



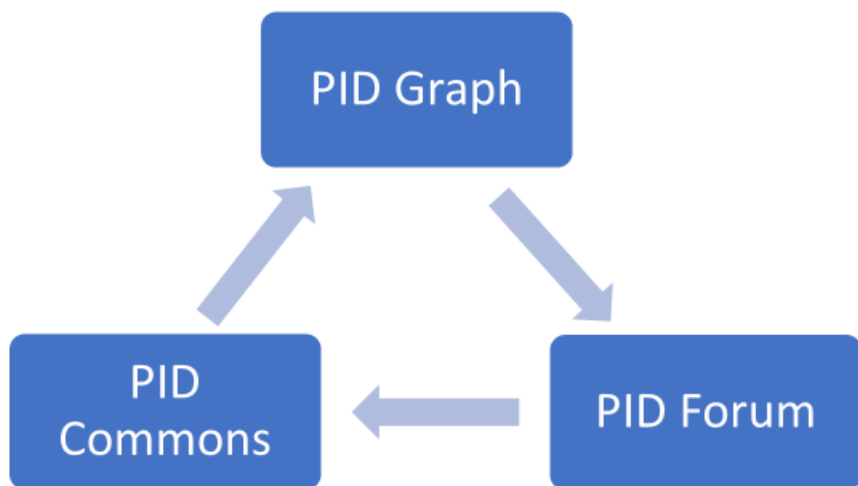
Services to support FAIR data

Vienna, 24 January 2019

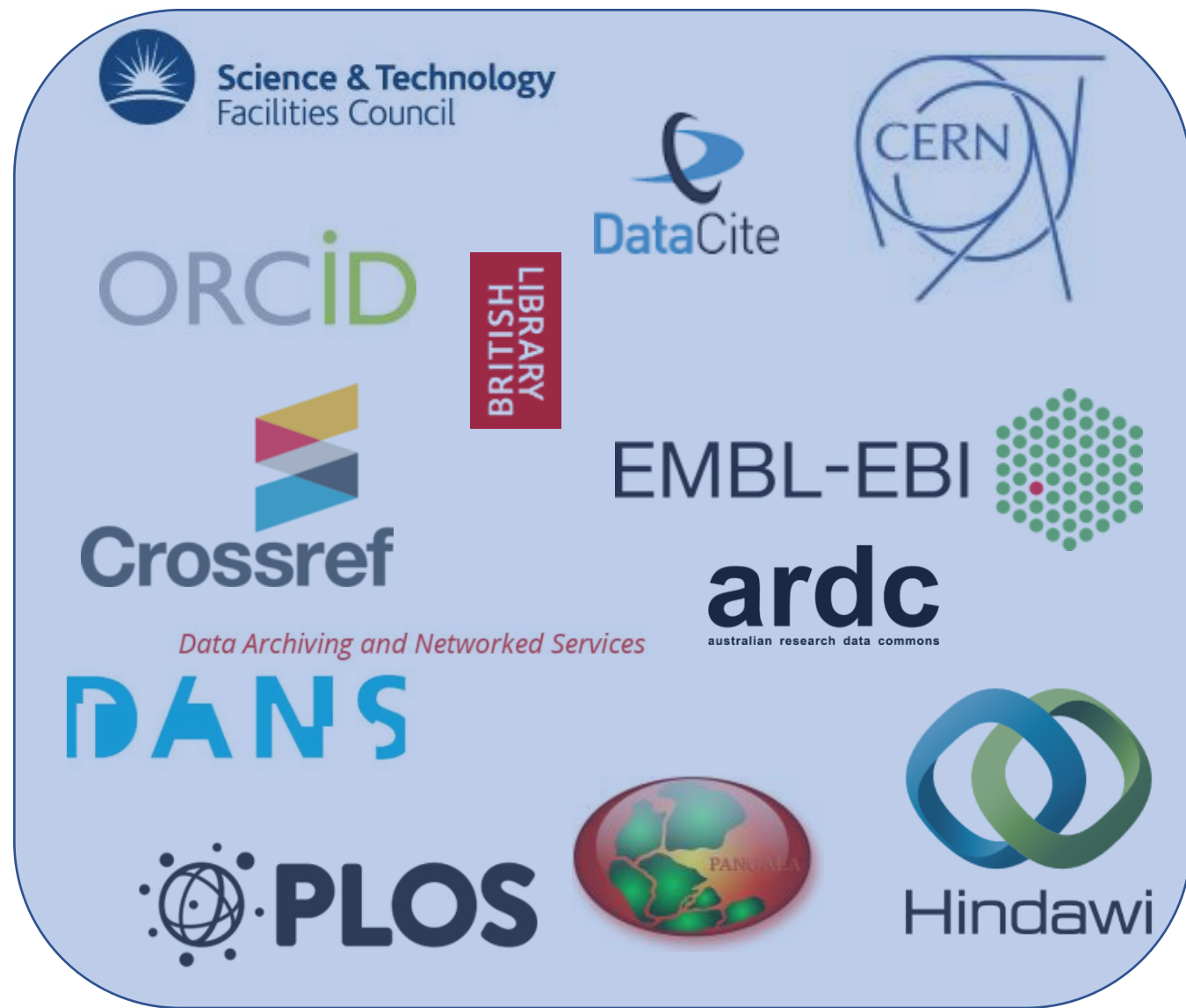
Simon Lambert
UKRI-STFC
FREYA Project Coordinator

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 !



Here a few persistent identifiers ...

10.1016/j.cell.2019.02.028

10.5286/ISIS.E.24089792

0000-0001-5633-2459

9783642168086

With a little more context ...

Published: April 04, 2019 • DOI: <https://doi.org/10.1016/j.cell.2019.02.028>

Authors Artemis Lavasa (CERN, orcid.org/0000-0001-5633-2459)
Sünje Dallmeier-Tiessen (CERN, orcid.org/0000-0002-6137-2348)
Stephanie van de Sandt (CERN, orcid.org/0000-0002-9576-1974)
Ioannis Tsanaktsidis (CERN, orcid.org/0000-0002-1567-3676)

Full Record Details

| | |
|----------------|---|
| Persistent URL | http://purl.org/net/epubs/work/39168516 |
| Record Status | Checked |
| Record Id | 39168516 |
| Title | Experiments with sparse Cholesky using a sequential task-flow implementation |

Public release date: 30 October 2015

Principal Investigator: *Professor Andre Strydom*
Experimenter: *Dr Devashi Adroja*

DOI: 10.5286/ISIS.E.24089792

• EUDAT Service ToU

- <https://hdl.handle.net/11304/e43b2e3f-83c5-4e3f-b8b7-18d38d37a6cd>

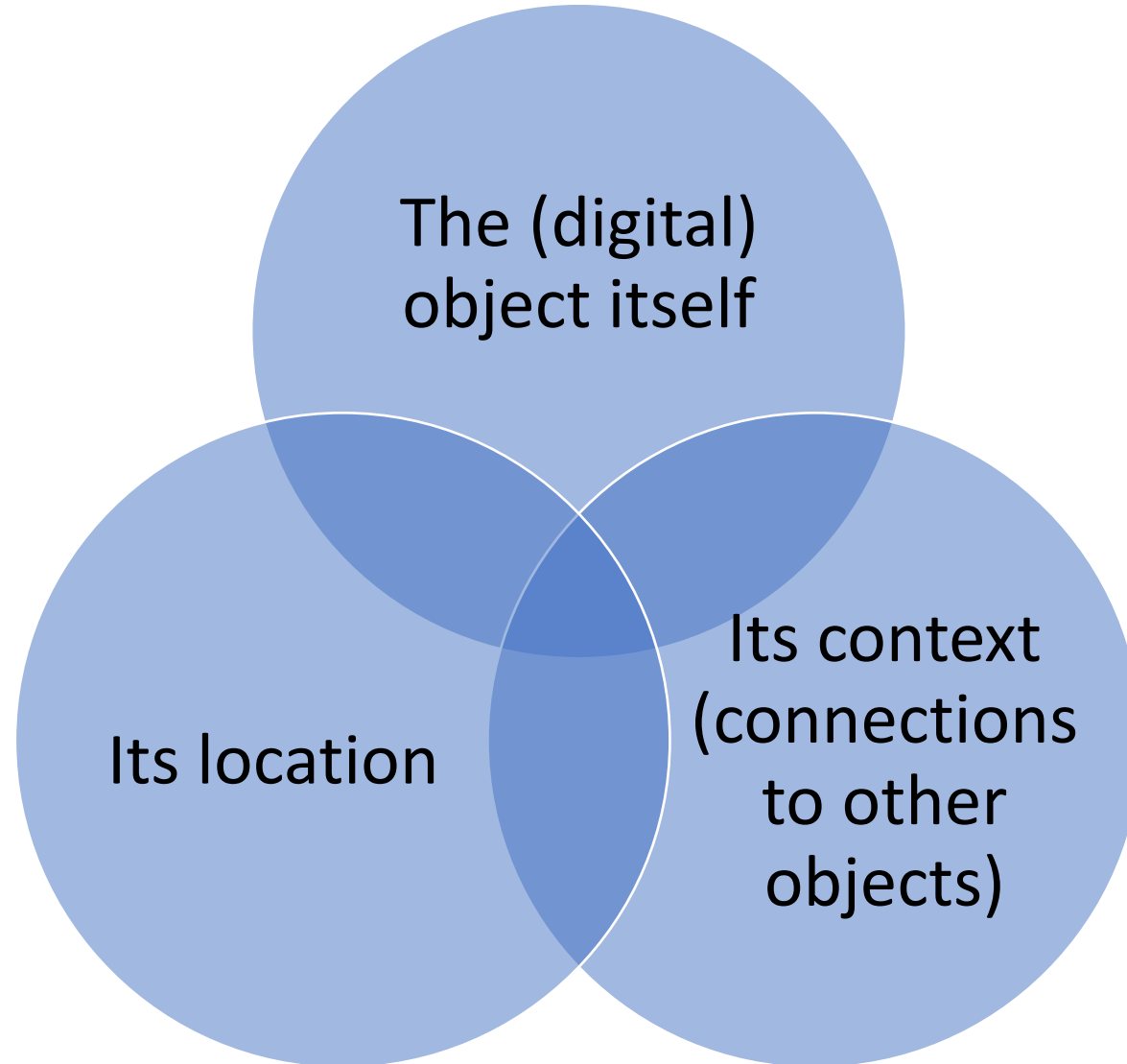
| | |
|-------------------|--------------------------------------|
| Title | Advanced digital preservation |
| Author | Giaretta D |
| Publisher: | Springer, |
| Pub date: | 2011 |
| ISBN: | 9783642168086 |

Where are the services?

- At minimum: issue PIDs and resolve them (and keep on doing that in the future)
- What more can PIDs offer?

- What kind of services are we talking about?
- “Technical” services (through API or user interface) vs. “human” services (e.g. training)
- Attributes of a service:
 - Something definite
 - ... expected to be sustained
 - ... of value to external parties
 - ... made available to them in a definite way

What can services relate to?



Where are the PID services in FAIR?

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.

Where are the PID services in FAIR?

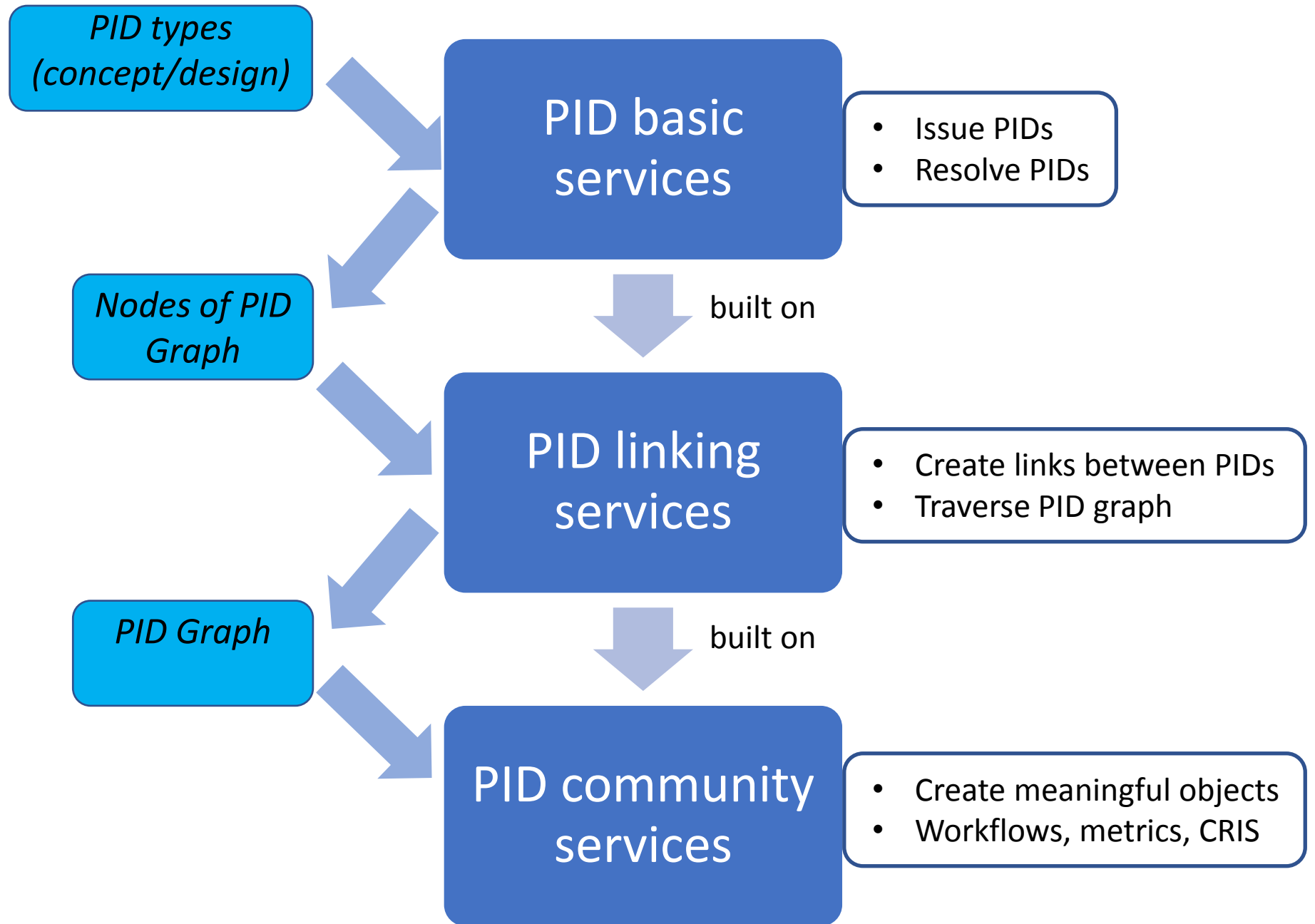
TO BE INTEROPERABLE:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

TO BE RE-USABLE:

- R1. meta(data) have a plurality of accurate and relevant attributes.
- R1.1. (meta)data are released with a clear and accessible data usage license.
- R1.2. (meta)data are associated with their provenance.
- R1.3. (meta)data meet domain-relevant community standards.

What are the services in FREYA?



Some user stories

Bases for
building PID
graphs to
satisfy needs

As a facility, I would like to track the published output related to the instruments provided. Also, i would like to be able to evaluate their impact.

As a funder, I want to track down the outcomes and beneficiaries of PhD studentship awards that I granted. There are many possible questions to be answered (with the help of PID graph):

- whether the PhD studentship actually ended in thesis, how to find and how to cite this thesis,
- what organizations benefited from PhD during or soon after the PhD research period, e.g. by hiring the PhD that I sponsored,
- who co-funded or otherwise supported the PhD research,
- what artefacts (papers, data, software, samples, instruments, ...) can be identified that either contributed to the PhD research or are the PhD research outcomes.

This issue is related to [#35](#) and in part to [#68](#) (as facilities are frequent supporters of PhDs).

Some user stories

As a repository manager, I want to get notified of new citations of datasets hosted in my repository, so that I can inform the authors.

As a software author, I want to be able to see the citations of my software aggregated across all versions, so that I see a complete picture of reuse.

"As a researcher at a scientific research facility, I want to authenticate the provenance of my soil sample by linking it to a specific research site, and to access information about the historic treatment of that site."

Services “within” PID infrastructure

- Example: DOI metadata provenance
- Available through API endpoint
 - **prov:wasGeneratedBy**. The unique identifier for the activity making changes to a DOI record.
 - **prov:generatedAtTime**. Timestamp of the activity.
 - **prov:wasDerivedFrom**. The DOI for which the changes are being tracked.
 - **prov:wasAttributedTo**. The client or provider account responsible for the changes.
 - **action**. Can be either create, update or delete.
 - **version**. Version number for the DOI record.
 - **changes**. Changes made to DOI metadata, broken down by attribute, with both old and new value.

Services “on top of” PID infrastructure (= “PID superstructure”?)

Example of
extended
PID graph

