



RDA Open Science Graphs for FAIR Data Interest Group

Martin Fenner, DataCite

research data sharing without barriers
rd-alliance.org

Case statement for interest group has been submitted in June, the group is not yet endorsed

1. Successful Research Data Graph BoF at RDA Plenary Philadelphia
2. Group page at <https://www.rd-alliance.org/groups/open-science-graphs-fair-data-ig>
3. Session submitted for RDA Plenary Helsinki
4. Intended Co-chairs are Paolo Manghi (OpenAIRE), Amir Aryani (Researchgraph.org), Wouter Haak (Elsevier) and Martin Fenner (DataCite)

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.

Possible approaches to address user story

1. Collect this information manually, including web searches, reaching out to all researchers at the institution, etc.
2. Pay someone to get this information. In this case a commercial Current Research Information System (CRIS) such as Symplectic Elements, Pure or Converis.
3. Help define standards and build infrastructure that dramatically reduces the effort required to collect this information. This is what the partners in the RDA Open Science Graphs Interest Group are trying to achieve.

1. Build on the outcomes and broaden the challenges of the Data Description Registry Interoperability (DDRI) and Scholarly Link Exchange (Scholix) RDA Working Groups
2. Investigate the open issues and identify solutions towards achieving interoperability between services and information models of Open Science Graph initiatives
3. The aim is to improve FAIRness of research data, and more generally FAIR*-ness of science, by enabling the smooth exchange of the interlinked metadata overlay required to access research data at the meta-level of the discovery-for-citation/monitoring and at the thematic level of the discovery-for-reuse.

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

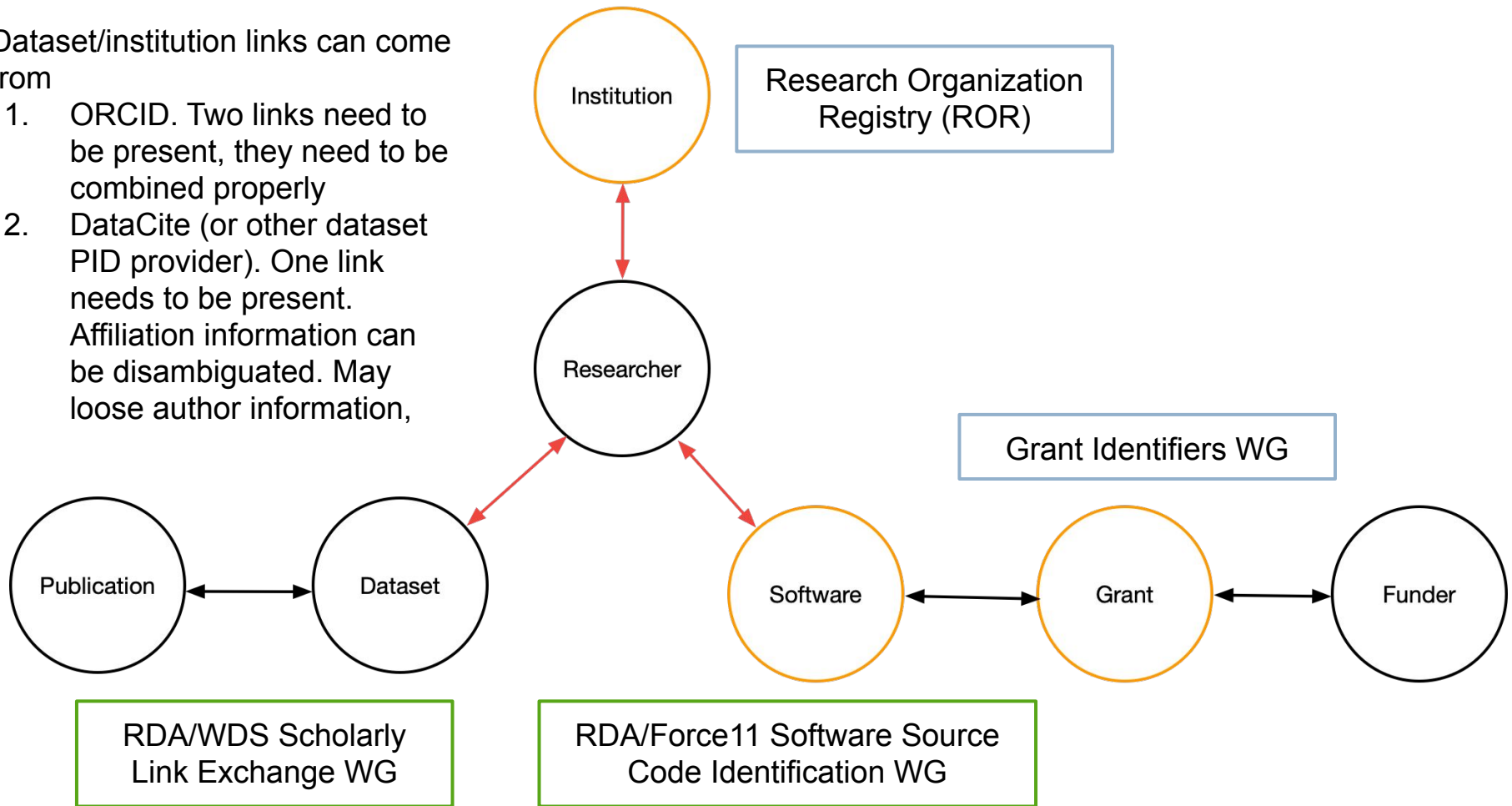
1. PID Graph is a graph of scholarly resources that all have persistent identifiers (PIDs), and metadata that describe connections to other PIDs.
2. The FREYA project has identified the most important use cases, has defined the standards used for the technical architecture, and has deployed a first implementation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



PID Graph 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 lose author information,



User Story Next Steps

1. **August:** support affiliation identifiers in DataCite metadata for datasets and software (and everything else)
2. **August:** extract all dataset/affiliation, software/affiliation (and publication/affiliation) links from DataCite metadata and send them to Event Data service, and include these links in PID Graph query API
3. **From August:** align with ORCID how these links can best be shared and included in PID Graph query API
4. **From August:** adoption, adoption, adoption
5. **October:** align with other Open Science Graph initiatives at RDA Plenary Helsinki
6. **November:** present available data at next euroCRIS meeting, FREYA deliverable describing organization identifiers in PID Graph