

Big Data Study Group

November 9, 2016

Li-Min Chang

Department of Accounting and Information Technology
National Chung Cheng University



Editorial

Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research

Ritu Agarwal

Center for Health Information and Decision Systems, Robert H. Smith School of Business, University of Maryland,
College Park, Maryland 20742, ragarwal@rhsmith.umd.edu

Vasant Dhar

Center for Business Analytics, Stern School of Business, New York University, New York, New York 10012,
vdhar@stern.nyu.edu

We address key questions related to the explosion of interest in the emerging fields of big data, analytics, and data science. We discuss the novelty of the fields and whether the underlying questions are fundamentally different, the strengths that the information systems (IS) community brings to this discourse, interesting research questions for IS scholars, the role of predictive and explanatory modeling, and how research in this emerging area should be evaluated for contribution and significance.

Agarwal, R., & Dhar, V. (2014). Editorial—Big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research*, 25(3), 443-448.

Introduction

- A search on Google in the middle of August 2014 for the phrases “**Big data**,” “**Analytics**,” and “**Data science**”
 - Yielded 822 million, 154 million, and 461 million results, respectively
- Major journals (Management Science, MIS Quarterly) have commissioned special issues on these topics.
- A new journal called Big Data, launched just over a year ago, is already seeing thousands of downloads of articles.
(Source: <http://www.liebertpub.com/overview/big-data/611>)
- **Big data** is possibly the most **significant “tech”** disruption in business and academic ecosystems since the meteoric rise of the Internet and the digital economy.

Introduction

(914) 740-2100 [Contact](#) [Support](#) [Cart](#)

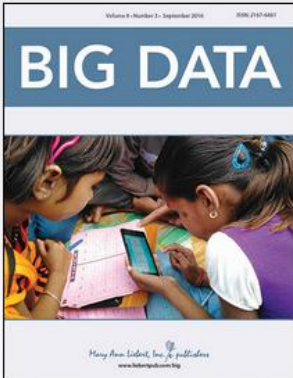
Enter your keywords 

R Tools for Visual Studio

Turn Visual Studio into a powerful R development environment.


Try it for FREE >

[Home](#) [Publications](#) [Resources](#) [Librarians](#) [Press](#) [Advertise](#) [PAIR DEVICE](#) [MY LIEBERT CONNECT](#)



Volume 8, Number 3 • September 2016 ISSN 2167-6461

BIG DATA

Mary Ann Liebert, Inc. publishers
www.liebertpub.com/big

[Read Online](#) [Subscribe/Renew](#) [For Authors](#)

Big Data

Editor-in-Chief: Vasant Dhar

ISSN: 2167-6461 • Online ISSN: 2167-647X • Published Quarterly

Current Volume: 4

Latest Impact Factor* is 1.696

*2015 Journal Citation Reports® published by Thomson Reuters, 2016

[LOGIN](#)

MY LIEBERT CONNECT

YOUR EMAIL:



Sign up to stay informed—receive email notifications for this journal

Free eBook! Healthcare Analytics on the Microsoft Cloud Platform

[Download the latest top-read articles](#)



Publish open access in this Journal.
[LiebertOpenAccess.com](#)



Featured Press Releases and Editorial Content

- Novel Tensor Mining Tool to Enable Automated Modeling Described in Big Data...[Read More](#)



About This Journal:

- [Overview](#)
- [Editorial Board](#)
- [Manuscript Submission](#)
- [Author Services Portal](#)
- [Read Online/TOC](#)
- [News Releases](#)
- [Featured Content](#)
- [Subscribe/Renew](#)
- [Recommend this title to your Library or Institution](#)
- [Reprints, Rights, & Permissions](#)

Introduction

- We posed five questions to frame our thinking about the domain:
 1. Are big data, analytics, and data science, as being described in the popular outlets, **old wine in new bottles or is it something new?**
 2. What are the strengths that the information systems (IS) community brings to the discourse on **business analytics**? In other words, **what is our competitive advantage?**
 3. **What are important and interesting research questions and domains** that may “fit” with on-going research in our community?
 4. To what extent **should robust prediction prowess** be used as a criterion in evaluating data-driven models versus current criteria that favor “explanatory” models without subjecting them to rigorous tests of future predictability?
 5. As editors and reviewers, **how should we evaluate research in this domain?** What constitutes a “significant” contribution?

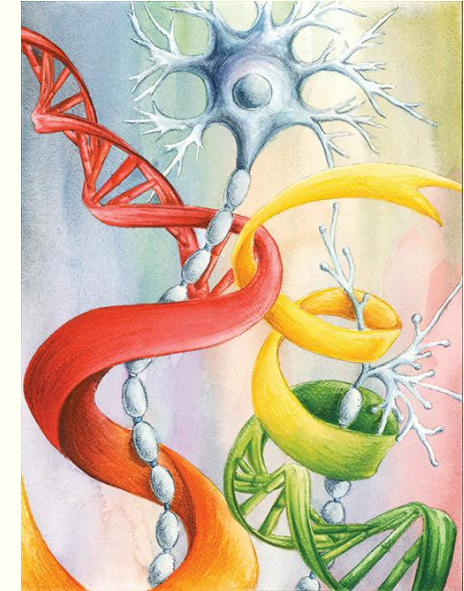
Introduction

- **We offer a synthesis of our collective reflection on the five questions.** (Galit Shmueli and Martin Bichler)
 1. On the Novelty of Big Data, Analytics and Data Science
 2. A Comparative Advantage for IS?
 3. On Research Questions
 4. Big Data, and Prediction vs. Explanation
 5. Evaluating Knowledge Claims Based on Big Data

On the Novelty of Big Data, Analytics and Data Science

1

- to understanding the pathology of **Alzheimer's** disease across millions of patients
- to observing **consumer response** to different marketing offers in large scale field experiments
- **turbocharged by machines** becoming a lot smarter through better algorithms, and by information technologies that enable people and things to be inherently instrumented for observation and interaction that feeds the algorithms
- younger people in a specific region of the world are becoming **increasingly diabetic** and then conjecture and test whether the trend is due to specific habits, diet, specific types of drugs



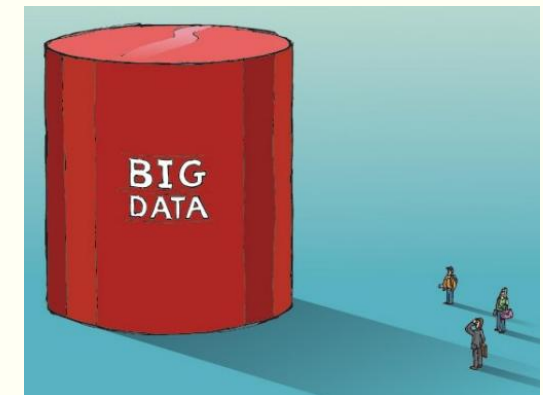
On the Novelty of Big Data, Analytics and Data Science

1

- **New challenging problems and inquiry** also lead to research on better algorithms and systems.
- We are seeing the emergence of **more powerful algorithms** and **better knowledge representation schemes** for making sense of all of this heterogeneous and fragmented information.
- **Text and image** processing capability are one frontier of research.
- **Networks**, such as those created by connections between individuals and/or products.
- To address this type of “**social interference**,” newer algorithms are required that support valid sampling and estimation of treatment effects (Ugander et al. 2013). (“relational” and “networked”)

A Comparative Advantage for IS?

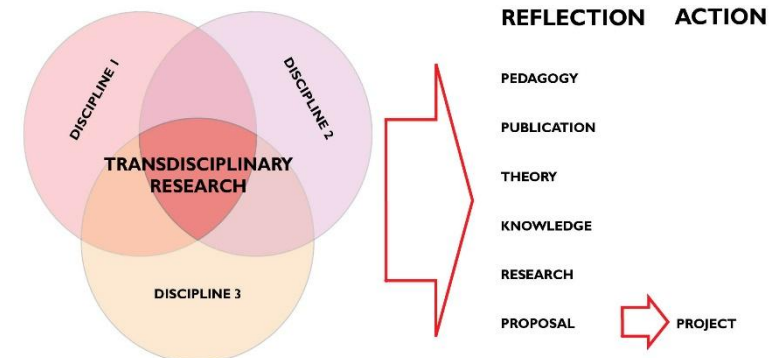
- The discipline of “MIS” emerged when computers enabled the automation of **business processes** and the digital capture of **business transactions**.
- Understanding how to design and implement systems to provide the “**right information to the right person at the right time**” was the *raison d’être* of IS programs and defined much of early IS research.
- It is not unreasonable to claim that IS scholars started off with a comparative advantage with **respect to big data**: we knew how to store, manage, process data; and about the complexity of data structures.
- **The challenges** associated with the infrastructure needed for handling the volume of data being generated today.



A Comparative Advantage for IS?

- Our research has been alternately praised and **criticized for being too cross-disciplinary.**
- IS scholars have **invoked theories** from the fields of economics, sociology, psychology, and political science to name a few, and studied phenomena such as electronic markets, consumer behavior, crowdsourcing, information security, and online retailing from a diversity of perspectives.
- This cross- and transdisciplinary nature of IS research to date positions us **uniquely to exploit the big data opportunity.**

TRANSDISCIPLINARY RESEARCH CENTER
MODEL D1

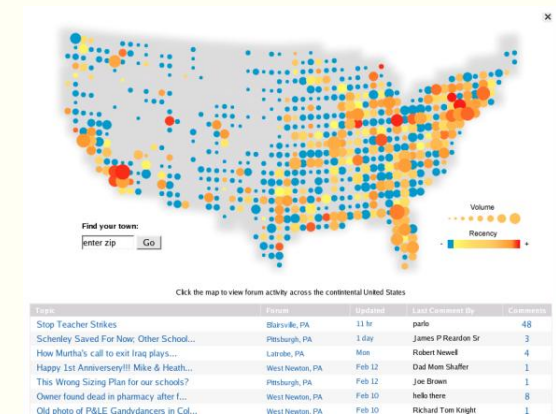


On Research Questions

- The wealth of possibilities enabled by big data is too many to enumerate in a brief commentary.
- First, the ability to observe and measure **micro, individual-level data** on a comprehensive scale.
 - ✓ diminishes social inequities
 - ✓ exploring labor markets
 - ✓ examining the actions of workers on social networking and employment sites
 - ✓ the quantified self-movement with sensors tracking exercise
 - ✓ information during daily living
 - ✓ producing a measurable effect on the incidence of disease or health problems

On Research Questions

- Second, following the call to focus on the **transformational aspects** of IT.
(Lucas et al. 2013)
 - ✓ the effects of technology mediation on learning outcomes, learner satisfaction
 - ✓ online courses
 - ✓ blended learning
- Third, this is an area where perhaps IS research has a robust set of studies to build on, is big data research in the context of **social networks and marketing outcomes**.
 - **Geo-coded social media** interactions coupled with extensive demographic and socio-economic data to **quantify** how networks affect
 - micro (individual behavior),
 - meso (organizational value), and
 - societal outcomes (economic and social value).



Big Data, and Prediction vs. Explanation

- The scathing review of **Google flu trends** (Lazer et al. 2014) (that uses search terms to predict the incidence of flu)
 - with respect to its accuracy as compared with estimates produced by the Centers for Disease Control and Prevention
- The ethical debate ignited by experiments conducted on **Facebook**.
 - Facebook study raises the specter of widespread experimentation on the Internet without adequate **protection of individual rights and privacy, ironically.**
 - the experiment is a great example of where causal claims can be made with some confidence

Big Data, and Prediction vs. **Explanation**

- **Prediction as an initial basis for theory building** also has particular value in a world where patterns often emerge before the reasons for them become apparent. (Dhar 2013)
- The scientific process is predicated on a cycle of **hypothesis generation, experimentation, hypothesis testing, and inference...**
- **Big data** are as, and increasingly more so, **useful at the hypothesis generation stage** as they are at the hypothesis testing phase.
- Scientific progress does not rely on a single study, and **big data-based studies offer the promise of novel discoveries.**



Evaluating Knowledge Claims Based on Big Data

5

- Some of it may **use new variables and measures** enabled by the digital capture of social and economic activity
- User generated context creatively combined with external data sets such as that obtained from the **Census Bureau or the Bureau of Labor Statistics** ([Big data center](#)).
- **There is no simple answer to this question** and, to a large extent, our standards and methods of evaluation will inevitably evolve as our experience with such research grows.
- **Fit, Interestingness, Rigor, Story, Theory (Agarwal 2012).**
- We expect that questions that challenge conventional wisdom and intuition or those that pose a hitherto unexplored puzzle.



Closing Thoughts

- This is an exciting time to be an IS researcher and to think beyond IS to science in general.
- Big data still aims in large part to deliver the right information to the right person at the right time in the right form.
- The IS discipline has been thinking and researching questions at the intersection of technology, data, business, and society for five decades and should leverage its thought leadership to become a centerpiece of education, business, and policy.



Thank you

