Open Data, Open Source NRN Systems for Global Science Decision-Making: Let’s Compete!

Katy Börner, Cyberinfrastructure for Network Science Center
School of Informatics and Computing
Indiana University, USA

VIVO Conference, Austin, TX
August 8, 2014

Visit Mapping Science exhibit at http://scimaps.org to see 100 S&T maps.
LinkedIn: World's Largest Professional Network
https://www.linkedin.com

ResearchGate GmbH
https://www.researchgate.net
Launched in May 2008, has a user base of 2.3 million scientists worldwide in Aug. 2013
Academia.edu—Share Research

http://www.academia.edu

Launched in Nov 2010.

Exemplary Expert Networking Service

Google Scholar

http://scholar.google.com

Launched in Nov. 2011. Has “unknown” number of profiles.

Exemplary Expert Networking Service
# Expert Networking Services—60 More

This table provides general information for each tool, name, developing institution, external links to information, whether the code is *Open Source* and known adopters of the software.

<table>
<thead>
<tr>
<th>Research Networking Tool</th>
<th>Link to Product Page</th>
<th>Developer/Owner</th>
<th>Open Source</th>
<th>Adopters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative Partnership Profile System</td>
<td>Collaborative Partnership <a href="http://en.wikipedia.org/wiki/Comparison_of_Research_Networking_Tools_and_Research_Profiling_Systems">🔗</a> and Research Excellence at UT Arlington <a href="http://en.wikipedia.org/wiki/Comparison_of_Research_Networking_Tools_and_Research_Profiling_Systems">🔗</a></td>
<td>University of Texas at Arlington</td>
<td>Yes</td>
<td>UT Arlington, UT Pan American, University of North Texas, UT Health Science Center, UT El Paso, UT San Antonio, UT Tyler, UT Health Science Center, University of North Texas, UT Dallas, UT Health Science Center at Tyler, Texas Christian University, (plans to add Gulf Coast Consortia, Rice University, Baylor College of Medicine)</td>
</tr>
</tbody>
</table>

Let’s Compete!
Let’s Compete for
The best system/functionality that truly addresses the insights **needs decision makers really have**:

- **Foresight + Context**: What areas are emerging, merging, dead? Where will I achieve highest return on investment of time, money, compassion?
- **LOCAL Strategic Decision Making**: What is the unique value proposition of my institution/region? What strengths and gaps exist and how can we overcome challenges and embrace opportunities?
- **GLOBAL Networks and Insight**: Who are the leading experts, institutions, regions? How are they connected and what R&D do they perform?


My guess?
**Winners will be systems that create/use de-facto standards for**

- Easy to map/crosswalk and cross-search data structures
- Plug-and-play data cleaning, pre-processing, analysis modules
- Easy to customize and use user interfaces (incl. visualizations, data exports)

**Likely, winning systems will**

- Promote (partially) open data and open code
- Easy harvesting and ingest of major publication datasets, e.g., MEDLINE, Elsevier, Reuters, SciELO, others.
- Inter-platform compatibility—VIVO, Profiles, SciVal Experts, Loki.
- Be part of federated search tools, Direct2Experts, CTSAconnect, and SciVal Community, etc.
- Support/use DOIs, author identifiers, e.g., ORCID, SciENcv

**Do you want to level up to the next challenge?**
Let’s Start With Search

Find experts
• At *my* institution
• In the U.S.
• On Earth
That work on topic “X.”

Let’s try this for ‘microbiome’ research.

Example: Search for ‘Microbiome’
Using existing systems to identify expertise on ‘Microbiome’

- CTSA DIRECT2Experts a pilot national network to demonstrate interoperability among research-networking platforms: http://direct2experts.org
- NSF Award Search: http://www.nsf.gov/awardsearch

To support
- Foresight + Context
- LOCAL Strategic Decision Making
- GLOBAL Networks and Insight
Inter-Institutional Collaboration Explorer

This visualization shows information about “collaborative publications” found at 2 or more Researcher Networking websites.

The idea that institutions don’t work together and that biomedical research is conducted in silos is not true. Researchers, even when separated by great distances, are in fact willing to work together, and this visualization demonstrates that they often do.

Contact: Nick Benik (nbenik@gmail.com), Harvard Medical School, Boston, MA. URL: http://scite.hackerceo.org/VIVOviz
Inter-Institutional Collaboration Explorer
The outer solid colored arcs represent the 11 institutions. The size of the arc is proportional to the number of collaborative publications found on the site. The inner colored bands represent the number of collaborative publications found between the two institutions that each band connects. Clicking an institution's arc will hide any bands not connected to that institution and will display a timeline of when that institution's collaborative publications were written.
69 NSF projects found.
They were awarded to 53 unique institutions.

<table>
<thead>
<tr>
<th>Institution</th>
<th>Awards</th>
<th>Sum of Awarded AmountToDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Illinois at Urbana-Champaign</td>
<td>3</td>
<td>3,263,277</td>
</tr>
<tr>
<td>University of Texas at Austin</td>
<td>3</td>
<td>2,496,169</td>
</tr>
<tr>
<td>University of Wisconsin-Madison</td>
<td>3</td>
<td>2,143,438</td>
</tr>
<tr>
<td>Cornell University</td>
<td>2</td>
<td>2,000,000</td>
</tr>
<tr>
<td>Duke University</td>
<td>2</td>
<td>1,998,143</td>
</tr>
<tr>
<td>Gordon Research Conferences</td>
<td>2</td>
<td>1,758,524</td>
</tr>
<tr>
<td>Marine Biological Laboratory</td>
<td>2</td>
<td>1,734,624</td>
</tr>
<tr>
<td>Michigan State University</td>
<td>2</td>
<td>1,700,000</td>
</tr>
<tr>
<td>SUNY at Stony Brook</td>
<td>2</td>
<td>1,508,535</td>
</tr>
<tr>
<td>University of Delaware</td>
<td>2</td>
<td>1,342,609</td>
</tr>
<tr>
<td>University of Maryland at Baltimore</td>
<td>2</td>
<td>1,019,366</td>
</tr>
<tr>
<td>University of Nebraska-Lincoln</td>
<td>2</td>
<td>997,214</td>
</tr>
<tr>
<td>Florida Agricultural and Mechanical University</td>
<td></td>
<td>950,000</td>
</tr>
<tr>
<td>Washington State University</td>
<td></td>
<td>900,000</td>
</tr>
<tr>
<td>Bigelow Laboratory for Ocean Sciences</td>
<td></td>
<td>842,000</td>
</tr>
<tr>
<td>University of Hawaii</td>
<td></td>
<td>822,919</td>
</tr>
<tr>
<td>TERC Inc</td>
<td></td>
<td>779,220</td>
</tr>
<tr>
<td>Michigan State University</td>
<td></td>
<td>741,685</td>
</tr>
<tr>
<td>University of Nebraska-Lincoln</td>
<td></td>
<td>726,806</td>
</tr>
<tr>
<td>Marine Biological Laboratory</td>
<td></td>
<td>723,528</td>
</tr>
<tr>
<td>University of Miami School of Medicine</td>
<td></td>
<td>704,365</td>
</tr>
</tbody>
</table>

69 NSF projects found. They were awarded to 53 unique institutions.
537 NIH awards found on 7/19/2014.
544 NIH awards found on 7/20/2014. Omitted 34 records with missing values.
U.S map and world map of #projects per geolocation.

1,406 projects with 4,326 participating countries (96 unique) found
Co-authorship network for the department of Information Systems
Source: Melbourne Research Windows. Contact Simon Porter simon.porter@unimelb.edu.au

Bimodal network of search terms and researchers extracted from research profile search results to show the University’s capability in Disaster Management to the Government
Contact: simon.porter@unimelb.edu.au

Top MeSH Disease Concepts Appearing in PubMed Publications by the University of Michigan Medical School. Links connect concepts where 100+ authors published about both concepts within the span of their careers.

Contact: Jeffrey Horon, J.Horon@elsevier.com
Top MeSH Disease Concepts Appearing in PubMed Publications by the University of Michigan Medical School. Links connect concepts where 100+ authors published about both concepts within the span of their careers.

This visualization revealed that animal disease models were central to disease research at U-M which encouraged additional thought and attention to animal husbandry, animal expenses, and core/shared services overall.

Contact: Jeffrey Horon, J.Horon@elsevier.com

P30 Member Collaborations – Sponsored Project Co-Participation and Co-Authorship Network. Used in successful P30 funding application. Shows the PI’s relationships with various P30 members, conveying that the PI was not only the formal center of the group but also the informal center and the person who exhibited the highest betweenness centrality. Contact: Jeffrey Horon, J.Horon@elsevier.com
Please register your institution’s researcher networking system at http://nrrn.cns.iu.edu
Next: NRN Competition Design

Let’s Learn From TREC
The Text REtrieval Conference (TREC), co-sponsored by the National Institute of Standards and Technology (NIST) and U.S. Department of Defense. TREC has the following goals:
• to encourage research in information retrieval based on large test collections;
• to increase communication among industry, academia, and government by creating an open forum for the exchange of research ideas;
• to speed the transfer of technology from research labs into commercial products by demonstrating substantial improvements in retrieval methodologies on real-world problems; and
• to increase the availability of appropriate evaluation techniques for use by industry and academia, including development of new evaluation techniques more applicable to current systems.
Let’s Learn From TREC

TREC is overseen by a program committee consisting of representatives from government, industry, and academia. For each TREC, NIST provides a test set of documents and questions. Participants run their own retrieval systems on the data, and return to NIST a list of the retrieved top-ranked documents.

NIST pools the individual results, judges the retrieved documents for correctness, and evaluates the results.

The TREC cycle ends with a workshop that is a forum for participants to share their experiences.

http://trec.nist.gov

Feel free to let me know if you are interested to be involved in designing a TREC-like competition for NRN systems!