STFC case study: PhD research graph

Vasily Bunakov (STFC UKRI, FREYA project)
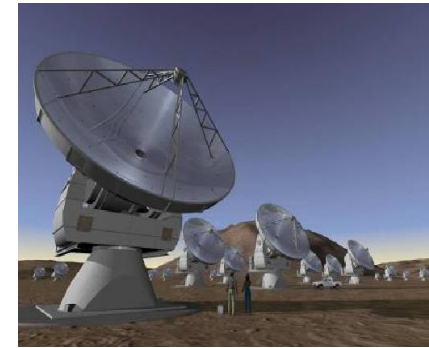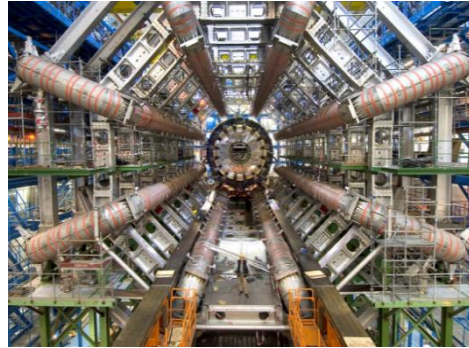

RDA UK and FREYA workshop, July 2019

Science and Technology Facilities Council and its research facilities

STFC funds and operates large scale instruments for the UK and visitor researchers in:
- physics, astronomy
- chemistry, materials
- biology, medicine

STFC research facilities:
- ISIS neutron and muon source
  www.isis.stfc.ac.uk
- Central Laser Facility
  www.clf.stfc.ac.uk

- Diamond Light Source
  (co-owned by STFC and Wellcome Trust)
  www.diamond.ac.uk

How we develop pilots in FREYA: with user stories

Tracking down PhD studentship outcomes, beneficiaries, co-funders and supporters #69

New issue

⊘ Open   **vasilyb** opened this issue on 14 Aug 2018 · 1 comment

vasilyb commented on 14 Aug 2018 · edited ▾   ···

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).

🏷 ⊞ **vasilyb** added `researcher` `STFC` `data` `instrument` `article` `person` `organization` `software` `funder` labels on 14 Aug 2018

🏷 ⊞ **uschindler** added the `PID Graph` label on 22 Aug 2018

Assignees

No one assigned

Labels

`British Library`
`DataCite`
`PID Graph`
`STFC`
`WP3`
`article`
`data`
`funder`
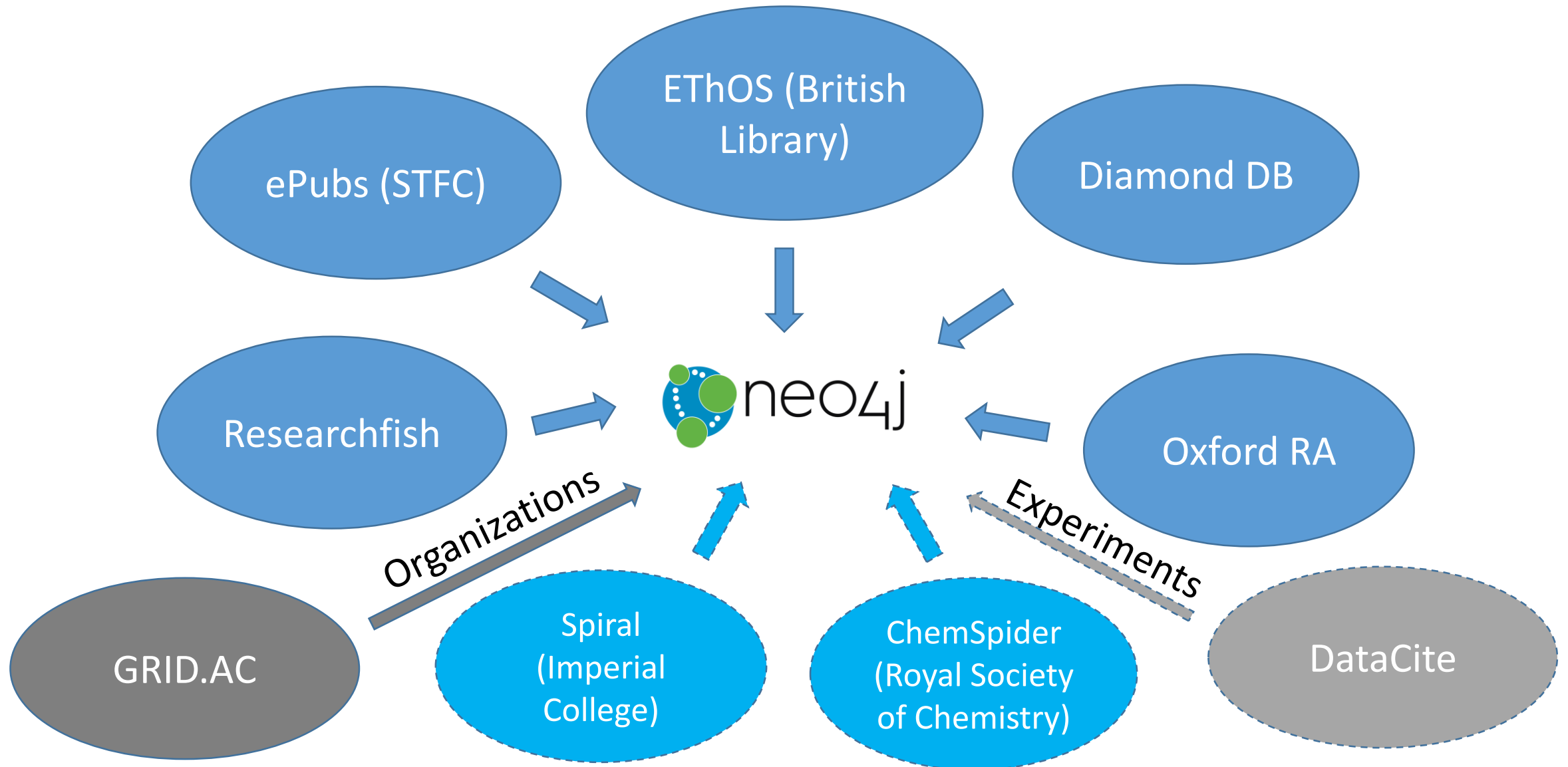`instrument`
`organization`
`person`
`researcher`
`software`
`user story`

PhD research use case as an illustration of case-driven approach in general

- **Collaboration:** British Library and STFC are the FREYA partners and operate repositories that can be used for data integration

- **Validation** of new PID services and recommendations – for Organizations and Instruments – and supplying feedback for their improvement

- **Demonstration** of PID graph value in a disciplinary context

- **Integration** of a disciplinary graph in a common PID graph (federation of services) via reasonable interfaces

Data sources to integrate

Enrichment and cross-repository harmonization of records as an incentive for building a knowledge graph with as much use of PIDs as possible

MATCH (ethos:EThOS_Thesis)-[r:sameThesisAs]-(x) WHERE ethos.Funders IS NULL RETURN count(ethos)	MATCH (ethos:EThOS_Thesis)-[r:sameThesisAs]-(x) WHERE ethos.Funders IS NOT NULL RETURN count(ethos)
454	150

Cases where STFC sponsored a PhD research (via monetary funding or via facilities' grants-in-kind) but EThOS "Funders" is empty

Where STFC is mentioned as a funder, another issue is observed: EThOS "Funders" is currently a free-text, so STFC can be referred to as:
 "Science and Technology Facilities Council (STFC)"
 "Science and Technology Facilities Council"
 "Science & Technology Facilities Council"
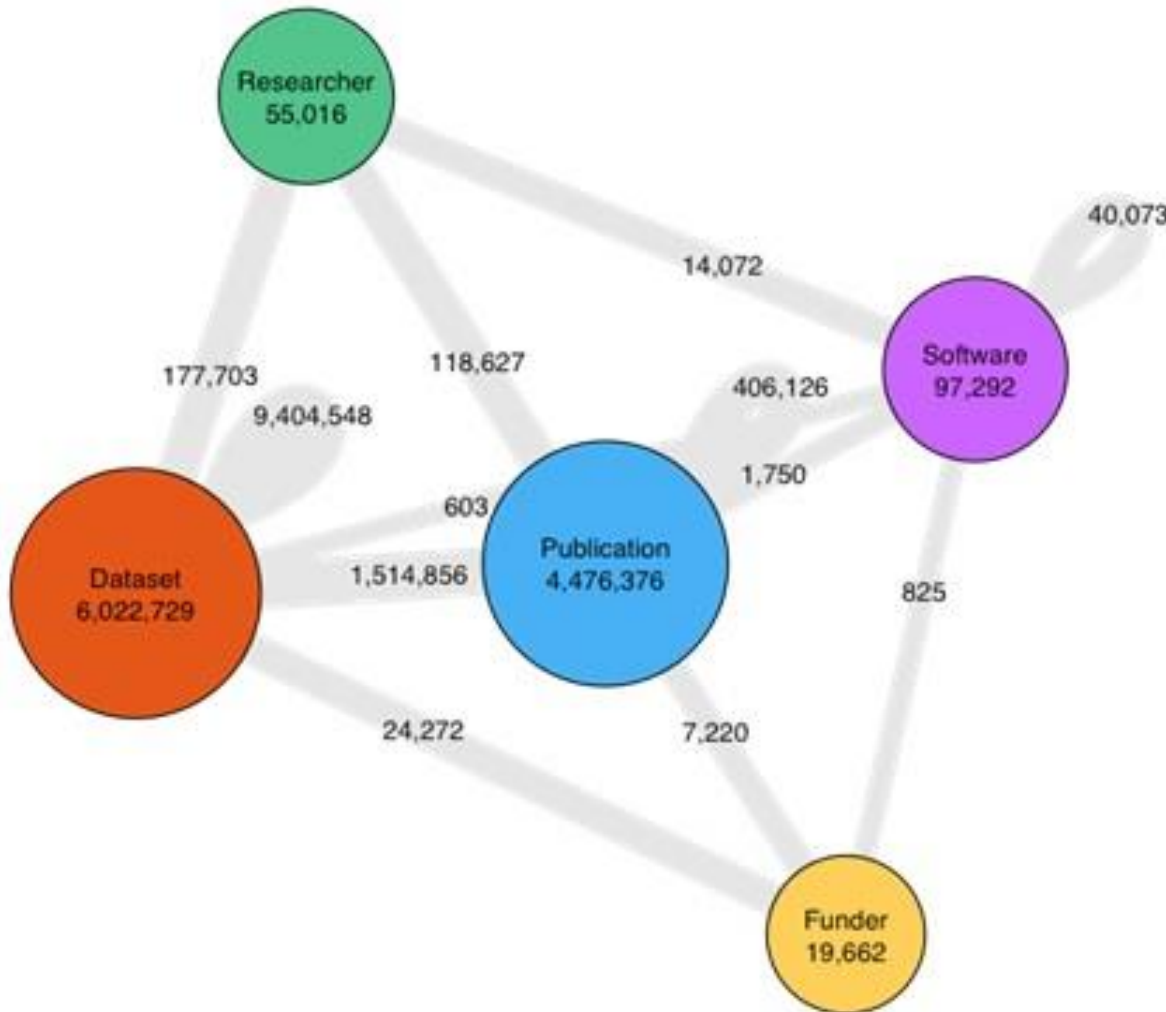 "Science and Technology Facilities Council (Great Britain) (STFC)"
 "STFC"

Imperial College PhDs
who experimented on STFC facilities

Another graph example
(with connections to EThOS and ChemSpider)

Next step in data integration: extension of the PhD research graph with chemical information

- Match more ChemSpider records with EThOS records, to build a bigger "seed" PhD research graph

- Align this effort with the forthcoming upgrade of the EThOS IT infrastructure

- Devise a new workflow for chemical information recording by PhDs in a few good UK universities

- Sounds like a dedicated project that a UK funder could be interested in

Next step in service development:
inclusion of the PhD research graph
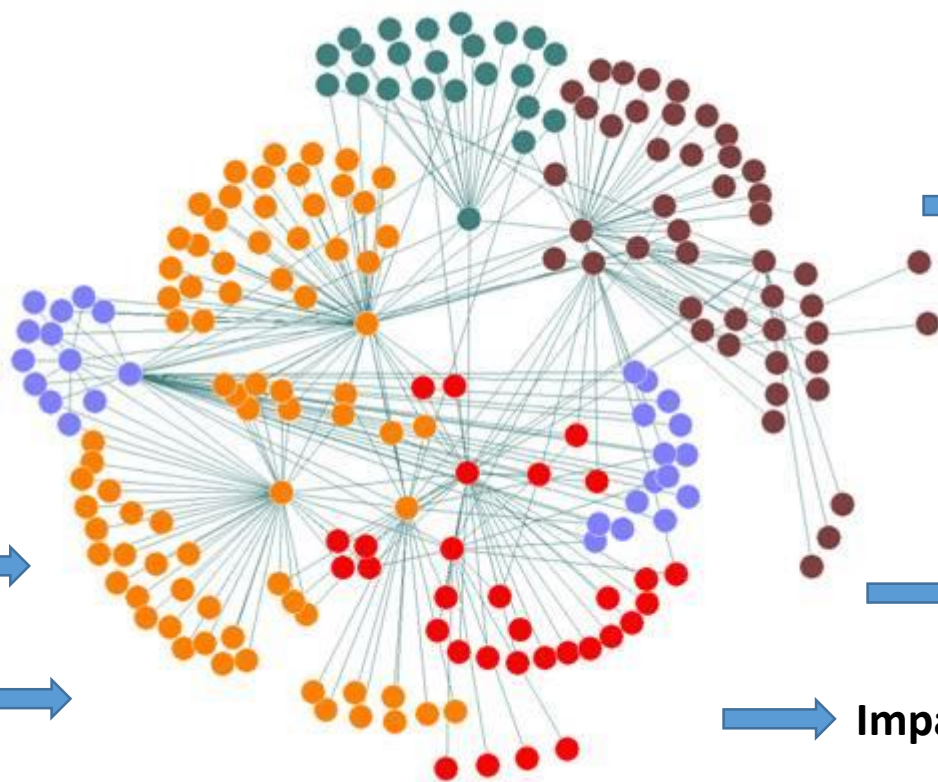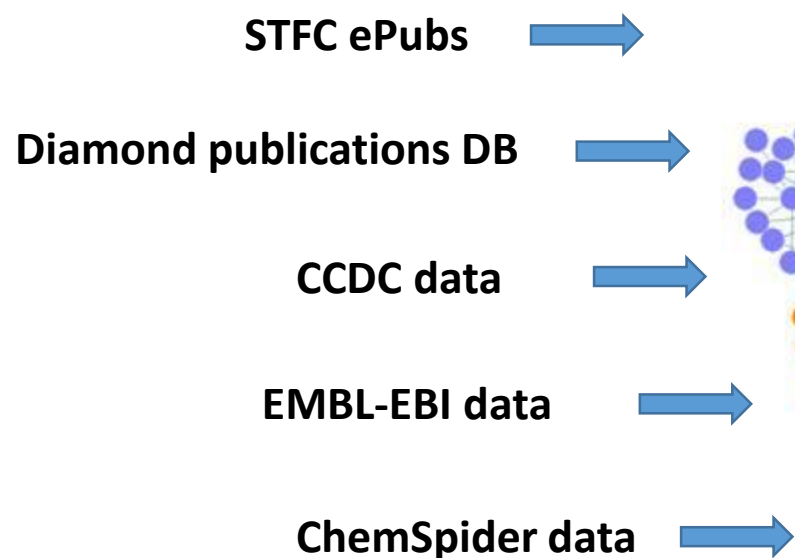# in a federation or services with a common GraphQL API
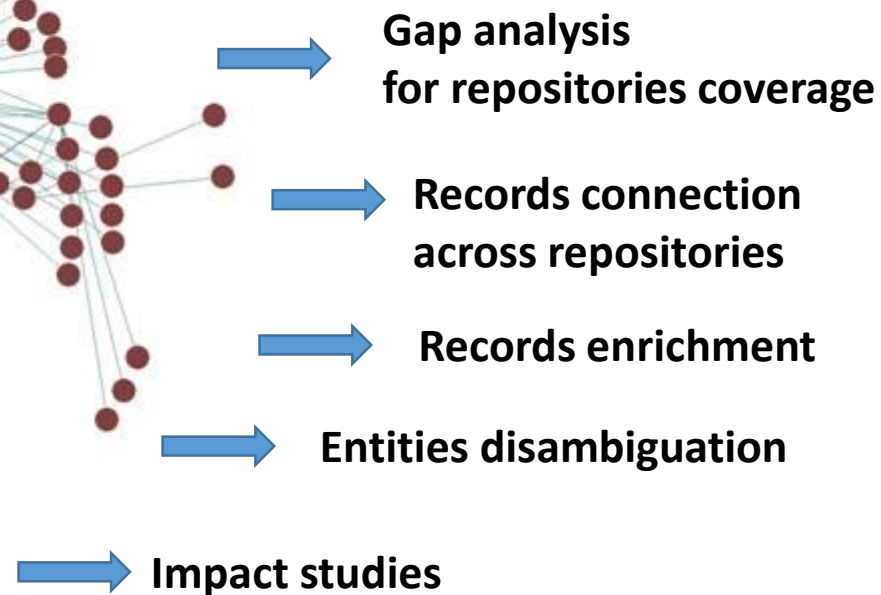

See Martin Fenner's blogpost:
https://doi.org/10.5438/bv9z-dc66

See also Jupyter notebooks:
https://github.com/datacite/notebooks

See also the section of PID Forum devoted to PID graphs: https://www.pidforum.org/c/pid-graph

PID graph as a (new kind of) infrastructure for Open Science

Possible data sources

STFC ePubs →

Diamond publications DB →

CCDC data →

EMBL-EBI data →

ChemSpider data →

Possible uses

→ Gap analysis
for repositories coverage

→ Records connection
across repositories

→ Records enrichment

→ Entities disambiguation

→ Impact studies

Use case connections to the RDA groups

- RDA Persistent Identification of Instruments WG (results adoption phase)

- Open Science Graphs for FAIR Data IG (community review and then the actual works of the group)

- Research data needs of the Photon and Neutron Science community IG (intermittent activity)

UK dimension:
RDA works in the national context

- Champion certain RDA outcomes and RDA ongoing effort in our organizations and disciplinary communities

- Offer RDA UK as a platform for networking with the EU and other international projects and initiatives

- Educate UK policy makers, and support continuous engagement with them

- Use RDA UK as a launchpad for the UK-funded projects

Thank you!

Web: www.project-freya.eu
Email: info@project-freya.eu
Twitter: @freya_eu

PID Forum: www.pidforum.org