



"This is the peer reviewed version of the following article which has been published in final form at <http://dx.doi.org/10.1002/pra2.2016.14505301014>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving."

**McCoy, C., Marcinkowski, M., Sawyer, S., Sanfilippo, M.R., Meyer, E.T. and Rosenbaum, H. (2016) 'Social informatics of data norms', *Proceedings of the Association for Information Science and Technology*, 53 (1), pp. 1-4.**

## ResearchSPAce

<http://researchspace.bathspa.ac.uk/>

This version is made available in accordance with publisher policies.  
Please cite only the published version using the reference above.

Your access and use of this document is based on your acceptance of the ResearchSPAce Metadata and Data Policies, as well as applicable law:-

<https://researchspace.bathspa.ac.uk/policies.html>

Unless you accept the terms of these Policies in full, you do not have permission to download this document.

This cover sheet may not be removed from the document.

Please scroll down to view the document.

# Social Informatics of Data Norms

**Chase McCoy**

Indiana University  
1320 E. 10th Street  
Bloomington, IN 47405-3907  
chamccoy@indiana.edu

**Michael Marcinkowski**

Bath Spa University  
Corsham Court, Corsham  
Chippenham, SN13 0BZ  
m.marcinkowski@bathspa.ac.uk

**Steve Sawyer**

Syracuse University  
344 Hinds Hall  
Syracuse, NY 13244  
ssawyer@syr.edu

**Madelyn R. Sanfilippo**

Indiana University  
1320 E. 10th Street  
Bloomington, IN 47405-3907  
mhomuth@indiana.edu

**Eric T. Meyer**

University of Oxford  
1 St Giles  
Oxford OX1 3JS United Kingdom  
eric.meyer@oii.ox.ac.uk

**Howard Rosenbaum**

Indiana University  
1320 E. 10th Street  
Bloomington, IN 47405-3907  
hrosenba@indiana.edu

## SPONSORSHIP: SIG-SI, SIG-STI

### ABSTRACT

Big data has been widely promoted across disciplines and sectors for its potential to enhance lives and promote knowledge discovery. However, challenges arise at all stages of the data lifecycle due to the complexity of interactions between data and the contexts within which they are collected and managed, which has implications for interpretations of this data and eventual use of information and the creation of knowledge products from these data. Starting from the perspective of social informatics, this panel will discuss: the reciprocal relationships between data and context; specific challenges in distinct stages of data generation, data repository implementation, data curation, data use, and data reproducibility; and the implications of these challenges and their potential solutions for both social informatics research and society in general.

{This is the space reserved for copyright notices.}

*ASIST 2016, October 14-18, 2016, Copenhagen, Denmark.*

[Author Retains Copyright. Insert personal or institutional copyright notice here.]

### Keywords

Big data, social informatics, information institutions, challenges

### INTRODUCTION

In the contemporary information society, big data generation and use are increasingly prevalent, and as a result, subject to scholarly and popular attention. Data are products of the context from which they were gathered or generated, as well as of the constraints on their collection, curation, storage, and use. Data also importantly shape repositories and products of their analysis, reflecting the mutual interactions between data and context.

It is important to consider the factors that shape data and the use of data, given the implications of big data and analytics to impact society and knowledge discovery. The dominant sentiment is that big data will promote progress and have positive influences, however, given the complexity of the contexts in which data is collected, processed, stored, used, and acted upon, it is probable that there are challenges posed to the positive outcomes sought, both practical and theoretical.

Specific concerns have been raised about: resistance to change; thoughtlessness, disorganization, and lack of institutions surrounding data processes; structural development; stakeholder needs and values; and social, political, and ethical issues surrounding the use of big data. As a result, various challenges can be predicted and identified. It is necessary to turn a critical perspective toward data and data analytic approaches, in order to better understand norms surrounding data at all stages, from production to use. This will allow society to capitalize upon potential benefits and minimize obstacles to appropriate use, moving forward.

Social informatics—as an interdisciplinary perspective from which to analyze relationships between people, technology, and context—represents an ideal perspective from which to critically assess relationships between data and society, given potential challenges, concerns, and implications surrounding the use of data in various contexts.

#### **DISCUSSION AGENDA**

This panel will be organized in three distinct segments. First an introduction to a social informatics perspective on big data will be provided, discussing how context shapes and is shaped by data and data practices. Specific interactions and challenges arising from these mutual interactions will be discussed briefly at various stages, reflecting ongoing projects and research with respect to distinct stages of the data lifecycle. Second, a moderated discussion on questions pertaining to the implications of these challenges for knowledge creation and societal enhancement will take place. This discussion among panel participants will specifically address: how do challenges in big data create barriers in specific contexts of use; to what extent do existing challenges faced in the use of big data bias knowledge creation or limit benefits of information use; and what potential solutions to challenges would promote societal enhancement and benefits within the lives of individuals? Third will be an opportunity for the audience to both pose questions to the panelists and join the discussion.

#### **PANELIST CONTRIBUTIONS**

This panel will discuss norms surrounding big data, specifically including challenges to the processes of institutionalization of norms, from a social informatics perspective. Panelists will consider institutions—policies, values, and norms—that govern various stages of the data lifecycle and data issues, including: data generation, data infrastructure, data curation, data use, data methodologies, data reproducibility, and critical data studies. Data sharing, use, and consumption patterns will be discussed with respect to various stakeholder perspectives. Challenges and support for sharing data with medical implications between user groups will be presented. Meta-analysis comparing

institutionalization of traditional quantitative approaches to data-analytic approaches in multiple social science domains will be employed to examine limitations to use of big data and the values that shape research outcomes. Furthermore, discussion will reflect epistemology and theoretical development in critical and empirical data studies. Structure and strategies employed in social and organizational data consumption will be critically analyzed. Data reproducibility will be discussed within multiple contexts. Without improving our understanding of the challenges faced within the domain of big data, the potential cannot be realized; it is necessary to examine how social and technical factors have mutually shaped outcomes at each stage of the data life cycle in order to better design and use data initiatives.

**Rosenbaum** will discuss organizational, political, ethical and social issues that surround the creation, dissemination and use of big data from the perspective of social and organizational informatics. The focus will be the ways in which the integration of big data is changing organizational structure, culture, and work practices.

**Marcinkowski** will discuss the continuing theoretical development of social informatics as it engages the rise of big data through critical and empirical research practices. Where early social informatics research investigated the rise of computerization, there exists today an analogous rise of the collection and use of data. As big data comes to affect the landscape of social informatics research, new understandings of social informatics' relationship to data, its approach to empirical research, and its understanding of agentive structures are necessary.

**Meyer** will discuss his OECD-funded work on data sharing among dementia researchers, and a new related project on automation in the UK health sector (which includes dementia as a case study). These projects together shed light on the context of medical data from multiple perspectives: medical researchers, medical practitioners, patients, and carers, among others.

**Sawyer** will discuss the nouveau data-rich, emphasizing concerns about thoughtlessness in data consumption patterns. This critical perspective will reflect challenges in use, as well as issues associated with interpretation and implications of big data.

**McCoy** will discuss his work on data literacy among university instructors in educational data science programs that consider teachers as practitioner researchers in the data analytics process. This work highlights the importance of

addressing data literacy training when integrating practitioners into the data analytics production process. .

**McCoy and Sanfilippo** will introduce their research on institutions structuring data analytics in social sciences, in comparison to institutions structuring more traditional quantitative research approaches. These challenges associated with use in scholarly knowledge creation will speak to issues of validity, reliability, and reproducibility, as well as implications for decision-making based on empirical research.

#### **PANELIST QUALIFICATIONS**

**Michael Marcinkowski** is a post-doctoral research assistant at Bath Spa University working on an Arts and Humanities Research Council-funded project studying the development and reception of “Ambient Literature,” a novel form of literary practice which combines new media approaches to literature with pervasive technologies. He earned his PhD at the College of Information Sciences and Technology at Penn State University, studying the implications of the availability of large-scale collection of data on the work of designers in the area of online education. Marcinkowski's research focuses on the philosophical and theoretical investigation of the relationship between the epistemic conditions of big data and the practices of science and technology design. His work on big data and social informatics has appeared in *JASIST* and *Computers in Human Behavior*, and has been presented at numerous conferences, including iConferences and Learning Analytics and Knowledge.

**Chase McCoy** is a doctoral student in Information Science in the School of Informatics and Computing at Indiana University. He is interested in approaching the data ecology within higher education from a social informatics perspective as it relates to the emergence of data analytics functionalities as components of educational ICTs, and the implications these technologies have on the educational environment. In addition, his work has examined the data literacy barriers of educators when they are repositioned as practitioner researchers in educational data science.

**Eric T. Meyer** is Associate Professor and Director of Graduate Studies at the Oxford Internet Institute, University of Oxford, where he has been on the faculty since 2007. Meyer's work focuses on shifts in work, knowledge creation, and interactions when digital technologies replace their previously non-digital counterparts. His research in this area has included studies of data sharing in dementia research, the use of digital images in biology, digital information practices in the sciences and humanities, and uses of data for public policy. His research, involving over 20 projects, has been funded by a range of organizations

including the Sloan Foundation, OECD, the Health Foundation, the UK's ESRC, Jisc, Nesta, and others. In 2015, his book (with co-author Ralph Schroeder) *Knowledge Machines: Digital Transformations of the Sciences and Humanities* was published by The MIT Press.

**Howard Rosenbaum** is Professor of Information Science in the School of Informatics and Computing at Indiana University and has been on the faculty since 1993. Rosenbaum studies social informatics, ebusiness, and online communities, has published in a variety of information science journals and presented at ASIS&T, iConferences, and elsewhere. In 2005, he published "Understanding and Communicating Social Informatics: A Framework for Studying and Teaching the Human Contexts of Information and Communication Technologies" with Steve Sawyer and the late Rob Kling. He has recently published two books, “Social Informatics: Past, Present, and Future” with Prina Fichman and “Social Informatics Evolving with Fichman and Madelyn SanFilippo.. He has been recognized often for excellence in teaching and for the innovative use of technology in education, most recently receiving the 2011 Thomson Reuters Outstanding Information Science Teacher Award from ASIS&T among other awards.

**Steve Sawyer** is on the faculty of Syracuse University's School of Information Studies and a research fellow at the Center for Technology and Information Policy. Steve's research focuses on the uses of information and communication technologies and changing forms of work and organization. This research builds from detailed field-based studies of work such and Sawyer has studied distributed scientific collaborators, software developers, real estate agents, police officers, organizational technologists, and other information-intensive workers in complex institutional settings. His work is published in a range of venues and supported by funds from the National Science Foundation, IBM, Corning, and a number of other public and private sponsors. Prior to returning to Syracuse in 2008, Steve was a founding faculty member of the Pennsylvania State University's College of Information Sciences and Technology. Sawyer earned his Doctorate in Business Administration from Boston University in 1995.

**Madelyn R. Sanfilippo** completed her doctoral studies in Information Science at Indiana University, Bloomington's School of Informatics and Computing in 2016, defending a dissertation entitled “An Unequal Information Society: How Information Access Initiatives Contribute to the Construction of Inequality”. Her work fundamentally addresses social and political issues surrounding information and information technology access, applying a social informatics perspective, particularly as relate to

unequal outcomes regarding interactions between policies, institutions, and information. She will begin as a postdoctoral Research Scholar at the Information Law

Institute at New York University's School of Law in September 2016.