# Notes on Operations
## Visual Representation of Academic Communities through Viewshare

**Violeta Ilik**

*This paper discusses how the Viewshare web application was used to generate and customize unique, dynamic views of data about faculty members in a large public university, specifically their areas of research and other data such as PhD granting institutions, location of the PhD granting institutions, Virtual International Authority File (VIAF) authority records, and gender. Viewshare, as a visualization platform, enabled the author to discover the departments' strengths and consider how the results could be used to benefit the library, students, and specific departments. Viewshare also enabled the author to show patterns and trends with graphics instead of volumes of text.*

**Violeta Ilik** (violeta.ilik@northwestern.edu) is Digital Innovations Librarian, Galter Health Science Library, Feinberg School of Medicine, Northwestern University Clinical and Translational Sciences Institute, Chicago.

The library's mission is closely intertwined with the university's mission, and librarians need to respond to the challenges that the research landscape is facing. Borgman states that the role of libraries in research institutions is evolving from a focus on reader services to a focus on author services.[1] Luce suggests that libraries are becoming part of new hybrid organizations, which will emerge as a result of tackling new support paradigms in the university system. Further, Luce advises that in the emerging paradigm of collaborative partnerships, libraries should emphasize proactive outreach and engagement by taking a role as conveners among the different stakeholders.[2] In a similar argument, Lougee explains that libraries must be able to constantly adapt to the changing landscape of scholarship and technology, especially as these two aspects of research interact.[3] While the library's role has traditionally been to build collections supporting faculty research activities, it is now apparent that libraries need to adapt to the new ways of conceptualizing research, specifically shaping and disseminating that research. Libraries need to position themselves in terms of a larger strategic process, becoming proactive and innovative rather than reactive.

Within this context, this paper describes how Texas A&M University (TAMU) Libraries moved towards action and innovation by testing a free, open-source visualization platform, Viewshare (http://viewshare.org), together with linked data principles, to visualize university research strengths, research outputs, collaborative relationships, and other characteristics of the campus research environment based on publicly available data about TAMU faculty. This project started as an experiment and a learning experience on the author's own time. It looks towards the future adaptation of library systems to the changes in academia:

it highlights authors and researchers, encourages and directs collaborative partnerships, and provides a source of interaction and innovation for library outreach and engagement. Subject selectors can use this tool for collection development as it presents the research focus of the faculty members in an interactive and dynamic way.

## Literature Review

This paper discusses linked data, the challenges of adopting linked data in libraries, visualization as a way to interactively display data, and open source platforms based on linked-data principles. Few authors have discussed similar projects at other institutions. None of the existing projects mentioned below have utilized the Viewshare platform used in this project.

When Tim Berners-Lee introduced the Semantic Web, he coined the term "linked data." He claimed that the Semantic Web implies more than putting documents on the web; instead, it is about making links to enable a person or a machine to expand data and knowledge.[4] This is a departure from the original web, which has conceptually been document-centric and static. According to Allemang and Hendler, the main concept behind linked data is to support a distributed web at the level of the data rather than at the level of presentation (e.g., documents).[5] Instead of having one web page point to another, one data item can point to another, using global references called Uniform Resource Identifiers (URIs). Miller and Westfall noted that linked data is not about stopping the current development of databases or database systems, but that this technology aims to leave data where it resides to provide the opportunity to connect in new ways and integrate data to solve a particular problem.[6] They further note that linked data is beyond the current capabilities of the World Wide Web.

## Adopting Linked Data in Libraries

Alemu et al. state that the road towards adoption of linked data in libraries is not without challenges.[7] MARC format has been extensively used and understood as the basis for both current library management systems and legacy metadata. It has a document-centric metadata structure; the data cannot survive in an environment where an actionable data-centric format is needed. The author notes that while libraries are aware of MARC's limitations, alternative formats, such as XML, have not been acceptable replacements.

A second challenge is the terminological disparity that exists between library and web-based standards.[8] Alemu et al. cite Wallis, who recommends that the library and the linked data community work in concert to bridge such

differences to facilitate the reusability and extensibility of library data by outside users.[9] Wallis also argues that initiatives to develop library standards, such as Resource Description and Access (RDA) and the Functional Requirements for Bibliographic Records (FRBR), should cater to simplicity while exploiting the metadata richness that is possible through the use of linked data.

Alemu et al. note that a third and important challenge is the complexity of linked data technologies.[10] The authors state that it is imperative that linked data technologies be made relatively easy to learn and use and comparable in simplicity to creating HTML pages during the early days of the web. As things currently stand, linked-data technologies are generally too complicated for people outside the linked data community to use. For a wider adoption to occur, anyone with the basic skills of website design should be able to create a page based on linked data standards.

Visualization provides a way to explore data in an interactive way regardless of the platform or tool used. Tools and platforms that are not built on linked-data principles also have the capability to present data interactively. The difference is that the data is not linked to other sources of data that might provide further insight into the subject. Heer and Shneiderman state that although the increasing scale and availability of digital data provides an extraordinary resource of information, users must be able to make sense of it to pursue questions, uncover patterns of interest, and identify (and potentially correct) errors.[11] As the authors note, multiple linked visualizations often provide clearer insights into multidimensional data than do isolated views.

Researchers at the University of Colorado-Boulder conducted a project in 2012 that demonstrates the use of Semantic Web technologies in a library. Lindquist et al. found, in working with an online heritage collection, that semantically enriched metadata and intelligent user services expose the complex, often nonlinear relationships, among topics, people, and places that are buried within the sources. This particularly occurs when data and services draw on ontologies and other specialized vocabularies that impart meaning to these concepts and the relationships among them in any given historical domain. They further note that linked data encourages the development of intelligent applications that are easy to use because they present the user with a range of options for analyzing and visualizing the data. The authors conclude that through linking related concepts by using a specialized vocabulary and enabling semantically rich services, they hope to empower users to find and use online primary sources efficiently and effectively.[12]

Schreur points out that "linked data has the potential to change most aspects of the universe of information creation and exchange. As a primary purveyor of information, the academy will be at the nexus of this revolution."[13] He further reiterates a call to libraries for reform and adaption: "True

beginnings do not happen often and revolutions can be swift and unexpected. Libraries must be leaders in this revolution. Information creation and exchange is the raison d'être of the academy. The time has come for a pivotal change in the entire information ecosystem and libraries cannot afford to let history simply repeat itself."[14]

## Research Community Tools and Applications

This paper explores the use of a visualization tool for displaying publicly available information about academic communities. To develop an initial proof of concept for a linked-data project, the author chose Viewshare (http://viewshare.org), an open source platform based on linked data principles that enables users to upload their data in various formats, share it with the community, and reuse other users' data. Viewshare has not previously been used for visualization of academic communities." Although some of the tools discussed below also use linked data to present directory information, the author chose Viewshare for an initial project because the data resides on the web. There was no need to request access to TAMU Libraries' server space, and the work would incur no expenses other than the author's time.

Heer and Sinderman noted that Viewshare enables a meaningful analysis in which users develop insights about significant relationships, domain-specific contextual influences, and causal patterns.[15] As Algee et al. suggest, there is an emerging consensus that tools that support this kind of exploratory process are valuable to a range of disciplinary perspectives.[16] They note that Viewshare has the ability to iteratively explore, compare data trends, and engender the accidental wisdom that comes from visualizing collections in new ways.

Perhaps the most significant linked-data directory project is VIVO (www.vivoweb.org), which creates a virtual life-science community to aid faculty, researchers, and students to discover common interests and make connections. This community organizes and presents information on people, research, and educational activities using an entity-relationship ontology model. VIVO has made possible the visualization of academic communities through open source applications. As stated on the VIVO home page (http://vivoweb.org), after initial installation, the developer populates the tool with researcher interests, activities, and accomplishments, and VIVO enables the discovery of research and scholarship across disciplines at that institution. VIVO supports browsing and a faceted search function for rapid retrieval of desired information, both encouraging natural discovery and allowing specific research. VIVO's developers are exploring the possibility of providing access not only to the Virtual International Authority File (VIAF) (http://viaf.org) but to other controlled vocabularies that exist as linked data sets. Devare et al. noted that a virtual community such as VIVO could serve as a model to explore synergies with peer institutions, museums, foundations, and research consortia to provide access to academic information on a national scale.[17]

Wolski et al. and Krafft et al. describe how the VIVO platform collects appropriate metadata from research collections within the university through customized feeds from the various university content management and corporate systems.[18] The system exposes this data to library discovery tools and other research information federations.[19]

Harvard Catalyst (http://catalyst.harvard.edu) is an open source tool for research networking that connects people by combining basic directory information with expertise keywords.[20] OpenScholar (http://theopenscholar.org) and BibApp (http://bibapp.org) provide for interactive research communities. In addition, all major commercial providers of scholarly content are involved in developing or are already running visualization tools (Elsevier's product, SciVal; ProQuest's product, Pivot; Symplectic Elements, to name a few).

## Background

TAMU has, 3,800 faculty, researchers, and advisors. It is home to more than 50,000 students and is the sixth largest university in the United States. It currently ranks among the top twenty universities nationally, with its research valued at more than $705 million annually.[21] TAMU Libraries is a Name Authority Cooperative Program (NACO) participant, and this project began as part of an effort to create name authority records for faculty members to contribute to the Library of Congress (LC) Name Authority File (NAF) and the TAMU Libraries' local name authority file. It gradually expanded into a visualization project when the author wanted to experiment with creating a dynamic and interactive view of data about TAMU faculty members to create various views with the available data. While the intention is to eventually include all faculty, the decision was made to start with a single department, TAMU's Department of Mathematics, one of the university's largest departments, with seventy-five tenure track and tenured professors, twenty-five visiting faculty, and twenty-nine lecturers. The goal was to create views of the department, its research areas, and faculty members' PhD granting institutions while also determining how many name authority records were needed.

### Data Collection

Data about all tenured and tenure track faculty were included in the project. The data is publicly available and presented no privacy, copyright, or other compliance issues. It was entered manually into a spreadsheet with data types varying

from static to linked data vocabularies. This step required sixteen hours of labor. Although the data was publicly available, it was not in formats that allowed automatic harvesting.

At the beginning of the project, only static textual data was collected, such as the names of faculty members, their research areas, PhD granting institutions, the date the PhD was granted, and date of hire. This data is available from the faculty directory website and was easy to collect. As the project expanded, and in consultation with members of the Department of Mathematics, the author decided to include additional data to enrich the data set: master's granting institution, bachelor's granting institution, PhD location that contains the latitude and longitude for the geographic place, links to their Department of Mathematics web page, personal web pages, a link to the Mathematics Genealogy Project, and a link to VIAF.

The OCLC Authority Record Number (ARN), which represents the LC NAF, was later included for verification purposes. Use of the LC NAF and the VIAF was considered and accepted, as it is best practice to reuse existing semantic vocabularies, even though they both offer limited coverage of names in the specific domain of mathematics (as mathematicians more often write scholarly articles than monographs). Inclusion of the Open Researcher and Contributor Identifier (ORCID) was considered. A careful examination of the number of faculty members with ORCID ids found that an insignificant number of individuals were registered, and the author decided that this data type was unnecessary for the prototype. Latitude and longitude data were collected from the GeoNames geographical database based on the corporate name for the PhD granting institution. The remaining data was collected from other sources such as the department page and personal faculty web pages. Some of the data (e.g., date of hire) came from the publicly available faculty directory.

### Data Normalization and Standardization

Once collected, the author developed standards for recording data in the spreadsheet. The name column was populated by entering last name, first name; department name was entered as Mathematics, and the college name was entered as established in the LC NAF; research areas were entered as found in the Library of Congress Subject Headings (LCSH); corporate names of bachelor's, master's and PhD granting institutions were entered as established in the LC NAF; PhD date and date of hire were entered in a "YYYY" format which is an ISO 8601 standard; PhD location was entered as a decimal value for latitude and longitude; department page, home page, image, and the link to the VIAF were entered as URLs (see figure 1).

Because normalizing and standardizing the data would help to show patterns, cleaning the data was an essential step

| College | College of Science |
| Department | Mathematics |
| Additional appointment | |
| Name | Aguiar, Marcelo |
| Research area | Algebra |
| Department page | http://www.math.tamu.edu/~maguiar/ |
| Home page | http://www.math.tamu.edu/~maguiar/ |
| VIAF | http://viaf.org/viaf/54453604 |
| ARN | 4552762 |
| PhD granting institution | Cornell University |
| Master granting institution | |
| Bachelor granting institution | Universidad de Uruguay |
| Mathematics Genealogy Project | http://genealogy.math.ndsu.nodak.edu/id.php?id=25171 |
| image | http://www.math.tamu.edu/directory/photos/aguiar_m.jpg |
| PhD date | 1997 |
| PhD location | 42.4406, -76.4969 |
| gender | male |
| Hire date | 2001 |

**Figure 1.** Sample Data Properties

before importing it into Viewshare. For example, if "algebra" was entered as a research area, algebra with a lower case "a" and also with an upper case "A" would be counted as two separate entries, although it is clearly the same entry. Additionally, there were entries with misspellings or numerical data entry errors. To address these issues, the author utilized OpenRefine (formerly known as Google Refine) (http://openrefine.org) to clean up the data used for this experimental project. OpenRefine (http://openrefine.org) is a free tool for working with messy data and transforming it from one format to another. This tool enabled a fast and efficient cleanup of the data.

### Data Ingestion into Viewshare

Data ingestion into Viewshare is a simple process. One can import data in different formats, such as spreadsheets in XSL format, XML, Dublin Core (DC) data from an Open Archives Initiative (OAI) end point, and some instances of ContentDM (Version 4 only).[22] Viewshare transforms the data from rows and columns to Resource Description Framework (RDF), the data model that underlies linked data. After ingestion, data can be quickly and easily visualized in various ways. Data was manually entered into the spreadsheet in the XSL format and ingested. Immediately after ingestion, Viewshare enables users to visualize data

**Figure 2.** List View Settings Options

using a drag-and-drop view-building workspace. Viewshare's open data principles allow multiple users to create different views of the same collection dataset.

### Viewing Data with Viewshare

After importing the collected data, the author utilized and explored options to choose layout, preview, add facets and views, and pick which features to display in the interface. Considering the number of options, creating the views required a negligible amount of time (less than two hours).

Viewshare allows the insertion of widgets, such as tag clouds based on research area data, lists of research areas, lists of faculty names, and a search window where one can search the data. There are also widgets that enable users to add a logo, slider, range, or text to enhance the visualization of the data set. See the Viewshare site for the TAMU Math Department at http://viewshare.org/share/1a848a62-d6fa-11e2-8aa1-4040e007d488/ for more information.

The default view provides a list of person records sorted by research area. See figure 2 for the options for list view. In the List View Settings, the label was set to Research area with the data displayed alphabetically in ascending order. In the List Lens Settings, the Title field was set to display the name of the faculty member, linked to his/her Department of Mathematics page. The person record shows all the collected data (or properties) for an individual with the exception of the Authority Record Number (ARN) from OCLC.

A second list view for PhD year was created. In this view, the Title field displays the faculty member's name, linked in this instance to the Mathematics Genealogy Project, a database that shows a mathematician's PhD thesis title, advisor, and affiliated graduate students tracing relationships among researchers through history. Some faculty members at TAMU are descendants of famous mathematicians such as Gauss, Euler, and his advisor Bernoulli, and others trace their academic genealogy to the 14th century. As with the default view by Research area, the author decided to display

all collected data except the Authority Record Number (ARN) from OCLC in the list by PhD year.

The next constructed view displays the PhD-granting institutions in a map with "PhD granting institution" as the label. In this view, Latitude/Longitude is the location of the PhD institution (see figure 3). The Zoom Level is set to "auto" to provide a full world map. Colored balloons help to visually distinguish multiple institutions that are close to each other on the map. In the Map Lens settings, "Title" is the faculty member's name, and the link is to the faculty member's home page.

It should be noted that initially the augment feature was used to generate the coordinates needed to display the location of the PhD granting institution. As stated on Viewshare's User's Guide site, Viewshare can augment or transform some types and forms of information into the proper format. Viewshare does not change existing data in the file during the augmentation process; it adds columns of new data to the file. Data can be augmented when loading or editing the data in the Viewshare tool and before creating views.[23] Out of seventy five PhD granting institutions, sixty-one had their values augmented through the Viewshare platform. It was then decided to collect the latitude and longitude value from the GeoNames geographical database. This decision made it possible to have all the values included in the data set which provides for complete map of PhD granting institutions.

The Timeline view (see figure 4) visualizes the length of time between when the individual's PhD was granted to when he or she was hired by TAMU. Each line is labeled with a person's name and links to the person record. There are two bands for time units with the top band set to year and the bottom band set to decade. Again, colors are used to distinguish the various institutions. In the Timeline Lens setting, the Title set to the faculty member's name and includes a link to the faculty member's home page.

The PhD Gallery view is sorted alphabetically by PhD granting institution. The List Lens Settings include the property image; the image comes from the Department of Mathematics' website. The name below the image links to the faculty member's VIAF record if available; if not, the link defaults to the default view list of person records sorted by research area. The link to VIAF links directly into WorldCat Identities, the LC NAF, and the International Standard Name Identifier (ISNI). In addition, the VIAF record for each faculty member may be viewed as an RDF record. As stated on the OCLC website, WorldCat Identities has a summary page for every name in WorldCat (currently there are about 30 million names), including named persons, organizations, and fictitious characters. The WorldCat Identities page "include[s] a list of most widely held by libraries, works by and about the identity, a list of variant forms of name the identity has been known by, a FAST tag cloud of places,

topics, etc. closely related to works by and about the person, links to co-authors, and more."[24]

The LC NAF "provides authoritative data for names of persons, organizations, events, places, and titles. Its purpose is the identification of these entities and, through the use of controlled vocabulary, to provide uniform access to bibliographic resources."[25]

ISNI (www.isni.org) is an International Standards Organization (ISO) standard (ISO 27729) that identifies the public identities of parties and serves as a tool for disambiguating public identities. While ORCID was the preferred choice for this data set, lack of use by study participants made inclusion unnecessary. However, future experiments using this data set will, most likely, include ORCID data because of TAMU's plans to actively promote ORCID. Additionally, a TAMU Libraries team (including the author) was awarded a grant from ORCID to assign identifiers to all TAMU faculty members.

A gallery called Research Area was created to sort faculty members alphabetically by research area. The title is set to the faculty member's name with a link to his or her departmental home page, which lists his or her publications. Many of the publication links go to the preprint versions of papers or to information on arXiv.org, a well-known archive that provides an e-print service for mathematics, physics, computer science, quantitative biology, quantitative finance, and statistics.

### Exploring the Views

The visualization of this data brought to light interesting relations and connections, enabling the author to fully realize new interpretations of data. Simple keyword searching reduces data to only the keyword being entered into the search box. For example, if one enters the term "group" in the search widget, all available views will reduce to display data containing the keyword entered. In this case, the result shows four faculty members, three of whom have "Combinatorial Group Theory" as their research area and one faculty member with research area "Group Representations" (see figure 5). When "Combinatorial Group Theory" is selected from the Research Area widget list, all the widgets and the selected view display data in relation to the research area selected (see figure 6).

When the "Combinatorial Group Theory" research area is selected, one can examine the various views and explore the data about each faculty member associated with the selected research area. Figure 7 demonstrates how the PhD granting institution map displays data associated with the three faculty members whose research area is "Combinatorial Group Theory."

A user may examine all the views and check the data about a specific faculty member. The PhD granting
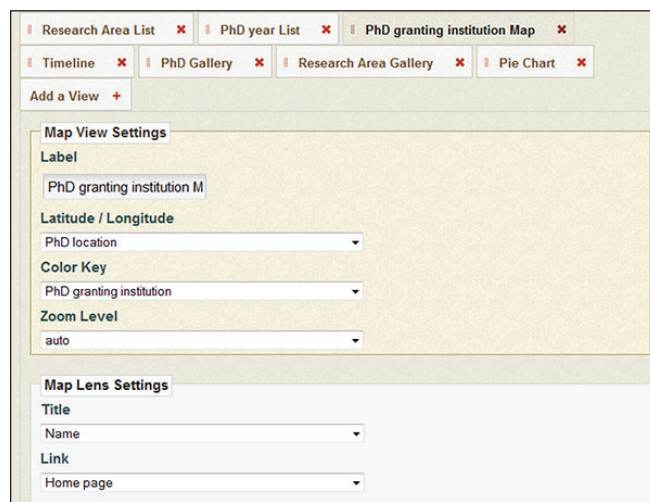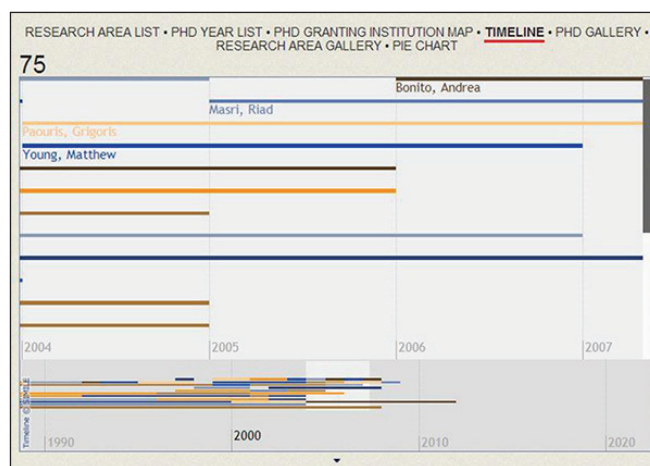


**Figure 3.** Map Settings



**Figure 4.** Timeline View: The Length of Time from When the Individual's PhD was Granted to When He/She was Hired by TAMU

institution map takes users directly to the location of the granting institution for that faculty member. If the user clicks on the pin located on the map, he can see all data about the selected faculty member displayed on that specific view. Similarly, by clicking on the timeline view, the user can see the year a specific faculty member received his PhD and the year he was hired by TAMU. The timeline view enabled the author to see differences in past and more recent hiring practices. Beginning in the year 2000, more faculty members were hired each year than in any year before. That trend continued until 2009, when the university faced budget cuts. No faculty members were hired during 2010–2011, and only one new faculty member was hired in 2012.

The pie chart view, when displaying the research area, provides a breakdown of all research areas in percentages
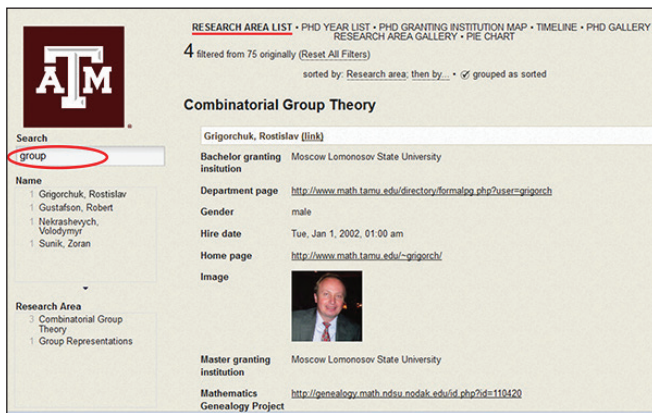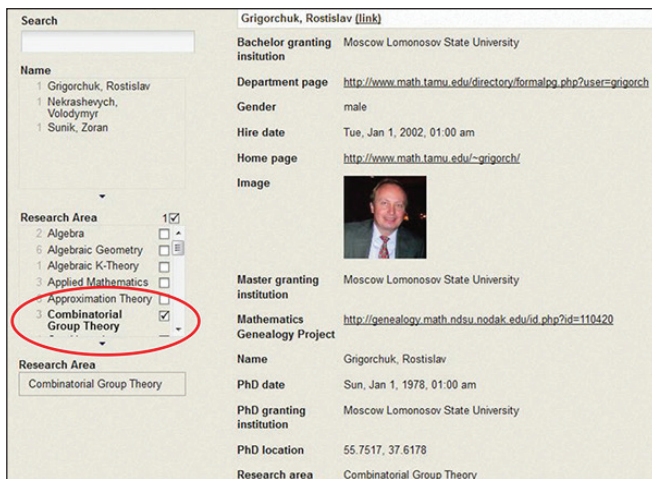
**Figure 5.** Keyword Searching



**Figure 6.** Selecting Research Area from othe Widget List

and number of faculty. From the pie chart view, the author discovered that the most represented research area in the Department of Mathematics at TAMU is Partial Differential Equations. Faculty members in charge of the design and content of the Department of Mathematics website were surprised to learn that Operator Theory is no longer the most prevalent area of research. This has immediate implications when recruiting graduate students and promoting the department's strengths, especially as the Viewshare tool is available for public use. Future incarnations of the Viewshare tool can be embedded in the LibGuides created by TAMU Libraries subject selectors for their liaison departments. Subject selectors noted the importance of this project for collection development as they perceive it as a useful tool in determining the research focus of academic departments.

Viewshare's pie chart view includes properties such as gender, PhD date, PhD granting institution, research area, master's granting institution, and bachelor's granting institution. When organized by gender, the pie chart shows that roughly one tenth of the faculty members are female. There are just eight female faculty members in a total of seventy-five faculty. The pie chart view clearly shows that Partial Differential Equations is the most represented research area in the TAMU Department of Mathematics.

### Exporting the Data

Data can be exported from Viewshare in various formats—RDF/XML, JSON, or semantic wikitext— for reuse. The views created by this project, which display data about the Department of Mathematics, are publicly available to anyone who uses Viewshare, and any user can download the data in a format suitable to their needs. One can also generate an HTML view. When the list views are exported in HTML, they can be used to create webpages with the information available from the list view in question, stylized to each user's preference. Figure 8 represents a snapshot of the HTML view from the Research Area list view. The HTML page was generated in Adobe Dreamweaver with only minimal customization: an added background image.

### Lessons Learned

The directory of faculty members was created relatively quickly by a librarian from the Cataloging Department. Sixteen hours of work was needed to collect all the data for faculty members from the Department of Mathematics. If the same rate was used per faculty member, it would take 600 hours to collect information for all faculty members on campus. If it were possible to populate Viewshare with data from the university's Research Information System office, it would shorten the time needed for this project, perhaps by two thirds. Only VIAF and the coordinates for the PhD granting institution might require particular attention. After examining the websites of multiple departments, the author discovered that not all departments publicly share the same information about their faculty members. Information about PhD granting institution and PhD date is absent in some cases, and not all departments provide individual home pages for their faculty.

Subject selectors suggested that faculty gender should be excluded from future projects as it may be perceived as a privacy issue. This concern was raised in relation to possible transgendered faculty members. Because of this and other possible privacy concerns, future projects will include an opt out/in survey so that faculty members may choose whether to share their information. Visualizing the academic community at TAMU will enable library patrons, students, faculty members, and other stakeholders to find information about the faculty as a whole, for example, insight into the

interdisciplinary work in which members of the Department of Mathematics are involved. It may be possible to visualize dual appointments.

To produce clean, interactive displays of data through the creation of various views, the data must be normalized. Simple removal of white trailing spaces, capitalization issues, and spelling mistakes were completed using OpenRefine. It was also necessary to replace the names of the institutions with the authoritative form as available in the LC NAF, and we intend to continue this practice. These two essential steps of normalizing data and using authoritative forms of names enabled us to see the patterns and trends among the faculty members from TAMU's Department of Mathematics.

As previously mentioned, we discovered that Partial Differential Equations is now the most prevalent primary research area. According to the faculty members that reviewed the Viewshare representation of data about the Department of Mathematics, Operator Theory was previously the most prevalent research area. Analysis of the data reveals that, as new faculty members were hired and others retired, the main research area for the department shifted. However, the reason for this perceived change may be because only the primary research area for each faculty member was collected instead of all research areas. This was a limitation of the project and it will be addressed in the future. A future project will include as many research areas as each faculty member shares through a survey or as many as are provided in the university's Research Information System.

GeoNames will be used from the beginning of the future large-scale project as we discovered that not all values of corporate names for PhD granting institutions are augmented through the Viewshare platform. The margin of error is not significant, as 81 percent of the PhD granting institutions had their location values augmented correctly, but it is desirable to have all the location values in the data set. Faculty members from the Department of Mathematics received their PhD degrees from institutions located either in North America or Europe (see figure 9).

Additionally, we discovered that almost half of the faculty members lacked Name Authority Records (NARs). As a NACO participant, TAMU Libraries has the capacity to create the remaining NARs. For nonmonographic publications, online research IDs will be essential for linking out to faculty publications. One solution is to use ORCID. If members of the Department of Mathematics had registered for ORCIDs, the Viewshare views would have been more complete. It is a goal going forward to establish ORCIDs for all TAMU faculty members.

In this pilot project, only one of the few existing URI-based vocabularies and ontologies was incorporated and used, VIAF. LCSH, Lexvo (URI referenced controlled list of characters, words, terms), DBpedia, and GeoNames
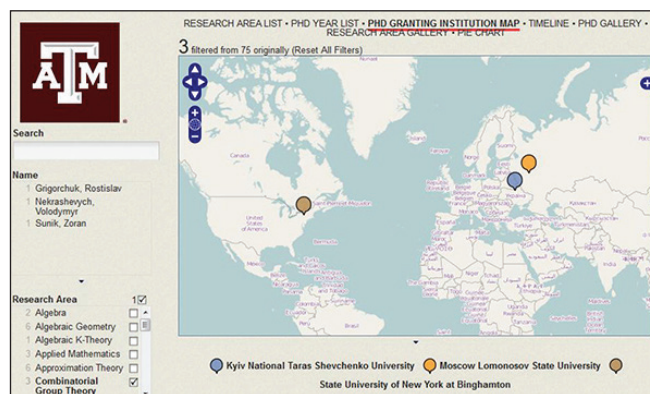


**Figure 7.** PhD Map View for Faculty Members with Research Area "Combinatorial Group Theory"



**Figure 8.** Generated HTML View of Research Area List: Algebraic Geometry (6)

are being considered as potential additions for future development. LCSH was consulted when normalizing the research areas represented within the faculty members of the Department of Mathematics.

## Future Large-Scale Project

Creating dynamic, interactive views of data describing Department of Mathematics faculty members was the first step towards a large-scale project that will create data visualizations for all TAMU academic departments. A team was identified to work on the large-scale project. As each department is unique and has its own specialties, we are aware that visual representation will pose new issues and research questions, yet we also anticipate that new departmental strengths will be discovered. The goal of the large-scale

**Figure 9.** PhD Granting Institutions for Faculty Members from TAMU

project is to expose hidden possible relationships between faculty members from different departments and facilitate collaboration and connection. Enabling researchers to find possible collaborators from different departments will create a stronger institution with increased opportunities for competitive national grants. Having one department represented in the pilot project did not provide an opportunity to see the relationships between departments. The relationships of these faculty members formed through their involvement in various interdisciplinary institutions that exist on TAMU's campus is also not apparent. As Borgman concludes, the boundaries are blurring between the sciences and the humanities. This blurring urgently calls for outreach and organization.[26] At TAMU, the libraries are responding by creating tools and methods to bring various researchers from our institution together, creating possible ground for new research.

The Dean of the TAMU Libraries and his management team have expressed interest in promoting the project to the university, and we expect full support from university administrators to further pursue this project. Installing a visualization platform that will provide for discoverability of faculty research output is our priority. To gain full support of the university administration, the software must be installed on our servers to provide for easy customization since Viewshare's web application does not support full customization. Deciding which platform to use depends on the university administration's support since the plan is to include data from external sources, such as the registrar's office, the Vice President for Research's office (VPR), the Dean of Faculties' office, the Research Information System office, and the human resources office. To provide access to data from those external sources, the project needs a platform with the capability to harvest external data. We are currently experimenting with VIVO on a test server and have previously experimented with BibApp.

In the summer of 2013, this prototype included all TAMU faculty. When this project concludes, we will have contributed to the enhancement of the University's brand profile and impacted the development of research for the University, individual researchers, and research groups. This initiative also provides a rich, internal discovery mechanism for faculty, current and future, plus graduate students and the general public. It will enable researchers, administrators, and students to obtain a meaningful snapshot of a given investigator's productivity and reach. Perhaps most useful, data collected for the Viewshare project can be easily ingested in different open source platforms, such as VIVO, based on linked data principles and reused for purposes not even considered by the author.

## Conclusion

This paper describes how all aspects of this experimental project, including the role played by the university's library professionals, could empower users to effectively find and use online primary resources about faculty members. The initial Viewshare project created interest in further development by both library and university administrators, who are now willing to invest computing resources and manpower toward expanding it.

Developing this Semantic Web–based service for collecting research data highlights the importance of reusing and exposing research data that resides in university websites and databases. Visualization of university research strengths, research outputs, collaborative relationships, and other characteristics of the campus research environment were presented. Siloed research content across the university should be discoverable through the aggregation of data from a range of scattered university systems, and the libraries can take the lead in these tasks as experts in constructing controlled vocabularies, personal name authorities, and corporate name authorities. The changing face of the research environment in the university system should not be ignored by libraries; we must respond and adapt to the changing landscape.

### References

1.  Christine L. Borgman, "The Digital Future is Now: a Call to Action for the Humanities," *Digital Humanities Quarterly* 3, no. 4, (2009), accessed May 3, 2013, http://digitalhumanities .org/dhq/vol/3/4/000077/000077.html.
2.  Richard E. Luce, "A New Value Equation Challenge: The Emergence of eResearch and Roles for Research Libraries," in *No Brief Candle: Reconceiving Research Libraries for*

*the 21st Century* (Washington, DC: Council on Library and Information Resources, 2008): 42–50, accessed May 3, 2013, www.clir.org/pubs/reports/pub142/luce.html.

3. Wendy Lougee, "The Diffuse Library Revisited: Aligning the Library as Strategic Asset," *Library Hi Tech* 27, no. 4, (2009): 610–23.

4. Tim Berners-Lee, "Linked Data," last modified June 18, 2009, accessed May 1, 2013, www.w3.org/DesignIssues/LinkedData.html.

5. Dean Allemang and James Hendler, *Semantic Web for the Working Ontologist, Effective Modeling in RDFS and OWL* (Waltham, MA: Morgan Kaufmann, 2011), 6.

6. Eric Miller and Micheline Westfall, "Linked Data and Libraries," *Serials Librarian* 60, no. 1–4 (2011): 17–22.

7. Getaneh Alemu et al., "Linked Data for Libraries: Benefits of a Conceptual Shift from Library-Specific Record Structures to RDF-based Data Models," *New Library World* 113, no. 11–12, (2012): 549–70, accessed May 3, 2013, http://conference.ifla.org/past/ifla78/92-alemu-en.pdf.

8. Thomas Baker et al., "Library Linked Data Incubator Group Final Report," report, W3C Incubator Group, October 25, 2011, accessed May 3, 2013, www.w3.org/2005/Incubator/lld/XGR-lld-20111025.

9. Richard Wallis, "Library of Congress to Boldly Voyage to Linked Data Worlds," *Data Liberate* (blog), November 2, 2011, accessed May 24, 2013, http://dataliberate.com/2011/11/library-of-congress-to-boldly-voyage-to-linked-data-worlds.

10. Alemu et al., "Linked Data for Libraries," 6.

11. Jeffrey Heer and Ben Shneiderman, "Interactive Dynamics for Visual Analysis," *Communications of the ACM* 55, no. 4 (April 2012): 45–54.

12. Thea Lindquist et al., "Leveraging Linked Data to Enhance Subject Access in Online Primary Sources: A Case Study of the University of Colorado Boulder's World War I Collection Online" (paper presented at the 2012 World Library and Information Congress, Helsinki, August 14, 2012) accessed April 30, 2013, http://conference.ifla.org/past/ifla78/117-lindquist-en.pdf.

13. Philip Evan Schreur, "The Academy Unbound," *Library Resources & Technical Services* 56, no. 4 (October 2012): 227.

14. Ibid, 237.

15. Heer and Shneiderman, "Interactive Dynamics for Visual Analysis," 45.

16. Lauren Algee, Jefferson Bailey, and Trevor Owens, "Viewshare and the Kress Collection: Creating, Sharing, and Rapidly Prototyping Visual Interfaces to Cultural Heritage Collection Data," *D-Lib Magazine* 18, no. 11 (2012): 3.

17. Medha Devare et al., "Connecting People, Creating a Virtual Life Sciences Community," *D-Lib Magazine* 13, no. 7–8 (2007), accessed May 3, 2013, www.dlib.org/dlib/july07/devare/07devare.html.

18. Malcolm Wolski, Joanna Richardson, and Robyn Rebollo, "Shared Benefits from Exposing Research Data," *Proceedings of the IATUL Conferences. Paper 5* (2011): 1–11, accessed May 5, 2013, www.bg.pw.edu.pl/iatul2011/proceedings/ft/Wolski_M.pdf; Dean B. Krafft et al., "VIVO: Enabling National Networking of Scientists" (paper presented at the meeting of the WebSci10: Extending the Frontiers of Society On-Line, 2010), accessed May 3, 2013, www.bibsonomy.org/bibtex/287a568555fcc35532e9384337c1ce68a/jaeschke.

19. Wolski et al., "Shared Benefits from Exposing Research Data," 1.

20. "About Harvard Catalyst," Harvard Catalyst, the Harvard Clinical and Translational Science Center, accessed April 29, 2013, http://catalyst.harvard.edu/about.

21. "A&M at a Glance," TAMU University website, accessed January 30, 2013, www.tamu.edu/about.

22. "Viewshare User Guide," Viewshare, accessed May 26, 2013, http://viewshare.org/about/userguide/#s3.1.

23. Ibid.

24. "WorldCat Identities," OCLC, accessed January 30, 2013, www.oclc.org/research/activities/identities.html.

25. "Library of Congress Names," Library of Congress, accessed January 30, 2013, http://id.loc.gov/authorities/names.html.

26. Borgman, "The Digital Future is Now."