Executive Summary

Research Graph VIVO Cloud Pilot

Many VIVO implementers find collecting, mapping, and loading data into VIVO to be quite difficult. For example, data on publications, grants, and datasets produced by an institution's faculty can be difficult to find and disambiguate. Understanding the ontologies used to describe data in VIVO and mapping faculty data to those ontologies involves a steep learning curve. Also, transforming the data to a linked data format, such as VIVO RDF, has proven difficult for most implementers due to gaps in skills and knowledge. These barriers have prevented organizations from joining the VIVO community and adopting the technology that enables access, discovery, and analysis of scholarship data.

Research Graph is an integrated network of information about researchers, their publications, grants, and datasets, across global research infrastructures such as ORCID, DataCite, CERN, CrossRef, and funders such as National Institutes of Health (NIH). For example, when provided "seed data," such as a simple list of researchers, Research Graph will identify publications, grants, and/or datasets related to those researchers and represent the information in a graph. These are referred to as "first order" connections. Research Graph is also capable of identifying and linking collaborators of the people in the "first order" data and linking their publications, grants and datasets. These collaborator links are referred to as "second order" connections.

Figure one shows the outcome of the augmentation process for the data from the National Computational Infrastructure in Australia.
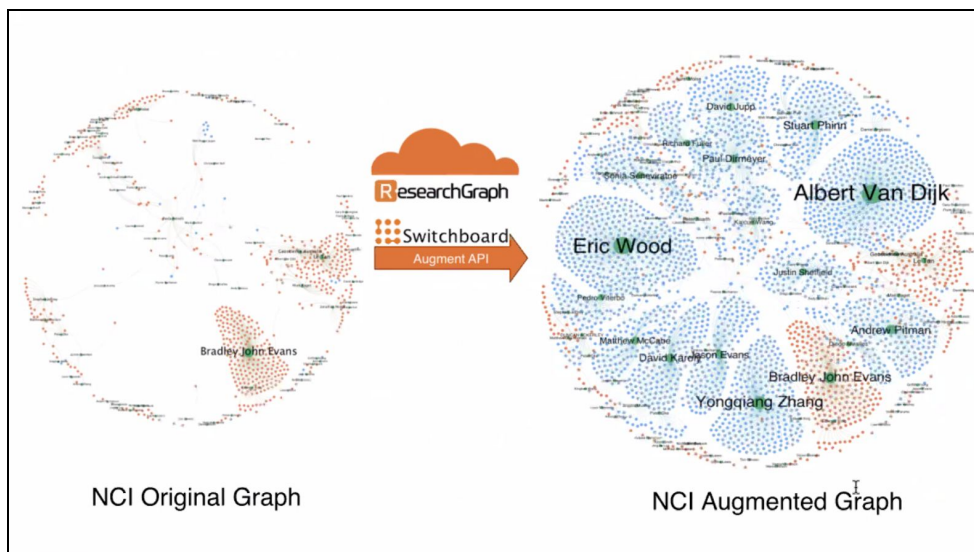

Figure One: Visualization of seed data before and after augmentation via Research Graph.

A recent collaboration between VIVO and Research Graph[1] developed a repeatable process for using seed data to build first and second order graphs, and to export, transform, and load those graphs in VIVO RDF format to a hosted VIVO instance. We believe 1) Repositories and Research Institutes, 2) Semantic Web Sites of government and research organisations, and 3) Current VIVO Sites that wish to enrich and augment their data can benefit from the collaboration between VIVO and Research Graph. The Cloud Pilot will have participants representing these three types of organizations.

Pilot participants will be asked to pay a participation fee of between $1,000 and $4,700 USD depending on the data retrieval and preprocessing assistance needed (provided at cost by Research Graph).

The Cloud Pilot will be executed in two phases over a three month term, running from February 2018 to April 2018. This will provide suitable results for a presentation at the 11th RDA Plenary Meeting in Berlin, Germany, March 21-23, 2018

Conclusion

The project will determine the value and potential of a long term collaboration between VIVO and Research Graph in the form of new services that could reduce barriers for organizations that want to find, disambiguate, transform, and map research data.

Please contact Erin Tripp, Business Development Manager at DuraSpace with questions about this project at etripp@duraspace.org.

---

[1] Conlon, Michael, and Amir Aryani. "Creating an Open Linked Data Model for Research Graph Using VIVO Ontology," July 24, 2017. https://doi.org/10.4225/03/58ca600d726bd.