Chances and challenges of creating a research information platform for the Berlin University Alliance

Fadwa Alshawaf, Claudia Adam, Rolf Guescini,
Florian Kotschka, Malte Dreyer
Humboldt University of Berlin, Germany
fadwa.alshawaf@hu-berlin.de

Abstract

This paper describes some of the challenges we face while developing a research information platform for the Berlin University Alliance (BUA). Since its founding in 2018, the BUA has put its focus on promoting the networking of researchers and their research activities and connecting many research groups within the clusters of excellence as well as beyond. The platform aims at presenting structured, transparent, categorized, and linked information about researchers and their research activities to improve the discoverability of expertise, connect researchers to their work across disciplines and boundaries, and facilitate new research collaborations. The platform is established using the open-source, web-based VIVO software, which uses semantic web techniques to connect research outputs, organizations, people, things, and research activities.

Keywords: *VIVO, research information, semantic web, Berlin University Alliance*

1 Introduction

Research information is an institutional asset that should be utilized to improve its visibility, research quality and range, and attract research cooperation. Research information systems gather distributed research information from administration and research units to enable an organized view of the research infrastructure, services, and outputs [1]. They can be implemented as specialized databases or as web platforms that combine databases with expert profiles such as the VIVO-based system [2].

According to the report published by the European commission in 2021, the research process is undergoing digital transformation, hence a reformed research assessment system is necessary [3]. This includes valuing cooperation, the acquisition of project money, and open access.

One of our endeavors to facilitate the discovery and cooperation between individual researchers and across disciplines, is implementing an online portal for research information for the Berlin University Alliance (BUA). The BUA was founded 2018 of four Berlin universities; the Humboldt-Universität zu Berlin, the Freie Universität Berlin, the Technische Universität Berlin, and the Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and HumboldtUniversität zu Berlin. The vivo-based research information platform aims at facilitating the digital connectivity and discovery of research and researchers of the BUA. An added value of our project lies in the search and navigation functions, which through the modeling of data and building of ontologies, can suggest and present relevant information that is otherwise not easily found. If we connect and relate BUA institutes, researchers and research activities, particularly the projects, using ontologies, this allows easier and faster discovery of organizations, people and research topics. The online portal provides a single point

of access for information on scholarly activities across disciplines and beyond the boundaries of each organization. The VIVO software is a member-supported, open source, semantic, web-based tool.

All information within VIVO is represented in the RDF data format and standard vocabulary via web ontologies with tools to edit existing ontologies and write new ones. The system links lists of various types of entities: People, activities, courses, events, organizations, grants, research and locations. It also has tools to record, edit, report, search, browse, visualize, analyze, adjust, and connect. The data can be ingested to the system manually and automatically using locally-stored files, organization databases (HR), and online information systems (publication aggregation). VIVO's search returns faceted results for rapid retrieval of the desired, more specific information across disciplines. The research information should be documented in standard format such as the Common European Research Information Format (CERIF) or research core dataset format [4].

2 Chances and Challenges

Information about researchers, research projects and publications (so-called “research information”) at the Berlin University Alliance (BUA) can be found online in various places. On the websites of individual institutes, for example, there are lists of contact information of employees, information about research projects and publications. However, the online distributed information is rarely linked together. For example, research information on websites of projects, funding organizations, ORCID profiles, or social networks may not be linked to the contributing researchers. The open-source software VIVO uses Semantic Web technologies to link information, which makes the information easier to find as well as the colocation of related information.

Using the VIVO software, data from existing systems can be automatically harvested. In this way, research information from different sources can be aggregated in a meta-platform. In addition, the software enables a search function so that the platform acts as a search engine to search for information, aggregated already from other systems ("One Stop Shop"). Since the project aims to develop a research information platform within the BUA, the described functions are even more effective; not only can information from different sources be aggregated and searched for one organization, but across the alliance.

The open-source software VIVO provides diverse analyses and visualization tools that go beyond the capabilities of existing research information systems. Functions such as the visualization of the co-author network or the Map of Science can facilitate and enrich reporting as well as research management.

The browse and faceted search functions of the platform facilitates to search for and find relevant information and people - including researchers - and thus promote networking and exchange within one discipline and across different disciplines.

While the platform has substantial benefits for the alliance, it also has to overcome some challenges. The quality of the mapped data is an important success factor for the platform; therefore, the first step for the project team is to obtain an overview of the data available at the individual organizations and which data can be used for the platform. It is important to decide how the data should be accessed, cleaned, harmonized, mapped, and kept up-to-date.

One of the benefits of research information records is the ability to analyze the data to provide indicators and metrics that support making informed decisions at leadership levels. This however requires complete records or at least an adequate coverage such that the analytics are representative and trustworthy. We here are faced with the barrier of personal data protection. Preparing a data protection concept for the platform is one fundamental challenge, because so far, there are diverse concepts within the BUA for the use of researchers’ data. In order to ensure and document compliance with the European General Data Protection Regulation (GDPR), an information security and data protection concept for

the platform must therefore be developed in the first phase of the project. This is a hindrance we face in developing this platform because even if the information are openly available on organizations websites or online aggregation systems, we need to ask for a consent to import and display the information on the platform.

Another issue that should be mentioned within this context is the creation of representative taxonomies and ontologies that match the German academic system, especially within the context of the BUA. At the same time, ontologies for disciplines and interdisciplinary fields should be built. VIVO uses a collection of ontologies to represent and connect researchers to their research activities. In computer science, an ontology refers to how to name and relate types, properties and relationships that exist in a particular domain such that they are interpretable in this specific context not only by human beings but also by machines. The VIVO ontology represents researchers by connecting them to their activities and accomplishments, resources, and institutions. This is done using a set of types and relationships (properties) that represent researchers and the entire context in which they work. These ontologies model organizations, concepts (research areas), persons, publications, grants, and others. The ontology requires precise identification of people, organizations, and works, and the correct relationships between these three.

3 Conclusions

In this paper, we gave a brief description of the VIVO-based research information platform that is created to showcase expertise and research activities within the Berlin University Alliance. The platform uses semantic web technologies to best connect researchers and their work to make them easily discoverable within the system and in search engines. This should encourage and facilitate cooperation, main transparency, and build bridges across disciplines.

4 Acknowledgement

Thanks to the Federal Ministry of Education and Research (BMBF) and the state of Berlin under the Excellence Strategy of the Federal Government and the Länder for funding this project.

5 References

[1] Ebert, B., Tobias, R., Beucke, D., Bliemeister, A., Friedrichsen, E., Heller, L., Herwig, S., Jahn, N., Kreysing, M., Müller, D., & Riechert, M. (2016). *Forschungsinformationssysteme in Hochschulen und Forschungseinrichtungen. Positionspapier*. Version 1.1. Zenodo. https://doi.org/10.5281/zenodo.45564

[2] Mitchell, S., Chen, S., Ahmed, M., Lowe, B., Markes, P., Rejack, N., Corson-Rikert, J., He, B. and Ding, Y. (2011). *The VIVO ontology: enabling networking of scientists.*

[3] European Commission, Directorate-General for Research and Innovation, *Towards a reform of the research assessment system: scoping report*. (2021) Publications Office. https://data.europa.eu/doi/10.2777/707440

[4] Biesenbender, S. and Hornbostel, S., 2016. *The research core dataset for the German science system: Developing standards for an integrated management of research information*. Scientometrics, *108*(1), pp.401-412.