

Creating an open linked data model for Research Graph using VIVO Ontology

Michael Conlon, Duraspace, mconlon@duraspace.org, orcid.org/0000-0002-1304-8447

Amir Aryani, Australian National University, amir.aryani@anu.edu.au, orcid.org/0000-0002-4259-9774

Session Type

- Presentation

Abstract

Research Graph is an open collaborative project that builds the capability for connecting researchers, publications, research grants and research datasets (data in research). VIVO is an open source, semantic web platform and a set of ontologies for representing scholarship. To provide interoperability between Research Graph data and VIVO systems we modelled the Research Graph metamodel using the VIVO Integrated Semantic Framework. To evaluate the mapping, we used the model to connect figshare RDF records to data collections in Research Data Australia using Research Graph API. In addition, we are working toward loading Research Graph data into a VIVO instance. VIVO provides a search capability, and pages for first class entities in the Research Graph model -- researcher, dataset, grant, and publication. The result provides a visualisation solution for co-authors, co-funding, timeline, and a capability map for finding expertise related to concepts of interest. The resulting linked open data will be made freely available and can be used in other tools for additional discovery.

Conference Themes

- Supporting Open Scholarship, Open Data, and Open Science
- Repositories of high volume and/or complex data and collections
- Managing Research Data, Software, and Workflows
- Integrating with the Wider Web and External Systems

Keywords

Open research outputs, linked open data, research discovery

Audience

Librarians, research administrators, technologists, decision makers, project leaders, evaluators, government agency representatives, and strategists should find the presentation informative regarding existing data resources and goals of Research Graph, the goals and capabilities of

the VIVO platform, and through collaboration, the resulting capability to create and use large quantities of connected metadata regarding scholarly outputs for research discovery.

Background

The presentation will address central issues in open scholarship -- collection, representation, and distribution of open metadata records and how to connect this information. The data is large-scale, and the work involves integrating data from many sources. The purpose of the work is to make scholarly works connected and discoverable. Toward building such connections we are creating interoperability between VIVO systems and the large scale Research Graph database created as part of a project in Research Data Alliance.

VIVO [1] originated at Cornell University and was developed as an open source enterprise system with Funding by the US National Institutes of Health 2009-2012. VIVO uses a collection of ontologies, including the VIVO Integrated Semantic Framework (VIVO-ISF [2]) to represent scholarship. VIVO is in use at more than 150 sites in 26 countries. VIVO data regarding scholarship can be queried using the W3C standard SPARQL query language, and has been pooled into a global search capability, CTSAssearch, hosted by the University of Iowa.

Research Graph¹ initially was created as part of the Research Data Switchboard project. This project addresses the problem of cross platform discovery by connecting datasets together across multiple registries on the basis of co-authorship or other collaboration models such as joint funding and grants. This work was derived from outcome of the DDRI² Working group of the Research Data Alliance [3]. The group had participants from Australian National Data Service (ANDS), Dryad (US), CERN InspireHEP (Switzerland), figshare (UK), da|ra and GESIS (Germany), Data Curation Unit (Greece), OpenAIRE (European Infrastructure), ORCID, and DataCite. The participants in the group have provided substantial metadata records including publications, datasets, researcher information and grant records that are currently available in a form of graph database hosted on the AWS cloud. At the time of writing this paper, the hosted graph database on AWS holds more than ten million nodes.

The authors have extensive experience with VIVO and Research Graph respectively. Creating and loading linked open data from Research Graph is a straightforward process -- this is part of the message of the presentation -- the scholarly ecosystem is changing, data is becoming widely available, tools for managing such large scale data are available, the costs of creating, maintaining and providing such data are falling. The consequences will be shown -- open national scale data available as linked open data for discovery.

In this talk, we demonstrate the new extension of Research Graph that maps the graph model to the VIVO ontology; in addition, we demonstrate how this mapping enables the interoperability

¹ <http://researchgraph.org>

² <https://rd-alliance.org/groups/data-description-registry-interoperability.html>

between figshare RDF records and scholarly records in Research Graph database collected from non-RDF repositories.

Content

The presentation will include a high level introduction to the goals and work of Research Graph, and the goals and work of the VIVO project. We will discuss why it is of interest to bring these two efforts together to create linked open data for Research Graph and what has been gained by making the data available through a VIVO instance. The presentation will be accessible to all attendees of Open Repositories. We will not cover technical issues in depth. We will present technical issues when they are of interest in achieving the overall goals of the collaboration.

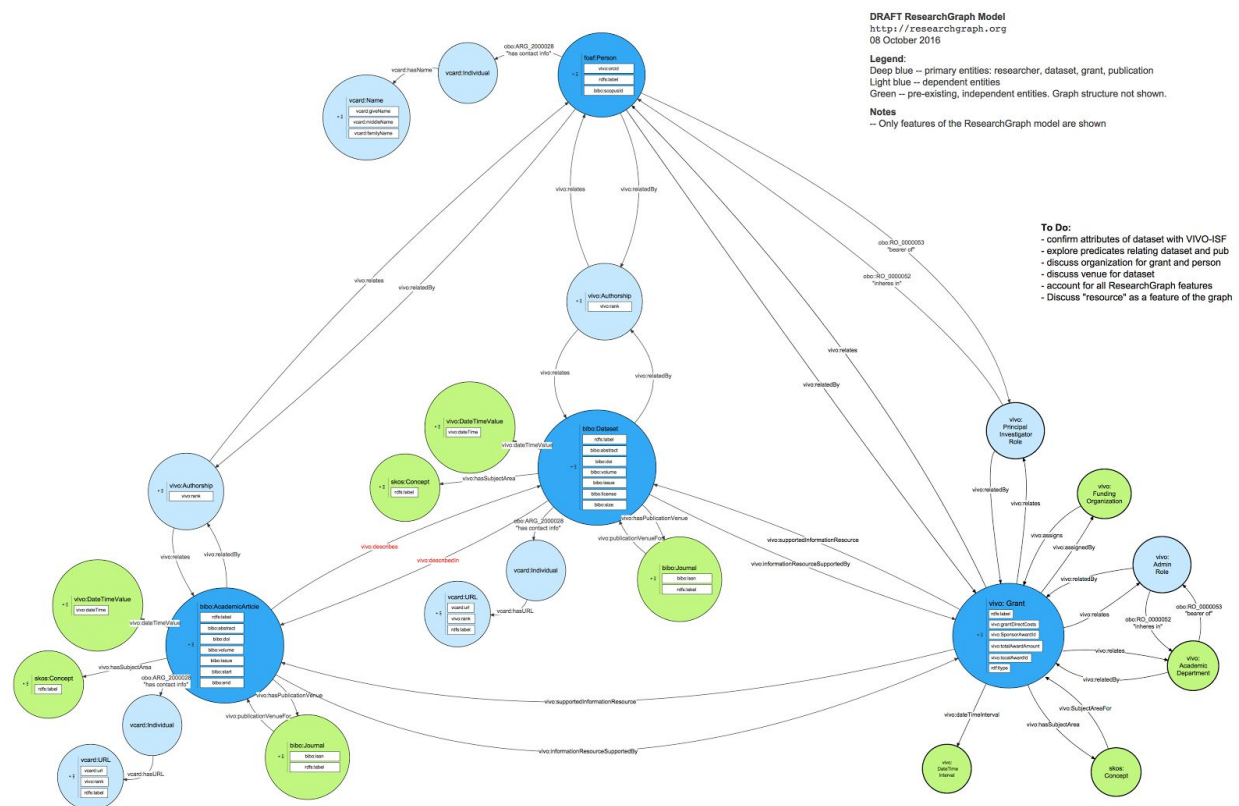


Figure 1. Preliminary high-level model of entities in Research Graph using VIVO-ISF. Additional modeling will including all Research Graph data and object properties for us in loading Research Graph data to a VIVO instance.

Note regarding this work. This work is in progress. To date, we have completed a preliminary model of the Research Graph data, demonstrating the feasibility of full modeling, and of hosting the data in a VIVO.

The preliminary high-level model is shown below. VIVO-ISF supports deep modeling of person (scholar, investigator), work (publication, dataset, many other types), grant (project, contract, other forms of funding). The figure shows the clusters of attributes related to each of the

primary entities, as well as the connections between entities, modeled as object properties in VIVO-ISF. Additional data properties will encompass the complete Research Graph data model.

We anticipate that very little will need to be added to the ontologies to represent the relationships and attributes of interest in Research Graph. This helps validate the ontologies by applying them to real-world use cases.

Next steps in the work include generating a complete data model for Research Graph, and using that data model to create RDF and load it into a VIVO instance. Creation of RDF and loading to a VIVO instance are straightforward processes using existing tools.

We have tested this model on figshare RDF records. A current mapping for figshare is available in GitHub [4]. Combining this mapping with the Research Graph API has enabled us to find connections between figshare records, and data collections in Research Data Australia and other Australian repositories.

Conclusion

Open metadata regarding research outputs and scholarly activities can be assembled at the national level and made available as linked open data using existing tools and data models. The resulting large-scaled, open, linked data can be used for research discovery. The Research Graph data currently includes a larger number of scholarly communication records from Australian repositories and international collaborators. In this work, we have created an interoperability between VIVO systems and Research Graph, and as part of the presentation we will demonstrate this interoperability using figshare records published in VIVO RDF.

References

1. Börner, K., Conlon, M., Corson-Rikert, J., Ding, Y. (eds.) *VIVO: A Semantic Approach to Scholarly Networking and Discovery*, Morgan-Claypool, 2012. p. 1-175.
2. OpenRIF: VIVO integrated semantic framework data standard github repository. Web site (accessed on Nov 20, 2016), <https://github.com/openrif/vivo-isf-ontology>
3. Amir Aryani, Data Description Registry Interoperability WG: Interlinking Method and Specification of Cross-Platform Discovery, Research Data Alliance, doi:10.15497/RDA00003
4. Crosswalk figshare:VIVO-RDF to Research Graph XML. Website (accessed on Nov 28, 2016), <https://github.com/researchgraph/crosswalks/tree/master/figshare.com>