CTSA Research Networking Evaluation Guide

CTSA Research Networking Affinity Group

I. Executive Summary

Institutional participation in Research Networking (RN) can have various meanings. At its core, it is about using web-based tools to discover and use research and scholarly information about people and resources. These Research Networking tools (RN tools) serve as knowledge management systems for the research enterprise through creation of expertise profiles for faculty, investigators, scholars, clinicians, and facilities. They are different from search engines such as Google in that they access information in databases and other data not on web pages. They differ from 'human connector' social networking systems such as LinkedIn or Facebook in that they represent authoritative or institutional compendia of data rather than individually asserted information, making them more reliable. Additionally, human connector systems, as opposed to RN tools, are typically relationship-based, do not have analytical capabilities, depend on serendipity, tend to replicate the same relationships over and over, have limited storage capacity, and are difficult to sustain without constant manual data entry. RN tools address these weaknesses.

The power behind the Research Networking approach is that it:

- connects institution-level resources, University enterprise systems, national research networks, publicly available research data, and restricted data harvested into the expertise profiles;
- enables more rapid discovery and recommendation of researchers, expertise, and resources;
- supports the development of new collaborative science teams to address new or existing research challenges;
- provides tools that support research, such as CV generation; and,
- facilitates evaluation of research, scholarly activity, and resources, especially over time.

This is particularly important because productive cross-disciplinary research collaboration is an essential feature of a robust translational research enterprise, and collaborative research is receiving proportionally increasing amounts of federal funding.

There are three different dimensions of creating a successful and sustainable initiative in support of research networking. First, there are new collaborative models and opportunities that guide and enable research networking not just institutionally, but nationally and internationally. Second, crucial to success, there are institutional financial, policy, and other commitments and decisions that need to be made when investing in a system and in participating in networks of such systems. Third, there are technical underpinnings and activities that must exist in order to implement RN tools.

Every institution has a unique perspective on how to engage in the internal and external relations that are an essential part of research networking, each has its own institutional culture, and each is at a different stage. Therefore, we frame our discussion as a maturity model and treat each institution as being somewhere along an evolutionary trajectory of maturity.

New Collaborative Models and Opportunities

Research networking tools enable a wide array of applications. As an example of research networking, consider the apparently simple query, "What investigators are studying genes implicated in breast cancer?" For a researcher well established in a social network of collaborators, the answer might well appear obvious. But what of someone outside that network or of investigators unknown to the networked researcher? A reasonably accurate response to this query could involve the use of an RN tools integrating a variety of data representing the connections between breast cancer and the genes implicated to date, between the genes and the relevant publications, and between the publications and their authors. Access to sufficient institutional and linked open data, data that are semantically structured and made publicly available, can attack this and other important problems in advancing research, such as:

- Discovery of potential collaborators, e.g., complementary investigators who can fill missing translational roles, particularly for cross-disciplinary research or the discovery of non-intuitive matches;
- The formation of disease or issue specific research teams that cut across traditional organizational or relationship barriers that can develop new methodologies for translational research;
- Intelligence gathering to understand topical or institutional funding trends;
- Support of virtual teams undertaking science, including the ability to include public and private data to accomplish research goals;
- Robust analytics and visualization to conduct network analysis of team science to improve our understanding of the nature of translational research and examine research trends;
- Identification of partners who have access to community groups or particular populations to support ongoing clinical trials and translation;
- The creation of digital vitae and other supporting documentation for grant applications;

Moreover, research networking and the use of sophisticated RN tools is a nascent effort, so we can expect new, unanticipated uses as well.

Institutional Commitments

In order to maintain the value of research networking, institutions must eventually make systemic and ongoing commitments to see ongoing value. Information about investigators and other translational staff is rich but constantly changing – new grants and publications are awarded, new roles are taken on, and people change institutions and research foci. Maintaining current data about researchers for use for the above applications requires institutional support to ensure reliability and completeness and to reduce redundant effort in data collection. Institutions already manage rich information, through faculty profiling or tenure and promotions processes, that can be reused for research networking and by most RN tools. By leveraging these data for research networking, an institution can advance translational research without separately creating and managing information that already exists elsewhere. In some cases, however, the value an RN tool can encourage investigators to update profile information as well, so information needs to flow both directions. For organizations that represent more than one institution (e.g., a state university system), there is a greater challenge of aligning multiple

entities with RN tools. Some important institutional considerations, discussed below, are financial, administrative, policy, and governance considerations:

Financial considerations. Investing in a RN tool solution, even if open source, involves one-time and ongoing costs, including personnel, to manage the systems, hardware, software, and data. In many cases, partnerships with research program teams, bioinformaticians, and/or librarians can provide the leadership partner with offices of research administration and/or research development to identify efficient approaches. By combining redundant efforts across campuses into more strategic investments, funding may be identified and pooled from existing programs. Institutions need to be comfortable with the return on this investment in terms of opportunities created, improved support of research and research administration, ability to evaluate and measure progress, and economies of scale. These costs include ongoing investments to create and sustain content, maintenance of IT systems and support personnel, management costs to keep RN projects vital, and outreach and training costs. These costs can be lowered by leveraging infrastructures that perform some of these roles and by eliminating redundant systems.

Administrative considerations. Central offices of research, provosts' offices, and/or human resources are typically the stewards of information about faculty members, and may be the logical group to manage an RN tool solution. Partnership with those entities, and support from the leaders of all those areas is critical to a sustainable strategy. These entities are concerned with institutional assessments of faculty performance yet these data can, without too much additional cost, be leveraged to support the research enterprise resulting in research development, improved grant competitiveness, reduced administrative overhead, and better translational science. Administrators and faculty must be comfortable with the processes and controls around the release of profile data and their ability to retain its integrity for other purposes.

Policy considerations. Institutions that manage information about faculty and staff already have methods and policies around how these data may be used and what they may be used for. In addition, faculty councils play an important role in determining the appropriateness of use. Several important considerations that need to be understood for each institution are:

> **Some institutional data are inherently sensitive or private**. Mechanisms need to be put in place to ensure that these data are kept private (or kept out of an RN tool altogether, e.g., social security numbers), and policies established for the release of personal attributes. In some cases investigators are also students, so their information is covered by FERPA. Coordination is recommended among administrators who manage faculty and staff personnel information, faculty organizations, offices of research administration, and data stewards for institution information. Each of these groups has a vested interest in these data, though their interests are not always aligned.
>
> **Trust needs to be established**. Processes need to be established by which individuals or persons in certain roles are allowed to add or modify information, with appropriate audit trails. These typically exist for administrative systems, and processes for research networking systems may likely need to align with these established processes. There must be clear policies about who is authorized to modify what data through what tools or processes. One potential advantage RN tools provide is the ability to give individuals good control over the content, and accuracy of that content. If connected to institutional systems, this can improve accurate while providing a single access point for faculty and staff. The key concern is identifying authority for self-updating institutional profile

information, with appropriate institutional checks and balances for accuracy, while still managing information spaces for additional data that are not institutionally managed, presumably in a research networking tool.

Operational and governance issues. Relationships need to be established by which there are ongoing checks and balances for the appropriate use of information. For research administrative purposes, these are already established, however, research networking represents a new set of uses that require additional consideration even if handled by the same governance process. Faculty will want input in these decisions. There are typically administrative processes for managing information about faculty and staff, and with careful consideration about what role research networking tools might play in this process, there are opportunities to create data intake and maintenance processes that will ensure a robust, single entry, information space for faculty and staff that can be reliable for a number of administrative and research functions.

Adopting Research Networking. Institutional evaluation and commitment is essential to successfully participating in research networking, however, one does not need to fully commit financially, administratively, and in terms of policy in order to begin to see some benefits from research networking. Institutions can participate a step at a time. The Capability Maturity Model Integration (CMMI), for example, lays out 5 maturity stages for an organization. Using the CMMI framework allows a particular institution to assess itself relative to the model, and establish an appropriate path forward.

Level 1: Identification. Evaluate and understand the needs driving research networking, champions in research and administrative areas. Evaluate where expertise profile data may be found (maybe incomplete or under several authorities). Extract some approved profile data sets and develop some initial pilot projects to evaluate value and institutional strategy for next steps. Evaluate existing research networking tools to determine the best solution (see 'Software Applications' below). Identify and procure the necessary resources (financial, management, FTEs, and hardware) for implementation and maintenance and a staged implementation plan.

Level 2: Initial. Installing an RN tool or portal to a research networking system and doing a one-time extraction of public faculty information from an institutional repository (such as a faculty annual report system) into that system, combined with NIH grants (RePORTER) and biomedical publications (PubMed). This will allow a local demonstration of the capabilities of research networking and participation in pilot projects such as direct2experts.org. This has fairly low cost and commitment, and can deliver short-term benefits.

Level 3: Managed. The above, with regular updates to the RN tool by faculty or administrators or that are 'pushed' from institutional repositories and public databases such as PubMed and RePORTER, thus providing periodically updated information.

Level 4: Defined. The exchange of data between the RN tool and institutional and external system(s) is regularized so that faculty members can update information at any point and it is updated centrally. This provides a sustainable and proactive process to maintain accurate and rich data. This is typically combined with adoption of tools that provide greater functionality for administrative and research support processes such as automatic CV and web page generation.

Level 5: Optimizing. This phase includes all of the above, but also now integrates institutional business processes for network analyses of team science activities,

predictive analytics, and prospective grant opportunity assessment. We don't think any institution is at this stage yet.

Technical Underpinnings and Activities

Successful research networking depends on linked open data. Linked open data is a semantic web infrastructure for sharing and connecting data to demonstrate relationships among them. Using linked open data to represent extended information about the clinical and translational research community, including investigators, research staff, industry, government, and community groups, and their research activities, including resources, data, methods, and trials, can be a particularly powerful way to advance research. Were the data necessary for the above uses readily available on web sites, or easily represented or combined via commercial search engines, much of this effort would not be necessary. Neither of these conditions, however, exists. In order to reap the benefits of employing RN tools, certain technologies lay an important foundation. These are:

Ontology standards describe data in common terms and allow it to be mapped across many institutions and applications. The VIVO project (supported with NIH ARRA funding), for example, has developed a comprehensive ontology that represents a broad array of information that can be attached to a particular individual, as well as ways to describe it, which is important for protecting sensitive data. This ontology is being promulgated as a standard and being aligned with other, existing, biomedical ontologies.

Architecture standards structure the data and make it available to software applications. Sometimes referred to as middleware, these software frameworks permit the exchange of data among systems and its representation so that it can be easily utilized by a number of different software applications. In the VIVO framework, for example, data are represented as Resource Data Framework (RDF) triple stores, are automatically exposed (as appropriate) to software applications as SPARQL endpoints, and made interpretable through XML markup. As these data are available as linked open data, they can be used by any number of software applications as another data set. Linked open data is an infrastructure for sharing and connecting data. Based in the semantic web, these data are stored as triple stores and exposed publicly on the Internet so they can be consumed by semantically based queries, increasing flexibility and access.

Software applications provide an array research networking services, including systems to auto-ingest data from institutional systems, administrative tools to manage profiles and generate reports, search tools for discovery, analytical tools for business intelligence, and tools for investigators that directly support research networking tasks such as collaborator discovery, CV generation, etc. These RN tools provide a rich variety of functions in various combinations. To the degree that these applications adhere to the architectural and ontology standards, they can participate in national and international research networking interoperability.

We are listing information about research networking systems last because we firmly believe that the right solution needs to be first based on institutional need, situation, and culture. We do, however, acknowledge the power of a small institutional pilot and its power for demonstrating value. One example of a national research networking demonstration is direct2experts.org. The systems participating in that pilot (with an example institutions) are:

- VIVO (Florida): http://vivo.ufl.edu

- SciVal Experts (Michigan): http://www.experts.scival.com/umichigan/
- Harvard Catalyst Profiles (Harvard): http://profiles.catalyst.harvard.edu
- Loki (Iowa): http://www.icts.uiowa.edu/Loki/
- Community Academic Profiles (Stanford): http://med.stanford.edu/profiles/
- HUBzero (Indiana): http://www.indianactsi.org
- LatticeGrid (Northwestern): http://latticegrid.feinberg.northwesterhn.edu

A more substantial list of research networking tools was developed by Holly Falk-Krzesinski Pamela Shaw, and Laura Wimbiscus-Yoon at Northwestern University. A continuously updated version can be found on Wikipedia at:
http://en.wikipedia.org/wiki/Research_networking_tool_comparison.

Corresponding Author:

William K. Barnett, Ph.D.
Indiana University and Indiana Clinical and Translational Sciences Institute
barnettw@iu.edu, +1 (812) 856-3038

Contributors:

Griffin Weber; Harvard Medical School; Boston, Massachusetts, USA
Mike Conlon; University of Florida; Gainesville, Florida, USA
David Eichmann; University of Iowa; Iowa City, Iowa, USA
Warren Kibbe; Northwestern University; Chicago, Illinois, USA
Holly Falk-Krzesinski; Northwestern University; Chicago, Illinois, USA
Michael Halaas; Stanford University School of Medicine; Menlo Park, California, USA
Layne Johnson; University of Minnesota; Minneapolis, Minnesota, USA
Eric Meeks; University of California, San Francisco; San Francisco, California, USA
Donald Mitchell; Stanford University School of Medicine; Stanford, California, USA
Titus Schleyer; University of Pittsburgh; Pittsburgh, Pennsylvania, USA
Sarah Stallings; University of Colorado, Denver; Aurora, Colorado, USA
Michael Warden; Elsevier; Ann Arbor, Michigan, USA
Maninder Kahlon; University of California, San Francisco; San Francisco, California, USA

References:

Börner, K., Contractor, N., Falk-Krzesinski, H.J., Fiore, S.M., Hall, K.L., Keyton, J., Spring, B., Stokols, D., Trochim, W., and Uzzi, B. (2010). A Multi-Level Systems Perspective for the Science of Team Science. **Science Translational Medicine** 2, cm24.

Contractor, N.S., and Monge, P.R. (2002). Managing Knowledge Networks. **Management Communication Quarterly** 16, 249-258.

Contractor, N.S., Wasserman, S., and Faust, K. (2006). Testing multitheoretical, multilevel hypothesis about organizational networks: An analytic framework and empirical example. **Academy of Management Review** 31, 691-703.

Cressman, D., Holbrook, J.A., Lewis, B.S., and Wixted, B. (2011). **Understanding the Structure of Formal Research Networks** (Vancouver, BC, Simon Fraser University).

Disis, M., and Slattery, J. (2010). The Road We Must Take: Multidisciplinary Team Science. **Science Translational Medicine** 2, 22cm29.

Falk-Krzesinski, H., Shaw, P.L., and Wimbiscus-Yoon, L. (2010). Comparative Matrix of Research Networking Tools. In **National VIVO Conference: Enabling National Networking of Scientists** (Queens, NY).

Falk-Krzesinski, H.J., Börner, K., Contractor, N., Fiore, S.M., Hall, K.L., Keyton, J., Spring, B., Stokols, D., Trochim, W., and Uzzi, B. (2010). Advancing the Science of Team Science. **Clinical and Translational Sciences** 3, 263-266.

Fried, N.F., and Gibbons, S. (2005). Understanding Faculty to Improve Content Recruitment for Institutional Repositories. In **D-Lib Magazine**.

Gewin, V. (2010). Collaboration: Social Networking Seeks Critical Mass. **Nature** 468, 993-994.

Guimerà, R., Uzzi, B., Spiro, J., and Amaral, L.A.N. (2005). Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance. **Science** 308, 697-702.

Huang, Y., Contractor, N., and Yao, Y. (2008). CI-KNOW: Recommendation based on Social Networks. In **The Proceedings of the 9th Annual International Digital Government Research Conference** (Digital Government Society of North America), pp. 27-33.

Jones, B.F., Wuchty, S., and Uzzi, B. (2008). Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. **Science** 322, 1259-1262.

Kibbe, W. (2010). **LatticeGrid**.

Pober, J.S., Neuhauser, C.S., and Pober, J.M. (2001). Obstacles facing translational research in academic medical centers. **FASEB J** 15, 2303-2313.

Schleyer, T. (2011). Conceptualizing and Advancing Research Networking Systems. **ACM Transactions on Computer-Human Interaction** Accepted for Publication.

Schleyer, T., Spallek, H., Butler, B.S., Subramanian, S., Weiss, D., Poythress, M.L., Rattanathikun, P., and Mueller, G. (2008). Facebook for scientists: requirements and services for optimizing how scientific collaborations are established. **J Med Internet Res** 10, e24.

Stewart, D. (2011). Enterprise Content Management in Three Easy Questions. In **Gartner Blog Network** (Stamford, CT, Gartner).

Stewart, D.L. (2010). Collexis Lays the Groundwork for VIVO Integration. In **Connected Knowledge**, D.L. Stewart, ed. (Portland, OR Oregon Health & Science University).

Stewart, D.L. (2010). Knowing What Your Know: Expertise Discovery & Management - Part 1. In **Connected Knowledge**, D.L. Stewart, ed. (Portland, OR Oregon Health & Science University).

Stewart, D.L. (2010). Sustaining the Collaborative Enterprise. In **Connected Knowledge**, D.L. Stewart, ed. (Portland, OR Oregon Health & Science University).

Wieder, B. (2011). Academic-Reference Firm Offers $10,001 for Best New Research Tool. In **The Wired Campus** (Washington, DC, The Chronicle of Higher Education).

Wuchty, S., Jones, B.F., and Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. **Science** 316, 1036-1038.