

# Open Science Graphs for FAIR Data Interest Group

## *Case statement*

### **Co-chairs:**

- Amir Aryani (Research Graph Foundation)
- Martin Fenner (DataCite)
- Wouter Haak (Elsevier, Mendeley Data)
- Paolo Manghi (OpenAIRE Infrastructure - Institute of Information Science and Technologies, CNR, IT)

## **Mission**

The goal of the Open Science Graphs Interest Group (OSG WG) is to build on the outcomes and broaden the challenges of the Data Description Registry Interoperability (DDRI) and Scholarly Link Exchange (Scholix) RDA Working Groups to investigate the open issues and identify solutions towards achieving interoperability between services and information models of Open Science Graph initiatives. 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. Such “FAIR-ness” and “interlinked-ness” provide strong support for research integrity and research innovation which in turn underpin significant social environmental and economic benefits.

## **Objectives**

Open Science is urging scientists, communities, institutions, and policymakers to refine and adopt methodologies, practices, and tools for publishing research products, beyond the scientific article, including research data, software, digital experiments, etc. The ultimate goal is to achieve transparency and reproducibility of science. As a consequence of this trend, researchers are depositing into scholarly communication data sources the metadata and files relative to all these products, together with semantic links between them, and towards other relevant entities, such as those kept in registries for authors, organizations, and data repositories (e.g. ORCID, ROR, re3data.org). De facto, Open Science publishing practices materialize a distributed/federated/de-centralized and global Open Science Graph. Needless to say, there is a great interest to contribute to and/or consume this Graph for sharing, discovering, and monitoring Open Science. To address this, several initiatives are aggregating targeted subsets of such sources to build specialized Open Science Graphs, subsets of the global Open Science Graph, capable of serving specific user needs: Google Scholar, Microsoft Academics, Scopus, FREYA PID Graph, Research Graph Foundation, OpenAIRE Research Graph, Open Knowledge Graph, Human Brain Project Knowledge Graph, as well as the

CERIF graphs built via CRIS systems are just a few of the real-case graphs being built and consumed out there.

Clearly, FAIRness of research data strongly relies on the success and diffusion of such graphs, both at cross-discipline level and at the thematic level. Research data (as well as software, or any research object) can be contextualized, thus maximizing its value and ability to reuse, and be reachable via navigation from other related objects. Nonetheless, research data value and related scientific reward may be derived from its constantly updated context, relying on a network of citations and usage statistics. The general architecture and use cases have been studied by the RDA Data Description Registry Interoperability (DDRI) Working Group. Besides, the RDA/WDS Scholarly Link Exchange (Scholix) Working Group has added significantly to our understanding of the subset of the research data graph connecting data and literature. Both working groups have done extensive work on the implementation of services and community adoption and are today in maintenance mode. However, other projects continue to work on these outcomes, including the OpenAIRE Research Graph, the Scholix Explorer graph, Research Graph, and the FREYA PID Graph. Driven by such motivations, the co-chairs of the aforementioned WGs organized a BoF on “Research Data Graphs” at the RDA Plenary Conference in Philadelphia, to check on the general interests and possible commitments on this topics, which resulted on this IG case-statement and proposal.

The goal of the Open Science Graphs Interest Group (OSG WG) is to build on the outcomes of DDRI and Scholix RDA Working Groups to investigate the challenges and identify solutions towards achieving interoperability between services and information models of Open Science Graph initiatives. The aim is to improve FAIRness of research data, and more in general 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 and at the thematic level of the discovery-for-reuse. Three main challenges can be identified as the worth of investigation:

- 1) **Build a community** of Open Science Graph initiatives working together in the context of RDA with a focus on FAIR data.
- 2) **Analyse the state of the art** in this domain, by making synergies with the tens of initiatives today building Open Science Graphs and provide an overview of current research data graph activities to frame a definition and classification of such graphs;
- 3) **Study the foundations of an information model**, a lingua franca, that would enable the realization of an interoperability layer facilitating the exchange of information between graphs;
- 4) **Discuss the ideal services, protocols, and APIs required to exchange graphs**, query graphs, navigate graphs in both aggregation scenarios and federated access scenarios.
- 5) **Identify one or more dedicated RDA Working Groups** to tackle/address relevant challenges.

## Participation

We envisage participation from the following kinds of actors:

- Open Science Graphs information models (e.g. Scholix.org, CERIF, Research Graph, [OpenAIRE](#) and PID Graph information models)
- Open Science Graphs providers targeting specific end-users:
  - *Thematic graphs* for scientists
  - *Citation graphs* for scientists, communities, institutions, policymakers, funders
  - *Monitoring graphs* for institutions, policymakers, funders
- Open Science Graphs consumers: scientists, communities, institutions, policymakers, funders

### Already Committed

- RDA Data Description Registry Interoperability (DDRI) Working Group (Benjamin Zapilko)
- RDA Scholarly Link Exchange Working Group (Wouter Haak)
- OpenAIRE Research Graph (Paolo Manghi)
- FREYA PID Graph (Martin Fenner)
- Research Graph Foundation (Amir Aryani)
- Thematic Graph (Mark Parsons)
- Open Research Knowledge Graph (Markus Stocker)
- CERIF/EuroCris (Jan Dvorak)
- DBPedia/Wikipedia (Daniel Mietchen)

## Interaction Mechanisms

OSG IG members will interact by means of a dedicated D4Science Virtual Research Environment (web site, mailing lists, online file system) and publish documents in Zenodo.org, in a dedicated Collection. The co-chairs will meet virtually three times a year (between plenaries), while members will regularly meet at the RDA plenaries twice a year. Other virtual meetings will be organized to address specific topics by sub-groups of interested members.

## Outcome

The Open Science Graphs for FAIR Data IG will be considered successful if:

- The main Open Science Graphs initiatives have started to form a community and have identified common goals and challenges.
- The challenges identified above will translate in recommendations and possibly in RDA Working Groups where standards, protocols can be defined and implemented by the committed members of the IG.
- Adequate visibility, participation, and uptake of the recommendations will be demonstrated within RDA and beyond.

## References

- <https://rd-alliance.org/groups/data-description-registry-interoperability.html>
- <http://researchgraph.org/>
- Most recent Research Graph publication: <http://doi.org/10.1038/sdata.2018.99>

- <https://rd-alliance.org/groups/rdawds-publishing-data-services-wg.html>
- <https://rd-alliance.org/groups/rdawds-scholarly-link-exchange-scholix-wg>
- <http://www.scholix.org/>
- Relevant Scholix publication <https://doi.org/10.1045/january2017-burton>