Visualization in VIVO:
A case study in how VIVO data and technology can be used
August 24, 2011, 8:30 am – 12:00 pm

Facilitators
- Katy Börner, Cyberinfrastructure for Network Science Center, Indiana University
- Chintan Tank, Cyberinfrastructure for Network Science Center, Indiana University
- Chin Hua Kong, Cyberinfrastructure for Network Science Center, Indiana University


Intended Audience
Some of the people who might be interested in this workshop are:
- Anyone interested in the visualization capabilities of VIVO, and the visualization work of the Cyberinfrastructure for Network Science Center.
- Software developers interested in an overview of the VIVO environment from a technical perspective, and a more in-depth exploration of VIVO through the lens of visualization development.
- Librarians and Science Administrators interested in gaining a deeper understanding of how VIVO works, exploring some of what it is already capable of, and understanding what it could be used for in the future.
- Companies that plan to offer value-added services for VIVO.
- Researchers that would like to utilize VIVO data in their scientific work.

Content Overview
VIVO is an excellent system for creating and managing faculty and researcher profiles, but it is also capable of much more. VIVO’s ability to obtain and interlink data from a variety of high quality sources, including institutional systems of record and online databases, coupled with VIVO’s use of open semantic web technologies, makes VIVO an exciting and powerful data platform. This platform makes VIVO itself capable of many new features, and also makes it possible for outside companies and researchers to develop their own VIVO applications and perform their own analyses using VIVO data.

This hands-on workshop aims to explore the possibilities of VIVO technology and data by way of the Cyberinfrastructure for Network Science Center’s VIVO visualization work. It starts with an overview of VIVO’s architecture, and explores how the architecture makes it possible to expand and build on VIVO. We then describe the Cyberinfrastructure for Network Science Center’s visualization work, including an overview of the visualizations created for VIVO thus far. Taking a closer look at the VIVO visualizations, we explore their individual motivations and features, and how we use VIVO to make these visualizations possible. We conclude with a hands-on component where participants are guided through the creation of a simple visualization using live data from VIVO instance. The workshop concludes with a general question-answer session.
Last but not least, we will showcase different data analyses and visualizations of VIVO data at the individual, institution, and national level such as:

- **Individual level.** Statistics and ego-centric scholarly networks on VIVO Profile pages.
- **Institutional level.** Analyses and visualizations of funding intake and publication output for departments and centers accessible via the VIVO Index page. Download of relevant data in tabular and network formats for further analysis using the Network Workbench tool.
- **National level.** Visualization of VIVO installations and their profile holdings together with web page access and general VIVO information requests. Plus, services that use VIVO URIs to access data across different VIVO instances.

The workshop concludes with a general question and answer session.

---

**Workshop Attendees**

Registered by Aug. 5, 2011

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Carolyn</td>
<td>Eckhardt</td>
<td>Senior IT Analyst</td>
<td>Duke Health Technology Solutions (DHTS)</td>
<td>Durham</td>
</tr>
<tr>
<td>2</td>
<td>Jeff</td>
<td>Erickson</td>
<td>IT Specialist</td>
<td>National Institute of Health</td>
<td>Bethesda</td>
</tr>
<tr>
<td>3</td>
<td>Holly</td>
<td>Fals-Krzesinski</td>
<td>Director, Research Team Support</td>
<td>Northwestern University</td>
<td>Chicago</td>
</tr>
<tr>
<td>4</td>
<td>Tricia</td>
<td>Gallagher</td>
<td>Research Informatics Analyst, Lead</td>
<td>Emory University</td>
<td>Atlanta</td>
</tr>
<tr>
<td>5</td>
<td>Jeffrey</td>
<td>Harper</td>
<td>SciVal Consultant</td>
<td>Elsevier, Inc.</td>
<td>New York</td>
</tr>
<tr>
<td>6</td>
<td>David</td>
<td>Jetter</td>
<td>Product Manager</td>
<td>Inform International</td>
<td>Albany</td>
</tr>
<tr>
<td>7</td>
<td>Layne</td>
<td>Johnson</td>
<td>Translational Science Information Scientist</td>
<td>University of Minnesota</td>
<td>Minneapolis</td>
</tr>
<tr>
<td>8</td>
<td>Craig</td>
<td>Knoblock</td>
<td>Research Professor</td>
<td>USC Computer Science Dept</td>
<td>Marina del Rey</td>
</tr>
<tr>
<td>9</td>
<td>Carlos</td>
<td>Liu</td>
<td>Facility Manager</td>
<td>GENE CENTER, HUNTER COLLEGE</td>
<td>NEW YORK</td>
</tr>
<tr>
<td>10</td>
<td>David</td>
<td>Lyons</td>
<td>IT Specialist</td>
<td>US EPA</td>
<td>Durham</td>
</tr>
<tr>
<td>11</td>
<td>John</td>
<td>Mark Okenbloom</td>
<td>Digital Library Planner</td>
<td>University of Pennsylvania</td>
<td>Philadelphia</td>
</tr>
<tr>
<td>12</td>
<td>Victoria</td>
<td>McGovern</td>
<td>Senior Program Officer</td>
<td>Burroughs Wellcome Fund</td>
<td>Durham</td>
</tr>
<tr>
<td>13</td>
<td>David</td>
<td>Palmer</td>
<td>Scholarly Communications Team Leader</td>
<td>The University of Hong Kong</td>
<td>Hong Kong</td>
</tr>
<tr>
<td>14</td>
<td>Simon</td>
<td>Porter</td>
<td>University of Melbourne</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Sean</td>
<td>Thomas</td>
<td>Program Manager, Scholarly Repository Services</td>
<td>Massachusetts Institute of Technology (MIT)</td>
<td>Cambridge</td>
</tr>
<tr>
<td>16</td>
<td>Geerdt</td>
<td>Van Groote</td>
<td>Senior researcher</td>
<td>Flemish government, Economy, Science &amp; Innovation department</td>
<td>Brussels</td>
</tr>
<tr>
<td>17</td>
<td>Ann</td>
<td>Vega</td>
<td>Scientific Applications Manager</td>
<td>US EPA</td>
<td>Cincinnati</td>
</tr>
<tr>
<td>18</td>
<td>Michael</td>
<td>Warben</td>
<td>SciVal Consultant</td>
<td>Elsevier, Inc.</td>
<td>Saline</td>
</tr>
<tr>
<td>19</td>
<td>Hal</td>
<td>Warren</td>
<td>Director, Publishing Innovation</td>
<td>American Psychological Association</td>
<td>Washington</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Workshop Schedule - Part I

Social Network Visualizations (Katy Borner)
- Visualization Types and Levels
- Exemplary User Needs
- Existing VIVO Visualizations

Analysis & Visualization of VIVO Data (Chin Hua Kong)
- Using Science of Science Tool (http://sci2.cns.iu.edu)
- Using Gephi (http://gephi.org)

10 min break

Workshop Schedule - Part II

Visualization-in-depth (Chintan Tank & Chin Hua Kong)
- Map of Science
- Temporal Graph
- National Researcher Networking

VIVO Visualizations (Chintan Tank)
- Accessing VIVO data
- VIVO Architecture
- How we use VIVO data: The Visualization Pipeline

15 min break
Workshop Schedule - Part III

Guided Tour (Chintan Tank)
- Create a new VIVO Visualization

Outlook (Katy Borner)
- Planned VIVO Visualizations
- National Level Visualizations

Q&A

Social Network Visualizations
- Visualization Types and Levels
- Exemplary User Needs
- Proposed VIVO Visualizations
## Type of Analysis vs. Level of Analysis

<table>
<thead>
<tr>
<th></th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–10,000 records)</th>
<th>Macro/Global (10,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical Analysis/Profiling</strong></td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of USA, all of science.</td>
</tr>
<tr>
<td><strong>Temporal Analysis (When)</strong></td>
<td>Funding portfolio of one individual</td>
<td>Mapping topic bursts in 20-years of PNAS</td>
<td>113 Years of physics Research</td>
</tr>
<tr>
<td><strong>Geospatial Analysis (Where)</strong></td>
<td>Career trajectory of one individual</td>
<td>Mapping a state’s intellectual landscape</td>
<td>PNAS publications</td>
</tr>
<tr>
<td><strong>Topical Analysis (What)</strong></td>
<td>Base knowledge from which one grant draws.</td>
<td>Knowledge flows in Chemistry research</td>
<td>VxOrd/Topic maps of NIH funding</td>
</tr>
<tr>
<td><strong>Network Analysis (With Whom?)</strong></td>
<td>NSF Co-PI network of one individual</td>
<td>Co-author network</td>
<td>NSF’s core competency</td>
</tr>
</tbody>
</table>

## Type of Analysis vs. Scale of Level of Analysis

<table>
<thead>
<tr>
<th></th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–10,000 records)</th>
<th>Macro/Global (10,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical Analysis/Profiling</strong></td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of USA, all of science.</td>
</tr>
<tr>
<td><strong>Temporal Analysis (When)</strong></td>
<td>Funding portfolio of one individual</td>
<td>Mapping topic bursts in 20-years of PNAS</td>
<td>113 Years of physics Research</td>
</tr>
<tr>
<td><strong>Geospatial Analysis (Where)</strong></td>
<td>Career trajectory of one individual</td>
<td>Mapping a state’s intellectual landscape</td>
<td>PNAS publications</td>
</tr>
<tr>
<td><strong>Topical Analysis (What)</strong></td>
<td>Base knowledge from which one grant draws.</td>
<td>Knowledge flows in Chemistry research</td>
<td>VxOrd/Topic maps of NIH funding</td>
</tr>
<tr>
<td><strong>Network Analysis (With Whom?)</strong></td>
<td>NSF Co-PI network of one individual</td>
<td>Co-author network</td>
<td>NSF’s core competency</td>
</tr>
</tbody>
</table>
Individual Co-PI Network
Ke & Börner, (2006)

Mapping Indiana’s Intellectual Space

Identify
- Pockets of innovation
- Pathways from ideas to products
- Interplay of industry and academia
Mapping Topic Bursts

Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.


Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions


Research questions:
1. Does space still matter in the Internet age?
2. Does one still have to study and work at major research institutions in order to have access to high quality data and expertise and to produce high quality research?
3. Does the Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research institutions?

Contributions:
- Answer to Qs 1 + 2 is YES.
- Answer to Qs 3 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.
Mapping Transdisciplinary Tobacco Use Research Centers Publications

Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

Zoss & Börner, forthcoming.

Supported by NIH/NCI Contract HHSN261200800812

Mapping the Evolution of Co-Authorship Networks

Needs-Driven Workflow Design using a modular data acquisition/analysis/modeling/visualization pipeline as well as modular visualization layers.
Social Network Visualizations

- Visualization Types and Levels
- Exemplary User Needs
- Proposed VIVO Visualizations

Exemplary User Needs

- **Individual level.** Researchers would like to enter data once and then use it to print cv’s, annual summary reports, find team members & mentors, render web pages to “become effortlessly visible” in support of collaboration and research.

- **Institutional level.** Campus level officials need to pool (expertise) resources for major grant applications, understand research strengths and trends of different units as part of competitive landscape analysis, advertise their institution to recruit and retain students and faculty.

- **National level.** Funding agencies and others need to understand who is working on what topic(s), what research areas/expertise centers are emerging, or who is funding/supporting a certain topic/expert team.
Exemplary User Needs - Researchers

- **Authors**—need to select promising research topics, students, collaborators, and publication venues to increase their reputation. They benefit from a global view of competencies, reputation and connectivity of scholars; hot and cold research topics and bursts of activity, and funding available per research area.
- **Editors**—have to determine editorial board members, assign papers to reviewers, and ultimately accept or reject papers. Editors need to know the position of their journals in the evolving world of science. They need to advertise their journals appropriately and attract high-quality submissions to increase the journal’s reputation leading to higher quality submissions.
- **Reviewers**—read, critique, and suggest changes to help improve the quality of papers and funding proposals. They need to identify related works that should be cited or complementary skills that authors might consider when selecting project collaborators.
- **Teachers**—teach classes, train doctoral students, and supervise postdoctoral researchers. They need to identify key works, experts, and examples relevant to a topic area and teach them in the context of global science.
- **Inventors**—create intellectual property and obtain patents, thus needing to navigate and make sense of research spaces as well as intellectual property spaces.
- **Investigators**—scholars acquire funding to support students, hire staff, purchase equipment, or attend conferences. Research interests and proposals have to be matched with existing federal and commercial funding opportunities, possible industry collaborators/sponsors.
- **Team Leads and Science Administrators**—need to evaluate performance and provide references for current or previous members; report the progress of different projects to funding agencies.

Social Network Visualizations

- Visualization Types and Levels
- Exemplary User Needs
- Proposed VIVO Visualizations
Existing VIVO Visualizations

- **Graphic Design Visualizations** that show VIVO team.
- **Individual level.** Sparkline statistics; ego-centric scholarly co-author and co-PI networks on VIVO Profile page, Temporal Comparison page.
- **Institutional level.** Analyses and visualizations of funding intake and publication output for departments and centers accessible via the Temporal Comparison page and Map of Science page. Download of relevant data in tabular and network formats for further analysis using MS Excel, the Science of Science Tool or Gephi.
- **International level.** Visualization of VIVO installations and their profile holdings together with web page access and general VIVO information requests.
This workshop covers a rather small piece of the entire VIVO project effort.

<table>
<thead>
<tr>
<th>Type of Analysis vs. Level of Analysis</th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–10,000 records)</th>
<th>Macro/Global (10,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical Analysis/Profiling</strong></td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of USA, all of science.</td>
</tr>
<tr>
<td><strong>Temporal Analysis (When)</strong></td>
<td>Funding portfolio of one individual</td>
<td>Mapping topic bursts in 20-years of PNAS</td>
<td>113 Years of physics Research</td>
</tr>
<tr>
<td><strong>Geospatial Analysis (Where)</strong></td>
<td>Career trajectory of one individual</td>
<td>Mapping a state’s intellectual landscape</td>
<td>PNAS publications</td>
</tr>
<tr>
<td><strong>Topical Analysis (What)</strong></td>
<td>Base knowledge from which one grant draws.</td>
<td>Knowledge flows in Chemistry research</td>
<td>VxOrd/Topic maps of NIH funding</td>
</tr>
<tr>
<td><strong>Network Analysis (With Whom?)</strong></td>
<td>NSF Co-PI network of one individual</td>
<td>Co-author network</td>
<td>NSF’s core competency</td>
</tr>
</tbody>
</table>
Temporal Analysis (When) Temporal visualizations of the number of papers/funding awarded at the institution, school, department, and people level
**Topical Analysis (What)** Science map overlays will show where a person, department, or university publishes most in the world of science. (in work)

**Network Analysis (With Whom?)** Who is co-authoring, co-investigating, co-inventing with whom? What teams are most productive in what projects?
Geospatial Analysis (Where?) Where are what NRN instances and what data holdings do they have?

Hands-On (Chin Hua Kong)

- Analysis & Visualization of VIVO Data

Guided Tour (Chintan Tank)

- Create new VIVO visualization
Required tools and resources

- Make sure the following applications are installed on your computer
  - Java 1.5 or higher version. You can install the latest Java through [http://www.java.com](http://www.java.com)
  - The Science of Science tool (Sci2) can be downloaded at [http://sci2.cns.iu.edu](http://sci2.cns.iu.edu). The online tutorial is available at [here](http://sci2.cns.iu.edu).
  - The Gephi tool can be downloaded at [http://gephi.org](http://gephi.org). The user guide is available at [here](http://gephi.org).

- Download the workshop’s data package from [http://wiki.cns.iu.edu/display/PRES/VIVO+Presentation](http://wiki.cns.iu.edu/display/PRES/VIVO+Presentation)

- Make sure your computer has Internet access. The username is VIVO2 and password is vivoweb.

- We also have the documents and software available for you on USB, if you need it.

---

Analysis & Visualization of VIVO Data

- Using Gephi ([http://gephi.org](http://gephi.org))
Why do we care?

Creativity -> Alternatives

- Create different visualizations on the same data
Evolving the visualization technique

Zero code milestone!

Fast!

Can I create a visualization?!

YES YOU CAN.

Utilizing VIVO data

- Administration
  - Providing abstract view of the organization
  - Cleaning the VIVO data

- Research
  - Free data for research
  - Analyze and visualize data for publication

- Development
  - Analyze and understand the data
  - Create Mockups

- Personal
  - Better understand of your network
  - Analyze your own interests
Analysis & Visualization of VIVO Data

- Three-Step Visualization
- Sci²: Organization Hierarchy Visualization
- Gephi: Organization Hierarchy Visualization
- 1.5-Step Visualization

Three-Step Visualization

- Three simplified steps to create a visualization

1. Playing with SPARQL queries
2. Cleaning and Preprocessing the data
3. Visualizing the data
Analysis & Visualization of VIVO Data

- Three-Step Visualization
- Sci²: Organization Hierarchy Visualization
- Gephi: Organization Hierarchy Visualization
- 1.5-Step Visualization

Playing with SPARQL

- What is a SPARQL query?
  - An RDF query language
  - A key technology for semantic web where a query can consist triple patterns, conjunctions, disjunctions, and optional patterns.
  - W3C definition
  - SPARQL Tutorial

- Please visit the SPARQL query end-point of the University of Florida (UFL) at [http://sparql.vivo.ufl.edu/sparql.html](http://sparql.vivo.ufl.edu/sparql.html) (You can find this link in the Readme.txt file located in the data package)
Playing with SPARQL

- To retrieve 5 organization names in the UFL’s VIVO system, copy and paste the query from the `FiveOrganizationLabelsSPARQL.txt` file into the text box at the SPARQL endpoint. The following shows the content of the query.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX core: <http://vivoweb.org/ontology/core#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT (str(?orgLabel) as ?organizationLabel)
WHERE
{
    ?org rdf:type foaf:Organization .
}
LIMIT 5
```

Playing with SPARQL

- You can save the results in five different formats
  - XML file
  - JSON file
  - Plaint text (in a table format)
  - CSV file (Comma-delimited values)
  - TSV file (Tab-delimited values)
Playing with SPARQL

- Example of organization structure in UFL VIVO data

  ![Organization Structure Diagram]

- Sample results from SPARQL

  \[
  \begin{array}{|c|c|}
  \hline
  \text{OrganizationLabel} & \text{subOrganizationLabel} \\
  \hline
  \text{UFL} & \text{Stanford University} \\
  \text{Office of the President} & \text{Faculty Senate} \\
  \text{UFL} & \text{Department of Mathematics} \\
  \text{Office of Research} & \text{Office of the President} \\
  \end{array}
  \]

  \[
  \begin{array}{|c|c|}
  \hline
  \text{UFLOrganizationLabel} \\
  \hline
  \text{UFL} \\
  \text{Office of the President} \\
  \text{Faculty Senate} \\
  \text{Office of Research} \\
  \end{array}
  \]

  `OrganizationToSubOrganization.csv`  `UFLOrganizations.csv`

Playing with SPARQL

- Connect to the UFL’s SPARQL endpoint at [http://sparql.vivo.ufl.edu/sparql.html](http://sparql.vivo.ufl.edu/sparql.html)

- Retrieve all organizations to sub-organizations table by using the SPARQL query in the `OrganizationToSubOrganizationSPARQL.txt` file

- Save and rename the output file as `OrganizationToSubOrganization.csv` (The pre-queried result is available in the data package)

- Use the SPARQL query in the `UFLOrganizationsSPARQL.txt` file (showed as below) to retrieve the sub-organizations (descendants) of the University of Florida (UFL)

- Save and rename the output file as `UFLOrganizations.csv` (The pre-queried result is available in the data package)
Cleaning and Preprocessing

- Extract rows from the `OrganizationToSubOrganization.csv` where the organizationLabel appears in the UFLOrganizationLabel column of the `UFLOrganizations.csv`. You can do this by using Excel, Python, etc.

- Save the new file as `UFLOrganizationHierarchy.csv`. (The pre-processed result is available in the data package)

- Run Science of Science tool (Sci²)

- Select File > Load menu to load the `UFLOrgHierarchy.csv`

---

Cleaning and Preprocessing

- Extract the directed network by selecting Data Preparation > Extract directed Network. A window will pop up.

- Fill in the following parameters and click OK. A directed network file will appear in the Data Manager panel.

- Select the Data Preparation > Detect Duplicate Nodes menu and execute with default parameters
Cleaning and Preprocessing

- Select both Merge table and the directed network file in the Data Manager panel.

- Execute the Data Preparation > Update Network by Merging Nodes menu to remove the duplicated nodes.

![Image of Merge Table and Directed Network File]

Cleaning and Preprocessing

- Node’s indegree vs outdegree
  - Indegress: Number of edges pointing in
  - Outdegree: Number of edges pointing out

- Generate the in-degree attributes by using the “Updated Network” file and executing Analysis > Networks > Unweighted and Directed > Node Indegree.

- This will result a “Network with indegree attribute …” file.

- Generate the out-degree attributes by using the “Network with indegree attribute …” file and executing Analysis > Networks > Unweighted and Directed > Node Outdegree.

- This will result a “Network with outdegree attribute …” file.
Visualizing: GUESS

- Visualize the network by using the “Network with outdegree attribute …” file and execute the Visualization > Networks > GUESS

- A GUESS window will pop up as showed in the following image. Please pay attention to the Graph Modifier tab and the Layout menu bar. We will use this functionality to modified the visualization.

Visualizing: Layout

- Re-layout the nodes by executing Layout > GEM / SPRING / Pin Back or any combination of these layouts for your satisfaction.

- Here is the result by laying out with SPRING and then Pin Back.
Visualizing: Color

- Now is time to bring information up front by using the functionality in the Graph Modifier tab.

- Color the node that have more than 10 sub-organizations. Using the following setting and click Colour button. Choose a color from the color box.

- You can also choose to show the labels

Visualizing: Resize

- Resize the nodes linearly based on the number of direct sub-organizations. To achieve this, click on the Resize Linear button and fill in the parameters as following.
Visualizing - Highlight UFL

- Since UFL is the root of the hierarchy, it will not have a parent node. In other words, the indegree value of UFL is zero.

- Apply the condition to select the UFL node. Then resize and recolor the UFL node to increase its visibility. Finally, show the UFL node’s label.

Visualizing: Result

- Here is the visualization result for the UFL organization hierarchy.
Analysis & Visualization of VIVO Data

- Three-Step Visualization
- Sci²: Organization Hierarchy Visualization
- Gephi: Organization Hierarchy Visualization
- 1.5-Step Visualization

Preprocessing

- Since this section also visualizing the organization hierarchy, we will avoid re-doing the Playing with SPARQL query; and Cleaning and Preprocessing Data steps.

- However, we have to export the “Network with outdegree attribute …” file as graphml. To do this, right click on the “Network with outdegree attribute …” file from the Data Manager and select Save.

- Choose GraphML as output type and save as UFLHierarchyNetwork.xml. Change the .xml extension to .graphml.
Visualizing: Load Data

- Run the Gephi tool and load the `UFLHierarchyNetwork.graphml` through the File > open menu.

- There are three main tabs in Gephi: i) **Overview** tab for editing the visualization effect; ii) **Data Laboratory** tab for manipulating data; iii) **Preview** tab for polishing up the final image.

Visualizing: Layout

- Choose a layout from the **Layout** tab and press run. You might need to stop the run once you are satisfied with the layout. Here is the result of Fruchterman Reingold layout.

- You can pan and zoom in/out from the view by using the mouse right-click and scrolling.
Visualizing: Color and Resize

- Select the **Ranking** sub-tab under the **Overview** tab.

![Diagram showing color and resize settings](image)

**Visualizing: Color and Resize**

- Choose the **Nodes** tab as the edited target.

- The squared green box is the target’s property icons: circle (node color), diamond (node size), alphabet ‘A’ with circle (label color), alphabet ‘A’ with diamond (label size). To edit the node color, select the circle icon.

- Select **OutDegree** from the drop down box.

- Double click on the right slider of the color bar to bring up the color chooser.

- Adjust the color coding by using the Spline Editor.

- Click the **Apply** button to implement the changes.

- Repeat the above steps to edit the node size, label color and label size. Show the labels by clicking the black ‘T’ icon at the bottom of the **Graph** tab.
Visualizing - Highlight UFL

- Now, highlight University of Florida (UFL) in the organization hierarchy.

- To do this, select the UFL node in the Data Laboratory tab. Adjust the size and color properties of the node to increase its visibility.

Visualizing: Preview and Polish

- Select the Preview tab and click on the Refresh button to show the result.

- Use the Presets drop down box to change the node and edge view.

- You also can enable / disable the nodes and edges, even edit the properties through the Preview Settings tab.

- Then press the Refresh button again to show the new result.

- Continue editing until you are satisfied with the look and feel.

- Gephi supports export formats as PDF, SVG, CSV, gdf, graphml, and gexf.
Analysis & Visualization of VIVO Data

- Three-Step Visualization
- Sci²: Organization Hierarchy Visualization
- Gephi: Organization Hierarchy Visualization
- 1.5-Step Visualization
1.5-Step Visualization

- It is possible to reduce the three-steps visualization to 1.5 steps

VIVO does it for you!

Data in CSV, graphml, etc

---

10 min break
Visualization-in-depth

- Map of Science
- Temporal Graph
- National Researcher Networking (NRN)

Concept

- A visual inter-discipline interface for the VIVO publications analysis
- Activity of any organization, person, or university i.e. entity in a VIVO instance in the world of science
- 13 Disciplines and 554 Sub-Disciplines
**Features + Interactions**

- Shows the publication activity of an entity in a VIVO instance, overlaid on the map of science

- Table lists all 13 disciplines and 554 sub-disciplines i.e. science field
  - Hover over a line so that corresponding field’s location on map is outlined
    - Color of row based on corresponding discipline’s color
  - Display impact of entity in terms of # of pubs in a particular science field
  - Sort, Search and Download data used to render the table, as CSV

- Size of node on the map indicates % of activity in that field
  - An entity has fixed node area to distribute across the displayed nodes
  - Toggle discipline label
  - Slider used to control how many fields to show on map

- How many publications were properly mapped?

---

**Data Access**

**Client**
- After the initial request, downloaded code on client-side requests just the raw data having “entity to sub-discipline to count” dictionary

**Server**
- Checks cache for information for subject entity
  - If found, processes it and sends the dictionary in JSON format to the front-end
  - Else, it requests information from the main graph and builds the cache
  - Also send the JSONified dictionary to the front-end

**Client**
- Parses the JSON data
  - Derive dictionary for “discipline to count”
  - Populates the table
  - Renders the map of science
Code Libraries Used

- Google Maps API v3 – JavaScript mapping library\(^1\)
  - Provides a number of utilities for manipulating maps,
  - Adding content to the map through a variety of services, allowing creation of robust maps applications

- DataTables – JavaScript tabulating library\(^2\) for jQuery
  - Client-side library for nicely tabulating data with pagination
  - Multi-column sorting with data type detection
  - Instant filtering of rows

- jQuery - fast and concise JavaScript library\(^3\)
  - simplifies HTML document traversing, event handling, animating, and AJAX interactions
  - provides uniform behavior on all major browsers

---

Visualization-in-depth

- Map of Science
- Temporal Graph
- National Researcher Networking (NRN)

\(^1\) [http://code.google.com/apis/maps/documentation/javascript/](http://code.google.com/apis/maps/documentation/javascript/)
\(^2\) [http://www.datatables.net/](http://www.datatables.net/)
\(^3\) [http://jquery.com/](http://jquery.com/)
Concept

- Compare research activity in terms of publications and grants amongst peers
- Identify most fruitful blocks of time
- Create report for policy decision makers

![Graph showing Comparing Grants of Organizations & People in University of Florida](image)

Features + Interactions

- Display line graph of publication and grant activity across years
  - Switch between grant and publication parameter
  - Preserve selections between parameters
- List all sub-entities (organization and people)
  - Drill-down to get the activity comparison of all sub-entities for a particular sub-entity i.e. sub-sub-entities
    - E.g. Compare activity for all the sub-entities for College of Medicine
    - Similarly drill-up from a sub-entity
- Search for entity in the table on name, count and type
  - Pagination
- Download data used to render the table, as CSV
Code Libraries Used

- Flot – JavaScript plotting library\(^1\) for jQuery
  - Produces graphical plots of arbitrary datasets on-the-fly client-side.
  - Works in modern browsers including IE and on iOS platforms.

- DataTables - JavaScript tabulating library\(^2\) for jQuery
  - Client-side library for nicely tabulating data with pagination
  - Multi-column sorting with data type detection
  - Instant filtering of rows

- jQuery - fast and concise JavaScript library\(^3\)
  - simplifies HTML document traversing, event handling, animating, and AJAX interactions
  - provides uniform behavior on all major browsers.


Visualization-in-depth

- Temporal Graph
- Map of Science
- National Researcher Networking (NRN)
Concept

- Provide an overview of the adoption of National Researcher Network systems across the U.S. and also internationally.
- Show the evolution of the NRN systems through time since Jan, 2010.
- Support of scientific discoveries, technological breakthroughs, and the communication of research results to diverse stakeholders.
- Encourage future adoption and usage of NRN systems.

National Researcher Networking (NRN)

[Image of NRN systems]

http://nrn.cns.iu.edu
Features + Interactions

- Display the institution’s resources on the map
  - The resources are People, Publications, Patents, Funding, Courses
  - Turn on / off resources through the check boxes
  - Markers are area-size-coded by count

- Show the evolution of national researcher networks
  - Play – Playback month by month since January 2010
  - Pause – Pause the playback
  - Stop – Clear the playback and show latest resources status

- Browsing resources through the NRN sites
  - Navigate to the resource’s site by clicking on the resource’s marker
  - Hide the map for a full screen view of the resource’s site

Code Libraries Used

- Science of Science tool (Sci²)
  - An desktop tool for scientific analysis
  - Contains Yahoo! Geocoding plugin for retrieving geo-location information of the institutions

- Google Map API V3
  - A Javascript Maps Applications for both the Desktop and Mobile Devices
  - A free service, available for any web site that is free to consumers

- jQuery - fast and concise JavaScript library³
  - simplifies HTML document traversing, event handling, animating, and AJAX interactions
  - provides uniform behavior on all major browsers.

VIVO Visualization

- Accessing VIVO Data
- Architecture
- Pipeline Explanation

Accessing VIVO Data

- Read “The Semantic Web: An Introduction”
  http://infomesh.net/2001/swintro
Linked Open Data (via RDF or N3)

- Accessible to anyone on the Web.
- It can be a bit tedious to work with large amounts of data quickly/easily.

- N3 example:
  - http://vivo.iu.edu/individual/person25557/person25557.n3

- RDF example:
  - http://vivo.iu.edu/individual/person25557/person25557.rdf

SPARQL Endpoints

- Working with data is easier/faster (using SPARQL queries).

- But may not be accessible to everyone.
VIVO Visualization

- Accessing VIVO Data
- Architecture
- Pipeline Explanation

VIVO Visualization Architecture

1. User requests the visualization
2. Request is received by the VIVO application
3. Specific controller gets control
4. Controller delegates the control of flow to the handler of the requested visualization, if permission validated
5. The handler passes request information to the Query Handler
6. Query Handler queries the semantic web data store (cached or live)
7. Results of the query are converted into Java objects
8. Java objects are used to generate response in the requested format
9. Request handler renders the generated response
VIVO Visualization

• Accessing VIVO Data
• Architecture
• Pipeline Explanation

Map of Science Visualization Pipeline

- Breakdown of serving the map of science visualization request received at, http://vivo.iu.edu/vis/map-of-science/IndianaUniversity

- Short URL Request
- It has following parameters,
  - /vis – Short URL prefix
  - /map-of-science – Visualization type
  - /IndianaUniversity – URI of subject of the visualization
- Long form looks like
  http://vivo.iu.edu/visualization
  ?vis=map-of-science
  &uri=http://vivo.iu.edu/individual/IndianaUniversity
Map of Science Visualization Pipeline

- `/vis` – Short URL Visualization Controller assumes the control of flow
- Parse URL to get visualization type – Map Of Science
- Gets permission requirements for Map of Science visualization
- Validates it against the requesting user’s privileges

- After validation captured parameters are passed to the Map of Science Visualization
- Check cache for all models pertaining to “IndianaUniversity” entity
- If not present create models and store in cache
Map of Science Visualization Pipeline

- Models used are,

  - ORGANIZATION_MODEL_WITH_TYPES
    - no input uri
    - all sub-organizations recursively
  - ORGANIZATION_TO_PUBLICATIONS_FOR_SUBORGANIZATIONS
    - specific for an input uri
    - all publications for entity-associated people
  - ORGANIZATION_ASSOCIATED_PEOPLE_MODEL_WITH_TYPES
    - specific for an input uri
    - e.g. President of University
  - PEOPLE_TO_PUBLICATIONS
    - no input uri
    - all people associated with publication

- SPARQL query fired against the previous mentioned cache
- Using the query results create the java objects
  - Entity (for subject entity)
  - SubEntity (for child entities)
  - Activity (for publication info)
Map of Science Visualization Pipeline

- Java objects used to generate output
- FreeMarker object
  - HTML markup including CSS, JavaScript
- JSON
  - Used by JavaScript to render tables, map etc

Output from visualization (FreeMarker + JSON) sent to Response Handler
- Makes sure output formatted properly
  - E.g. For FreeMarker – compile HTML markup
15 min break
Hands-On (Chin Hua Kong)
• Analysis & Visualization of VIVO Data

Guided Tour (Chintan Tank)
• Create new VIVO visualization

Create a new VIVO Visualization

- Identify the data to be represented
- Research visualizations to represent the data
- Create SPARQL Queries to get the data
- Create Back-end module for a new visualization
- Create Front-end module that renders the data created by the back-end
- Wire together the pieces to get a complete working pipeline
Playing with SPARQL

- Get titles of all publications in the system for a person

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX bibo: <http://purl.org/ontology/bibo/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX core: <http://vivoweb.org/ontology/core#>

SELECT (str(?publication) as ?publicationURILit) (str(?publicationTitle) as ?publicationTitleLit)
WHERE {
  ?publication rdf:type bibo:AcademicArticle .
  ?publication rdfs:label ?publicationTitle .
}
```

Playing with SPARQL

- What are the results?

<table>
<thead>
<tr>
<th>Publication URI</th>
<th>Publication Title Label</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://vivo.ufl.edu/individual/n282">http://vivo.ufl.edu/individual/n282</a></td>
<td>Seismic evidence for small-scale...</td>
</tr>
<tr>
<td><a href="http://vivo.ufl.edu/individual/n1597">http://vivo.ufl.edu/individual/n1597</a></td>
<td>Small scale lateral shear velocity...</td>
</tr>
<tr>
<td><a href="http://vivo.ufl.edu/individual/n7858">http://vivo.ufl.edu/individual/n7858</a></td>
<td>Coexisting shear- and compressi...</td>
</tr>
</tbody>
</table>

...
Code Structure

- For visualization development
  - Front-End Module
    - `$VIVO_ROOT/productMods ($FRONT_END)`
      - `$FRONT_END/templates` - FreeMarker templates
      - `$FRONT_END/css` - Stylesheets
      - `$FRONT_END/js` - JavaScripts
      - `$FRONT_END/images` - Images
  - Back-End Module
    - `$VIVO_ROOT/src ($BACK_END)`
      - `$BACK_END/..//webapp/controller/visualization` - Controller
      - `$BACK_END/..//webapp/visualization` - Specific visualizations
Front-End Module

- Create folders for “Word Cloud” in templates, css, js folders
  - Copy js, css libraries in to newly created wordcloud folder

- Put code from the standalone page into a newly created ftl file
  - Change parts of code specific to integration with the back-end like inclusion of scripts, import of json data

- Difference between the standalone html page and FreeMarker file
  - Inclusion of external scripts (stylesheets & JavaScripts)
  - Access to back-end provided variables

- How does the magic happen?

Code Structure

- Back-End Module
Back-End Module

- Create package for “Word Cloud” along with other visualizations

- Query Runner
  - SPARQL Query
  - Java Objects

- Visualization Request Handler
  - Gets input from controller
  - Requests Query Runner to get data
  - Prepares data
    - FreeMarker config
    - JSON output

- Dependency Injection

Publication Title Word Cloud

VIVO Visualization Workshop

Publication's Title Word Cloud for Frazer, Tom K

QCLOUD vis is used. 103 total publications were considered.
Outlook (Katy Borner)

• Create new VIVO visualization
• Planned VIVO Visualizations
  • Empower Others to Visualize VIVO Data
  • Map of Science Comparison Visualization
  • Make VIVO Visualizations Useful for Other NRN, e.g., as Open Social gadgets
• (Inter)National VIVO Visualizations

Empower Others to Visualize VIVO Data:
Provide VIVO data ready for analysis, e.g.,

1. **Organizational hierarchy:** from root note (e.g., UFL) to persons.
2. **Bimodal network of people-organizations:**
   *Organization Name | Person Name(s) separated by “|”*
   WORKFLOW: Extract bimodal network, calculate indegree/outdegree, vis in GUESS with diff colors for diff node types and nodes area coded by degree, label high degree nodes.

3. **Bimodal network of people-funding awards for any organization node (e.g., SLIS or all of IU):**
   *Funding Award Title | $ amount | Person Name(s) separated by “|”*
   WORKFLOW: Extract bimodal network, calculate indegree/outdegree, vis in GUESS with diff colors for diff node types and nodes area coded by total award amount for funding or degree, label high degree/$ nodes.
4. **Bimodal network of people-publications for any organization node (e.g., SLIS or all of IU):**

   - **Paper Title | times cited | Person Name(s) separated by “|”**

   WORKFLOW: Extract bimodal network, calculate indegree/outdegree, vis in GUESS with diff colors for diff node types and nodes area coded by times cited for papers or degree, label high degree/#citations nodes.

5. **Bimodal network of people-courses for any organization node (e.g., SLIS or all of IU):**

   - **Course title | Person Name(s) separated by “|”**

   WORKFLOW: Extract bimodal network, calculate indegree/outdegree, vis in GUESS with diff colors for diff node types and nodes area coded by degree, label high degree nodes.

---

http://sci2.wiki.cns.iu.edu/display/SCI2TUTORIAL/8.2+Network+Analysis+and+Other+Tools has a table with 22 network layout algorithms.
Making VIVO Visualizations Useful for Other NRN, e.g., as Open Social gadgets

OpenSocial (http://opensocial.org) defines a common API for social applications across multiple websites. With standard JavaScript and HTML, developers can create apps that access a social network’s friends or update feeds.

We plan to make VIVO visualization available as Open Social Gadget for use in VIVO, Harvard Profiles, SciVerse.

- Thanks to Duke U, VIVO will soon have an Open Social Container
- Thanks to USF, Harvard Profiles has an Open Social Container.
- Elsevier’s SciVerse has a Open Social Container.
From Local to (Inter)National
See also National Search

VIVO Documentation, Code, and Data

Documentation and Code:
- VIVO Web Site: http://vivoweb.org
- VIVO Support: http://vivoweb.org/support
- VIVO Ontology: http://vivoweb.org/download#ontology
- Sourceforge for source code: http://sourceforge.net/projects/vivo

Workshop materials and slides
- http://wiki.cns.iu.edu/display/PRES/VIVO+Presentation
All papers, maps, tools, talks, press are linked from http://cns.iu.edu

CNS Facebook: http://www.facebook.com/cnscenter
Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience