



FREYA

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

Brian Matthews
Data Science and Technology Group
Scientific Computing Department
STFC

Persistent Identifiers



Long-lasting reference to a resource

- The referent of the identifiers "persists" over time
 - Does not change
 - Does not disappear
 - But could be moved
- The identifier is globally unique
 - Created under some authority
 - Metadata
- May be de-referenceable
 - Landing page
 - Possibly to the (digital) resource itself



Why PIDs?



- To know what you are talking about really is what you are talking about
 - Especially when they are outside your control
- Maintaining integrity as
 - Resources change, versions, rewrites
 - Resources move around
- Publishing into the scholarly record
 - Citation
 - Credit
 - Reproducibility
 - Reuse



PID Schemes



- DOIs
 - Publications
 - Data
 - ORCID

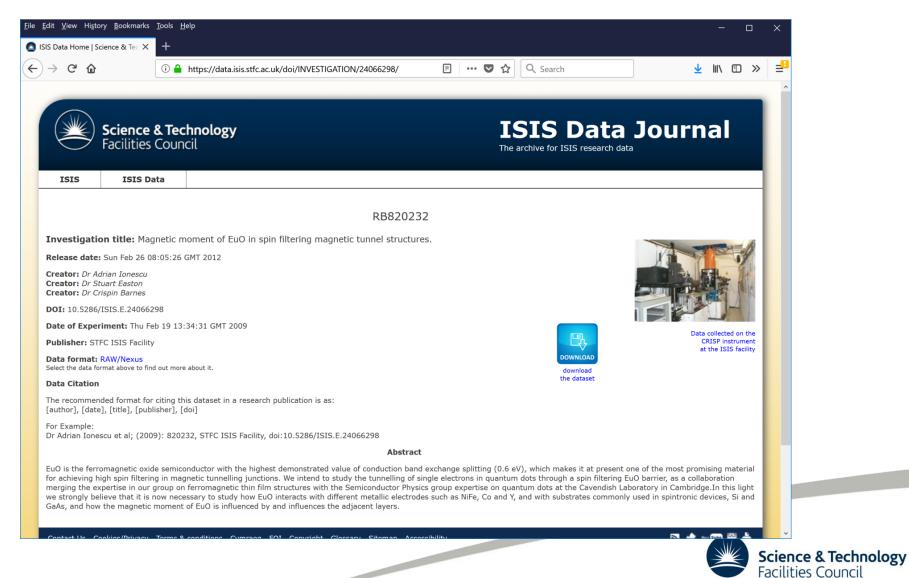
- ARK
- Handle
- PURL
- URN

- ResearcherID
- INCHI
- LSID
- ISNI
- ISBN
- ...



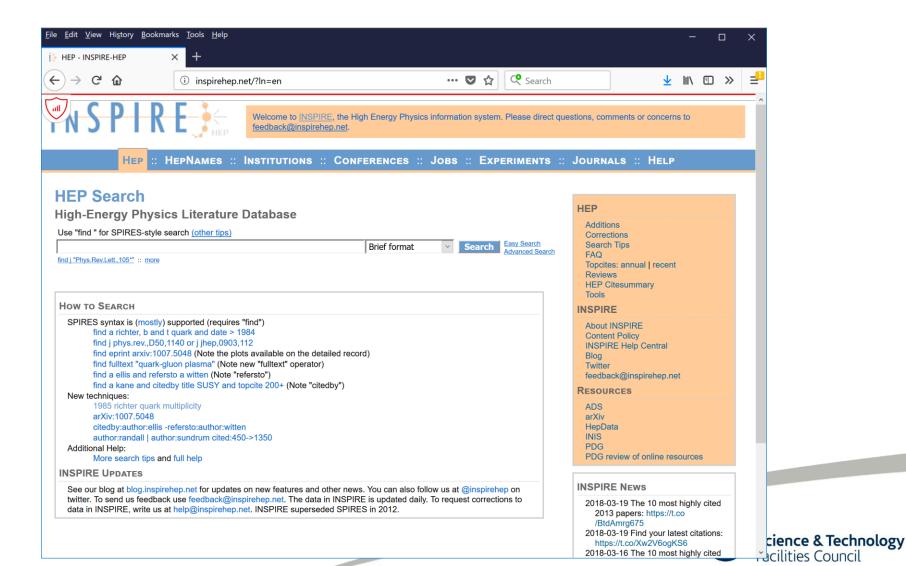
ISIS Landing Page





http://inspirehep.net/?In=en





The FREYA Project

- FREYA = persistent identifiers
 - "... iteratively extend a robust environment for Persistent Identifiers (PIDs) into a core component of European and global research einfrastructures"
 - Part of the European Open Science Cloud
- Builds on THOR
 - which in turn built on ODIN
- Started 1 December 2017











The FREYA partners































Three pillars of FREYA



An **output** of FREYA

PID Commons ...

- feeds requirements into PID Graph
- ensures population of PID Graph
- establishes governance and sustainability of PID Graph and services on it
- establishes rules of engagement for PID Graph

PID Graph

PID Graph ...

 gives benefits (through services) to communities of PID Forum

PID Commons

An **outcome** of FREYA

PID Forum ...

- convenes stakeholders
- feeds requirements for PID Commons

PID Forum

A working tool and an output of FREYA



New Services for existing PIDs



Making existing PID service providers more robust and useful

- Multiple Locations
 - Describing the criteria for picking one amongst several resource locations
- Direct Access to content
 - Exploring data resources that already provide content (DataCite, GitHub issue)
 - identify life sciences databases that provide links to content
- Link Checker
 - Looking for stale content
- Core Metadata
 - Cross PID types
 - Schema.org



New PID Types



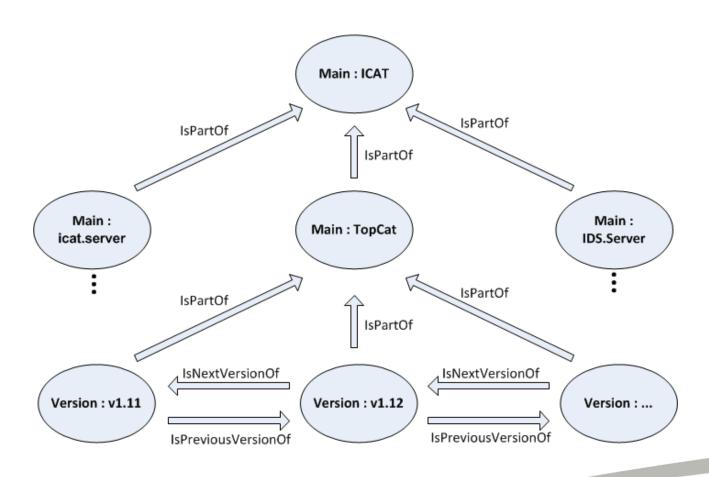
Develop services to support new PID types e.g.

- Funder
- Organisation
- Conference
- Project
- Experiments, Clinical trials
- Places
- Facility/Instrument/Equipment
- Software
- Sample, Physical Objects
- Licences
- Events
- These are at very different levels of development
 - Currently undertaking a review
- Common Metadata?



Example: Software



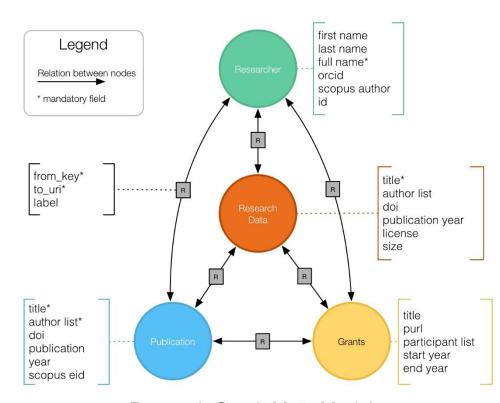




Building the PID Graph



- Developing a schema and tools to represent and use the PID graph
- Amir Aryani (ANDS)
 - http://researchgraph.org/
- Developed by:RDA Data Description Registry Interoperability WG
 - https://rdalliance.org/groups/datadescription-registryinteroperability.html
- Should take into account other work:
 - Research Objects
 - PROV-O
 - **DCAT**
 - **CFRIF**



Research Graph Meta Model

Version 2.0 (Aug 2016)



Implications for the PID Graph



- Provenance and attribution
- Possible new entities in research discourse e.g. "analysis"
- Usage patterns: formulating and testing research questions
- Borders between in-facility data management, in-community data handling, and cross-science pieces of scholarly communicable units

The PID Graph forms a superstructure, holding the dynamic, moving parts of the EOSC in place



What will FREYA be doing about it? ?

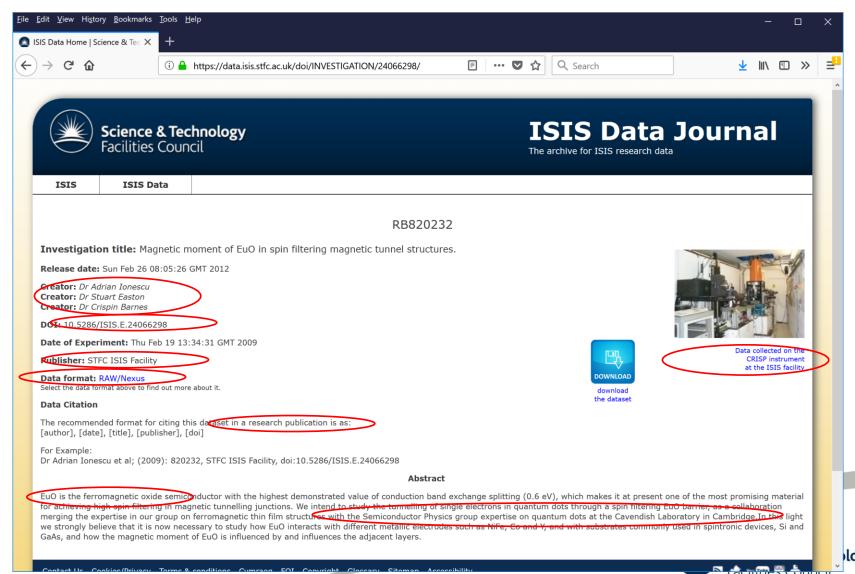


- FREYA's own disciplinary pilot applications
 - British Library, DANS: humanities and social sciences
 - CERN: high-energy physics
 - EMBL: life sciences
 - PANGAEA: earth and environmental sciences
 - STFC: facilities-based science with ISIS Neutron Source
- Each building a part of the PID Graph ...



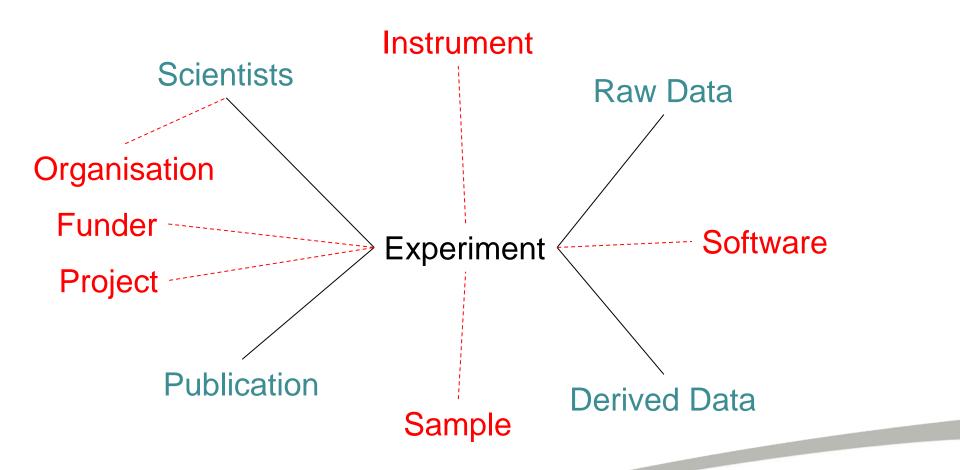
ISIS Landing Page





Building the PID Graph







STFC Use Case



- This is the service and user group
 - ISIS Data User and Data Management systems
 - ISIS User Office, ICAT Data catalogue, ePubs (institutional repository), eData (other data repository), SESC (software repository)
- This is what the service does now
 - Register user take their ORCID Id
 - Issue a grant of instrument time with locally unique ID
 - Record experiments and data collected on them
 - Issue DOIs for "data" (or is it experiments?) many years to develop a data policy which has supported this.
- This is the goal for the service during the next year (or beyond if necessary)
 - Access the Orcid system upload grant of instrument time to the Orcid record, extract relevant publications
 - Linking to our funder information GatewaytoResearch, ResearchFlsh funders, organisations
 - Make the notion of adding data collection datasets across experiments associated by the scientist.
 - Enrich the record equipment (instrument definition file), sample, software
 - Provenance record pulling it all together at an experiment level (but allowing people to pull together other ways).
- We need PIDs to enable...
 - Better connection between users, grants, data,
 - Finding papers which reference ISIS experiments/data/equipment
 - Links to equipment, software, samples
 - Aggregates pulling it all together at an experiment level



The PID Forum



- FREYA's PID Forum will be closely linked to the RDA
 - "The PID Forum is a stakeholder community, whose members collectively oversee the development and deployment of new PID types; it will be strongly linked to the Research Data Alliance (RDA)."
 - A foundation for the future PID Commons



Finally, this week



- Towards the PID Forum at the Research Data Alliance
- Berlin Plenary Meeting in March:
 - BoF with EOSC-hub and OpenAIRE-Advance
 - Wednesday 20 March
- Co-located event,
 - European Open Science Cloud-related projects: liaison with RDA and new contributions
 - Friday 22 March, 2pm.

