International Journal of Engineering Education Vol. 34, No. 3, pp. 1104–1119, 2018
Printed in Great Britain

0949-149X/91 $3.00+0.00

The Role of the Visualization Tool DIA2 in Supporting STEM Researchers, Educators, and Administrators*

JEREMI S. LONDON, ANN MCKENNA and WEN HUANG
The Polytechnic School, the Schools of Engineering, Arizona State University, Mesa, AZ 85212, USA. E-mail: jeremi.london@asu.edu, ann.mckenna@asu.edu, whuang82@asu.edu

ANDREEA MOLNAR
University of Portsmouth, Portsmouth, UK. E-mail: andreea.molnar@port.ac.uk

SAMANTHA N. CRUZ
Counseling Psychology Department, the College of Integrative Sciences and Arts, Arizona State University, Tempe, AZ 85287, USA. E-mail: samantha.n.cruz@asu.edu

The impact of federally supported STEM education research is difficult to document, because knowledge resulting from research is deeply embedded in complex networks of social interactions, structures and artifacts. DIA2 is a web-based search and visualization tool designed to make knowledge associated with NSF-funded projects more accessible to the STEM education scholarly community. No studies have communicated its value to the community it was designed to serve. This study investigates users' perceptions of DIA2's usefulness for carrying out teaching, research, and administrative duties. Using a qualitative interpretivist approach, researchers conducted semi-structured interviews with 89 principal investigators on NSF-funded grants. We used thematic analysis, and interpreted the results using the Distributed Cognition and the Affordances of Information and Communications Technology frameworks. Results indicated ten ways DIA2 is valuable for carrying out a spectrum of routine scholarly activities, including initiating research collaborations, preparing promotion and tenure documents, and informing strategic decisions. With increased accessibility to the knowledge DIA2 provides, we are better equipped to characterize the impact of federally supported STEM education R&D.

Keywords: data knowledge and visualization; knowledge sharing; learning via discovery; value of information

1. Motivation

The National Science Foundation (NSF) supports 24% of all federally funded research conducted in America's colleges and universities as part of its mission to "promote the progress of science" in the U.S. [1]. Projects in NSF's portfolio of science, technology, engineering and mathematics (STEM) education R&D have a scope of influence that addresses needs at various levels—from the individual to the systemic—and target a spectrum of needs that exist within the education ecosystem. Despite tremendous investments in STEM education throughout the existence of the agency, ongoing calls for "innovation" and "transformation" in higher education suggest that the greatest impact of NSF investments in STEM education research has not been realized, or at the very least is difficult to detect and define [2]. One premise of this study is that part of the reason the full potential of research has not been realized is because of limitations associated with the data revealing project outcomes. DIA2 is a cyberinfrastructure developed for the STEM community to make sense of NSF data that directly addresses some of the data challenges associated with NSF funded projects with hopes that the tool will be a resource to the STEM community as its members pursue evidence-based activities that will lead to lasting change in K-12 and higher education.

The next sections briefly summarize the need for a tool like DIA2, provides an overview of the tool, and how it fits within the larger context of ways scholars find information in a highly-digitized world. The results of this study reveal DIA2 users' perceptions of the usefulness of DIA2 in carrying out teaching, research, and administrative duties. The results indicate that this kind of cyberinfrastructure is a valuable tool that the STEM community can use to better leverage and increase the impact of NSF investments in R&D.

1.1 Challenges with data in the digital age

Data are an integral part of our digitized world and influences every aspect of our daily [3–5]. As the field of data mining and visualization continues to advance, the more common the term "big data" becomes [6]. Despite ongoing discoveries in the field of big data, the sheer amount of data produced each day makes it difficult for even the most sophisticated technologies and researchers to acquire, store, manipulate, and present data in meaningful ways.

* Accepted 5 January 2018.

More importantly, an ill-designed visualization may cause users to struggle with understanding its representation and even make wrong decisions based on what is presented [7]. This reality is important in this context, because the amount of data associated with each NSF-funded research project (e.g., written documents, raw data, artifacts) is a barrier that is crucial to making the most of outputs resulting from NSF investments in STEM education research. Furthermore, data associated with 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 articles) or within social structures (e.g., among communities of practice). Within the context of NSF investments in research, Principal Investigators (PIs) and those impacted at their institutions are key entities in the social structures.

PIs on NSF-funded projects are essential to promoting the progress of science. Because they are ultimately responsible for carrying out research and disseminating the findings and outputs, they hold valuable expertise on the topic 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. Thus, in many ways, the NSF PI community is what Cummings, Finholt, Foster, Kesselman, and Lawrence [8] refer to as a "virtual organization" minus the cyberinfrastructure to help them function more cohesively.

The development of a robust cyberinfrastructure dedicated to making the collective PI community's research activities and outputs more visible not only helps integrate the activities carried out by individual project teams, but it also supports the generation of new knowledge. Part of the reason such benefits can be realized is because the "dark data" [9] hidden within the networks is being codified in a way that is more readily accessible and beneficial to the members of the community. Moreover, an increased awareness of other community members' activities and research insights adds to the potential diffusion of knowledge and contributes to community building. The development of a robust cyberinfrastructure designed around NSF data was critical to uniting the activities of the broader STEM research community as part of increasing the impact of NSF investments in research. On the other hand, previous research indicates that proper visualization of others' or historical data could increase one's awareness of their behavior and have a positive impact on their own practices and results [10]. New tools are needed to address the changing work-related needs of scholars in the digital age.

1.2 Scholarly communications in the digital age

"Scholarly Communication," a term that has been used since the mid-1970s, often refers primarily to the process of publishing peer-reviewed research [11]. More recently, however, the term has been used more broadly to capture the range of processes in a researcher's everyday activities—including the creation, transformation, dissemination and preservation of knowledge related to teaching, research and scholarly works and products [12]. Although there might be some differences across disciplines (e.g., the difference between the science and quantitative social sciences and the more interpretive and qualitative disciplines) [13, 14], the most traditional and common method of scholarly dissemination is writing the research findings in a manuscript and getting them published in a peer-reviewed journal. Over the past two decades, however, there has been a discussion about issues with scholarly communication, which led to the widespread belief of a "crisis in publishing" and weaknesses in the peer-review system [15]. Considering the increasing price to publish in journals [16], the limited access to published research [17], [18, 19], and the growing concern of research getting "lost" due to the disappearance of publishers or journals, there has been a call for changes and innovations in scholarly communication based on the real needs and workflow of researchers and administrators [20–22].

Several studies have been conducted to investigate the status of scholarly communications in the digital age. Harley, Earl-Novell, Arter, Lawrence, and King [23] conducted a study with 160 interviewees across 45 institutions in seven different disciplines to investigate faculty's values and behaviors throughout the scholarly communication lifecycle (including career advancement, sharing, collaborating, publishing, resource generation, and engaging with the public). This study identifies the two biggest challenges for disseminating research faced by faculty: (1) the lack of time and (2) the need for "noise" filters (e.g., mechanisms to filter out low-quality work found both on the internet and in journal publications). When it comes to sharing work with other scholars and keeping up to date in a field, faculty reported they always communicate within a circle of trusted colleagues using informal email exchanges and Web 2.0 technology (e.g., personal website, blog) [13]. The channels researchers used to collaborate with one another, however, are rather conservative and narrow: collaborations often grow out of relationships forged in graduate school with peers or mentors. Technologies, such as Skype and password-protected wikis, are used to communicate during the collaboration.

New technologies (such as Web 2.0 social media tools, Wikipedia's model of crowd-sourced production, big data and cyberinfrastructure) were reported to transform how research is conducted, disseminated, and rewarded [24]. With aid from the huge volume of technology tools launched in the past decade, scholars have more options than ever to communicate their research results to wider audiences using more than the traditional channels. The world of scholarly communication is changing as a result of the new technology and innovative models. Research from Altmetric [25] shows the impact of research findings can extend beyond the traditional measures of peer-reviewed publications, because Web 2.0 technologies allow others to mention research insights in other mediums (e.g., personal websites, blogs). A recent study conducted by Poynter Research Center [24] indicate social media has been used by more researchers to communicate with audiences and share research than when more traditional methods were the only option available. Procter et al. [26] also reported researchers' increased use of Web 2.0 services to facilitate novel forms of scholarly communication.

Furthermore, Kramer and Bosman [27] identified 101 innovations in scholarly communication and their effectiveness on changing research workflow between 2000 and 2015. These 101 innovative tools were categorized into six phases of scholarly communication: discovery, analysis, writing, publication, outreach, and assessment. Typical workflows (e.g., traditional, modern, innovative, experimental) were modeled using these innovative tools. In short, this study highlights the increasing use of social discovery tools and scholarly social media as part of the typical scholar's workflow. Kramer and Bosman [27] identified the multidisciplinary and citation-enhanced databases as the most important development for the discovery phase, collaboration and data-driven for the analysis phase, more and better-connected researcher profiles for the outreach phase, and importance of societal relevance and non-publication contributions for the assessment phase. One of the more recent innovations that scholars have found to be particularly useful are scientific mapping tools. DIA2 is an example of such.

DIA2 is a web-based search and visualization tool that was designed to make knowledge about and results from NSF-funded projects more accessible to members of the STEM community. The purpose of this study is to explore DIA2 users' perceptions of the value to the scholarly community for whom it was developed. Such insights advance our understanding of how scholars can add a visualization tool to the repertoire of resources they use to engage in scholarly activities. More broadly, the findings discuss the characteristics a visualization tool that should be considered for inclusion among the growing number of scientific mapping tools that scholars use to communicate insights to one another. From a practical perspective, DIA2 can serve a model for cyberinfrastructure that functions as a central location for a distributed network of scholars (like PIs) interested in certain types of insights among that is oftentimes difficult to access because of the aforementioned barriers.

The remainder of this document describes relevant literature on the value STEM scholars have obtained from using scientific mapping tools, and the conceptual underpinnings of the question of interest in this study. After providing an overview of DIA2 and methods for collecting data, we will present DIA2 users' insights on its value for carrying out a variety of scholarly tasks. The findings will be discussed in light of the conceptual lens guiding this work and existing research. The implications for a variety of stakeholders will also be discussed.

2. Literature on the value of science mapping tools for scholarly activities

Science mapping technology has been used to effectively promote communication among scholars. Science mapping or bibliometric mapping, is a unified name for various data mining approaches which explore the connections among different scientific knowledges through the visual representation in the field of bibliometrics [28]. The connection establishment is based on the extraction and analysis of the critical information (e.g., researchers' name, affiliated institute, award, published article, high-frequency vocabulary in the text) from a large number of interrelated scientific literature and other supported metadata. This visual representation helps users to understand an overview of the structure of the scientific network in a specific discipline domain. Visual representations offered through science mapping technologies include treemaps, mosaic plots, ego-networks, and statistical graphics (e.g., plots, bar charts) [29]. For example, a circular diagram of article co-occurrence by category in engineering education research from 2005–2008 was developed by Jesiek, Borrego, Maura, Beddoes et al. [30]. In this diagram, the size of each node is aligned with the number of articles in the category the node represents, while the size of each line between two nodes reflects the times that the content of an article covers both of these two categories at the same time. With this visual, researchers can easily identify clusters of related research in the field of engineering education (e.g., project-based learning, capstone design, first-year engineering). Previous studies serve as concrete examples of how science mapping has also been applied to help researchers to understand the relationship between researches who study design thinking [31] and entrepreneurship education [32].

One of the main benefits of using tools that perform science mapping analysis is efficiency in processing massive amounts of data to garner useful insights. There are several examples in the literature on how such tools were used among STEM scholars. One tool that is commonly used is called CiteSpace II [33]. It is a Java application for visualizing emerging trends and patterns in scientific literature [34]. It uses co-citations of references as the mechanism to organize the science network. CiteSpace II researchers investigated its value to its users and found that the visual representation helps researchers explore trends of awarded grants; this is useful for writing more purposeful grant proposals [35]. Chen [35] argues that since the core of a cluster in the network represents the intellectual base of a field of study and the boundary corresponds to scientific frontiers, CiteSpace II can enhance users' ability to identify relationships between core and the boundary of a research space; this also improves the competitiveness of universities and research laboratories [35].

Apart from preparing to write competitive grant proposals, science mapping tools are also useful to understanding things that may be happening within an institution. For example, Folkstad and Hayne examined the differences between of their school's research funding based on the funding source (e.g., NSF, National Institutes of Health (NIH), Department of Education) and between two periods of time [36]. From the visualized Principle-Investigator social network created by Pajet [37] and the results of quantitative analysis, they gathered evidence on the impact of a new educational policy on interdisciplinary collaborations among researchers at the institution. Although Pajet is not a pure science mapping software, the data is based on grant proposals, and the data visualization mechanisms are similar to traditional science mapping software programs.

In addition to understanding what may be happening within an institution, science mapping tools are also useful for garnering insights about researchers around the world. Jigsaw [38] is a software program designed to extract entities from unstructured text, and present their relationships in different views [34]. One aerospace engineering researcher used it to compare two comparable standards published by the United States and the European Union. Each standard includes thousands of items. Jigsaw helped her to quickly identify similarities and relationships between items in each standard, which was hard to complete before programs like Microsoft Word [39]. Similarly, an intelligence analysist at a national laboratory found suitable post-doc and researcher applicants with Jigsaw from a number of resumes. From the extracted critical information from each resume, they could efficiently make connections between applicants' backgrounds and the vacant positions [39].

Lastly, Small, Boyack, and Klavans agree that the visual representation benefits the identification of emerging topics in science and technology fields [40]. However, the focus of their study was on the software's value to decision makers at funding agencies. RePORT, the software program they developed, helps NIH analyze more than 55,000 peer-review records of grant applications, and generate descriptions of the scientific structure in the biomedical domain as part of their annual funding evaluation [41]. Such tools are also useful for estimating researchers' productivity, which, in turn, helps funding agency make more informed funding decisions since they are more inclined to fund researchers who succeed in continuously publishing, or to fund opportunities for a broad set of researchers [42].

In summary, existing literature includes a variety of examples of how scholars and funding agencies have gotten value from science mapping tools. The use of advanced algorithms and visualization techniques leads to efficient ways to extract and reorganize massive amounts of data in a way that mitigates users' workload when carrying out a myriad of scholarly tasks. While DIA2 as developed for the STEM community of researchers, educators, and administrators, such an analysis has been formed for this tool. Moreover, its value compared to other tools in a similar classification is also unknown. Lastly, the studies that have been performed to date have focused primarily on research-related tasks (e.g., writing competitive proposals, identifying research networks). Thus, the value of these tools for carrying out teaching and/or administrative tasks has not been explored. This study seeks to fill this gap. This study focuses on one tool that was developed to promote scholarly interactions among STEM education researchers working in the digital age—a tool that was designed with a wide range of scholarly tasks and user groups in mind. The research design and results will follow a brief overview of DIA2.

3. Overview of DIA2

DIA2 [43] is a web-based science mapping platform for searching, viewing, and analyzing the NSF research portfolio. It was designed for the "casual expert" who has a high degree of training in their discipline yet has little-to-no training in advanced visualization and analytics [29]. Beyond the motivations for developing the tool, DIA2 has the potential to improve scholarly communication by expanding the reach and impact of NSF-funding based on the notion that providing 'affordance is innovation.' One premise of the technology is that learning new knowledge means absorbing networks of information, not individually reading thousands of documents [44]. To help users get actionable insights from NSF investments in R&D, DIA2 employs visualization algorithms not because of their novelty but because of their capability of producing familiar or self-explanatory representations to a casual expert.

DIA2 currently archives data from January 1973 to September 2016 (only data from 1995 is exposed), the total size of the data is 1 Terabyte. DIA2 uses a dashboard metaphor to accommodate the different needs of our diverse user groups (e.g., members of the STEM community, NSF personnel). The tool includes six widgets designed to meet the needs of the user groups. More specifically, the widgets allow users to look into NSF's organization structure (NSF Org Structure), perform keyword searches on concepts (Thesaurus Concepts), search for PI/co-PI on NSF-funded projects (People Explorer), explore NSF-related information about funded institutions (Institution Explorer), investigate NSF programs (NSF Program Explorer), and explore NSF proposal project details (Topic Explorer), respectively. A set of services is designed to be used as needed, including services to show the structure of the NSF organization, to visualize the collaboration network of PIs/coPIs, and to visualize the geographical distribution of awards. DIA2 also uses bar charts, pie charts, and time-series plots for visual representations of research portfolios for NSF programs, institutions, and PIs. Other publications stemming from this study include additional details about DIA2's design. Visit the DIA2 website to explore its functionality. The development of DIA2 is based on "user-centered" method [45, 46]. Target population, such as researchers, and NSF staffs, have been broadly participated in the software design, development, and evaluation phases. Specific methods, such as typical user interview, survey, focus group, observation, have been used to gather users' needs, and guide the design of the software [45–48]. The software's usability is also assessed through investigating representive NSF staffs after it is developed [49]. However, how the scholars use the platform in their daily work, and what is the value that DIA2 brings to them have not been explored before. Answering these questions can not only promote the diffusion of DIA2 in scholar communities, but also help the policy-makers to better understand scholar communities. Both of these two points can boost the boom of the scholar communities and further positive influence the country's progress in the STEM field

4. Research design

This study is part of a larger project that employs a user-centered design approach to developing a cyberinfrastructure that can serve as a central resource for the STEM community. An interpretivist theoretical lens guides this study. In short, interpretivist research includes social inquiry, through the use of naturalistic methods (i.e., interviews), that leads to evidence-based claims as a result of interpretations of individuals or groups' lived experiences [50, 51]. The research question guiding this qualitative study is: *In what ways do members of the STEM education community perceive DIA2 adds value?*

4.1 Conceptual underpinnings

Distributed cognition [52, 53], and an extension of Gibson's Affordances theory [54, 55] serve as the conceptual foundation for the interpretation of the results in this study on the value of DIA2 to members of the STEM scholarly community.

Cognition, the brain's ability to process information, is one of three steps of human information processing: perception, cognition, and action. The Distributed Cognition framework (referred to as "DCog" hereafter) is part of a wider movement within the contemporary cognitive science discipline to account for the role of the environment (social, cultural, and material) in shaping cognition [53]. More specifically, DCog affirms that: knowledge is not confined to an individual; the learner plays an active role in constructing knowledge; and tools within in the environment scaffold cognition as the learner interacts with them. In this context, the use of the word scaffold may mean the properties of a tool makes a task easier to complete, enables a task to be done more efficiently, or provides a means of accomplishing a task that could not have been accomplished without the tool.

Several researchers have noted and argued for the value of using the distributed cognition framework for information visualization research [52, 56, 57]. It is widely accepted among information visualization researchers that "The purpose of visualization is insight, not pictures. The main goals of this insight are discovery, decision making, and explanation" as in [58]. In an attempt to effectively design visualizations, information visualization researchers are beginning to ask questions about what insight

Table 1. Summary of Information and Communications Technology Affordances

Affordance	Description
Accessibility	ICT enables relatively easy access to immense amounts of information through a variety of mechanisms.
Speed of change	ICT mediates abundant and rapidly changing information.
Diversity	ICT offers exposure to a vast range of diverse and different experiences that can inform learning.
Communication and collaboration	ICT opens up new forms of dialogue and enable the potential for learning as a result of engagement with others.
Reflection	ICT encourages reflection and critical thinking as users engage in discussions over longer time frames than is possible with face-to-face communication, and as they access and build on archived materials from past discussions.
Multimodal and non-linear	ICT allows learner to move beyond linear pathways of learning to adopt more individualized strategies and pathways.
Immediacy	ICT dramatically increases the speed at which information can be exchanged over the web.
Surveillance	ICT allows users to extend their gaze and secure greater knowledge and control over others.

means, how it is measured, and how it is generated [59, 60]. Fundamentally, questions about insights are questions about human cognition. Thus, the Distributed Cognition framework is a useful model of cognition when interpreting findings that involve information visualizations [52, 56, 57].

In this study, DCog is a useful theoretical underpinning for articulating the role of DIA2 in constructing knowledge as part of completing tasks typically performed while engaging in scholarly work. Because the interaction between the users and the environment is an important part of DCog, it is important to note that the environment provides both affordances and constraints that influence human cognition and action. In light of this, it was valuable to use a framework that delineates some of the affordances of information and communication technologies (ICT), of which DIA2 is an example.

While the field of learning technologies continues to mature, the need for theoretical frameworks continues to grow [61]. This need motivated a study to understand how learning technology's "properties might be exploited in particular learning and teaching contexts" as in [54]. Conole and Dyke [54] built on Gibson [55] original Affordances Theory to articulate eight affordances of ICT. The summaries provided in Table 1 are adopted from Conole and Dyke [54] to provide succinct descriptions of the affordances of information and communications technologies.

Unlike with physical devices, where the properties and affordances of the tool might be evident and largely determined by the designer, affordances of networked technologies are not as apparent because they are hidden in software algorithms and may vary by user. Thus, is it important to rely on the users' perspective to communicate the value of a tool in which the affordances are hidden. The results of this study reveal users' perceptions of the usefulness of DIA2. The Affordance of Information and Communications Technology framework [54] enable scholarly interpretation of the results in light of similar tools.

4.2 Participants

In Fall 2014, all NSF-funded PIs on active grants associated with the Transforming Undergraduate Education in STEM/Course, Curriculum, and Laboratory Improvement (TUES/CCLI) programs were invited by email to participate in this study. Among the 790 PIs invited, eighty-nine people agreed to participate (11% response rate). Among the participants, 47% were male, and 53% were female. Tables 2 and 3 summarize the participants' area of disciplinary expertise (as defined by the areas inclusive of the division of undergraduate education), and tenure rank/position, respectively. (Note:

Table 2. Demographic Data: Disciplinary Expertise

Discipline	Proportion of Participants (n = 89)
Engineering	26%
Biology	15%
Physics	11%
Chemistry	11%
Computer Science	11%
STEM Education	11%
Other	15%

Table 3. Demographic Data: Position/ Tenure Rank

Position/Tenure Rank	Proportion of Participants (n = 89)
Full Professor	31%
Associate Professor	26%
Assistant Professor	12%
Researcher-only	2%
Instructor	8%
Administrator-only	6%
Administrator, Professor, or Instructor	15%

the disciplinary categories are determined by NSF, are a characteristic associated with each NSF-funded grant and are specified by the PI on the grant to denote the disciplinary focus of the proposed work.)

The results of this study reflect diverse perspectives. The representation of gender was nearly equal, with slightly more females than males. Eleven areas of disciplinary expertise were represented among the participants. Participants with engineering expertise were the most represented (26%); several disciplinary areas (i.e., astronomy, geology, economics) had one representative and are included in the "Other" category. Participants with expertise in mathematics (7%) and psychology (4%) were also included in the "Other" category. The STEM Education label includes people from Science Education, Math Education, and Engineering Education. We wanted proportional representation in the STEM disciplines NSF funds; and the participant representation does align with the disciplinary representation in the portfolio.

Lastly, the participants include administrators, tenure-track and non-tenure track members of the STEM education community. The three largest representations of these participants include Full Professors (31%), Associate Professors (26%), and people with some combination of administrative titles (20%). As it relates to the administrators, there were some who only had administrative titles and no teaching or research responsibilities (e.g., Director of a research center associated with a university), while others had administrative duties as well as teaching and/or research responsibilities (e.g., Full Professor & Department Chair). A range of academic institution types was represented among the participants as well (i.e., public and private universities; research-intensive universities; undergraduate colleges; community colleges; and research centers).

4.3 Data collection

In 2014, DIA2 researchers conducted interviews with members of the STEM research PI community over Skype to garner their reactions to the features in the Spring 2014 version of DIA2. After exploring the functionality, participants were asked: *Is the information presented useful or just interesting? If useful, in what way is it useful? Is there anything else you would like to see?* All interviews were audio-recorded and later transcribed.

4.4 Data analysis

Participants' responses to the three protocol questions of interest were analyzed using three rounds of coding by two researchers; and all rounds were performed using NVivo qualitative analysis software [62]. Codebook, as a compilation of codes and their explanations, was used to ensure coders work with consistent standards. Because of the size of the data set, the first round involved holistically coding [63] one third [64] of the transcripts. Holistic coding is an exploratory approach to coding that attempts to capture the basic themes in the data by absorbing large data chunks at a time (rather than line by line). Participants' response to each protocol question served as a chunk. This resulted in a list of 96 codes. The codes were collapsed into tentative themes surrounding the value of DIA2 and a "DIA2 wish list". The two coders met to discuss the big ideas that emerged from the data, and the contents of the preliminary codebook (i.e., coding guidelines, tentative codes) before moving to the second round of coding.

Thematic analysis [64, 65] was used as the second round of coding. Braun and Clarke [65] describe six phases of thematic analysis: (1) familiarize yourself with the data; (2) generate initial codes; (3) search for themes; (4) review themes; (5) refine and name themes; (6) produce the report. Once fifty of the ninety transcripts had been coded at least once, the coders met to discuss coding choices, discrepancies in selections, and ways to improve the clarity of the codebook. At this point, the codebook was relatively stable (i.e., no new codes were emerging), and the coders coded twenty new transcripts using the stable codebook as part of conducting interrater reliability analysis.

Cohen's Kappa [66] was used to measure the interrater reliability. Cohen's Kappa measures the agreement of raters who classify a set of items into mutually exclusive categories; in this case, "items" are a participant's response to each protocol question. The Kappa-value ranges from 0 to 1 where large numbers mean better reliability; values near or less than zero suggest that agreement is attributable to chance alone. Table 4 includes a list of how Kappa-values might be interpreted [67].

The Kappa-value indicating the extent of agreement between the two coders for every transcript is included in Table 5. Only four out of twenty transcripts coded in this round had a value less than 0.8, but all fell within the "substantial agreement" range. The overall average Kappa-value is 0.9,

Table 4. Interpretation of κ-values

Kappa (κ)	Interpretation
<0	Poor agreement
0.0–0.20	Slight agreement
0.21–0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–1.00	Almost perfect agreement

Table 5. κ-values Corresponding to the Interrater Reliability Analysis

Transcript ID	κ	Transcript ID	κ	Transcript ID	κ	Transcript ID	κ
51	0.92	56	1.00	61	0.97	66	0.96
52	0.78	57	1.00	62	1.00	67	0.89
53	0.91	58	0.82	63	0.70	68	0.94
54	0.86	59	0.73	64	0.96	69	0.95
55	0.98	60	0.92	65	0.78	70	0.96
Average κ -value: 0.90							

which means that the coders had almost perfect agreement by the time two thirds of the data had been coded at least once. The researchers met to discuss differences in coding and to refine the codebook one last time before finishing the coding.

With the final version of the refined codebook, one coder re-coded all 89 transcripts (including recoding the 50 transcripts leading up the stable version of the codebook) to ensure that the coding across all of the transcripts was consistent with final version of the codebook. This was the third and final round of coding. The final codebook includes 10 themes on the value DIA2 to users in the STEM community. Two frameworks helped with interpreting the participants' responses; they will be presented after the results, as part of the discussion of the findings.

5. Results on DIA2's value

The value of a data visualization tool relies on a thorough investigation of users' interactive graphical processes and the impact on their performance [68]. The STEM research community members who participated in this study find DIA2 useful for carrying out teaching, research, and administrative duties. In this study, the themes are written as value propositions—written to reflect the benefit to the user. The results of this study include one value proposition for carrying out teaching responsibilities, five for research-related activities, and four for administration. The section includes three types of results: a table summarizing the value of DIA2, detailed descriptions of each value proposition, and descriptive statistics associated with the value propositions. Participants also shared what changes they would like to see added to future designs of DIA2. These results will be discussed in the next section. Table 6 includes a summary of the DIA2 Value Propositions.

5.1 DIA2's value for teaching

The participants in this study discussed one particular way DIA2 adds value for carrying out teaching-related activities. Specifically, DIA2 adds value when looking for instructional resources related to a particular pedagogy. For example, it can be used to identify NSF-funded projects with outputs related to specific pedagogical resources that can be used in the classroom. One biology associate professor stated it this way:

> "I also might use it to try and find people—I always felt there were lots of good grants and people doing great work, but unless you go to a PI meeting, you don't really know about what's happening. It would be a way to look for resources that other people have developed."

Here is a brief exchange with an assistant professor doing biology education research that also sees

Table 6. Summary of DIA2 Value Propositions

Activities	Value Propositions
Teaching	DIA2 adds value when looking for instructional resources related to a particular pedagogy.
Research	DIA2 adds value as an alternative way to explore research activities apart from scholarly literature.
	DIA2 adds value when trying to understand facets of NSF.
	DIA2 adds value at all stages of writing an NSF proposal.
	DIA2 adds value when disseminating project outputs and helps understand dissemination patterns.
	DIA2 adds value as an instrument for conducting research.
Administration	DIA2 adds value when trying to present, understand and/or evaluate the research profile of an individual researcher.
	DIA2 adds value when gathering information and/or develop dashboards (metrics) to better understand people and activities at the institution.
	DIA2 adds value when developing local programs.
	DIA2 adds value during strategic planning.

value in using DIA2 as a springboard for guidance on developing instructional resources:

> Interviewee: "Let me see. Now I'm looking at this person. I wanna see what his award is. Oh, it's got the little abstract for his grant. That's kinda interesting."
>
> Interviewer: "Why is that interesting?"
>
> Interviewee: "Cause I didn't know that's what he was working on, and I wanted to—he's working on molecular animation, and something that I myself want to do in my course, to develop some animation… I guess what I would do next is I would go and see if he's been publishing anything, and then I might get in touch with him if it looked like something exciting was happening there."

DIA2 can also be used to determine if a particular pedagogy is being used or researched at your institution or within your geographic region. An associate professor of engineering discussed this possibility after exploring the geographical data visualization:

> "The other thing that I did was, of course, click on Alaska. Okay. There are no classes—there are no awards in Alaska that specifically trigger on the key word flipped classroom. That's of interest to me. My particular proposal did not—was not focused on flipped classrooms, but that tells me either that nobody is working on flipped classrooms in the state of Alaska or they're calling it by different things. It's not explicit, right? They're not necessarily doing it. That means that in Alaska maybe I'm the person who should go around and talk to people about flip classrooms. Who knows?"

Lastly, DIA2 can be used as a starting point for identifying and recruiting participants for a study focused on a particular pedagogy. While the first value statement focused on teaching, the next set of value propositions center around using DIA2 as part of conducting research.

5.2 DIA2's value for research

According to the participants in this study, DIA2 adds value at various stages of the research process—from exploring existing research to disseminating research outputs. The participants in this study shared five specific ways DIA2 adds value for conducting research. First, DIA2 adds value as an alternative way to explore research activities apart from scholarly literature. At a minimum, it is a resource that can be used to search and identify "who is doing what" and "who collaborates with whom." More importantly, the tool can be used to identify researchers or projects to explore when shifting to a new research area, starting a literature review, and looking for projects associated with the latest research to cite. DIA2's visualizations allow researchers to infer connections between research topics and NSF-funded projects (based on connections between people or geography), and aids in understanding the research landscape associated with an area of research. It is also a way to obtain basic information about a NSF-funded project of interest; basic information includes names of researchers involved and abstract summarizing the study. This information can serve as a basis for looking for publications related to a topic. Furthermore, it is a resource that provides evidence to support claims about trends in a research area.

Provided is an example of a chemistry professor who discussed several ideas associated with this theme:

> "Well, I think right away the picture showed me how much activity there was, and how connected the activity was. Essentially one, two, three, four, five, six, seven, eight—maybe nine investigations. I can see from the graph that they span the last 20 years. Although in the last decade there's more activity. It gives me the names of the people involved right away. If this were an area that I was highly interested in, this would give me an idea what people have been doing before. It would give me a quick way to see what sort of funded research activity has existed independent of what I might find in the literature. Sometimes people have done work and things don't show up in the literature, particularly in science education. You don't always know what journals to look in to find things. Because it's very broad compared with, say, organic chemistry, or analytical chemistry, which is a much narrower range of journals. . . .Yeah, so when I put—when I run my cursor over each of the little dots I can see who the PIs are. Interesting."

This value proposition focused on exploring scholarship more broadly; the second value proposition associated with research is focused on NSF, a federal funding agency supporting STEM research.

The second value proposition focused on research is DIA2 adds value when trying to understand facets of NSF. More specifically, DIA2 provides a window for discerning the structure of the funding agency and connections between levels of the organizational structure (i.e., directorates, divisions, programs). It is also a means for learning about the Agency's budget associated with various aspects of the organization and inferring the span of reach associated with NSF programs, divisions, and directorates. DIA2 can be used as a tool for understanding the projects in a NSF division, program, and program officers' portfolio of active and legacy grants. Such insights allow you to answer questions about how many grants were funded by a division and the amount of funding invested in a topic. Moreover, insights on NSF's funding trends provide one with the ability to infer the Agency's topical interest. With the data on trends, it is also possible to obtain a yearly breakdown of NSF investments in a topic, understand funding trends in light of economic shifts, and compare funding trends for different topics. Apart from using DIA2 to understand the NSF funding context, the tool is useful for grant writing as well.

Continuing with the connection to NSF, the third research-focused value proposition is DIA2 adds value at all stages of writing a NSF proposal. This includes, but is not limited to, exploring funding NSF mechanisms, conceiving a research proposal in light of what has already been funded, identifying people to connect with for advice on submitting a proposal to a particular program, or request a copy of their proposal in order to use it as a model for the proposal you plan to submit.

One chemistry associate professor mentioned how it would be particularly useful for finding collaborators if one is in a niche area:

> "Okay, so that now—see, this would be useful to find collaborators because the nanotechnology in undergraduate education is a very—it's a niche. It's not a big program, so it's hard to find out who's funded and what they're doing and who they're working with, so this would be useful for finding collaborators rather than just searching for individual people and the funding information."

As it relates to developing the contents of the proposal, DIA2 can be used to identify potential collaborators and partners based on shared interests as well as determine the typical characteristics (e.g., number of collaborators, reasonable budget) of grants comparable to the proposal you plan to submit. DIA2 can also aid in understanding how your research idea is situated within a broader research landscape. Furthermore, insight on what projects have been funded in the past helps with evaluating where to submit a proposal, assessing the likelihood of a proposal being awarded (by getting a sense of what kinds of ideas that seem "fundable"). The next value proposition associated with DIA2's value for research relates to after the study commences.

The fourth value proposition is DIA2 adds value when disseminating project outputs and helps us understand dissemination patterns. DIA2 facilitates networking, because, even before actually engaging in dissemination activities, DIA2 can be used to identify successful PIs to reach out to for advice on effective dissemination. Additionally, it is a tool that can be used to identify people who might be interested in the outputs of your project—whether that may be tangible products, pedagogy, workshops/training materials, etc. Moreover, the visualizations depicting connections between researchers associated with a research area help researchers infer efficient paths for making people aware of your work, communicating information, or propagating outputs.

An associate professor of biology shared the following during the interview:

> Interviewee: "I think network maps can be misinterpreted and can be misleading, but I think they can also be very informative. . . One of the first things I did yesterday when I logged on was I just pulled up one of the network maps. I was really curious who the big hubs were. Where those people I knew? Were they not people I know? Was there a way I could get into that network or hub?"
>
> Interviewer: "Can you expand a bit on your motivation for getting into a big hub?"
>
> Interviewee: "I think it could be a valuable way to make other people aware of what you are doing."

The last value proposition related to research is DIA2 adds value as an instrument for conducting research. For example, DIA2 can be used as a mechanism for performing social network analysis to explore social networks affiliated with NSF investment in R&D. It can also be used to visualize and analyze highly connected and fragmented networks associated with topics, NSF programs, within an institution, and between PIs at different institutions. Likewise, DIA2 can be used as a mechanism for assessing how easy or difficult it is to join a network and provide the preliminary information necessary for developing a strategy for joining it. Additionally, it is useful for creating a profile of a particular type of NSF award (e.g., CAREER, ADVANCE), and conducting comparisons of the impact of NSF divisions, programs, sets of institutions, and PIs.

A psychology assistant professor with an understanding of graph theory and network analysis shared thoughts with respect to the possibility of using DIA2 to figure out ways to get information through networks and to understand networks associated with NSF-funded projects in general:

> "When I look at these TUES projects, this global network is not very connected, right? Information can't flow through this network very well. There's this highly connected component here. When I look at it by concept and I turn the active awards to off, this is probably a healthier network, right? It's possible for information to get through this network in a short series of steps. The path link—there is probably an average path link that could get you to most places in this network, or at least a good number of them. When I turn the active awards only on, it becomes much more fragmented. What I can see is that the number of awards are being dominated by these highly connected components of people, right?"

An associate professor of chemistry gave a specific example of how it would have been useful when writing a review of a NSF program:

> "I just realized this would have been really useful for a—I wrote a review of an NSF program, actually, let me try that. . . It's really hard to find budget information about NSF programs, and so this—I don't need that information a lot, but I recently did need it, and it was really hard [laughter] to dig through NSF's web-

site, and the data was not entirely accurate anyway. This could be nice."

The last set of value propositions relate to DIA2's usefulness when carrying out administrative duties.

5.3 DIA2's value for administration

There are four value propositions related to carrying out a variety of administrative activities. Collectively, the information might be useful as part of promotion and tenure decisions, forming collaborations for projects apart from writing grants, identifying candidates for job openings based on their expertise and funding history, making recruitment and hiring decisions, and identifying potential advisory board members. The first of four in this set is DIA2 adds value when trying to present, understand and/or evaluate the research profile of an individual researcher. At a minimum, using DIA2 is an efficient way to identify individuals and networks of people with similar research interests. Additionally, the information DIA2 presents about a PI provides a meaningful way to present and understand their NSF awards portfolio—namely insights about their research interest, funding history, institutional affiliations, names of other people in their network, and the names of program officers who have funded their NSF awards. The information DIA2 presents also provide insights on how someone's career changes over time, indicators of people's productivity, and allows for inferences about the extent of an individual's influence in a research community. Apart from focused on a single individual, DIA2 can assist with seeing people in the context of their peers and for setting benchmarks of "success" at various stages along the tenure track.

One assistant professor of psychology reflected on how she would use the tool if she were on a promotion and tenure committee:

> "Yeah, it's certainly telling me a little bit about the impact of the researcher as a highly connected component of a particular network of researchers. I don't think that that's typically how impact is evaluated by a tenure and promotion committee, but I would certainly—if I was on a tenure and promotion committee, consider this to be evidence of impact, at least to the extent that I'm influential over other people with my ideas and I'm good at collaborating with other people."

While in an administrative role, it is important to set organizational goals and plans to achieve them. Oftentimes, this also requires establishing metrics for determining the extent to which goals have been met. In light of this, DIA2 adds value when gathering information and/or developing dashboards (metrics) to better understand groups of people and activities at the institution. One example of this includes using DIA2 for a quick assessment of the yearly breakdown of funding awarded to an institution or within a state as part of developing an institutional profile, including insights on the amount and types of people receiving NSF funding. It is also useful for inferring an institution's research foci and the areas of expertise among the researchers—based on the NSF-funded research conducted there. Additionally, it is possible to use DIA2 as a tool for beginning to map institutional change stemming from research activities. Apart from using DIA2 to understand research activities and people at your institution, it can also be used to obtain summative information about activities at comparable institutions. Said differently, the insights DIA2 provides can help with situating the work of your institutions in the context of similar programs. These are examples of the ways DIA2 can be used to understand groups of people and activities at the institution. Such insights are useful to a variety of people at academic institutions, such as faculty, administrators, job seekers, NSF personnel, and even public officials.

Provided is an example from the director of a research center discussing how the tool is useful for understanding the work of an institution in the context of similar institutions:

> "It's all about numeracy, right? This is simply showing me numbers. It's showing me award amount, and it's showing me number of PIs. One thing I can quickly do, and I'll do it while I'm on the phone, I can quickly compare universities. I'm gonna look up University of California-San Francisco. Now I'm looking at UC Berkeley versus UC San Francisco. I can see the number of PIs; the award amount. The University of San Francisco has $73 million and 99 PIs for NSF. Where UC Berkeley has $1.4 billion and 1,400 PIs. . . I think it can give you a good snapshot of the NSF's impact at a university and also, you can compare universities and see what it is. I'm very interested if this goes beyond universities. For instance, and stop me if I'm talking too much because this is pretty interesting, I'm trying to look at Museum of Science. . . I'm gonna do Science Museum Minnesota and Museum of Science which is in Boston. Let's see. Immediately from here, I can see these are very similar institutions. They have about $40 million of NSF money. Science Museum Minnesota has 61 PIs versus 47 at Museum of Science. It looks like they've got about the same amount of awards."

This is one of several examples participants shared on how DIA2 can assist with understanding the research activities of groups and institutions. The last two value propositions also focus on DIA2's value for completing administrative tasks but extend beyond addressing informational needs to be an integral part of planning and decision-making.

One part of administration at an academic institution is implementing programs and initiatives at your local institution. DIA2 adds value when devel-

Table 7. Descriptive Statistics Surrounding DIA2 Value Propositions

Number of Participants Who Stated DIA2 Adds Value	
. . . when developing local programs (A)	3
. . . during strategic planning (A)	4
. . . when disseminating project outputs and helps us understand dissemination patterns (R)	6
. . . when looking for people, research, or resources for the classroom or related to a particular pedagogy (T)	6
. . . as an instrument for conducting research (R)	12
. . . when trying to understand NSF (R)	33
. . . when gathering information and developing dashboards to better understand people and activities at an institution (A)	35
. . . when trying to present, understand, or evaluate the research profile of an individual researcher (A)	41
. . . as an alternative way to explore research activities apart from scholarly literature (R)	43
. . . at all stages of writing an NSF proposal (R)	60

oping local programs. A chemistry associate professor offered a concrete example of how DIA2 might help with developing a local program.

> "The other is, again, more for—I've been thinking a lot about, as well as the AAU, about cultural change at your university. This informs me that these people on this list are probably some like-minded people, and I can go ahead and try to form maybe a faculty learning community with these people, or things like that."

Apart from leveraging insights about social networks within an institution to form learning communities, DIA2 users can use it to identify groups of people to connect with when designing other campus programs. Using insights from DIA2's geographical visualizations, these programs might be developed not only for people at the home institution but also with the intention of inviting people from nearby institutions. At a minimum, it can be used to identify researchers to invite to give seminars on campus. While the focus of this value proposition was on local activities, the last value proposition focuses on long-term planning.

People operating at various levels of administration will find that DIA2 adds value during strategic planning. Several participants mentioned concrete examples of how they might use the tool in this way. For example, DIA2 can be used to identify comparable institutions to serve as models to learn from when setting state-level funding and research goals. Another mentioned the possibility of using the NSF funding trends as indicators of national research priorities, and one engineering professor and department chair mentioned using insights on NSF investments "to think about where to invest" their departments' future investments in research. Others talked about using the funding patterns associated with their institution to determine which NSF program officers to meet with during an upcoming visit to NSF. One participant noted how DIA2 might add value for strategic planning at NSF—namely, insights on the geographic distribution of funding can assist program officers at the Agency with determining where to focus future investments.

5.4 Descriptive statistics on DIA2's value

The next set of quantitative results summarizes which participants cited each value proposition. More specifically, Figure 1 provides a global perspective across all ten themes. The subsequent tables reveal patterns based on demographics (i.e., academic rank, discipline).

Figure 1 depicts how often a participant mentioned a particular value proposition. The letter in parenthesis indicates whether the value proposition was associated with teaching (T), research (R), or administration (A). Again, five value propositions relate to research activities, four to administrative responsibilities, and one relates to teaching. According to the figure, users tend to perceive that DIA2 adds more value for research and administrative tasks, rather than carrying out teaching responsibilities. Users find DIA2 most valuable throughout the process of developing an NSF proposal; very few people mentioned DIA2's usefulness as part of developing programs at their local institution.

6. Interpretation of the Results

Table 8 includes the detailed mappings. Gephi, an open source software [69], was used to generate this network graph. A force-directed algorithm [70, 71] was chosen to layout the data in order to help identify the important aspects of DIA2's value according to the framework. Value propositions are visualized as outlined circles and labeled using abbreviations. (The numbers correspond to the order in which they were presented in the results.) Affordance constructs are depicted using solid black nodes; the size of the label is proportional to the nodes' degree, or the number of connections.

There are a variety of interpretations to garner from this visualization. First, what should become apparent immediately is DIA2's ability to enable immediate access to rapidly changing information about NSF funded projects and to stimulate reflective thinking as a result of visualizing insights associated with this set of projects. This result highlights the importance of up-to-date information to

Table 8. Mappings Between Value Propositions and Affordances Framework Constructs

Abbreviation	DIA2 Value Propositions	Accessibility	Speed of Change	Diversity	Communication Collaboration	Reflection	Multimodal and non-linear	Immediacy	Surveillance
Teaching	DIA2 adds value when looking for instructional materials related to a particular pedagogy.	X	X	X			X	X	X
Research_1	DIA2 adds value as an alternative way to explore research activities apart from scholarly literature.	X	X			X	X	X	X
Research_2	DIA2 adds value when trying to understand facets of NSF.	X	X			X	X	X	
Research_3	DIA2 adds value at all stages of writing an NSF proposal.	X	X	X	X	X	X	X	
Research_4	DIA2 adds value when disseminating project outputs and helps us understand dissemination patterns.		X		X			X	X
Research_5	DIA2 adds value as an instrument for conducting research.			X	X	X	X	X	
Administration_1	DIA2 adds value when trying to present, understand and/or evaluate the research profile of an individual researcher.	X	X			X		X	X
Administration_2	DIA2 adds value when gathering information and/or develop dashboards (metrics) to better understand people and activities at the institution.	X	X			X	X	X	X
Administration_3	DIA2 adds value during strategic planning.					X	X	X	X
Administration_4	DIA2 adds value when developing local programs.					X	X	X	X

the research in the STEM field. It implies that the funding agencies and policymakers should strengthen the communication with scholarly communities and notify researchers the last information in time.

Furthermore, "Communication and Collaboration" and "diversity" are not the primary affordances of DIA2, although it does provide a starting place for identifying with whom to communicate and collaborate for research endeavors. Some popular Internet platforms, such as Google and Microsoft, may provide search engines for researchers, but they are not specialized in information retrieval related to scientific funding. Future research may explore how to integrate the advantages of DIA2 and other Internet platforms together as well as propose the best practice to better serve scholars' daily work.

Apart from connections to the framework, there are several connections between the value propositions the participants in this study shared and the current literature on scholarly communications in the digital age. For example, Harley et al. [23] concluded that the lack of time and "noise" filters are faculty members' biggest obstacles associated with disseminating research. DIA2 can help with addressing both of these challenges. More specifically, although DIA2 cannot give faculty more time for conducting research, it is a tool that can help faculty perform some routine activities more efficiently and, thus, free up more time for conducting research. Furthermore, because all NSF-funded projects go through a rigorous peer-review process before getting funded and DIA2 is a search and visualization platform presented insights about the NSF portfolio, the filtering faculty desire should be largely addressed, and there should be little "noise" associated with research DIA2 presents.

Another connection to existing literature relates to how scholars stay up-to-date in their field and tend to collaborate within a circle of trusted friends [13]. Participants' perception of DIA2's value as a starting place for establishing collaborations is highly informative. These results might add a nuance to previous findings [13]: scholars are more willing to connect with people outside of their trusted circles of peers or mentors if presented with reliable information regarding the degrees of separation between a familiar researcher and the potential collaborator (via DIA2's ego-centric circles) and visualizations depicting connections between people based on mutual interests. Based on the current findings, DIA2 may enhance scholars' networks of collaborators in novel ways.

7. Conclusion

DIA2 is a web-based tool that provides access to timely insights about NSF investments in STEM education research. This qualitative study includes the perspectives of STEM research community members on the value of DIA2 and reveals that this cyberinfrastructure is a valuable asset for carrying out teaching, research, and administrative duties. Tools like DIA2 represent a new, innovative genre of tools that can be added to the suite of traditional methods used to facilitate scholarly communications in the digital age. With use of the insights DIA2 provides, the STEM community is better equipped to leverage and magnify the impact of federally supported STEM education R&D.

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, NSF at a Glance, https://www.nsf.gov/about/glance.jsp, accessed 27 December 2017.
2. A. F. Mckenna, J. Froyd and T. Litzinger, The complexities of transforming engineering higher education: Preparing for next steps, *Journal of Engineering Education*, **103**(2), 2014, pp. 188–192.
3. The Economist, The data deluge: Businesses, governments and society are only starting to tap its vast potential, http://www.economist.com/node/15579717, accessed 27 December 2017.
4. Time, The Wireless Issue, *Time*, **180**(8), 2012.
5. J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. H. Byers, Big data: The next frontier for innovation, competition, and productivity, *Analytics*, June, 2011, p. 156.
6. S. Interests and P. Sciences, *Big Data: A Workshop Report*, 2012.
7. C. Demmans Epp and S. Bull, Uncertainty Representation in Visualizations of Learning Analytics for Learners: Current Approaches and Opportunities, *IEEE Transactions on Learning and Technologies*, **8**(3), 2015, pp. 242–260.
8. J. Cummings, T. Finholt, I. Foster, C. Kesselman and K. A. Lawrence, *Beyond Being There: A Blueprint for Advancing the Design, Development, and Evaluation of Virtual Organizations*, Office of Cyberinfrastructure, National Science Foundation, Arlington, VA, 2008.
9. P. B. Heidorn, Shedding Light on the Dark Data in the Long Tail of Science, *Library Trends*, **57**(2), 2008, pp. 280–299.
10. T. Auvinen, L. Hakulinen and L. Malmi, Increasing Students' Awareness of Their Behavior in Online Learning Environments with Visualizations and Achievement Badges, *IEEE Transactions on Learning Technologies*, **8**(3), 2015, pp. 261–273.
11. H. Morrison, Scholarly communication for librarians, *Chandos Information Professional Series*, Elsevier, 2009, pp. 103–134.
12. D. Harley, Scholarly communication: cultural contexts, evolving models, *Science*, **342**(6154), 2013, p. 80.
13. S. K. Acord and D. Harley, Credit, time, and personality: The human challenges to sharing scholarly work using Web 2.0, *New Media Society*, **15**(3), 2012, pp. 379–397.
14. W. Wei, *Scholarly Communication in Science and Engineering Research in Higher Education*, Routledge, 2013.
15. C. C. Bonwell and J. A. Eison, Active Learning: Creating Excitement in the Classroom, *ProQuest Dissertations and Theses*, 2008, p. xviii, 495.
16. The Guardian, Harvard University says it can't afford journal publishers' prices, http://www.theguardian.com/science/2012/apr/24/harvard-university-journal-publishers-prices, accessed 27 December 2017.
17. R. E. Abel, L. W. Newlin, Books, P, Francis, J, Edelman, H, Follett, and Oda R J R, Scholarly Publishing: Books, Journals, Publishers, and Libraries in the Twentieth Century, *Journal of Scholary Publication*, **34**(3), 2002, pp. 1–8.
18. J. L. Contreras, Confronting the crisis in scientific publishing: latency, licensing, and access, *Santa Clara Law Review*, **53**, 2013, pp. 491–560.
19. T. W. Mason and A. B. Western, The Engineer as Entrepreneur: Education for the 21st Century at Rose-Hulman Institute of Technology, *American Society of Engineering Education Annual Conference and Exposition*, 2001.
20. C. N. Davidson, The Futures of Scholarly Publishing, *Journal of Scholarly Publishing*, **35**(3), 2004, pp. 129–142.
21. J. Willinsky, The future of scholarly publishing, *Current Sciety*, **84**(7), 2003, pp. 873–874.
22. L. Withey, S. Cohn, E. Faran, M. Jensen, G. Kiely, W. Underwood, B. Wilcox, R. Brown, P. Givler, A. Holzman, K. Keane and N. A. Press, Sustaining Scholarly Publishing: New Business Models for University Presses The Association of American University Presses, *Journal of Scholarly Publishing*, **42**(4), 2011, p. 37.
23. D. Harley, S. Earl-Novell, J. Arter, S. Lawrence and C. J. King, The Influence of Academic Values on Scholarly Publication and Communication Practices, *Journal of Electronic Publishing*, **10**(2), 2007.
24. S. Gunelius, Study Finds Online Mentions of Scholarly Documents Affects Research Impact, http://aci.info/2015/02/27/study-finds-online-mentions-of-scholarly-documents-affects-research-impact/, Accessed 27 December 2017.
25. S. Gunelius, Blogs and Social Media Play Important Roles in Helping Scientists Engage the Public, http://aci.info/2015/02/16/blogs-and-social-media-play-important-roles-in-helping-scientists-engage-the-public/, Accessed 27 December 2017.
26. R. Procter, R. Williams, J. Stewart, M. Poschen, H. Snee, A. Voss, and M. Asgari-Targhi, Adoption and use of Web 2.0 in scholarly communications, *Physical and Engineering Science*, **368**(1926), 2010, pp. 4039–4056.
27. C. G. Wood and B. C. Leibovich, 101 Innovations in Scholarly Communication—the Changing Research Workflow, Force, 2015.
28. M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma and F. Herrera, Science Mapping Software Tools: Review, Analysis, and Cooperative Study Among Tools, *Journal of American Society for Information Science and Technology*, **62**(7), 2011, pp. 1382–1402.
29. K. Madhavan, N. Elmqvist, M. Vorvoreanu, X. Chen, Y. Wong, H. Xian, Z. Dong and A. Johri, DIA2: Web-based cyberinfrastructure for visual analysis of funding portfolios, *IEEE Transactions on Visisualization and Computer Graphics*, **20**(12), 2014, pp. 1823–1832.
30. B. K. Jesiek, M. Borrego, K. Beddoes, M. Hurtado, P. Rajendran and D. Sangam, Mapping global trends in engineering education research, 2005–2008, *International Journal of Engineering Education*, **27**(1), 2011, pp. 77–90.
31. A. L. Fleury, H. Stabile and M. M. Carvalho, An Overview of the Literature on Design Thinking: Trends and Contributions, *International Journal of Engineering Education*, **32**(4), 2016, pp. 1704–1718.
32. G. B. Da Silva, H. G. Costa and M. D. De Barros, Entrepreneurship in Engineering Education: A Literature Review, *International Journal of Engineering Education*, **31**(6), 2015, pp. 1701–1710.
33. CiteSpace II. http://cluster.cis.drexel.edu/~cchen/citespace/, accessed 27 December 2017.
34. C. Chen, Visual Analytics, *Mapping Scientific Frontiers: the Quest for Knowledge Visualization*, 2014, pp. 321–339.
35. C. Chen, CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature, *Journal of American Society for Information Science and Technology*, **57**(3), 2006, pp. 359–377.

36. J. Folkstad and S. C. Hayne, Visualization and Analysis of Social Networks of Research Funding, *2011 44th Hawaii International Conference on System Science*, 2011, pp. 1–10.
37. Pajet, http://vlado.fmf.uni-lj.si/pub/networks/pajek/, accessed 27 December 2017.
38. Jigsaw, http://www.cc.gatech.edu/gvu/ii/jigsaw/, accessed 27 December 2017.
39. Y. A. Kang and J. Stasko, Examining the Use of a visual analytics system for sensemaking tasks: Case studies with domain experts, *IEEE Transactions on Visualization and Computer Graphics*, **18**(12), 2012, pp. 2869–2878.
40. H. Small, K. W. Boyack and R. Klavans, Identifying emerging topics in science and technology, *Research Policy*, **43**(8), 2014, pp. 1450–1467.
41. K. W. Boyack, M.-C. Chen and G. Chacko, Characterization of the Peer Review Network at the Center for Scientific Review, National Institutes of Health, *PLoS One*, **9**(8), 2014, p. e104244.
42. J. P. A. Ioannidis, K. W. Boyack and R. Klavans, Estimates of the continuously publishing core in the scientific workforce, *PLoS One*, **9**(7), 2014.
43. DIA2, http://www.dia2.org/, Accessed 27 December 2017.
44. P. E. Bourne, T. W. Clark, R. Dale, A. De Waard, I. Herman, E. H. Hovy and D. Shotton, Improving The Future of Research Communications and e-scholarship, *Informatic-Spectrum*, **35**, 2012, pp. 56–58.
45. Q. Liu, M. Vorvoreanu, K. P. C. Madhavan and A. F. McKenna, Designing discovery experience for big data interaction: A case of web-based knowledge mining and interactive visualization platform, *International Conference on Design, User Experience, and Usability: Web, Mobile, and Product Design*, **8015**(4), 2013, pp. 543–552.
46. M. Vorvoreanu, K. Madhavan, K. Kitkhachonkunlaphat and L. Zhao, Designing for STEM Faculty: The Use of Personas for Evaluating and Improving Design, in *Digital Human Modeling: Applications in Health, Safety, Ergonomics and Risk Management*, 9745, Springer International Publishing Switzerland, 2016, pp. 369–380.
47. K. P. C. Madhavan, M. Vorvoreanu, N. Elmqvist, A. Johri, N. Ramakrishnan, G. A. Wang and A. McKenna, Portfolio mining, *Computer*, **45**(10), 2012, pp. 95–99.
48. A. McKenna, J. London, A. Johri, M. Vorvoreanu and K. Madhavan, Developing and advancing a cyberinfrastructure to gain insights into research investments: An organizing research framework, *Proceedings of the 122nd ASEE Annual Conference and Exposition*, 2015.
49. A. Molnar, A. F. McKenna, Q. Liu, M. Vorvoreanu and K. Madhavan, Using Visualization to Derive Insights from Research Funding Portfolios, *IEEE Computer Graphics and Applications*, **25**(3), 2015, pp. 6–12.
50. Y. S. Lincoln and E. G. Guba, *Naturalistic Inquiry*, Sage, Beverly Hills, CA, 1985.
51. M. D. Myers, *Qualitative Research in Business & Management*, SAGE Publications, 2008.
52. J. D. Hollan, E. Hutchins and D. Kirsh, Distributed cognition: toward a new foundation for human-computer interaction research, *ACM Transations on Computer-Human Interaction*, **7**(2), 2000, pp. 174–196.
53. E. Hutchins, *Cognition in the Wild*, MIT Press, Cambridge, MA, 1994.
54. G. Conole and M. Dyke, What are the affordances of information and communication technologies?, *Reserch in Learning Technology*, **12**(2), 2004, pp. 113–124.
55. J. J. Gibson, The theory of affordances, in R. Shaw and J. Bransford (eds), *Perceiving, acting, and knowing*, Lawrence Erlbaum, Hillsdale, 1977.
56. C. A. Halverson, Activity theory and distributed cognition: Or what does CSCW need to DO with theories?, *Computer Supported Cooperative Work*, **11**(1–2), 2002, pp. 243–267.
57. Z. Liu, N. Nersessian and J. Stasko, Distributed cognition as a theoretical framework for information visualization., *IEEE Transations on Visualization and Computer Graphics*, **14**(6), 2008, pp. 1173–1180.
58. S. K. Card, J. D. Mackinlay and B. Shneiderman, *Readings in information visualization: using vision to think*, Morgan Kaufmann Publishers Inc., San Francisco, CA, 1999.
59. C. North, IEEE Computer Graphics and Applications, *IEEE Computer Graphics and Applications*, **26**(3), 2006, pp. 6–9.
60. J. S. Yi, Y. Kang, J. T. Stasko and J. a Jacko, Understanding and characterizing insights: How Do People Gain Insights Using Information Visualization?, *Beliv 2008—Beyond Time and Errors: Novel Evaluation Methods for Information Visualization, A Workshop of the ACM CHI 2008 Conference*, Florence, Italy, April, 2008, p. 1.
61. G. Conole, Research questions and methodological issues, in J. Seale (eds), *Learning technology in transition: from individual enthusiasm to institutional implementation*, 2004.
62. P. Bazeley and L. Richards, *The NVivo Qualitative Project Book*, Sage Publications, Thousand Oaks, CA, 2009.
63. J. Saldaña, *The Coding Manual for Qualitative Researchers*, 3rd edn, Sage Publications Ltd, Washington, DC, 2016.
64. R. E. Boyatzis, *Transforming Qualitative Information: Thematic Analysis and Code Development*, Sage Publications, Thousand Oaks, CA, 1998.
65. V. Braun and V. Clarke, Using thematic analysis in psychology, *Qualitative Research in Psychology*, **3**(2), 2006, pp. 77–101.
66. J. Cohen, A coefficient of agreement for nominal scales, *Educational and Psychologicla Measurement*, **20**(1), 1960, pp. 37–46.
67. J. R. Landis and G. G. Koch, The Measurement of Observer Agreement for Categorical Data, *Biometrics*, **33**(1), 1977, pp. 159–174.
68. A. Shoufan, Z. Lu and S. A. Huss, A web-based visualization and animation platform for digital logic design, *IEEE Transactions on Learning Technologies*, **8**(2), 2015, pp. 225–239.
69. M. Bastian, S. Haymann and M. Jacomy, Gephi: An Open Source Software for Exploring and Manipulating Networks, *the Third International AAAI Conference on Weblogs and Social Media*, 2009.
70. T. M. J. Fruchterman and E. M. Reingold, Graph Drawing by Force-directed Placement, *Software-Practice and Experience*, **21**(11), 1991.
71. S. G. Kobourov, Spring Embedders and Force Directed Graph Drawing Algorithms, *Computer Sciences*, 2012, pp. 1–23.

Jeremi S. London, PhD, holds BS and MS degrees in Industrial Engineering, and a PhD in Engineering Education all from Purdue University. London is currently an Assistant Professor in The Polytechnic School, one of six Schools in the Ira A. Fulton Schools of Engineering at Arizona State University. Prior to her PhD, London 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. London worked in the National Science Foundation's Division of Undergraduate Education on research and evaluation projects related to the use of cyberlearning 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 and modeling tools to evaluate educational programs. She holds leadership positions in the American Society for Engineering Education.

Andreea Molnar, PhD, holds a BSc And MSc in Computer Science, a MSc in Modeling and Simulation, and a PhD in Technology and Enhanced Learning. Molnar is currently a Senior Lecturer at the University of Portsmouth in the United Kingdom. Prior to joining the University of Portsmouth, Molnar was a Postdoctoral Research Associate and Faculty Associate at Arizona State University; Research Fellow at Brunel University; Research Associate at City University-London; Research Intern at Telefonica I+D; and Software Developer at Silnet and Fortech SRL. Molnar has research interest in the use of technology to support teaching and learning.

Ann McKenna, PhD, 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 BS and MS degrees in Mechanical Engineering from Drexel University and PhD from the University of California at Berkeley.

Wen Huang, is a PhD Student in the Engineering Education Systems and Design program at Arizona State University. Prior to studying in Arizona State University, he worked in the Chinese Academy of Sciences and has taken part in many cyberlearning system development projects. He has interest in technology-supported STEM learning.

Samantha Cruz, MA, is a PhD student in Counseling Psychology at Arizona State University and previously earned her BA in Psychology and MA in Counseling Psychology. She has previously conducted research investigating the impact of diversity interventions on campus climate as well as the impact of racial discrimination on Latina/o adolescents' academic outcomes. Her research interests center on examining racial and ethnic disparities in education as well as resources to promote the academic success of students from marginalized backgrounds.