Research Graph: Connecting Researchers, Research Data, Publications and Grants using the Graph Technology

Nathaniel Lewis[1]

Jingbo Wang[2], Marta Poblet[3], Amir Aryani[4]
1 The University of Sydney, Australia, nathaniel.lewis@sydney.edu.au
2 The National Computational Infrastructure (NCI), Australia, jingbo.wang@anu.edu.au
3 RMIT University, Australia, marta.pobletbalcell@rmit.edu.au
4 Australian National Data Service, Australia, amir.aryani@ands.org.au

Abstract

In this BOF, we will discuss the challenge of connecting research information including linking researchers to their publications, datasets and research grants. The main focus of the BOF is on using the research graph and the Neo4j graph technology to link the scholarly works, and we demonstrate how these technologies have been implemented in The National Computational Infrastructure (NCI), Australian National Data Service (ANDS) and The university of Sydney using the open source software including Research Data Switchboard, Neo4j and Research Graph schema.

Description

Form 2014, the Data Description Registry Interoperability Working Group in Research Data Alliance has worked on a project called Research Data Switchboard. This project demonstrated how graph modelling and Neo4j graph technology can connect datasets across multiple platforms using the co-authorship and jointly funded project. The best metaphor for this solution is the SEE ALSO section in online bookstores, where customers are invited to look at other products by the same author, related topics or similar publishers. The following institutions have adopted the RD-Switchboard source code: NCI – National Computational Infrastructure, Australia, ANDS – Australian National Data Service, University of Sydney, Australia and National Institute of Informatics, Japan.

However, the early adopters of this project including NCI and The University of Sydney demonstrated that the functionality of the graph database can be a useful tool for identifying the connections between researchers, publications, research grants and research datasets (data in research). These connections can provide answers to questions such as
- What are the publications by researchers in my research group?
- What are the research collaboration networks inside and outside of my research institution?
- Who is working on a given research problem? For example, who works on stem cell research at my university?

As part of the early adopters' projects for implementing the RD-Switchboard we have identified challenges in the area of technical implementation of the graph database and legal interoperability of research metadata. In this BOF the speakers will talk about these challenges and also talk about the following opportunities
- Exploring the potential research reporting using the connected research information
- Application of this technology in the gap analysis of research metadata such as identifying missing connections

Finally, in this workshop we aim to talk about the cross-platform interoperability between research information systems and how they can be linked to ORCID, DataCite, Dryad, CERN and other international registries / repositories.

ABOUT THE SPEAKERS

Nathaniel Lewis is the Director of Research Reporting in the DVC Research portfolio at the University of Sydney. He is interested in using technology to uncover relationships in data and what research enterprises need to consider to prepare their systems and associated data management processes to support this discovery process.

Jingbo Wang is the Manager of the Data Collections Team at National Computational Infrastructure (NCI). NCI manages 36 data collections (10+ PB) categorised as earth system sciences, climate and weather model data assets and products, earth and marine observations and products, geosciences, terrestrial ecosystem, water management and hydrology, astronomy, social science and biosciences.

Marta Poblet is VC's senior researcher at RMIT University and an Associate Professor at the RMIT Graduate School of Business and Law. She is working on socio-legal and ethical aspects related to open research data and collaborates in research exploring legal interoperability issues of open data.

Dr Amir Aryani is the co-chair of the Data Description Registry Interoperability WG in Research Data Alliance and the project lead for the Research Data Switchboard. He is working in the capacity of a project manager for Australian National University (ANDS), and part of this role is to manage ANDS interoperability projects with international partners.