Linking International Databases to Build Strategic Academic Partnerships in Science Workshop

Meeting Summary

May 14-15, 2014

Co-organized by Indiana University and the U.S. Department of State
Co-hosted by the National Science Foundation (NSF) and the American Association for the Advancement of Science (AAAS)

This brainstorming workshop focused on catalyzing the development of a global knowledge platform designed to help universities build strategic international science, technology and innovation (STI) partnerships. Initial presentations covered the U.S. need for strategic global engagement and partnerships and the lack of readily available data and tools for U.S. and foreign universities; this mutual lack of information leads to a largely unfilled worldwide matrix of strong potential academic matches, with lost opportunities for U.S. scholars and institutions, the nation, and the world.

The participants explored the socio-technical feasibility of such a platform, now called PEGASCIS (the Platform for Enhancing Global Academic Strategic Collaboration in Science). The term “platform” was used in a broad sense to include a community of stakeholders, the organizational entity of the platform itself, a wide range of IT programs including matchmaking programs, a clearinghouse of relevant information to aid strategic decision-making, search and visual analytics functionality, and other functions that stakeholders might develop. The desire to sustain community ownership of inclusive and widely available data and tools led to preference for open source and open data options wherever possible.  

Social dimensions:

The participants concluded that social issues around platform establishment were at least as challenging as technical issues. Social challenges include:

- both the institutional and national need to respond to major changes in global landscape of higher education and STI;
- the need to establish trust and value to allow cooperation in the highly competitive U.S. academic arena;
- the strong tradition of independence among faculty and institutions that now makes a framework for unique (i.e., disambiguated) identifiers difficult;
- concerns about faculty effort required for, privacy issues with, and funding for platform participation;
- a weak tradition of using data for strategic academic decision-making;
- the need for short-term and long-term incentives/benefits for platform adoption; and

More information about the workshop can be found at: [http://cns.iu.edu/workshops/event/140514.html](http://cns.iu.edu/workshops/event/140514.html)

Prepared by Elizabeth Lyons with input from participants. The Meeting Summary should not be assumed to represent the views of the U.S. Department of State or the National Science Foundation. It is solely intended as an informal record of individual views expressed by the participants at the workshop.

The participants also benefited from a brief presentation by Sara Farley, Chief Operating Officer of The Global Knowledge Initiative (GKI), about GKI’s experience with hands-on fostering international partnerships in developing countries.
• the difficulty of sustaining focus and action in a huge “orphan space” where the leadership and funding roles of government/business/universities/NGOs in international academic science are unclear.

Technical dimensions:

The participants focused primarily on the matchmaking aspects of the platform, rather than on its contextual knowledge-sharing function (e.g., for lessons learned, country-specific STI priorities, funding for international STI collaboration\(^4\), or STI foresight\(^5\)). Presentations highlighted the current capabilities of National Researcher Networking (NRN) systems that support online faculty profiles (e.g., Scholars@Duke, an implementation of VIVO\(^6\)), provide value-added services (e.g., printing a CV formatted for grant submission or compiling and visualizing an institution’s research portfolio (e.g., Weill Cornell Medical College's map of science\(^7\) derived from its research publications), or aggregate information on international collaborations (e.g., using UCosmic\(^8\), moveon, or the NIH World RePORT).

Representatives from several NRN systems described the growing number of compatible information systems that use the VIVO set of relational definitions (e.g., its ontology) to expose their data and can now be used to run federated searches (e.g., to find a potential collaborator or institution with research complementarity using direct2experts or CTSAssearch) and visualize science activity.\(^9\)

Technical challenges include

• uncertain interoperability across proliferating programs and platforms;
• the need for expanded ontologies to cover all academic endeavors and international activities;
• the complexity of capturing university activities beyond publications and research grants;
• adding an international dimension to domestic efforts and/or harmonizing with efforts of other countries; and
• meshing with international development efforts so that the platform does not enlarge the digital divide between developed and developing countries.

\(^4\) For example, as presented by CRDF’s Newton’s List.
\(^6\) VIVO is an open source semantic web search and visualization application originally developed and implemented at Cornell. There are now many diverse activities associated with the VIVO project, across federal agencies, academic institutions, professional societies, for-profit publishers, and data providers in the U.S. and in other countries, as well as a variety of efforts with the semantic web and ontology development communities.
\(^7\) Weill Cornell Medical College’s map of science.
\(^8\) UCosmic, a consortium of over 20 institutions in more than a dozen countries, is an international open-source software initiative to enable comprehensive mapping of a wide range of global activities taking place across an institutional enterprise, enabling a university to leverage assets and achieve internationalization objectives.
\(^9\) These same NRN systems also now integrate with ORCID, which assigns unique identifiers to investigators. This will become increasingly important as publishers and funding agencies adopt ORCID and so reduce identifier ambiguity.
An issue common to all of these challenges is the need for agreed-upon standards across platforms and institutions, which can enable mutual cooperation without requiring top-down coordination.

Outcomes:
The workshop generated a number of specific outcomes, both in terms of easily achievable projects and general approaches:

1. Framework for a university profile system:
   In the absence of any global portal, institutional matchmaking between the over 4,000 U.S. institutions of higher education and the diverse array of foreign universities is daunting. The group delineated the basics of a university profile system that would use publicly available information to help both U.S. and foreign universities identify potential STI partners. Such a profile, with faceted (i.e., with user-chosen categories or filters) search and visualization capacity, could include
   - information about the institution (e.g., its Carnegie classification, degrees granted);
   - its research capacity (e.g., blended from NSF, NIH grants from federal databases, with publications from NRNs, and available at individual, department and/or school level);
   - its teaching (e.g., on-campus curriculum, MOOCs offered);
   - its innovation ecosystem (e.g., patents and research parks);
   - its medical capacities if applicable (e.g., hospitals, patient volumes); and
   - its honors (e.g., graduate program and international rankings, National Academies members, and Nobel Laureates).

   Equally important could be information about
   - languages taught (from on-line course catalogs);
   - number of outgoing and incoming Fulbright scholars\(^{10}\);
   - number and location of International Branch Campuses\(^ {11}\);
   - number and location of international development assistance projects\(^ {12}\);
   - number of foreign students and study abroad students\(^ {13}\);
   - percentage of graduates who joined the Peace Corps\(^ {14}\);
   - presence of an Engineering without Borders\(^ {15}\) chapter on campus; and
   - Department of Education International\(^ {16}\) awards such as Area Grants.

   Universities that use UCosmic or a similar system could also display country-specific information on

\(^{10}\) For example, Fulbright provides a list of top-producing institutions for Fulbright awards.

\(^{11}\) For example, from SUNY at Albany’s Cross-Border Education Research Team database.

\(^{12}\) For example, the USAID-funded AidData project can provide information on academic participants for some donors/projects.

\(^{13}\) For example, from the Institute of International Education’s Open Doors publications.

\(^{14}\) For example, Peace Corps announces top-producing institutions for Peace Corps volunteers.

\(^{15}\) For example, from Engineers without Borders website.

\(^{16}\) For example, the Department of Education website describes awards for international and foreign language education.
- alumni living abroad;
- foreign-born faculty;
- degree institution of faculty;
- international research or teaching projects;
- existing agreement with foreign universities; and
- international business alliances.

2. **Development of platform use scenarios:**
Workshop participants articulated basic scenarios for individual researchers, departments, institutions, and state university systems which might use the platform. In general, NRN tools help students find mentors, conference organizers find speakers, and patients find physicians. They help faculty promote their work, disclose conflicts of interest, advertise open positions in their labs, and manage their publication lists. They help institutions address challenging problems such as diversity in the workforce and how to allocate resources. They help funding agencies link research inputs to outputs and explain how investment in research benefits the public. PEGASCIS could help to add an international and strategic dimension to such use scenarios, for example,
- by allowing an institution to add a geographic focus to VIVO and so visualize in what countries its faculty had expertise, research and/or teaching;
- by using summaries of the scope of institutional activity in one location (and discipline) within a weighted system to compare themselves with others operating in the same geography (and discipline);
- by using national and international foresight reports to identify fields of high potential impact, and then using international federated searches and visualization tools to explore/match their own and foreign institutions’ research profiles to find synergistic possibilities.

3. **Discovery of obvious synergies and areas of easy progress:**
Workshop participants identified several opportunities for synergy, for example,
- developing an ontology for international activities and linking this ontology in UCosmic with the VIVO ontology so that it can be cross searched together with the other NRN systems at [http://nrn.cns.iu.edu](http://nrn.cns.iu.edu);
- adding country search terms and visualization to federated searches (e.g., in CTSAsearch, Harvard Profiles, Academic Analytics); and
- linking ORCID data with NRN systems (e.g., Harvard Profiles) in order to increase the networking analysis and visualization capacity of ORCID, thus helping it become a powerful global portal for finding expertise worldwide.

4. **Identification of entrepreneurial approaches to jumpstarting the platform:**
Among the approaches recommended were
- rapid prototyping of different programs;
- TREC-like coding competitions\(^\text{17}\);

\(^{17}\)TREC is the NIST-managed Text Retrieval Conference, a major driver in innovation in text retrieval systems.
• challenge grants for tools and solutions;
• use of collaborative software development approaches (e.g., student coding projects, GitHub, OuterCurve Foundation projects);
• development of multiple smaller modules in parallel; and
• engagement of highly innovative organizations and universities for faster progress.

5. Enumeration of additional benefits that the platform could deliver:
Participants expanded on the utility of a global knowledge commons that could
• facilitate matchmaking across many institutional strengths among domestic and international universities;
• strengthen metrics for analyzing science and international collaboration;
• better showcase institutional and country strengths (a key benefit to many U.S. institutions that are not well known internationally),
• enrich academic profiles for a wider range of academic endeavors;
• link contextual information (e.g., best practices, foresight, country profiles) with analytical tools;
• bolster community ownership of data, network and tools;
• diversify the range of institutions engaged in strategic international partnerships;
• link to similar efforts in other countries;
• contribute to understanding of patterns of global STI collaboration; and
• provide insight for science policy makers.

On the technical side, development of the platform could
• promote open data and open code;
• leverage related ongoing technical efforts for efficient use of funds, deliver transformative analytical power (e.g., by institution, geography, discipline, demography);
• bolster ontology development in non-medical fields;
• create definitions and standards for international activities;
• enable interoperability with extant resources; and
• encourage the use of universal identifiers (like ORCID) and archives (like PubMed).

Summary:
The consensus reached by the workshop participants was that there is tremendous potential to provide the data and tools needed for strategic academic decision making in S&T within a global context. Such efforts are timely and can leverage much work done by individual institutions (e.g., ExploreUGA), consortia (e.g., VIVO, UCosmic), states (e.g., NC Reach), businesses (e.g., Thomson Reuters, Elsevier, Academic Analytics), the U.S. government (e.g., Science of Science and Innovation Policy efforts), other countries (e.g., Brazil’s Lattes system) and international organizations (e.g., OECD).

18 For example, the OECD Science, Technology and Industry Scoreboard.
Next steps:
A follow-on Scoping Meeting is planned for July 21, 2014. The goal of that meeting is to conduct a scoping exercise to determine possible trajectories for PEGASCIS establishment. The participants will explore scenarios that clarify platform scope, organization, priorities, timeframes, and activities that meet stakeholder needs. The scoping can also include identifying organization(s) which could lead/host the consortium and approaches to linking with related efforts in other countries.
Participants:

Organizers:
Katy Börner, Indiana University
Elizabeth (Libby) Lyons, U.S. Department of State (on detail with support from the National Science Foundation)

Academic experts:
Paul Albert, Weill Cornell Medical College
Kiki Caruson, University of South Florida
David Eichmann, University of Iowa
Jane Gatewood, University of Rochester
Julia Trimmer, Duke University
Sally Crimmins Villela, State University of New York
Griffin Weber, Harvard University

Government representatives:
Brian Haugen, NIH
Jim Herrington, NIH
Derek Hill, NSF/SBE/NCSES
Bill Neufeld, NSF/EHR
Maria Zemankova, NSF/CISE

NGO representatives:
Neda Bostani, AAAS
Vaughan Turekian, AAAS
Brad Farnsworth, ACE

Business representatives:
Amy Brand, Digital Science
Jeff Clovis, Thompson Reuters
William Gunn, Elsevier/Mendeley