6-14-2015

Developing and Advancing a Cyberinfrastructure to Gain Insights into Research Investments: An Organizing Research Framework

Ann McKenna
Arizona State University

Jeremi London
Purdue University

Aditya Johri
George Mason University

Mihaela Vorvoreanu
Purdue University

cm@purdue.edu Krishna Madhavan

Developing and Advancing a Cyberinfrastructure to Gain Insights into Research Investments: An Organizing Research Framework

Dr. Ann F. McKenna, Arizona State University, Polytechnic campus

Ann F. McKenna is a Professor in the Ira A. Fulton Schools of Engineering and Director of The Polytechnic School at Arizona State University. Prior to joining ASU she served as a program director at the National Science Foundation in the Division of Undergraduate Education, and was on the faculty in the Department of Mechanical Engineering and Segal Design Institute at Northwestern University. Dr. McKenna received her B.S. and M.S. degrees in Mechanical Engineering from Drexel University and Ph.D. from the University of California at Berkeley. Dr. McKenna is also a Senior Associate Editor for the Journal of Engineering Education.

Dr. Jeremi S London, Arizona State University, Polytechnic campus

Dr. Jeremi London is a Postdoctoral Research Associate at Arizona State University. She holds B.S. and M.S. degrees in Industrial Engineering and a Ph.D. in Engineering Education, all from Purdue University. Prior to her PhD, she worked in quality assurance and logistics roles at Anheuser-Busch and GE Healthcare, where she was responsible for ensuring consistency across processes and compliance with federal regulations. For four consecutive summers (2011-2014), she worked in the National Science Foundation's Division of Undergraduate Education on research and evaluation projects related to the use of technology in STEM education. Dr. London masters mixed methods and computational tools to address complex problems, including: science policy issues surrounding STEM learning in cyberlearning environments; evaluation and impact analysis of federal investments in R&D; and applications of simulation & modeling tools to evaluate programs.

Dr. Aditya Johri, George Mason University

Aditya Johri is Associate Professor and Chair in the Applied Information Technology Department. Dr. Johri studies the use of information and communication technologies (ICT) for learning and knowledge sharing, with a focus on cognition in informal environments. He also examine the role of ICT in supporting distributed work among globally dispersed workers and in furthering social development in emerging economies. He received the U.S. National Science Foundation's Early Career Award in 2009. He is co-editor of the Cambridge Handbook of Engineering Education Research (CHEER) published by Cambridge University Press, New York, NY. Dr. Johri earned his Ph.D. in Learning Sciences and Technology Design at Stanford University and a B.Eng. in Mechanical Engineering at Delhi College of Engineering.

Dr. Mihaela Vorvoreanu, Purdue University, West Lafayette
Krishna Madhavan, Purdue University, West Lafayette

Dr. Krishna Madhavan is an Assistant Professor in the School of Engineering Education at Purdue University. He is Co-PI and Education Director of the NSF-funded Network for Computational Nanotechnology (nanoHUB.org which serves over 330,000 global researchers and learners annually). Dr. Madhavan was the Chair of the IEEE/ACM Supercomputing Education Program 2006. In January 2008, he was awarded the US National Science Foundation (NSF) CAREER award for work on learner-centric, adaptive cybertools and cyber-environments. He was one of 49 faculty members selected as the nation's top engineering educators and researchers by the US National Academy of Engineering to the Frontiers in Engineering Education symposium. Dr. Madhavan leads a major NSF funded effort called Deep Insights Anytime, Anywhere (DIA2) that attempts to characterize the impact of NSF and other federal investments in the area of science, technology, engineering, and mathematics education using interactive knowledge mining and visual analytics for non-experts in data mining. DIA2 is currently deployed inside the NSF and is already starting to affect federal funding policy. Dr. Madhavan also served as Visiting Research Scientist at Microsoft Research, Internet Services Research Group. His research has been published in Nature Nanotechnology, IEEE Transactions on Computer Graphics and Applications, IEEE Transactions on Learning Technologies, and several other top peer-reviewed venues. Dr. Madhavan currently serves as PI or Co-PI on federal and industry funded projects totaling over $20M.

Developing and Advancing a Cyberinfrastructure to Gain Insights into Research Investments: An Organizing Research Framework

Abstract

Although the National Science Foundation (NSF) funds approximately 24% of basic research conducted in U.S. colleges and universities [1], we know little about how NSF funding decisions have shaped the current research terrain. For instance, what new research areas have emerged and how have faculty collaborated over time? The Deep Insights Anytime, Anywhere (DIA2) project was precipitated by the need to better understand these issues and translate them into easy to understand insights for the STEM education community. We were also motivated to understand how research outcomes, particularly of STEM education projects, impact STEM education practice. As part of this project, the DIA2 project team has designed an information and visualization portal (http://www.dia2.org) that allows researchers and scientists to browse and search public data from NSF to understand the research terrain (including information about research on specific topics and researchers active in the area). There are many challenges associated with developing and using such a cyberinfrastructure, but also many potential advantages for practitioners, researchers, and policymakers. In this paper we discuss the research opportunities provided by DIA2 and present the research framework guiding the DIA2 project—a description of the three major themes/areas of research for the study. The paper summarizes the research questions and research activities corresponding to each of the themes, presents next steps, and based on our findings, highlights the value of DIA2 to members of the STEM education community. These concentrated efforts can not only help us better understand the impact and landscape of STEM education research, but this study can also serve as a framework for other large scale cyber-enabled research projects.

Motivation

As part of its mission to "promote the progress of science" in the United States, the National Science Foundation (NSF) invests billions of taxpayers' dollars in basic and applied research each year (NSF, n.d.). In fact, NSF supports 24% of all federally funded research conducted in U.S. colleges and universities (NSF, n.d.). NSF funds research that has the potential to effect changes at various levels—from the individual to the systemic. The outcomes of these investments reveal that NSF-funded research leads to a variety of advances in knowledge, products, and networks. Unfortunately, the full potential of the research outcomes resulting from NSF-funded work has not been characterized; part of the reason is because of limitations associated with the data that reveal these outcomes.

In our project, we directly tackled the challenges related to data that are embedded in any large organization such as NSF, including data cleaning and translation of data into useful information, and in this paper we first discuss these aspects of our project. We then discuss the current system, and then describe the research questions that guide the project. We believe that a crucial aspect of the project has been the research advances that it has catalyzed across a range of interdisciplinary topics. This paper concludes with insights on how this study adds value to the STEM education community, and other researchers interested in a process for developing a research framework for their cyber-enabled research endeavors.

Data Challenges and Cyberinfrastructure Development

In such a digitized world, data and knowledge has inevitably become a big part of our everyday lives [2,3]. It affects everything from the way we learn, to the way we work and play [4]. As the field of data mining and visualization continues to grow and make significant advances on unchartered territory, the more familiar term "big data" refers to this general area [5]. Despite ongoing advances in big data and analytics, the sheer amount of data produced each day makes it difficult for even the most sophisticated technologies and researchers to get a handle on it, make sense of it, and for others to use the results in meaningful ways. As it relates to this context, the amount of data and information associated with each NSF-funded research project (e.g., written documents, raw data, artifacts) is one of the problems associated with making the most of outputs resulting from NSF investments in research. Additionally, the data and information stemming from NSF-funded projects come in various forms, is not well organized, and is stored in disparate places—some of which include physical locations (e.g., on the pages of scholarly publications) or within social structures (e.g., among communities of practice).

Principal Investigators (PIs) on NSF-funded projects are critical to advancing the progress of science in the U.S. Since they are ultimately responsible for executing the research and sharing the resulting insight and outputs, they possess valuable expertise on the topics of interest in their study. Additionally, they are inadvertently members of a highly distributed organization of PIs and STEM researchers—an organization in which knowledge is deeply embedded in a network of diverse interactions, structures, and artifacts. In many ways, the current NSF PI community is what Cummings, et al. [6] refer to as "virtual organization" minus the cyberinfrastructure to help them function more cohesively. Interactive Knowledge for Engineering Education Research (IKNEER, www.ikneer.org) is a cyberinfrastructure that was designed to unite the activities of the engineering education research community [7]. The broader NSF PI community needed a cyberinfrastructure to support it, and IKNEER was its prototype.

Deep Insights Anytime, Anywhere (DIA2, http://www.dia2.org) is a web-based visual analytics platform that allows users to search, visualize, and analyze the NSF's portfolio of past and current research awards to create actionable knowledge and inform decisions [8,9]. DIA2 is being developed by a multidisciplinary team of experts in STEM education, user-centered design, data mining, visualization and system integration, diffusion of innovations, and social media integration and optimization. DIA2 is a central resource project designed for the STEM community. It characterizes and graphically represents data embedded in the knowledge structures in a way that is beneficial to the community. At its most fundamental level, DIA2 helps provide insights about the networks resulting from NSF investments—such as who has expertise on what, and who collaborates with whom.

The development of a robust cyberinfrastructure dedicated to uniting the NSF PI community not only helps integrate the activities carried out by individual project teams, but also supports the generation of new knowledge—because the "dark data" [10] hidden within the networks is being codified in a way that is more readily accessible and beneficial to the members of the community. Furthermore, an increased awareness of other community members' activities and research insights adds to the potential diffusion of knowledge and contributes to community building.

Currently, DIA2 is in its last year of the five-year NSF grant supporting it. As a result of this, much of the DIA2 research is well underway and beginning to wrap up. The remainder of this paper presents the research framework guiding the DIA2 project. It summarizes the research questions and research activities corresponding to the three research thrusts of the framework, and summarizes next steps. This paper concludes with highlights of the value of DIA2 to members of the STEM research education community, and how this study can serve as a model for other cyber-enabled research projects.

DIA2 Research Framework and Research Activities

DIA2's research framework includes three thrusts, which are summarized in Figure 1.

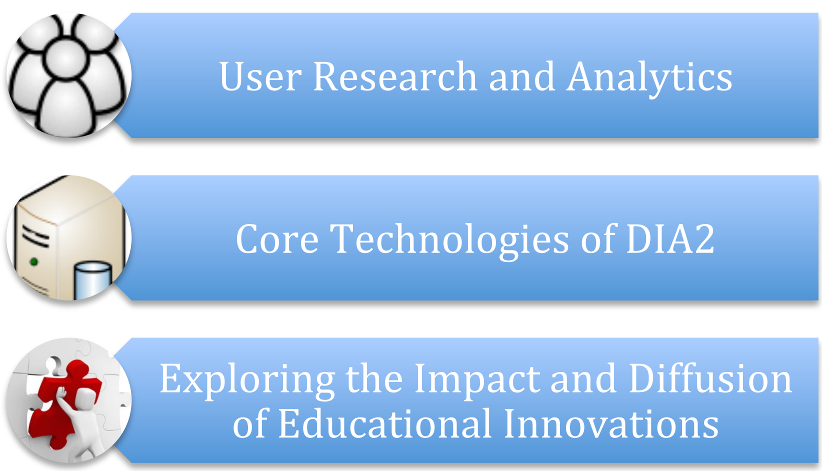

Figure 1. DIA2 Research Framework's Three Foci

The first of the three DIA2 research thrusts is ***User Research and Analytics.*** The users of DIA2 are inclusive of those internal to NSF such as program officers, and those who are members of the broader STEM community (e.g., researchers, practitioners, administrators). With these user groups in mind, the DIA2 team has conducted extensive research to understand the users' needs, goals, and preferences that would help them better perform their work while conducting research, teaching, or executing policy.

Research questions associated with this emphasis include:

1. What are the needs, goals, and preferences of the intended DIA2 users?

2. How should DIA2 be designed to satisfy the needs, goals, and preferences of the intended users?
3. How can digital traces of DIA2 users' activities reveal how they use DIA2?

As part of this research, the DIA2 research team conducted user studies with NSF personnel (i.e., program officers, science assistants) and members of the broader STEM community (e.g., PIs, university administrators). Such research led to the development of user profiles, use cases, and design requirements that informed the development of the core DIA2 technologies and their current and prospective use [11-14].

The second of three DIA2 research thrusts in the research framework is the ***Core Technologies of DIA2***. This part of the research includes the use of advanced data mining and visualization methodologies to develop the DIA2 technology. This includes gathering, storing, and indexing the dark data on NSF investments in STEM research. As part of this work, DIA2 researchers develop visualizations to represent the data based on insights that emerge from the user research. Moreover, this part of the research includes developing the DIA2 interface that runs inside a web browser.

Research questions associated with this emphasis include:

1. How can complex data mining and visualization techniques be optimized to understand trends in NSF data and lower the barrier of entry for a broad range of NSF stakeholders?
2. How do research topics and NSF community members scatter and converge through the use of algorithms?
3. How can traditional bibliometric and scientometric analyses be used to validate DIA2 results?

The development of DIA2 is well underway. When developing DIA2, the goal is to design interfaces that are easy to use and easy to understand. The DIA2 design philosophy is: "No training. No manuals." DIA2 is currently available for use (www.dia2.org). Many DIA2 users comment on how "intuitive" it is. Figures 2 and 3 are screenshots of the latest version of DIA2 (images captured January 2015). Figure 2 shows the pop-up users are presented with before launching DIA2, while Figure 3 is an example of what is presented when exploring a PIs' NSF-funded work and their collaborators.

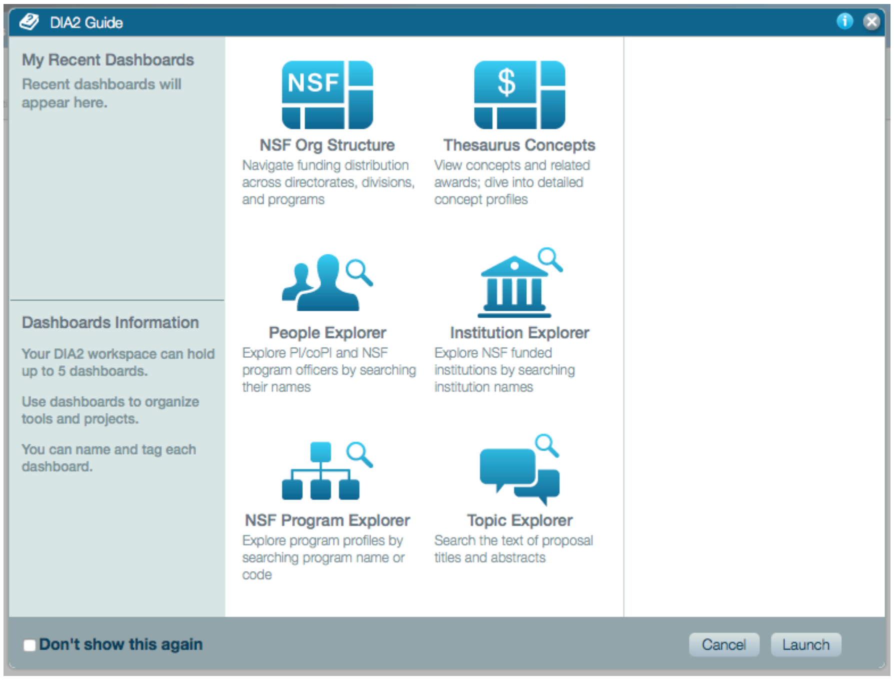

Figure 2. “DIA2 Guide” Presented Before Launching DIA2

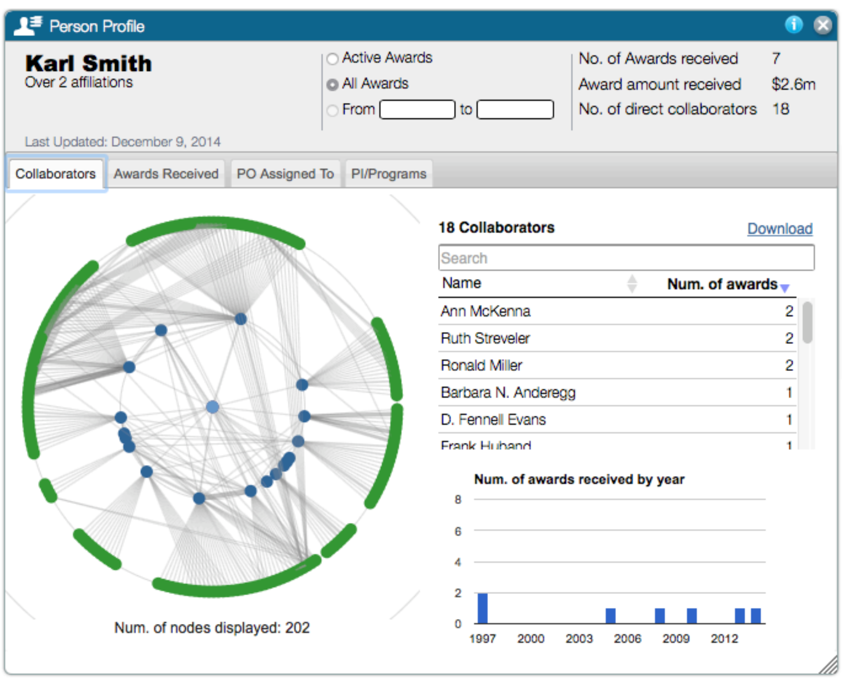

Figure 3. Screenshot Resulting from DIA2’s “People Explorer” Widget

As the core DIA2 technology continues to be developed, the *User Research and Analytics* research strand continues to include usability testing and heuristic evaluations of the DIA2 interface. Furthermore, the DIA2 research team is analyzing the digital traces of users’ activities to understand how people are using DIA2.

The last of the three research thrusts of the DIA2 research framework is: ***Exploring the Impact and Diffusion of Educational Innovations.*** This part of the research focuses on understanding the value of DIA2 to the community, and understanding the insights that result from using DIA2.

Research questions associated with this emphasis include:

1. How do DIA2 users understand and interpret the visualizations DIA2 presents?
2. In what ways does DIA2 add value to the STEM research and education community?
3. What is a meaningful model of the impact in the context of STEM teaching and scholarship?

In some ways, aspects of this research focus directly complement the user studies. The user research informed the design of DIA2, and provided meaningful insights on what potential users would want to see, while this strand of research activities focuses on how people are actually using it now that DIA2 has been developed. Over 100 interviews have been conducted with members of the STEM community, in addition to interviews conducted with NSF staff. The interviews were completed in the fall of 2014 and preliminary findings are emerging. For example, many commented on how DIA2 is useful to find collaborators, and to explore the funding landscape of NSF awards. These were activities that the project team had anticipated would be useful for the community. However, some new insights we are gaining from the interviews include the extent to which PIs and administrators would use DIA2 to inform promotion and tenure activities, or that job seekers would use it in preparation for going on an interview at an academic institution. Formal analyses are underway of the interview data and detailed reports and findings are forthcoming.

This research area also focuses on understanding the impact and diffusion of other innovations besides DIA2. The team is using DIA2 to answer some of the questions that motivated the study—such as, how can trends in the use of pedagogical resources or research methodologies enhance our understanding of what "impact" means in the context of teaching and scholarship?

In short, a research framework that guides the DIA2 research project includes three research thrusts: User Research and Analytics, Core Technologies of DIA2, and Exploring the Impact and Diffusion of Educational Innovations. Research activities in each of these areas are starting to conclude as this project is in its fifth year of the five-year grant supporting this work. Several publications have resulted from this work thus far and are included as references within the paper.

Highlighting the Value of DIA2 to STEM Education Research Community

DIA2 is a NSF-funded central resource that researchers, teachers, administrators, graduate students and NSF personnel find valuable for a myriad of reasons. Researchers who have used DIA2 discuss how DIA2 makes it easier to narrow down the list of NSF programs to target when developing a proposal. DIA2 also makes it easier to quickly understand the research expertise and experience of potential collaborators. Teachers using DIA2 mention how simple it is to find people at their institution who may be developing resources or conducting research on pedagogy they are interested in. Administrators discuss the value DIA2 brings to enhancing their understanding the research experience and funding history of candidates applying for faculty positions, and how a candidates' background may/not align with the department's strategic vision. Graduate students find DIA2 useful while on the academic job market because it helps identify a set of institutions to explore openings at because they have researchers who share similar research or pedagogical interests. NSF personnel use DIA2 to understand the projects in their funding portfolio, and to inform future funding decisions. These are some of the preliminary insights on how DIA2 adds value to the community, but the greatest impact of DIA2 is realized in how users continue to leverage the knowledge and networks resulting from what DIA2 presents.

DIA2 Research Framework as a Model for other Cyber-Enabled Research

Members of the STEM education research community may also find value in using the DIA2 research framework as a model for conceptualizing their cyber-enabled research projects. Essentially, the research framework includes three parts, which are depicted in Figure 4.

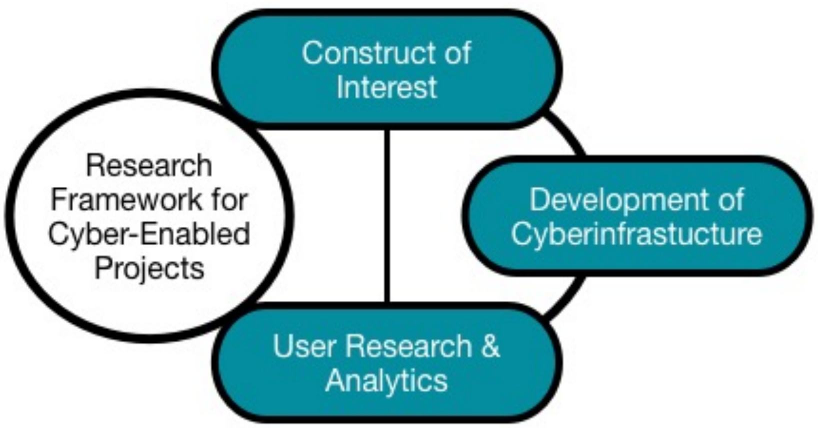

Figure 4. Research Framework for Cyber-Enabled Projects

The execution of the research project involves iterative cycles between three parts: a construct or phenomenon of interest; user research and analytics to inform development and understand usage, and the development of the cyberinfrastructure. Additionally, this type of research requires a multidisciplinary team of people with expertise in areas such as: user-centered design, data mining, visualization and system integration, and the construct of interest (e.g., STEM education, diffusion of innovations).

In this study, for example, DIA2 was designed to be a central resource for the STEM community. The phenomenon of interest was impact and the landscape of STEM education research resulting from NSF's investments in it. Since the goal was to provide a tool that would be valuable to both NSF personnel and members of the STEM community, both groups' perspectives were included in the user research and in the ongoing analysis of the value of DIA2. Insight on each user groups' needs, goals, and preferences, as well as data on how people are actually using the tool is coupled with expertise on data mining and visualization to inform the development of the tool. Now that the tool is in the latter stages of development, the information resulting from it is contributed to new advances on the construct that motivated it. Thus, there is an iterative cycle between the construct of interest, user research, and technology development.

Summary

Despite millions of dollars of federal investments in STEM education research each year, we know little about how funding decisions have resulted in a research landscape. At the same time, there are significant challenges that come with compiling, cleaning, indexing, interpreting, and presenting the variety of data associated with NSF investments in STEM education research. The DIA2 project was precipitated by the need to better understand these issues and translate them into easy to understand insights for the STEM education community. Data collection is mostly complete for each of the research areas. However, in many ways this work will be ongoing for the life of DIA2. In particular, this project intends to not only produce research findings to inform the community, but to also provide a lasting tool/resource that can be used to continue to advance the STEM education and research enterprise, and to share a research framework that others engaged in cyber-enabled research might find useful. We anticipate several forthcoming publications for each of the research areas once data analyses are finalized.

Acknowledgments

This work was supported by the U.S. National Science Foundation awards TUES-1123108, TUES-1122609, and TUES-1123340. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. NSF. About the National Science Foundation, <http://www.nsf.gov/about/> (n.d.).
2. Manyika, J. et al. Big data: The next frontier for innovation, competition, and productivity. (McKinsey Global Institute, Washington, D.C., 2011).
3. The Data Deluge (2010). The Economist, http://www.economist.com/node/15579717
4. The Wireless Issue. (2012). *Time, 180*.
5. The National Academies Press. Big Data: A Workshop Report. (The National Academies Press, 2012).
6. Cummings, J. N., Finholt, T. A., Foster, I. T., Kesselman, C. & Lawrence, K. A. Beyond being there: A blueprint for advancing the design, development, and evaluation of virtual organizations., (Office of Cyberinfrastructure, National Science Foundation, Arlington, VA, 2008).
7. Madhavan, K., Xian, H., Vorvoreanu, M., Jesiek, B., Johri, A., Wang, A., Wankat, P. (2010) Interactive Knowledge Networks for Engineering Education Research (iKNEER).
8. Madhavan, K. et al. Portfolio Mining. Computer 45, 95-99 (2012).
9. Madhavan, K. et al. DIA2: Web-based Cyberinfrastructure for Visual Analysis of Funding Portfolios. IEEE Transactions on Visualization and Computer Graphics 20, 1823-1832 (2014).
10. Heidron, P. B. Shedding the light on dark data in the long tail of science. Library Trends 57, 280-299 (2008).
11. Chowdhury, B.T., & Johri, A. (2014). U.S. Graduate Engineering Students' Perceptions of and Strategies Towards Acquiring External Funding for their Education. *International Journal of Engineering Education, 30*(5), 1136-1144.
12. Johri, A., Teo, H. J., Lo, J., Dufour, M. & Schram, A. Millennial Engineers: Digital Media and Information Use Among Engineering Students. Computers in Human Behavior 33 (2014).
13. Liu, X., Wang, G. A., Johri, A., Zhou, M. & Fan, W. Harnessing Global Expertise: A Comparative Study of Expertise Profiling Methods for Online Communities. Information Systems Frontiers, 13 (2013).
14. Liu, Qing, Vorvoreanu, Mihaela, Madhavan, Krishna, and McKenna, Ann F. (2013). *Designing Discovery Experience for Big Data Interaction: A Case of Web-Based Knowledge Mining and Interactive Visualization Platform*, In Design, User Experience, and Usability. Web, Mobile, and Product Design, Lecture Notes in Computer Science, Springer-Verlag Berlin Heidelberg, v 8015, 543-552