Advances in Engineering Education

SPRING 2016

Guest Editorial: Data Sharing in Engineering Education

ADITYA JOHRI
George Mason University
Fairfax, VA

AND

MIHAELA VORVOREANU

AND

KRISHNA MADHAVAN
Purdue University
West Lafayette, IN

In recent years, there is widespread recognition of the need to share research data among stakeholders to maximize the impact and utilization of publicly funded resources and to improve the validation and reproducibility of research work. Consequently, there have been many calls by funding agencies and other entities to make research more "open". Many steps have been taken in this direction. Several agencies that fund research now require that data be shared among researchers. While some others have not made data sharing a requirement, steps have been taken towards this end goal (e.g. NSF's *Data Management Plan* and *Public Access Plan*[1]).

Sharing research data has multiple advantages and in some disciplines it is hard to make scientific progress without sharing data. In natural sciences - particularly in physical and biological sciences – this is often the case. High energy physicists have made data sharing a requirement to maintain any level of research currency in their field. Even in engineering disciplines sharing data is a pre-requisite for advancements in the field. For example, geo-informatics and hazard prevention rely on researchers sharing data at multiple levels. Sharing research data is a complex endeavor that touches on many sensitive issues such as data privacy, confidentiality, safety of data, management and access, and policy. Furthermore, the propensity of researchers to share data and to use data shared with them depends to a large extent on the norms of the community and the disciplinary culture of research. For those interested in learning more about data sharing we refer them to our ASEE article (Allameh et al., 2015) that covers some significant background on this topic.

No doubt, recent increase in interest on data sharing is also an artifact of increased use of information technology in research. As a growing number of scientific methodologies and processes become digitized, an increasing amount of digital data also gets created with the potential to increase

1 http://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf

data sharing activities. This shift also signals an opportunity to share data through easier-to-use infrastructures that can be scaled. Furthermore, if data is in a digital format, the process becomes easier. Although data sharing has become common across many scientific disciplines, it is still far from the norm in engineering education. In order to better understand the extent of data sharing, specifically the practices of those who shared data we invited a range of scholars across the spectrum to discuss their ideas and experiences related to data sharing in the form of case studies. We also invited a few cases from outside the engineering education community but which exemplify data sharing practices, especially as related to learning and education and social science data. All papers underwent rigorous peer review process and the finalized volume consists of eight papers, plus a ninth paper by the guest editors and a colleague in addition to this introduction.

In their paper, Walther et al explore the challenges and opportunities for sharing qualitative data within engineering education research. The authors argue that interpretive research methods are increasingly being adopted by the community and therefore it is essential to discuss some of the epistemological assumptions that undergird the call to data sharing. They also discuss possibilities for secondary analysis. They argue that interpretive data provides unique challenges such as the richness of available and potentially relevant explicit details, the context- and purpose-dependent relevance of such details to the primary study and the sometimes tacit nature of factors that significantly shape the data set. The unintended benefit though of trying to share interpretive data for collaborative research is that the purposeful attention to and documentation of the research context and process improves overall quality. This is a potential benefit for the engineering education community as data sharing can offer an interesting perspective on how we, as a research community, communicate quality considerations and efforts in the context of traditional publications.

Cheville presents insights from the perspective of a researcher who has spent time at NSF as a program officer in his article to explicate how the policy and research communities view data sharing differently. He also discusses how data sharing can be something around which these two communities can coalesce. He argues that the role of data sharing is growing in a policy landscape with increased value being placed on program evaluation but there are legal, regulatory, and ethical constraints to sharing data. He outlines three value propositions for data sharing: help tell stories, create usable ontologies, and support networks. Finally, he concludes by discussing the question of data at scale and ethical and philosophical issues that underlie any future attempts to share data.

Adams et al. discuss their experience of designing a shared dataset related to design thinking. The intended purpose of the process of creating the dataset and the shared data was to catalyze discussions among researchers. From their experience, the authors synthesize the following lessons learned for data sharing: It is important to engage various stakeholders early and often, adopting an iterative human-centered approach to designing data sharing systems; it is important to pay attention to misalignments across the data sharing system; Include data use agreements as part of the human subjects review process; Be open to the ways data sharing can create learning partnerships; and keep in mind that constraints of large scale data sharing may require new approaches to translating data into shareable data.

The paper by Trevelyan discusses the challenges of sharing data just within a confined environment of a research group. The barriers in this case spring from aligning student interests with the larger project and ensuring that data collected across sites is compatible in terms of analysis. The paper outlines the challenges of maintaining sufficient qualitative data consistency and quality in student research projects so that their data can inform a much broader analysis. The paper also illustrates how student contributions extended well beyond the gathering of research data.

Ohland and Long reflect on their experience of designing, populating, and using one of the largest datasets of its kind within the engineering education community. The Multiple-Institution Database for Investigating Engineering Longitudinal Development or MIDFIELD dataset has a history spanning more than 16 years. They reflect on the process of building trust around data collection and data sharing through the use of a clear set of policies for how data are handled in publications; reciprocity so that partner institutions benefit before and after they submit data to the project; and, flexibility in terms of accepting data in institution's native format so that the burden for an institution to provide data is significantly reduced. They argue that adopting established data management practices enhances scholarly trustworthiness, improves research access, and lessens the burden of maintaining the dataset. Finally, they suggest that having a plan for expanding participation, expanding access, and sustaining the database are all important strategies that promote data sharing.

The paper by Brogan et al. discusses the development and implementation of a system that generates and then analyzes a lot of data. Although the system does not directly create data for research as we traditionally understand it, the tool is targeted more towards education and learning about watershed systems. It is precisely the kind of system that can have a data layer easily designed on top of it. It is an example of an engineered system that serves dual purposes – technical research coupled with educational practice and research. This project demonstrates the promise that the generation of digital data traces holds for improving engineering learning through reflection by students who use the system as well as engineering educators who have designed and implemented it.

In their paper Gilmore, Adloph & Millman present an overview of Databrary (databrary.org) a digital library for videos they have created with the explicit purpose of supporting data sharing among researchers. Their primary target population is researchers working on learning and developmental issues where video data is commonly used. Videos of learners across a range of settings provide a rich form of data but since videos contain information about participants' identities, it is challenging to share the data. Researchers also face technical challenges that have to do with the large size of video files, diversity of formats, and incompatible software tools. Their article describes how they were able to overcome barriers to sharing of research videos and associated data and their project can serve as a model for other similar efforts.

Toye et al reflect on their experiences with the Academic Pathway Study (APS) research program within NSF's Center for Advancement of Engineering Education (CAEE). From 2003 – 2010 as part of this effort researchers collected a multitude of longitudinal as well as episodic data sets of varying types. In their paper Toye et al discuss the complexities of working on shared datasets within the context of a large project team. In particular, the paper describes how they scaled up as they added new expertise and as the interests of the project team changed over time. The APS leadership recognized early on that effective management of its data collection, data sharing and data reuse would be essential to the success of their collaborative research effort, and ensured that they had guidelines in place for data sharing and re-use.

AUTHORS

Aditya Johri is an Associate Professor in the Department of Information Sciences & Technology in the Volgenau School of Engineering, George Mason University, Fairfax, VA, USA. He studies the use of information and communication technologies (ICT) for learning and knowledge sharing, with a focus on cognition in informal environments. He also examines the role of ICT in supporting distributed work among globally dispersed workers and in furthering social development in emerging economies. He is a co-editor of the Cambridge Handbook of Engineering Education Research (CHEER), Cambridge University Press (2014). He can be reached at johri@gmu.edu. More information about him is available at: http://mason.gmu.edu/~johri

Mihaela Vorvoreanu is an associate professor in Computer Graphics Technology at Purdue University, where she leads the Human Centered Design and Development concentration. Her research focuses on user experience design, as it specifically applies to creating useful and useable systems. Dr. Vorvoreanu leads the UX research and design on DIA2.com, an NSF-funded project for knowledge mining and visualization. Other research interests include UX education and the impact of new communication technologies such as social media on culture, society, and organizations.

Krishna Madhavan is an associate professor in the School of Engineering Education. In 2008 he was awarded an NSF CAREER award for learner-centric, adaptive cyber-tools and cyber-environments using learning analytics. He leads a major NSF-funded project called Deep Insights Anytime, Anywhere (http://www.dia2.org) to characterize the impact of NSF and other federal investments in the area of STEM education. He also serves as co-PI for the Network for Computational Nanotechnology (nanoHUB.org) that serves hundreds of thousands of researchers and learners worldwide. Dr. Madhavan served as a Visiting Researcher at Microsoft Research (Redmond) focusing on big data analytics using large-scale cloud environments and search engines. His work on big data and learning analytics is also supported by industry partners such as The Boeing Company. He interacts regularly with many startups and large industrial partners on big data and visual analytics problems. He was one of 49 faculty members selected as the nation's top engineering educators and researchers by the U.S. National Academy of Engineering to the Frontiers in Engineering Education symposium.