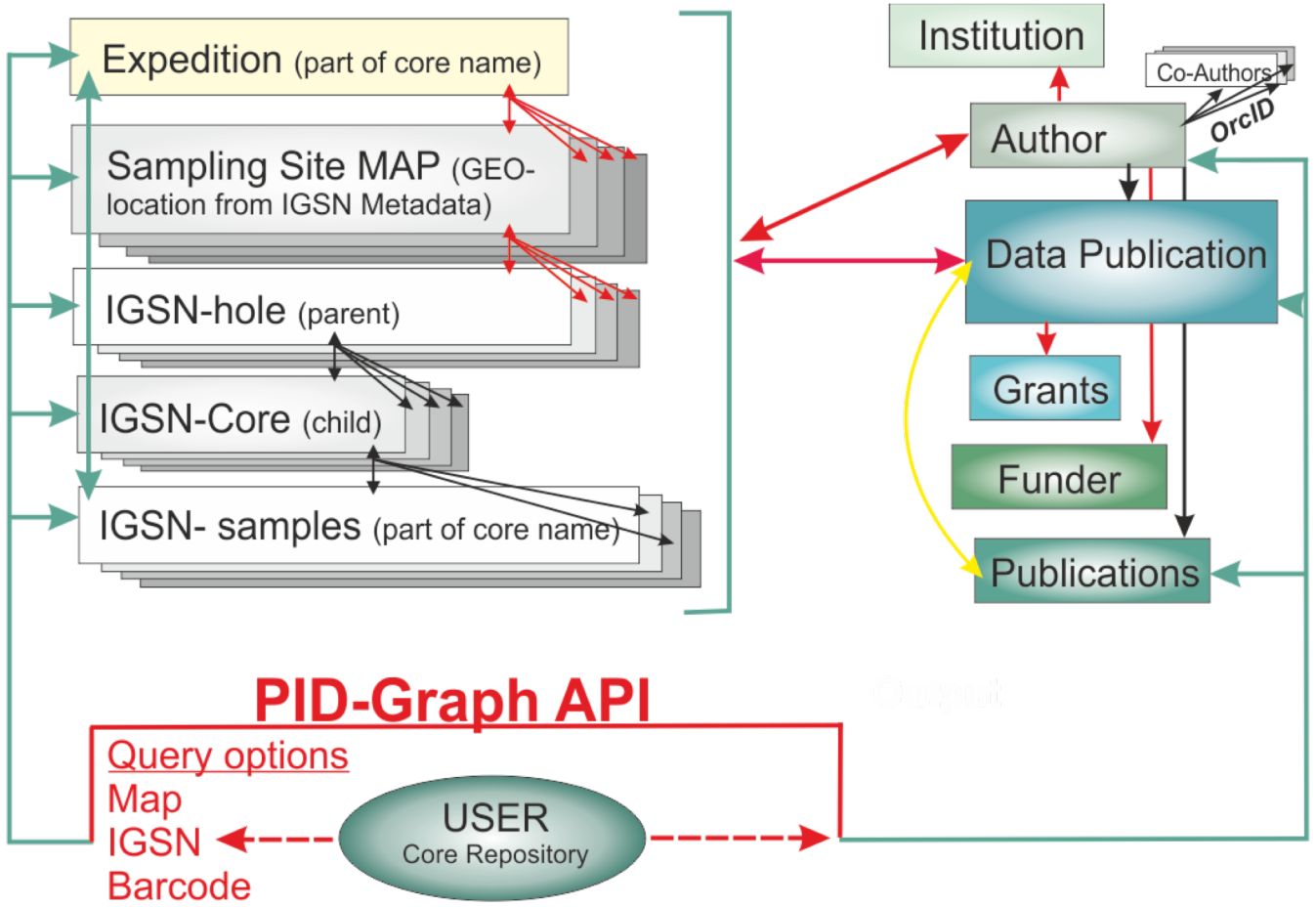
2) **Extending the function to include information from external resources (exemplified by PANGAEA-DataCite links)**

Sample Use Case (IGSN)



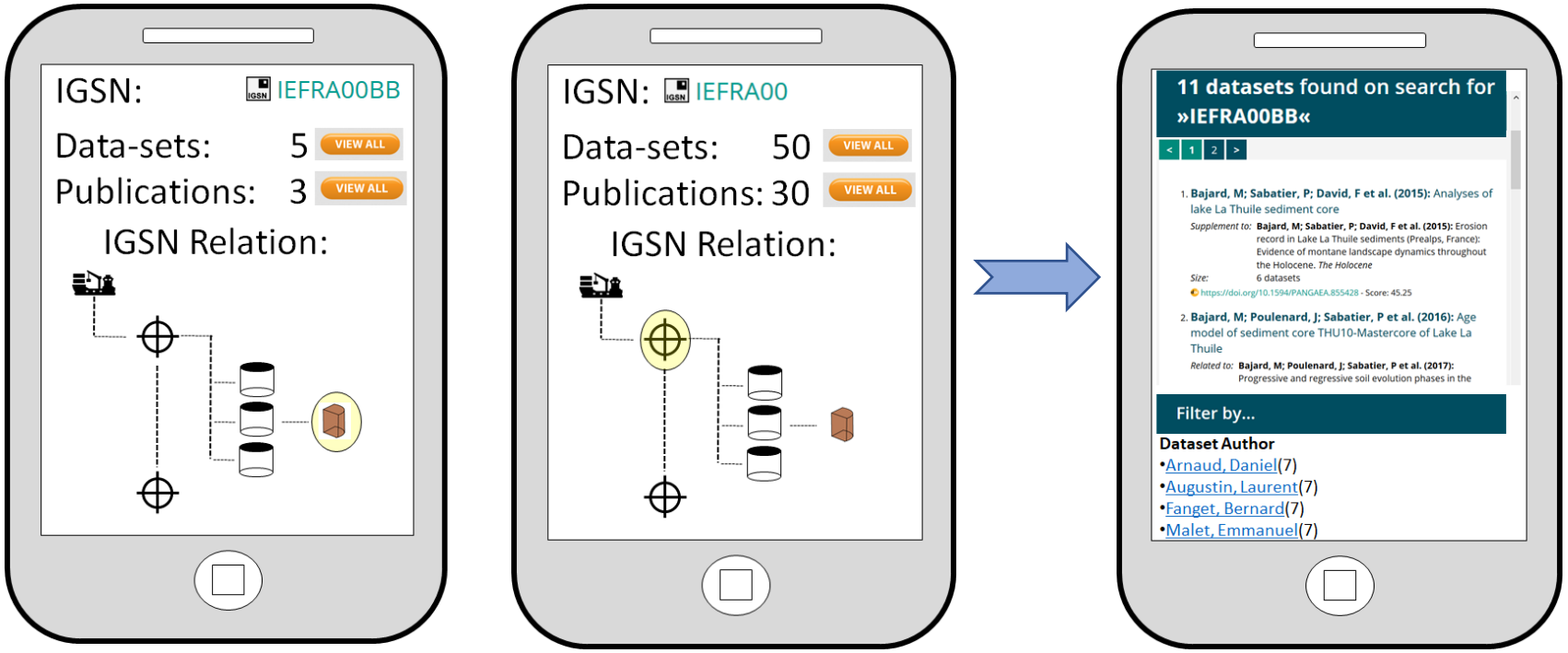
3) **Expand on the existing recommendation system in PANGAEA to allow for recommendations related to specific PIDs**

Sample Use Case (IGSN)



4) **Build user interface - Software application allowing easy PID Graph access for users**

Sample Use Case (IGSN) – User Interface



Expedition Site Bore hole Sample

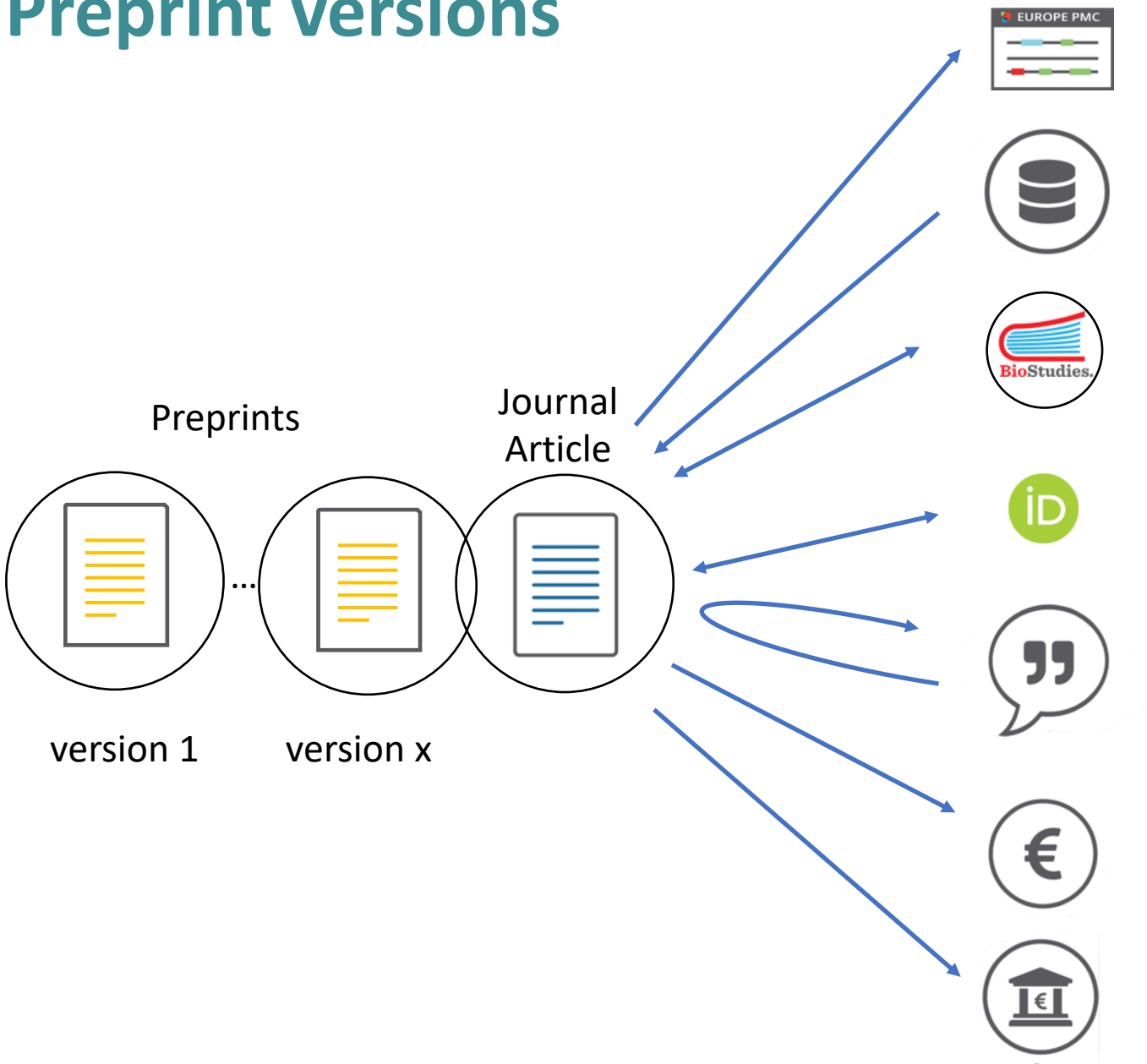
Demonstrator #2
Extending the literature-centric graph

Preprint

35M abstracts
5M FT articles

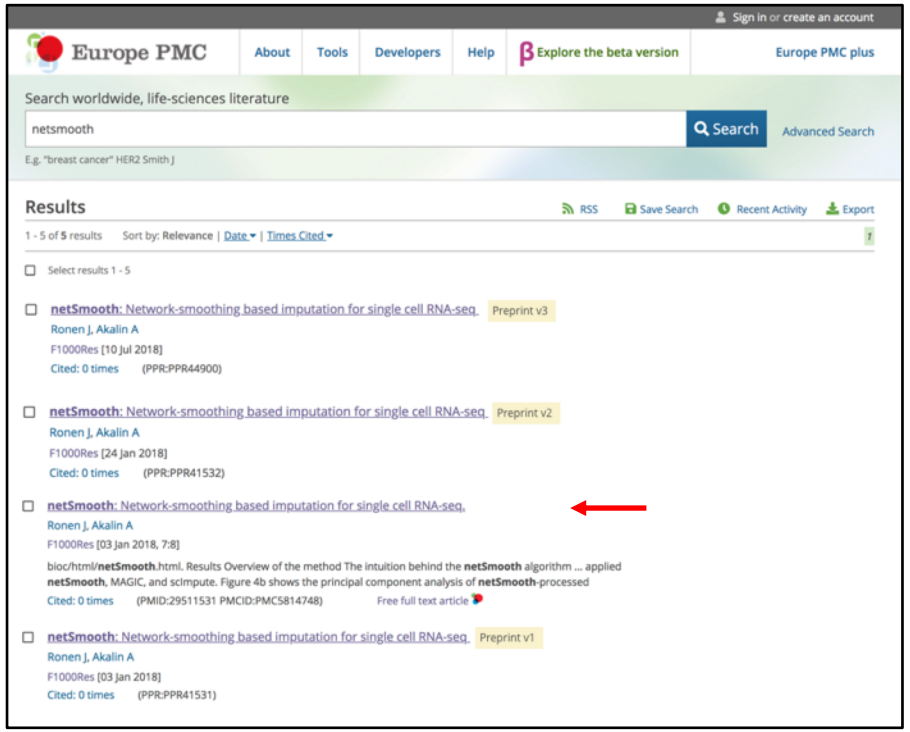
95 M datasets
1M ORCID Ids
8M funding acknowledgements

Preprint versions

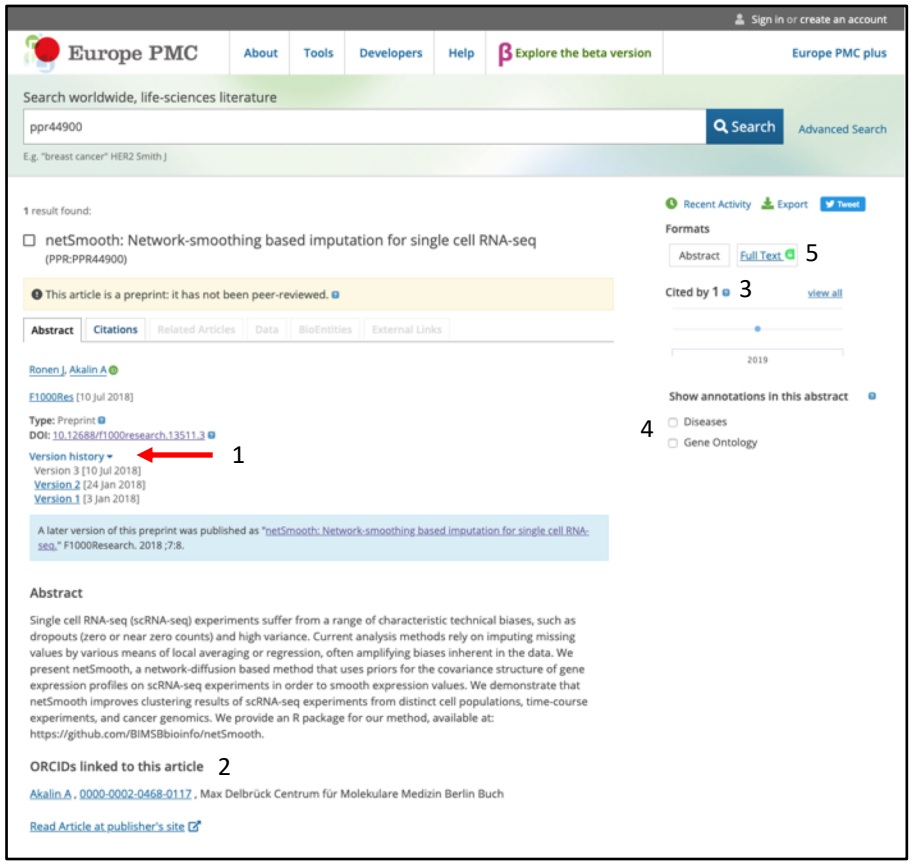


Exposing preprint version information

Search results page

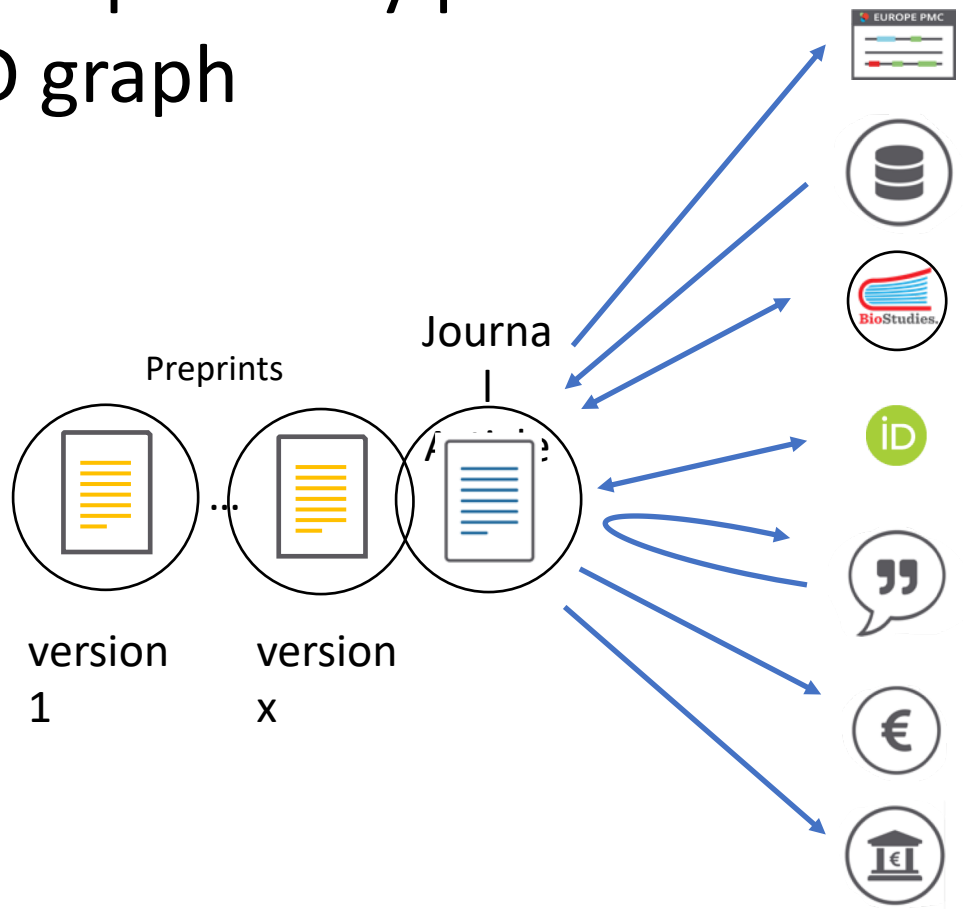


Individual preprint record



Summary

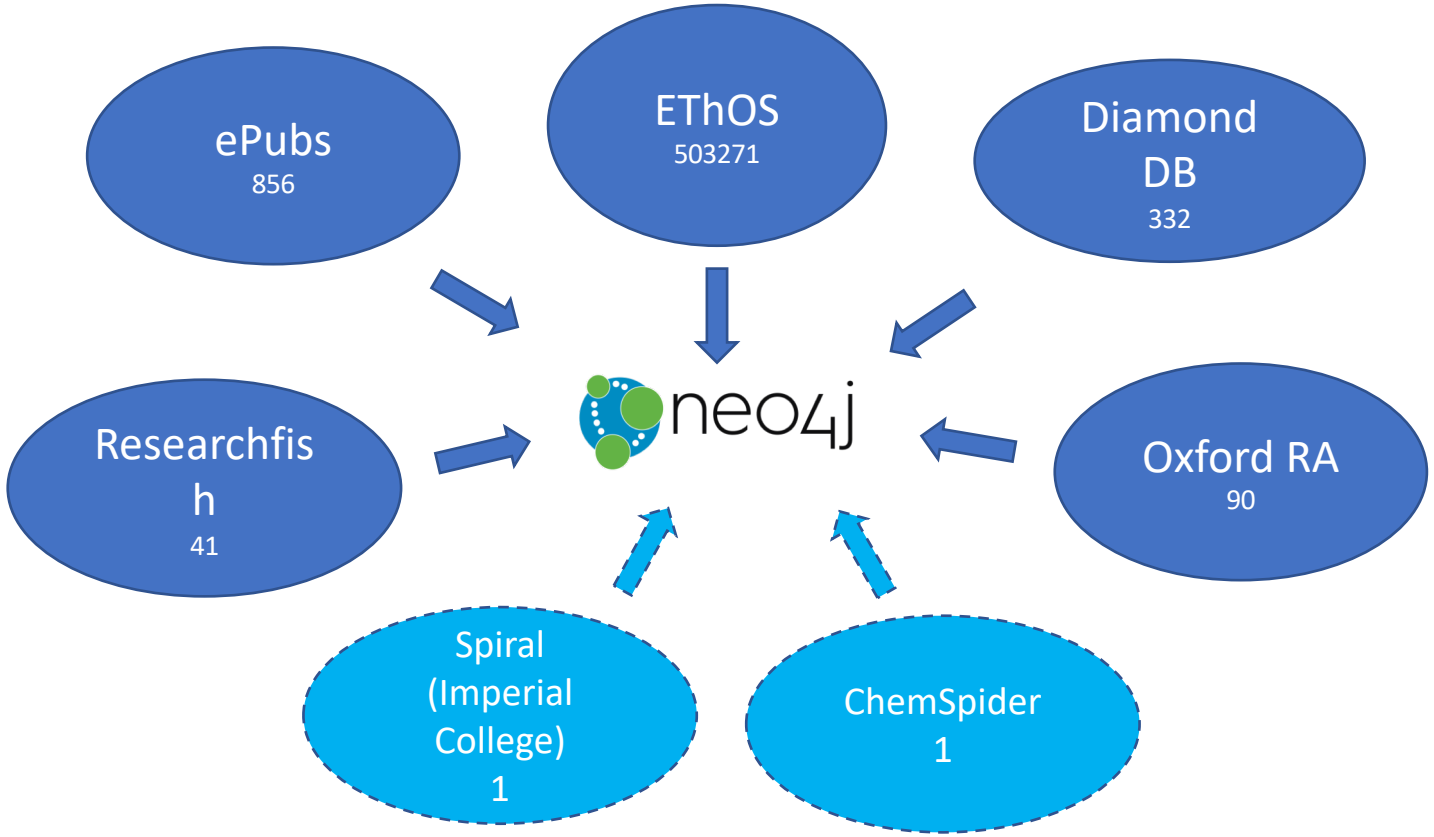
- Extending provenance
- Providing multiple entry points
  to the PID graph



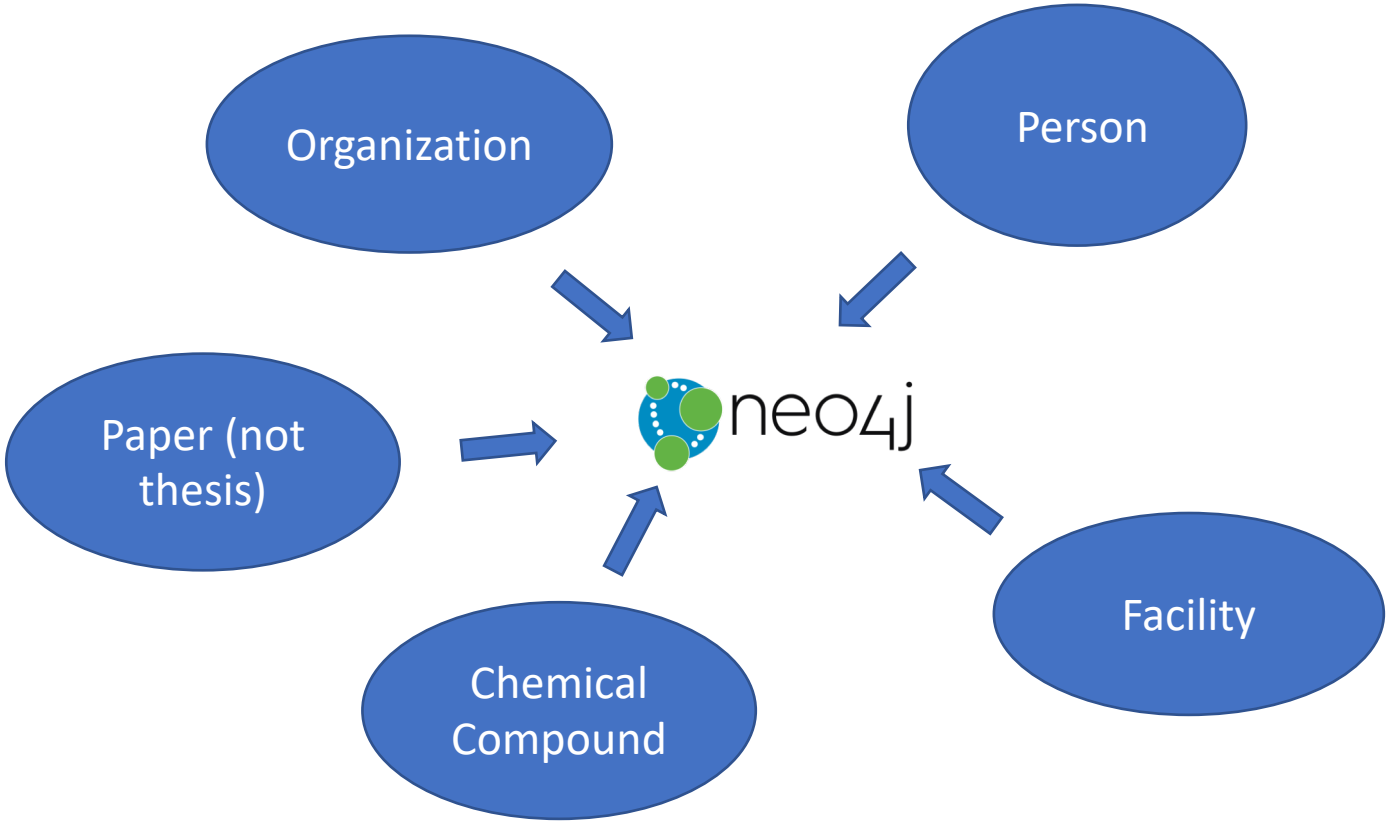
Demonstrator #3
Graphs representing PhD research

Case developed by STFC and British Library
using EThOS theses repository, STFC institutional repositories,
GRID.AC and a few other data sources

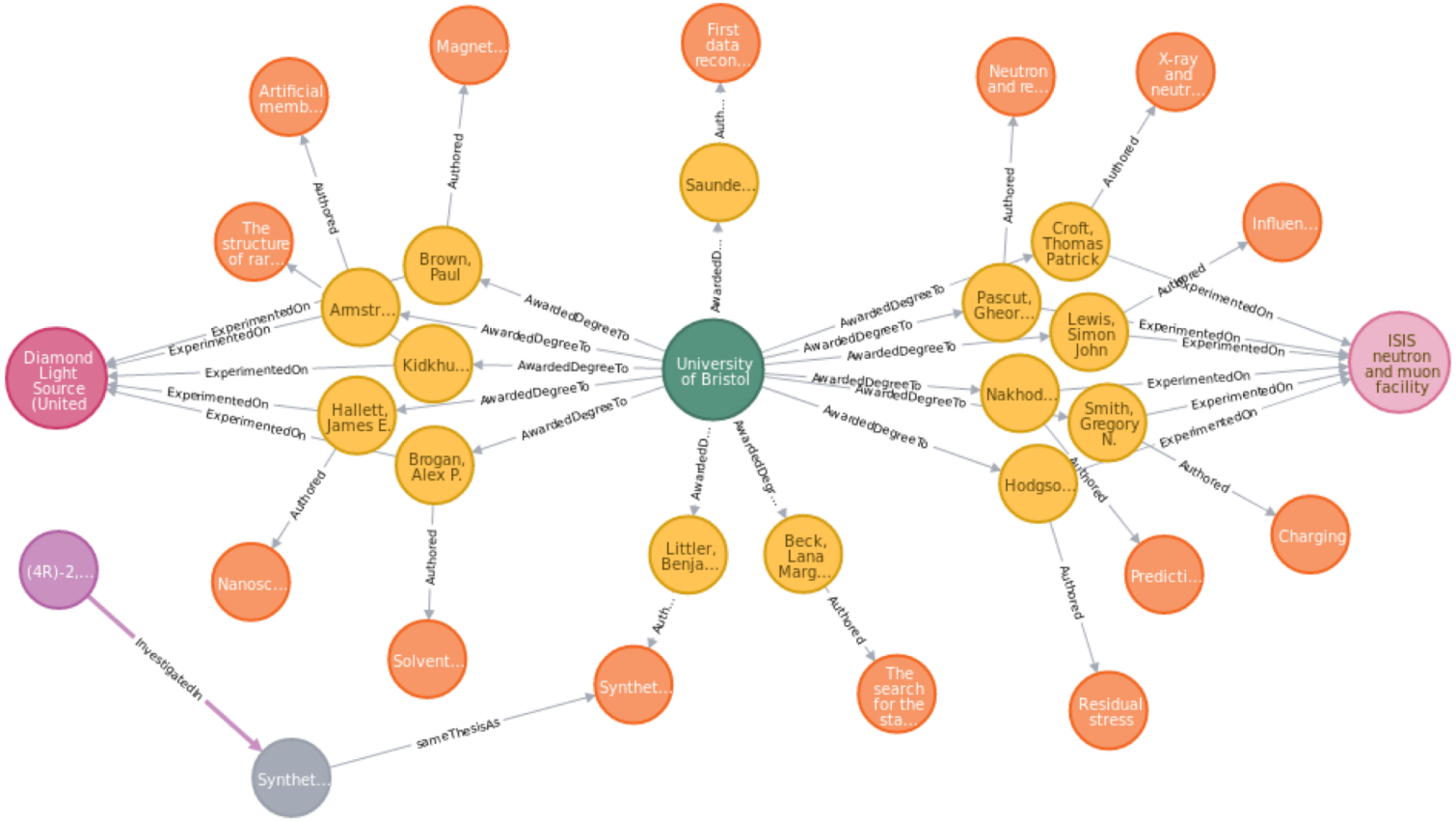
Various data sources have been used to populate the graph database (dark blue: automated ingestion, light blue: manual ingestion to illustrate specific cases)



More node types created, with PIDs assigned where available at source



Example of subgraph that represents University of Bristol PhD researchers, facilities they used and their research outcomes



Requirements gathering for new PID types

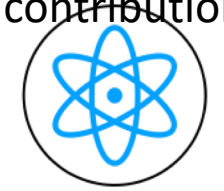
Christine Ferguson

EMBL-EBI

User stories as drivers

"As a <role>, I want <capability>, so that <benefit>"

As an institution, I want to track the outputs of all affiliated researchers.
This concerns papers, data, code and their impact (citations), but also contributions to specific conferences.



"As a core facility provider, I want to track usage of my facility so that I can demonstrate its value."

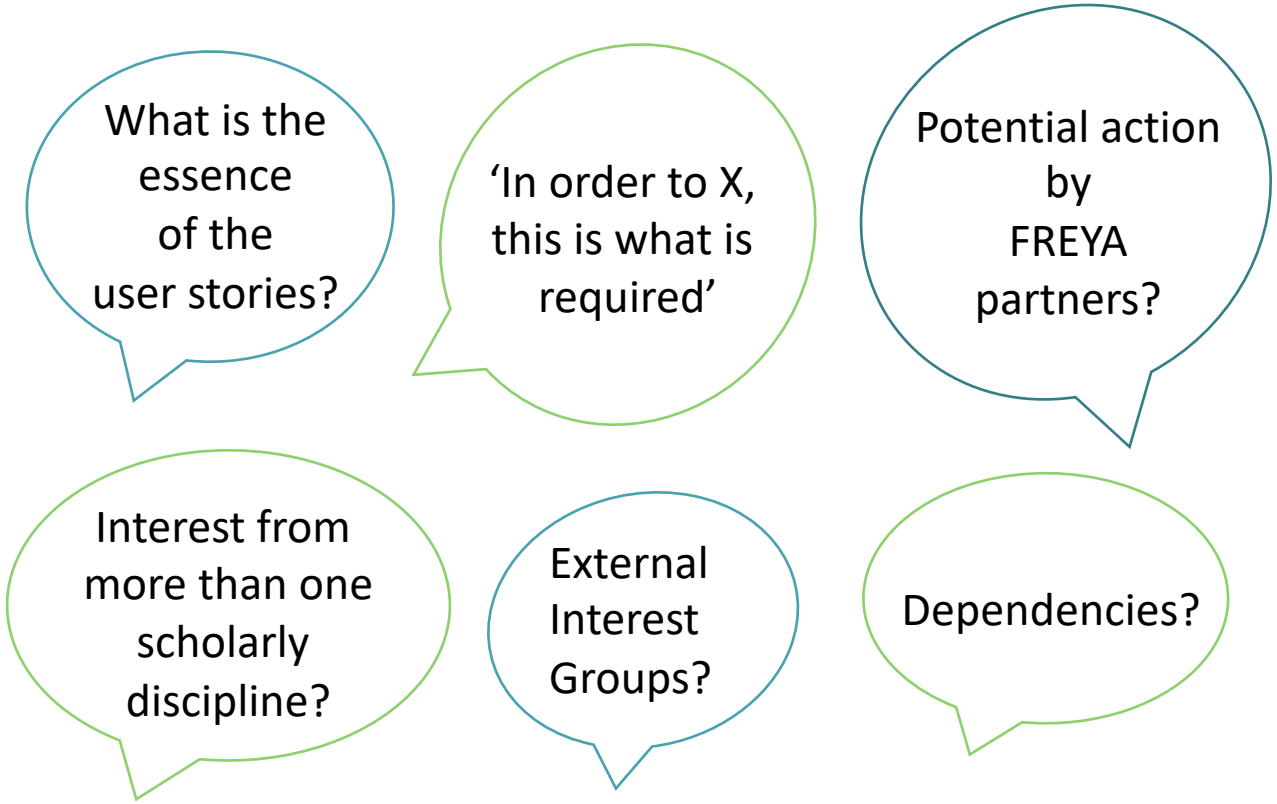
"As a funder, we want to be able to identify who (including orgs and individuals) benefitted from a given grant, for boosting management info and for looking at impact"

As a museum curator, I would like to history of the placement and display of museum items, where it has been stored and the atmospheric conditions of those storage locations over time.

Entities mapped to FREYA partners

- facilities
- grants
- organisations
- Data Management Plans
- software
- conferences
- instruments
- research campaigns
- physical samples and cultural artefacts

Deep dive of the user stories



Summary: http://doi.org/10.5281/zenodo.2649230

Readiness for building service prototypes

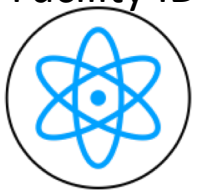
Entity	Possible for prototyping within the timeframe of FREYA	Lead partner for requirements gathering in this report
Instruments	yes	PANGAEA
Facilities	yes	UKRI-STFC
Grants	yes	EMBL-EBI
Organisations	yes	DataCite
Research campaigns	yes	PANGAEA
Software	No	DataCite
Data Management Plans	No	DataCite
Samples (physical)	No	BL
Samples (Cultural Artefacts)	No	BL
Conferences	No	CERN

Potential prototypes

Instrument IDs

RDA WG for instrument PIDs.
Plans to include PIDs for sensor related instruments on German research vessels.

Facility IDs



UKRI-STFC: facility vs instrument vs funding organization
Considering the identifier type/landing pages for a facility ID.
Linking PhD theses to a facility

Organization IDs

ROR ID is an open, community governed identifier for research organisations
https://ror.org
MVR now available

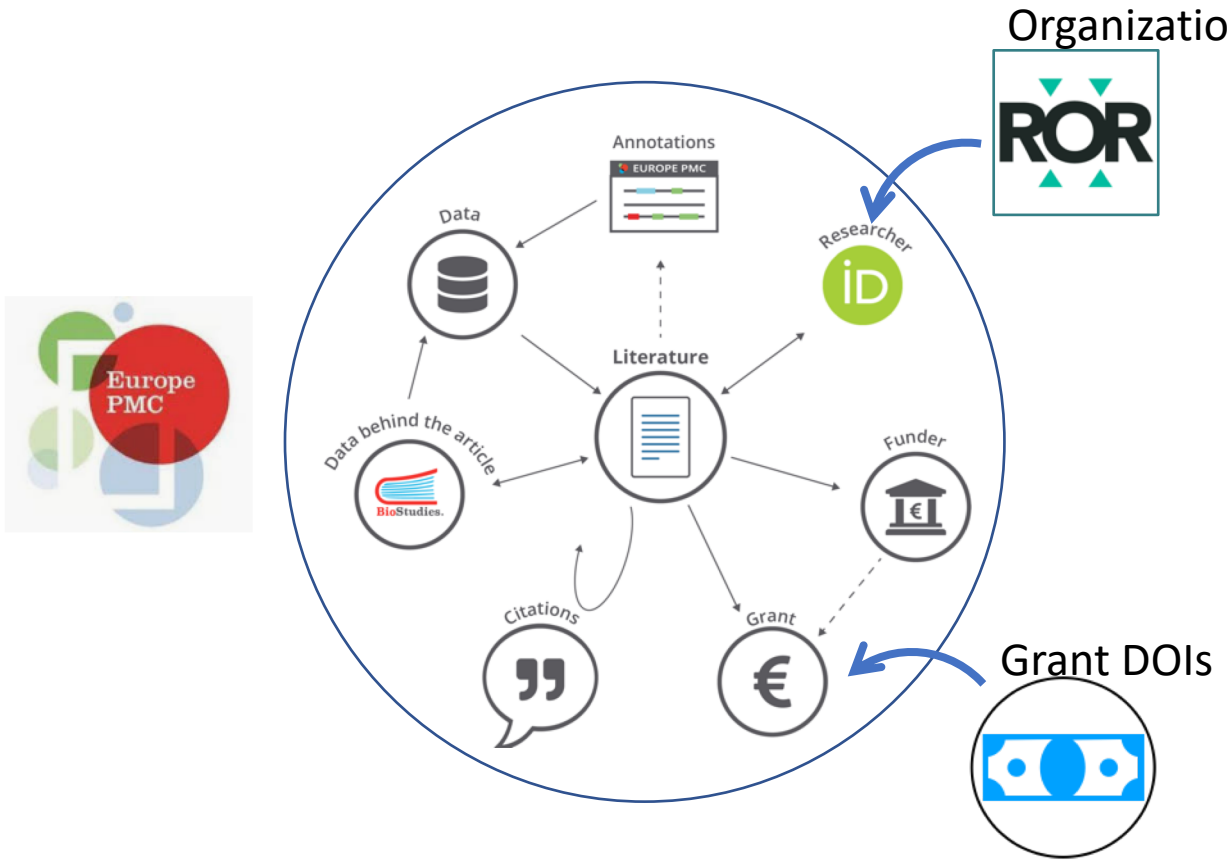
Grant IDs

DOIs assigned by Crossref; working with Europe PMC who will host the landing pages and starting with Wellcome Trust grants.

What FREYA is not prototyping

Entity and PID (if known)	Possible for prototyping within the timeframe of FREYA	Lead partner for requirements gathering in this report
Instruments on German research vessels - Instrument IDs	yes	PANGAEA
Facilities (photon and neutron sources) - Facility IDs	yes	UKRI-STFC
Grants - DOIs	yes	EMBL-EBI
Organisations - ROR-IDs	yes	DataCite
Research campaigns - Cruise IDs	yes	PANGAEA
Software - DOIs	No	DataCite
Data Management Plans	No	DataCite
Samples - IGSNs, BioSample accession numbers; RRIDs; ARKs	No	BL
Samples (Cultural Artefacts)	No	BL
Conferences	No	CERN

Integration into the PID graph



How to Engage with FREYA

Frances Madden
Research Identifiers Lead, British Library
9 May 2019

FREYA events

Europe

- EOSC projects
- E-infrastructures

Global

- International PID workshops
- PIDapalooza
- RDA IGs & WGs

Webinars

Watch past webinars on our YouTube channel!

Training

Upcoming: PID training events for different stakeholders

www.pidforum.org

Building a global PID community

- User stories
- PID best practices
- Training materials
- News, blogs, jobs
- PID events
- Discussions

If you're interested in PIDs, please join!

The PID Forum

Ambassador Programme

- Two way dialogue
- Engage with the community
- Amplify the project
- Insight into the project

27 Ambassadors
16 countries
Accepting new ambassadors now!
Arctic Ocean
Arctic Ocean
Greenland
Iceland
Finland
Sweden
Russia
Canada
United States
Mexico
North Atlantic Ocean
Ukraine
Kazakhstan
Mongolia
China
South Korea
Japan
Italy
Turkey
Iraq
Iran
Afghanistan
Pakistan
India
Thailand
Algeria
Libya
Egypt
Saudi Arabia
Mali
Niger
Sudan
Chad
Nigeria
Ethiopia
Kenya
DRC
Tanzania
Angola
Namibia
Botswana
Madagascar
South Africa
Indonesia
Papua New Guinea
Australia
New Zealand
Venezuela
Colombia
Brazil
Peru
Bolivia
Chile
Argentina
South Pacific Ocean
South Atlantic Ocean
Indian Ocean

How to Engage with FREYA

Our website:	www.project-freya.eu
The PID Forum:	www.pidforum.org
Twitter:	@freya_eu
Get in touch:	info@project-freya.eu