

# VIZ–VIVO: Towards Visualizations-driven Linked Data Navigation

Muhammad Javed<sup>1</sup>, Sandy Payette<sup>2</sup>, Jim Blake<sup>3</sup>, Tim Worrall<sup>4</sup>

Cornell University Library (CUL)  
Cornell University  
Ithaca, NY 14850, United States  
{mj495<sup>1</sup>, sdp6<sup>2</sup>, jeb228<sup>3</sup>, tlw72<sup>4</sup>}@cornell.edu

**Abstract.** Scholars@Cornell is a new project of Cornell University Library (CUL) that provides linked data and novel visualizations of the scholarly record. Our goal is to enable easy discovery of explicit and latent patterns that can reveal high-impact research areas, the dynamics of scholarly collaboration, and expertise of faculty and researchers. We describe VIZ-VIVO, an extension for the VIVO framework that enables end-user exploration of a scholarly knowledge-base through a configurable set of data-driven visualizations. Unlike systems that provide web pages of researcher profiles using lists and directory-style metaphors, our work explores the power of visual metaphors for navigating a rich semantic network of scholarly data modeled with the VIVO-ISF ontology. We produce dynamic web pages using D3 visualizations and bridge the user experience layer with the underlying semantic triplestore layer. Our selection of visual metaphors enables end users to start with the big picture of scholarship and navigate to individuals faculty and researchers within a macro visual context. The D3-enabled interactive environment can guide the user through a sea of scholarly data depending on the questions the user wishes to answer. In this paper, we discuss our process for selection, design, and development of an initial set of visualizations as well as our approach to the underlying technical architecture. By engaging an initial set of pilot partners we are evaluating the use of these data-driven visualizations by multiple stakeholders, including faculty, students, librarians, administrators, and the public.

**Keywords:** Data Driven Documents (D3), Visualizations, Linked data, RDF, VIVO, User Interface, Web of Data (WoD)

## 1 Introduction

As stewards of the scholarly record, Cornell University Library (CUL) has consistently leveraged new and emerging technologies to improve access and discovery to scholarly resources and to preserve and archive them for future generations. A new project of CUL is *Scholars@Cornell*<sup>1</sup>, a data and visualization service built upon a semantic, linked data knowledge-base that represents the record of scholarship produced by Cornell faculty and researchers. One key question we set out

<sup>1</sup> <http://scholars.cornell.edu/>

to answer was how can visual mediation help users navigate the rich semantic data that represent the scholarly record? We have developed *VIZ-VIVO*, as an extension to the VIVO<sup>2</sup> framework to enable exploration of the scholarly record through a configurable set of data-driven visualizations that leverage relationships between people (e.g., faculty/researchers), their affiliations (e.g., academic departments or colleges), and their research outputs (e.g., journal articles, books, datasets). Our goal is to enable easy discovery of both explicit and latent patterns that reveal high-impact research areas, patterns of scholarly collaboration, and expertise of faculty and researchers.

In this paper, we discuss how we bridged the chasm between our backend of semantically rich data in VIVO triplestore, with a fresh frontend user experience that takes advantage of new developments in the dynamic world of web visualizations. We present our ongoing work on integration of D3 visualizations into the VIVO frontend pages. The visualizations-driven approach provides efficient overviews of the huge network of interconnected resources. D3 visualizations are intuitive for the users to interact and offer the ability to visualize, filter and navigate. The word cloud (presenting the *keywords* from the scholarly works) is used not only to visualize the domain expertise of a researcher, but also to provide access to the actual VIVO page of such scholarly work. The concept network map and sunburst visualizations are used to present the person-to-subject-areas map and inter-unit/cross-unit collaborations, respectively. Additionally, both traditional metrics (i.e. citation count, journal ranking.) and alternate metrics [8, 9] are used to illustrate the impact of a particular scholarly work. By engaging our initial of pilot partners at Cornell University, we are currently evaluating the use of these data-driven visualizations by multiple stakeholders, including faculty, students, librarians, administrators, and the public.

This paper is structured as follows: In Section 2 we discuss the user viewpoints that we intended to accommodate with the VIZ-VIVO visualizations. In Section 3, we discuss how we conceived of particular visualizations that could be driven by linked data navigation. We highlight specific examples showing sub-graphs from the underlying VIVO data can result in particular visualizations in the user interface layer. In Section 4, we discuss our methodology for integrating visualizations into the VIVO framework. Our approach involved developing a new component and API to be used with the core VIVO software and working with users to evaluate usability and functional suitability of visualizations. A short evaluation is given in Section 5. Related work is presented in Section 6 and we end with some discussion.

## 2 VIZ-VIVO: Motivation and Viewpoints

In our work at Cornell University Library, the primary entity of interest is *scholarship*, of which people and organizations are, by definition, both the creators and consumers. From this perspective, attention is focused on aggregate views

---

<sup>2</sup> <http://www.vivoweb.org>

of scholarship and our visualizations become the entry points into a rich graph of knowledge that can be explored interactively. Using VIZ-VIVO capabilities, a user can begin with a Cornell-wide perspective of subject areas, move to views of particular research topics, and discover faculty members in context, along with traversing lateral pathways to collaborators across the university and beyond. While other faculty profiling systems exist, including other VIVO-based applications and as well as services such as ResearchGate<sup>3</sup>, these tend to use limited list-oriented metaphors that draw the user's attention to a primary entity – an individual, which is presented as a profile web page. Our selection of visual metaphors is motivated by enabling end users to start with the big picture, and encounter individuals (e.g., faculty members, collaborators) in context, within an interactive environment that can visually guide the user through the sea of scholarly data depending on the questions the user wishes to answer. While users can still view profiles or download data about individual entities, they are found within the process of exploration when using active graph visualizations. This is a paradigm shift in faculty profiling systems that is consistent with the mission of the library in supporting discovery of knowledge, by starting with macro views of scholarship upfront.

Among the user stories that inspired our Scholars@Cornell pilots, are how the scholarly data can answer questions from different user viewpoints. For example, how academic deans and department chairs identify where collaborations are happening and the impact of their unit's scholarly output? How can librarians use novel views of publication patterns to inform decisions about library holdings and preservation strategies. How can faculty and researchers have new ways to expose their work or locate other researchers with whom they might collaborate. Finally, how can university executives explore macro-level overviews of the university and its research to highlight high impact research areas and assist them in decision making related to invest in emerging research areas. Based on the above given viewpoints, we opt for a *bottom up approach* where we started from a set of questions. Following are some specific examples of such questions that different users would like Scholars@Cornell to answer.

- List of contributions/research output of a unit.
- List of contributions/research output of a researcher.
- List of potential collaborators (people/organizations).
- List of inter-unit/cross-unit collaborations.
- List of domain experts (for a specific subject area).
- Impact of a scholarly work.

### 3 Visualizations-driven Linked Data Navigation

Once user stories and the set of questions were outlined, next step was to realize what data we need to answer such questions and how such data can be presented. In regards to the presentation, we opt for Data Driven Documents

---

<sup>3</sup> <https://www.researchgate.net>

(D3) visualization [7]. They can bridge between the best of the two worlds (i.e. frontend and backend).

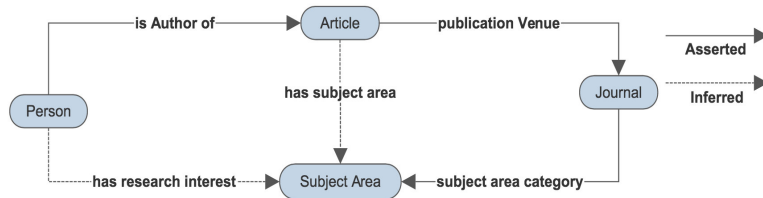
At the backend, we record data in a RDF triplestore that is modelled based on a standard, community-driven ontology – VIVO-ISF<sup>4</sup>. Having data in RDF not only allowed us to create links from local entities to the external resources but also to use the entity URIs as the local authorities, specifically for the persons, publications and the organizations. Creation of such URI-driven authority records is in alignment with the CUL’s current goals for digital preservation and open access. Further, use of ontology-driven data modelling approach allowed us to use inference mechanisms to realize the knowledge that is not explicitly given in the triplestore.

A picture is worth a thousand words. At the frontend, we use D3 visualizations. Visualizations permit a user to navigate through the large linked data network in an intuitive way. These are semantic-data driven visualizations that allowed us to move from list-driven data views to visualization-driven data views and navigation. Filters are used to narrow down the lists that consist of hundreds of entries. In addition to all other benefits, visualizations also give a fresh look and feel to the user interface.

In the following section, we discuss few of the visualizations and the questions they address.

### 3.1 Person-to-Subject-Area Map (*Department level*)

At the department level, the *concept map network* (see Fig. 2) presents a mapping between a person and his/her research interests. Different citation indexing applications (e.g. Web of Science, PubMed, arXiv) classify the publication venues (e.g. journals, conferences, workshops) in different subject categories. Therefore, all the scholarly works, published in a selected venue, receive the same category as applied to the venue. For example, if an article is published in a journal that is classified under the category of *polymer science*, the article is considered to be related to the subject *polymer science*. We made use of the same transitivity approach in order to represent the research interests of an author (see Fig. 1).



**Fig. 1.** Person-to-subject-area map inference

<sup>4</sup> <https://wiki.duraspace.org/display/VIVO/VIVO-ISF+Ontology>

Let, journal  $j$  be the publication venue for an article  $t$  and the person  $p$  is one of the author of the article  $t$ . If journal  $j$  is categorized under the subject area category  $c$ , we can infer that the article  $t$  can be classified under the subject area  $c$  and the person  $p$  has the research interest in the subject area  $c$ .

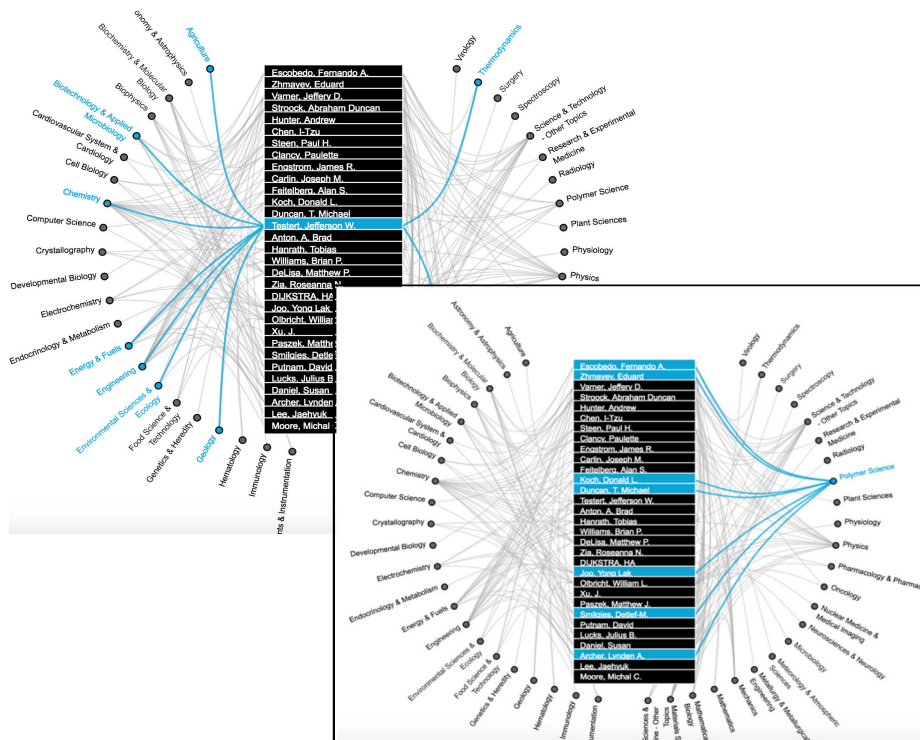


Fig. 2. Concept map network - presenting Person-to-Subject-area (P2S) map

The person-to-subject-area map is helpful for the identification of i) list of research interests of a person and ii) list of potential collaborators. The map not only demonstrates the overlap of research interests among different persons but also allows a user to navigate through the underlying linked data. A profile page link is given on every personal subject area map view. The link navigates the user to the persons’s profile page. On the other hand, subject area views direct a user to the subject area’s profile page, where one can learn what other entities (e.g. book chapters, dissertations, organizations, collections) are associated to the same subject area.

### 3.2 Domain Expertise of a Scholar (Person level)

The subject area terms given above in the P2S map are fairly generalized. These terms are useful to categorize the persons (and their research publications) and



$org_2$ ), article  $a$  will be considered as a collaborative work between organizations  $org_1$  and  $org_2$ . Later, if faculty member  $p_1$  changes his/her affiliation from  $org_1$  to  $org_3$  and faculty member  $p_2$  changes his/her affiliation from  $org_2$  to  $org_4$  (where  $org_3$  and  $org_4$  can be local or external organizations), article  $a$  will still be considered as a collaborative work between  $org_1$  and  $org_2$ .

The inter departmental and cross-unit collaborations were identified based on the *affiliations* extracted from the citation data of a scholarly work. For example, if citation data of an article describes multiple distinct affiliations for the co-authors, then it is a collaborative work. The reconciliation process for the *affiliation data strings* and the *person name strings* (coming from different citation data sources) to VIVO's person and organization URIs, is not discussed here and is out of the scope of this paper.

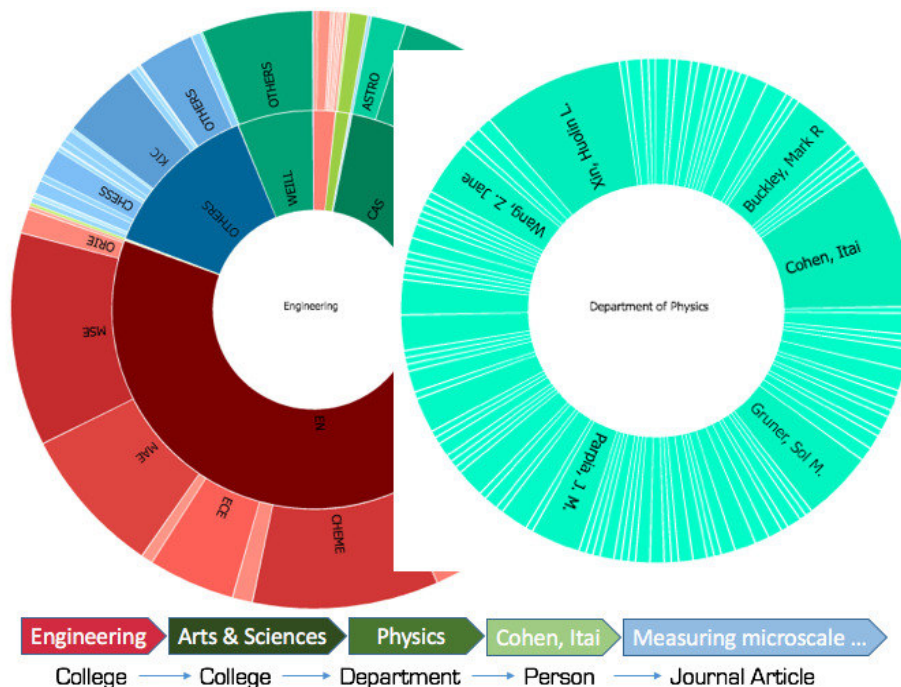


Fig. 4. Sunburst visualization - presenting inter-unit/cross-unit collaborations

*Sunburst* visualization (see Fig. 4) is used with the tooltips in order to present the inter-unit/cross-unit collaborations. Figure 4 presents an inter-unit/cross-unit collaboration view for the *College of Engineering (EN)*<sup>5</sup>. The collaboration view consists of three layers i.e., college-level layer (inner most), departmental layer (middle) and the person-level layer (outer most). To view the person-level layer, a user is first required to select the college of interest in the inner most

<sup>5</sup> <https://www.engineering.cornell.edu>

layer. To browse the collaborations within a unit, in this example EN-to-EN, one can select the *Engineering (EN)* in the college-level layer. If a user is interested in cross-unit collaborations, s/he can select a college other than *Engineering*. A user can start navigation from the college level layer and while traversing through the department (of his/her interest), can reach to a specific person who has co-authored some of the collaborative work. Clicking on a person's name section displays a tooltip that consists of the list of the (collaborative) work titles. User can further click on one of the work titles from the list in order to redirect to the specific profile page of such work.

## 4 Integration of D3 Visualizations in VIVO

Visualizations can be created in the browser using powerful JavaScript libraries like D3 or InfoViz. Such libraries create visualizations that are animated, responsive, and captivating. Server is only required to provide the data – a difficult task in itself. We envision a range of data sets that would be required for these visualizations. This will certainly include queries against the VIVO's triplestore and search index. It will also include data sets that require extensive processing and will likely be created by batch processes and cached for serving on demand. Existing APIs in VIVO framework are not flexible enough to handle these requirements. Further, the APIs do not provide the granularity of authorization that we require. If the JavaScript on the browser can obtain data, then the same data may be obtained by any client using HTTP, and used for any purpose.

### 4.1 Approach and the API Design

A data distribution API is created and is already been submitted to the VIVO community for adoption into the core. The API permits a site administrator to create a configuration file that enables requests for data from different sources. The JavaScript code on the browser requests a dataset by name, supplying qualifying parameters as appropriate. As each request type must be configured by a site administrator, a user cannot make unexpected or unauthorized requests. Requests may be restricted to logged-in users, or even to users who are logged in with a specific authorization level. Users without sufficient authorization will receive no data. It is important that the actual source of the data is hidden from the JavaScript request. For example, if a request takes too long to fulfill, the site admins can change the configuration, satisfying the request from a file of cached data, instead of a query in real-time.

On startup, the server loads a configuration file that associates data requests with the means of satisfying them. The configuration file includes the name of a Java class that will supply the data, along with any parameters that the class requires. The Java class implements a `Data Distributor` interface. This means that the API controller can ask what action the class satisfies and what MIME type should be assigned to the resulting data (see Fig. 5). The API controller



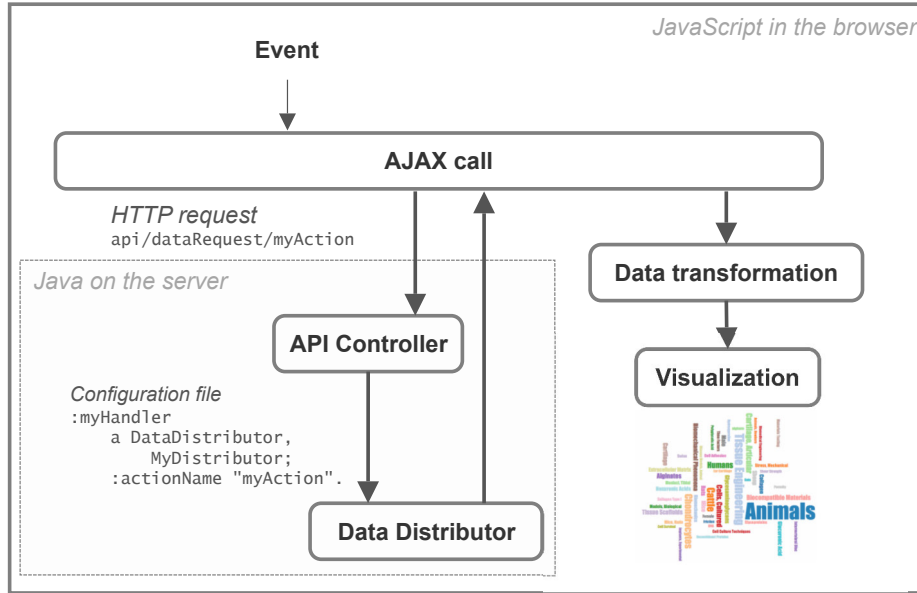


Fig. 5. The API design

asks the distributor class to produce the data and the controller routes the data to the user.

One of the powerful concepts in this design is that a Java class can be written as special-purpose code for a single use, or as general-purpose code which can be configured for each use case. We expect to see more general-purpose `DataDistributor` classes, such as:

- `SolrDataDistributor`, for queries against the search index.
- `HttpRelayDataDistributor`, for queries against services at other sites.

#### 4.2 Data Transformation

We divide the workload into two tasks i.e., i) designing and implementing the visualization and ii) providing the data. A transformation step was added to the data flow (see Fig. 5), and a JavaScript function named `loadVisualization()` was written to control the operation. The JavaScript code that creates the *key-word cloud* visualization is given below as an example.

```
<script>
$.ready(function() {
  loadVisualization({
    url : api/dataRequest/person_word_cloud,
    transform : transform_word_cloud_data,
    display : draw_word_cloud,
  });
});
</script>
```

In this example, the visualization is created when the page is loaded into the browser. The visualization loader requests the data with the action name of `person_word_cloud`. When the data is received, the visualization loader calls the JavaScript function `transform_word_cloud_data`, that transforms a SPARQL ResultSet to a JSON structure. Finally, the visualization loader calls the function `draw_word_cloud`, that populates the visualization with the transformed data.

## 5 Evaluation

To evaluate the proposed visualizations and the linked data navigation methodology, *journal articles*, as a category of scholarly work, was selected. Citations data for more than 11K journal articles, authored by the faculty members and researchers of college of engineering, were uploaded in the VIVO triplestore. A user-based evaluation approach was selected in order to empirically evaluate the usability and the functional suitability of the integrated visualizations.

### 5.1 Empirical Evaluation

We involved college administrators and librarian to evaluate the visualizations. College administrators highly ranked the zoomable collaboration view. Using different colors for representing distinct colleges, helps a user to understand the breakdown of collaborations between a unit and the collaborating units. The zoom in and out functionality helps to traverse, see visualization from different angles and to answer different questions. Such questions include “where cross-unit collaborations are happening?”, “which departments of a college are participating in these collaborations?”, “who are the collaborating faculty members and how often do they collaborate?”. Further, the visualization also presents the *outcomes* of these collaborations. The outcomes, such as co-authored articles, are presented under the tooltips where a user can click on any of the publication title in order to redirect to a specific page of such scholarly output. This is identical to the traversing of the linked data graph – from one node to the other. Furthermore, having impact view of a scholarly work is very helpful to evaluate the contributions and scholarly outcomes of a unit or person.

Librarians were very much attracted to the person-to-subject-area map and the keyword cloud view. Person-to-subject-area map is helpful for understanding the interests of a faculty member and in learning how these interests overlap with the others. Keyword cloud view is very useful especially for those (e.g. post graduate students) who are not only interested in learning about the research interests of a faculty member, but also in their scholarly works. For example, if a user has interest in the domain of *chemotherapy* and found this (or related) term in the keyword cloud, by clicking on the keyword s/he can view the list of publications, authored by the faculty member, on this topic.

## 6 Related Work

In last few years, a number of approaches for ontology and linked data visualization have been emerged [2, 4–6, 10, 12, 13]. A similar work on the representation of research expertise of a person is presented in [2]. In contrast to our methodology of usage of subject terms mentioned in the *keyword* category of an article, author proposed a method for the keyword extraction from the text of the document. Similar to our approach, author used word cloud view to present the expert profiling. In [3], author presents the preliminary results on a framework for the representation and visualization of an ontology, extracted from the text of a document. In [4], author presents a rule-based recommendation system that was subsequently developed to help the user choose a suitable ontology visualizer. In [5], author presents a tool named as Visualization Playground (VISU). VISU is a SPARQL interface that supports querying and visualizing the university data. The goal here was to create a flexible and user friendly tool for exploring and visualizing data from linked universities. *LinkDaViz* [10] is a framework that supports the automatic binding of linked data to the visualizations. The framework is based on the analysis of the input data structure and a visualization model in order to facilitate automatic bindings. In [9], author presented a model for assessment of research impact beyond citation analysis. Eysenbach [8] found that the number of tweets about a research article, within the first three days of an article’s publication, can predict which articles will be highly cited.

## 7 Discussion and Future Work

In the Scholars@Cornell project, we have developed VIZ-VIVO as a means of bridging the best of the two worlds. The back-end world leverages semantic technologies by providing an ontology-driven data model of the fundamental entities of scholarship (people, publications, organizations). Through automated feeds and selected curation, this data is instantiated in a VIVO triplestore to form a graph that can be traversed, visualized, and subjected to inference. The front-end world uses visualizations driven by sub-graphs of the entire corpus of this scholarly data. This has enabled us to move away from user experiences dominated by lists and directories (e.g., articles and people), towards a visual interactive experience that leverages the power of relationships in the data.

In this paper, we presented our approach on using D3 visualizations in combination with a VIVO-based semantic knowledge base of the scholarly records of Cornell University. Our work enables easy discovery of subject area and domain expertise of faculty members and researchers, as well as the discovery of patterns in the scholarly records. Our approach offers a mean of presenting linked data through dynamic web pages with visualizations that permit a user to navigate a large corpus of semantic data and visually explore interdisciplinary collaborations and emerging trends in scientific publications.

In the future, we will be working with users to explore and evaluate other types of visualizations based on the more complex questions they seek to answer

in VIVO. We will also evolve our data to express common categorization schemes for different disciplines to better support visualizations of research areas. Our user interface will evolve to make all data that underlies each visualization downloadable in common formats such as csv, json, and xml. Our continuing work will focus on making linked data easily discoverable, consumable, visualizable, and downloadable to bring the power of semantic data to users of VIVO and other linked data applications.

**Acknowledgements.** Other members of the Scholars@Cornell team contributed to backend systems development, data modeling, data feeds, and data curation. We especially acknowledge the efforts of Joe McEnerney, Jill Wilson, Jason Kovari, and George Kozak.

## References

1. Krafft, D.B., Cappadona, N.A., Caruso, Corson-Rikert, J., Devare, M., Lowe, B.J., & VIVO Collaboration. VIVO: Enabling national networking of scientists. Paper presented at the Web of Science Conference, Raleigh, NC, 2010.
2. Koperwas, J., Skonieczny, L., Kozłowski, M., Andruszkiewicz, P., Rybiński, H., Struk, W.: Intelligent information processing for building university knowledge base. *Journal of Intelligent Information Systems*, Springer, pages 1–23, 2016.
3. Dasiopoulou, S., Lohmann, S., Codina, J., Wanner, L.: Representing and visualizing text as ontologies: A case from the patent domain. In: *Proceedings of the International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data (VOILA 2015)*. CEUR-WS, vol. 1456, pp. 83–90 (2015)
4. Dudáš, M., Zamazal, O., Svátek, V.: Roadmapping and navigating in the ontology visualization landscape. In *19th International Conference on Knowledge Engineering and Knowledge Management*, pages 137–152. Springer, 2014.
5. Alonen, M., Kauppinen, T., Suominen, O., Hyvonen, e.: Exploring the Linked University Data with Visualization Tools. In *proceedings of the European Semantic Web Conference (ESWC)*, 2013.
6. Atemezing G.A., Troncy, R.: Towards a linked-data based visualization wizard. In *Proceedings of the 5th International Workshop on Consuming Linked Data (COLLD 2014)* co-located with the ISWC, 2014.
7. Bostock, M., Ogievetsky, V., Heer, J.: D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*. volume 17, number 12, pages 2301 – 2309, 2011.
8. Eysenbach, G.: Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research (JMIR)*, volume 13, issue number 4, 2011.
9. Sarli, C.C., Dubinsky, E.K., Holmes, K.L.: Beyond citation analysis: a model for assessment of research impact. *Journal of the Medical Library Association (JMLA)*, volume 98, issue number 1, pages 17–23, 2010.
10. Thellmann, K., Galkin, M., Orlandi, F., Auer, S.: LinkDaViz – Automatic Binding of Linked Data to Visualizations In *proceedings of International Semantic Web Conference (ISWC)*, 2015.

11. Abello, J., van Ham, F., Krishnan, N.: ASK-GraphView: A Large Scale Graph Visualization System. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, volume 12, number 5, 2006.
12. Bikakis, N., Liagouris, J., Krommyda, M., Papastefanatos, G., Sellis, T.: Towards Scalable Visual Exploration of Very Large RDF Graphs. In *proceedings of the European Semantic Web Conference (ESWC)*, 2015.
13. Frischmuth, P., Martin, M., Tramp, S., Riechert, T., Auer, S.: Ontowiki—an authoring, publication and visualization interface for the data web. *Semantic Web Journal*, volume 6, number 3, pages 215–240, 2015.
14. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L. and Su, Z.: Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data mining*. pages 990–998, 2008.
15. Osborne, F., Motta, E. and Mulholland, P.: Exploring scholarly data with Rexplore. In *International Semantic Web Conference (ISWC)*, pages 460–477, Springer Berlin Heidelberg, 2013.
16. Monaghan, F., Bordea, G., Samp, K. and Buitelaar, P.: Exploring your research: Sprinkling some saffron on semantic web dog food. In *Semantic Web Challenge at the International Semantic Web Conference (ISWC)*, vol. 117, pages 420–435, 2010.