FREYA PID Graph

Martin Fenner
DataCite Technical Director
https://orcid.org/0000-0003-1419-2405

PID Graph

PID Graph is the Open Science Graph that the EC-funded FREYA project is building

1. PID Graph is a graph of scholarly resources that all have persistent identifiers (PIDs), and metadata that describe connections to other PIDs.
2. The FREYA project has identified the most important use cases, has defined the standards used for the technical architecture, and has deployed a first implementation that can be explored, for example using Jupyter notebooks.
3. FREYA is working towards production infrastructure and client applications in 2020.
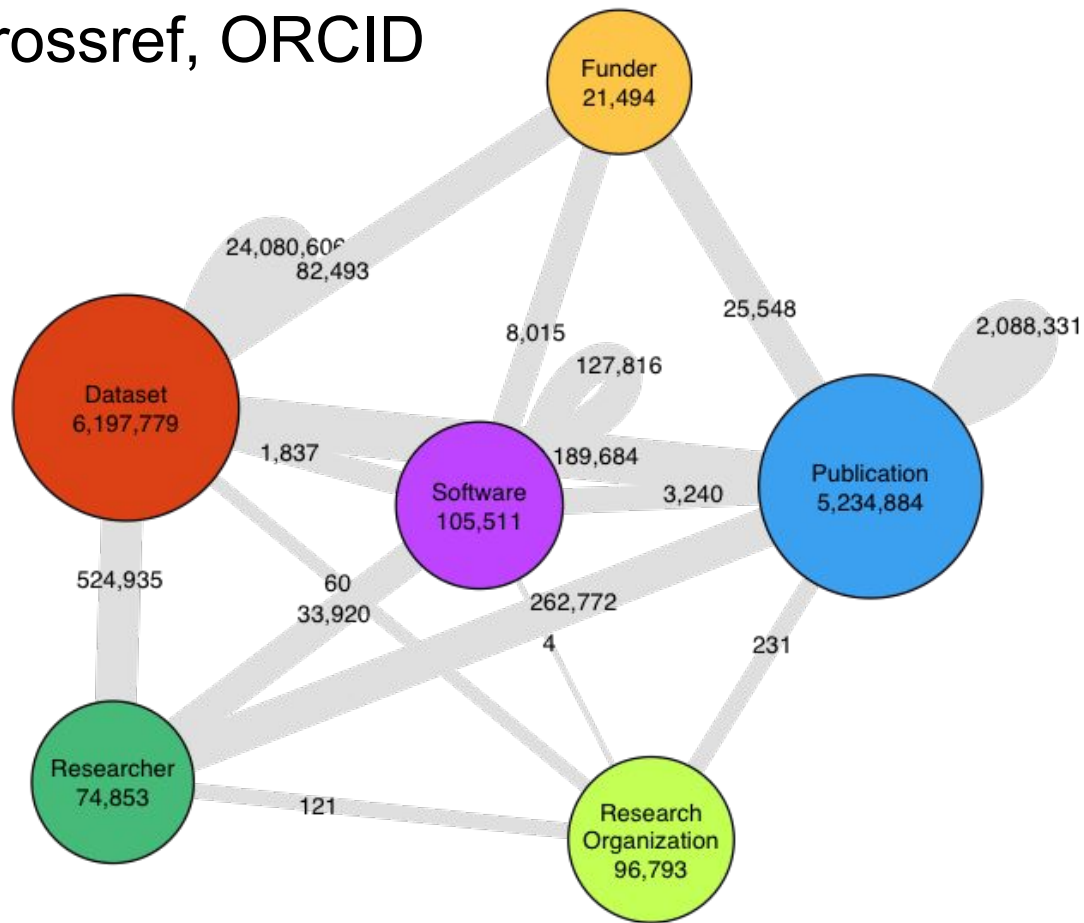
PID Graph Implementation

1. Use cases identified in August 2018 workshop as drivers
2. Technical architecture built around GraphQL, a mature and widely adopted API framework. Not yet used much in the scholarly community.
3. Jupyter notebooks as platform to analyze and visualize the PID Graph
4. Client applications by FREYA partners and the community to consume the PID Graph

PID Graph
Number of nodes and connections (17 September 2019)

Still missing: Crossref, ORCID

Funder 21,494

24,080,606
82,493

8,015
127,816

25,548

2,088,331

Dataset 6,197,779

1,837

189,684

Publication 5,234,884

Software 105,511

3,240

524,935

60
33,920

262,772

231

4

Researcher 74,853

121

Research Organization 96,793

https://doi.org/10.14454/3bpw-w381

Example user story

As a university administrator, I want to get a list of all datasets and software published by our researchers, so that I can get a comprehensive view of our research outputs

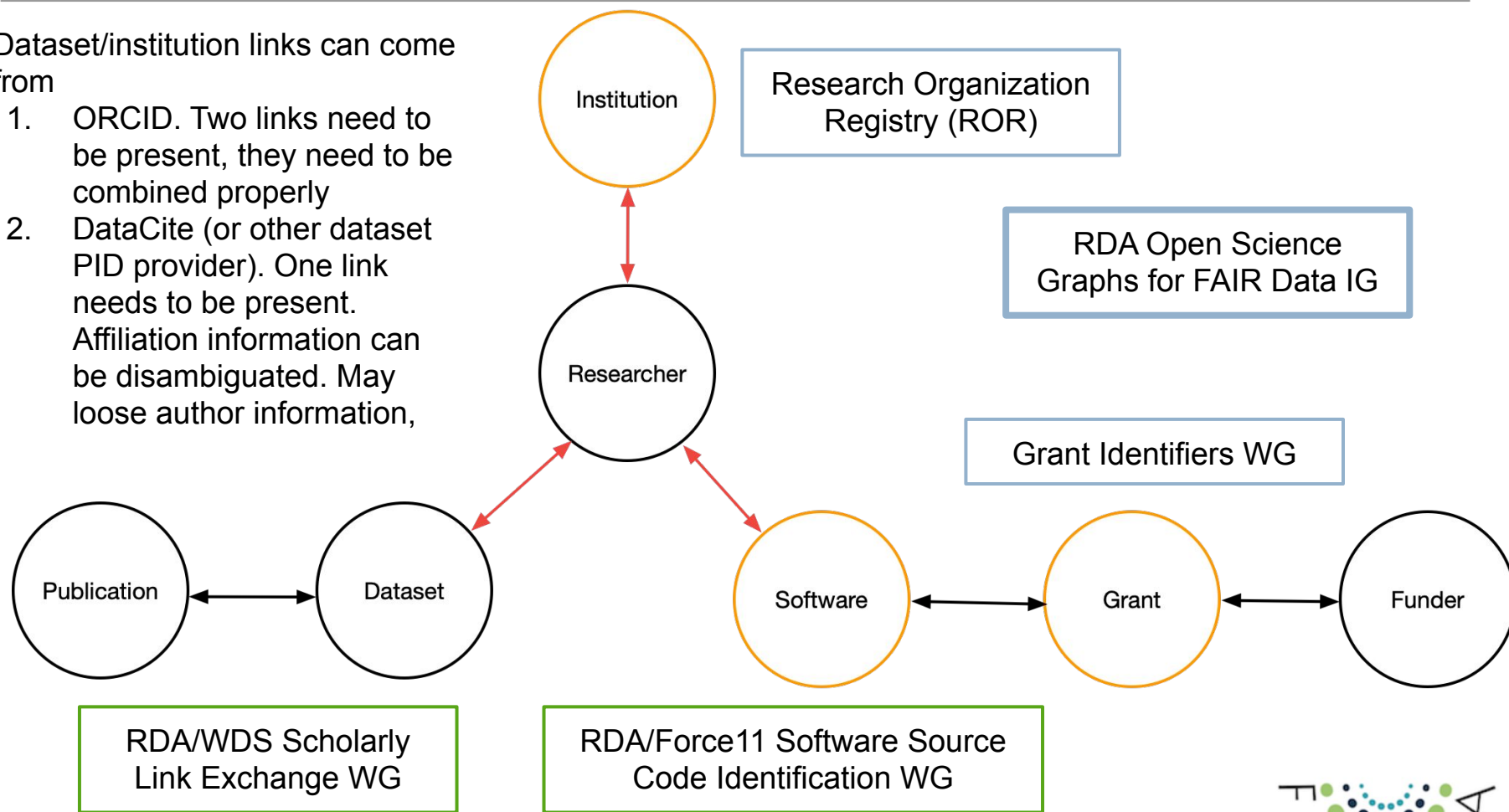If possible, can I also get all the data and software citations for these outputs.

And, please, also the funders and grants that supported these outputs.
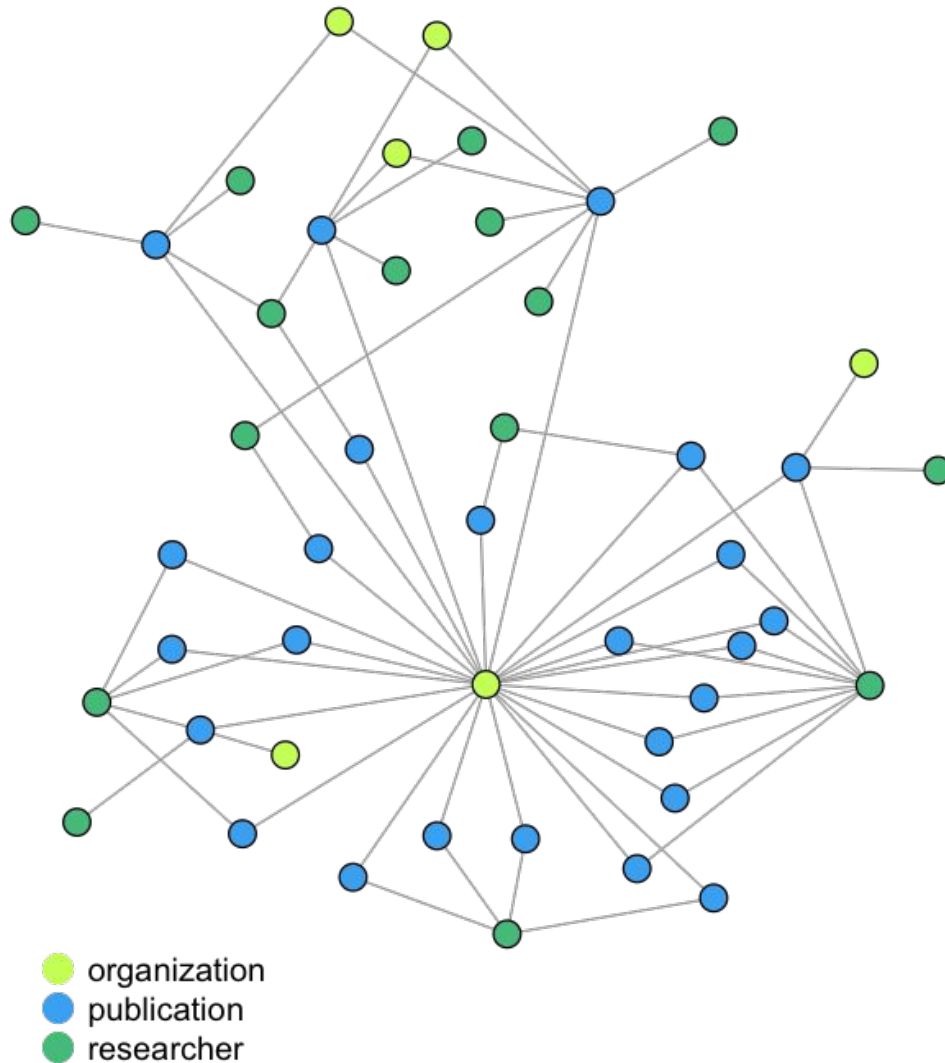
PID Graph for Example User Story

Dataset/institution links can come from
1. ORCID. Two links need to be present, they need to be combined properly
2. DataCite (or other dataset PID provider). One link needs to be present. Affiliation information can be disambiguated. May loose author information,

Institution

Research Organization Registry (ROR)

Researcher

RDA Open Science Graphs for FAIR Data IG

Grant Identifiers WG

Publication

Dataset

Software

Grant

Funder

RDA/WDS Scholarly Link Exchange WG

RDA/Force11 Software Source Code Identification WG

PID Graph for Organization DataCite

organization
publication
researcher

Open Science Graphs: Requirements

1. Relevant use cases
2. Relevant corpus of scholarly outputs
3. Openly available metadata for scholarly outputs describing relations to other outputs, researchers, institutions, grants, etc.
4. Technology for building and querying the graph, and for combining graphs from different sources
5. Clients and tools to access the graph
6. Coordination between the various Open Science Graph activities
7. Community adoption