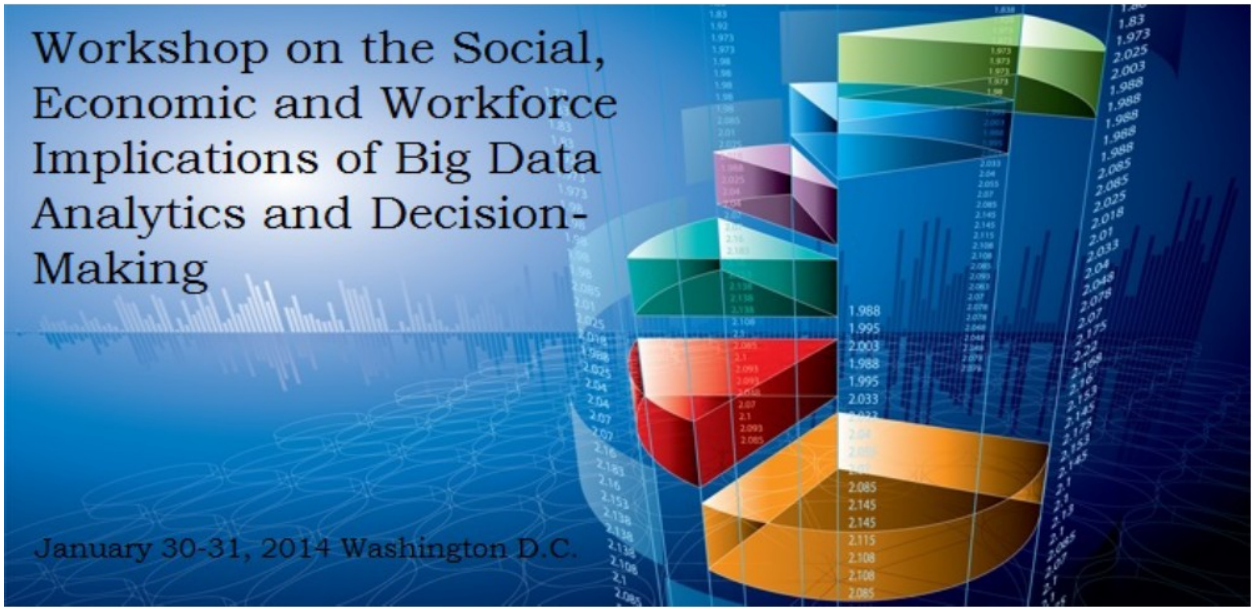
Big Data, Big Decisions for Science, Society, and Business

Report on a Research Agenda Setting Workshop

Funded by the National Science Foundation, Award #IIS: 1348929



M. Lynne Markus and Heikki Topi

Bentley University

September 2015

Foreword

The report from the workshop, "Big Data, Big Decisions for Government, Business and Society," makes a number of astute contributions. There is no need to replicate them in this foreword – they are in the report. What might be missed comes between the lines, where provocative points are made. Big Data means big opinions and big stakes. Those who think Big Data important want to be proven right, those who think Big Data a passing fad want Big Data to fade, and those who think Big Data will bring profound change hope for change. Big Data, like everything important, is political.

The science community is well-positioned to investigate Big Data, but there are serious challenges. Beyond the hyperbole that surrounds Big Data, there is as yet no clear definition of what it is, nor a comprehensive explanation of what it means. It might be too early to create either because the phenomenon is emerging. There are few reliable analogies to Big Data, which makes reasoning by analogy difficult. People like to say, "Big Data is like X," but find it hard to get beyond the observation that both Big Data and X are "really big." Many fans of Big Data are also fans of technology, and technological determinism is common. Critical perspectives on Big Data are seldom discussed outside of topics such as "privacy." These topics are themselves often complicated, suggesting how complicated the subject of Big Data can be. However, simply invoking such topics does not clarify much. Finally, it is widely known that some people have made huge fortunes by exploiting data. It is difficult to comprehend the effects of hoped-for economic gain on what people support or oppose.

There is consensus that Big Data might be very important. Some of what it brings could be genuinely new. There are early signs that Big Data can make a big difference in analysis and decision-making of many kinds. These signs are suggestive, not conclusive. Radical change could come from Big Data, but is not guaranteed, nor is it needed because even modest improvements could be very valuable. The history of technology teaches that important outcomes sometimes surprise. This presents a dilemma. On one hand, since outcomes cannot be foreseen precisely, some argue that everyone should get out of the way and let innovation bloom: a *laissez innover* position – let things happen as they will. On the other hand, some argue that outcomes that can be anticipated call for readiness. Many at this meeting believe some outcomes can be anticipated and handled in ways that accelerate innovation. Here is a short discussion of what can be anticipated and possible steps to deal with the concerns.

Technological change might happen faster than institutional change, yet institutional change might be important to the future of technological change. Problematic outcomes from Big Data activity might occur before institutional mechanisms (e.g. legal, regulatory, political) are in place to alleviate problems and increase benefits. How can this co-evolution be accelerated? History provides some guidance. Motor transport (automobiles, trucks) has had huge beneficial impact over more than a century. However, these benefits were dependent on evolution of other things: social conventions (e.g., which side of the road to drive on), regulations (e.g., speed limits), enforcement (e.g., traffic police), risk-mitigation (e.g., automobile insurance), complementary-asset funding (e.g., tolls and fees on vehicles, fuel, batteries, tires and other things to pay for roads). How best to meet these needs for motor transport is still being debated today. If Big Data is important, such evolution will probably occur. It makes sense to start thinking now about how to meet the needs. This requires study of what might be the case, and what to do if so. This should not retard useful experimentation and innovation. A lot has been learned about managing this trade-off. There are many options beyond letting whatever happens happen, or killing innovation with stifling regulations. High stakes require thinking through implications. It is a balancing act.

The balance can be difficult to effect because of knowledge shortfalls along the way. Two important cases from the annals of medicine illustrate this. Acetylsalicylic acid, ASA, also known by

the Bayer brand name Aspirin, was discovered in the bark of the willow tree in the 18th century and first synthesized in the late 19th century. It has valuable analgesic, antipyretic and anti-inflammatory properties. For decades no one knew how it worked. In the late 20th century researchers began to understand ASA's mechanism of action, and much was learned about physiological processes (e.g., the role of prostaglandins). Acetylsalicylic acid is one of the most important drugs in the world. Demands to "fully understand" its mechanism of action before allowing its use would have deprived countless people of its benefits. Yet research on the mechanism of ASA produced additional benefits.

On the other hand, the immunomodulator drug thalidomide was sold in the mid-20th century as a sedative or hypnotic. This inadvertently caused thousands of cases of severe birth defects. Stereochemistry was new at the time, and scientists did not know that isomers of thalidomide could be dangerous or therapeutic. Thalidomide has since become a powerful treatment for leprosy and certain cancers. Additional knowledge turned a "problem" drug into an "important" drug. Safeguards to protect pregnant women were needed, but so was additional research. The question of in-body racemization, or conversion to isomers, is still not understood and remains important with thalidomide. More research is needed to help with use of this powerful but potentially dangerous drug.

New ways of doing things can upset preparation mechanisms such as education. Discrete mathematics (e.g., probability, statistics, number theory) have become essential to computer science and other fields. These newer forms of mathematics are not "more important" than traditional continuous mathematics such as calculus. The best ways to teach mathematics continue to evolve. At the same time, calculus has a several century head start. There are consequences that must be recognized when "new" meets "old." Data have always been important in science. Big Data does not change that. But Big Data might require changes in preparation. For example, the traditional role of "theory" in research might be challenged. Large-scale analysis of data might produce outcomes beyond the reach of traditional models of theory and experimentation. This might already be underway in evidence-based medicine that relies heavily on the empirical rather than the theoretical. Data might lead theory in many cases. The issue is not either/or. Theory will remain useful, no matter how important Big Data becomes. However, the traditional methodological dominance of theory might be challenged by a Big Data revolution. The next generations of scientists must be prepared to deal with the opportunities of Big Data. This will require the modification and enhancement of methods training.

The history of technology has shown that obtaining "low-hanging fruit" in the early days of new capability is sometimes easier and sometimes harder than imagined. Few predicted the speed with which the Internet affected commerce. Early "benefit overruns" occur, and no one is saying the benefits came too fast and we should wait to enjoy them. At the same time, cost overruns are possible. Research on natural language translation by computers, underway in the 1950s, caused intelligent people – some later won the Nobel Prize – to predict fluent computer translation of natural language by the 1960s. Computer translation is better than it used to be, and much has been learned about natural language. Nevertheless, 60 years later there is still no fluent machine translation of natural language. Perhaps the 1950s vision of computer translation will take 100 years instead of 10 years to achieve. It will still be important, even if it takes longer than imagined.

Big Data could be as important as the steam engine. However, it is wise to remember what Lawrence Joseph Henderson proclaimed in 1917, "Science owes more to the steam engine than the steam engine owes to science." The full implications of Big Data remain unclear, but that is no reason to fall into traps that can be foreseen and avoided. Research on the implications and consequences of Big Data is needed to foresee and avoid those traps.

John Leslie King
University of Michigan and London School of Economics and Political Science

Table of Contents

Foreword i
Executive Summary iv
Introduction 1
What is Big Data? 2
What kinds of implications might Big Data have? 5
How significant might Big Data's implications be? 7
Summary 9
The Case for Big Data Implications Research 12
Categories of research about Big Data 12
Big Data application research 12
Big Data infrastructure research 13
Big Data implications research 13
Costs and controversies of Big Data implications research 15
Costs 15
Controversies about Big Data implications research 15
Summary 18
Big Data Implications Research Priorities 19
Science and science–technology policy 19
New data partnerships 19
Data metadata 20
Evolution of the academy 20
Individuals and everyday life 22
Privacy reimagined 22
Data amateurs 24
Benefits and burdens 25
Organizations and work 26
Corporate data scientists 26
Corporate data practices 29
Knowledge and expertise 31
Cross-cutting themes 34
Values and ethics 34
Data, algorithm, and decision quality 35
Knowledge and skill 35
Summary 36
Guidelines for Big Data Implications Research 39
Summary of Recommendations 42
Appendix A: Workshop Participants 44
Endnotes 46

Executive Summary

An NSF award was made for hosting a multidisciplinary workshop on the social, ethical, economic, and workforce implications of Big Data. Invited academics from various fields (including computer science, economics, ethics, information systems, organizational behavior, law, sociology, and public policy) met with members of the business community and representatives of research funding agencies and foundations in Washington, DC, in January 2014, to discuss research topics and priorities. The focus of the workshop was on research about the *implications and consequences* of Big Data, rather than on research on how to use or build large datasets, analysis techniques, or technology infrastructures. This report presents an agenda for a systematic program of research on Big Data's implications and consequences: The aim of Big Data implications research is to generate scientific evidence that can help individuals, businesses, and public policy makers maximize benefits of Big Data while minimizing its negative consequences.

Workshop participants defined Big Data to include not just data, but also data analytics and the individual and organizational decisions made on the basis of data. Big Data as a phenomenon is not confined to any one social or economic sector or type of data. Big Data has many different kinds of implications and potential consequences, including both benefits (e.g., scientific discovery, product and marketing innovations, corporate and government efficiency) and potential harms (e.g., invasion of personal information privacy and breaches of data security). Big Data also has implications for employment opportunities, educational needs, the validity and integrity of scientific research, the quality of social interaction and connectedness, the division of labor between humans and machines, and the likelihood and corrigibility of errors in data, algorithms, and decisions. Because of its growing pervasiveness, Big Data is a potentially transformative innovation. That is, it can lead to disruptive changes by displacing older socioeconomic activities and creating entirely new opportunities. Historically, disruptive transformations have brought many benefits but also some intractable problems.

The implications and potential consequences of Big Data are controversial topics. Polarization is already apparent. On the one hand, data enthusiasts insist that Big Data has great potential to contribute to the public good and that nothing should be done to inhibit innovation. Even discussing possible harms is seen as a threat to progress. On the other hand, data skeptics claim that the potential harms of Big Data are so significant that some innovations should be regulated or blocked.

Participants in the workshop charted a middle course between these extremes. Workshop participants concluded that systematic research on both the benefits and risks of Big Data is needed and beneficial for: 1) identifying and evaluating alternative technical and nontechnical strategies for reducing Big Data's actual or potential harms without

erecting barriers to the achievement of Big Data's benefits, 2) generating or testing new approaches for increasing Big Data's economic and public value through better tools and techniques, improved services and business models, enhanced work practices, and augmented educational programs, and 3) allaying unfounded public fears, through careful analysis of Big Data's realized or potential costs, benefits, and risks and their distributions across stakeholder groups.

Discussions covered three overlapping contexts of Big Data use: 1) *scientific research and science-technology policy*, including open government data and university-based research, 2) *everyday life*, including civic engagement with Big Data through citizen science and data "hackathons" or use of social media and fitness applications, and 3) *organizations and work*, including commercial applications of Big Data in engineering, marketing, and business operations, and the application and use of Big Data by data scientists and knowledge workers. Because of the focus on these three contexts, use of Big Data in the national security context was explicitly excluded from discussion at the workshop.

In the area of *science and science-technology policy*, workshop participants identified three high-priority research areas:

1) *Data-oriented partnerships*: What are the lessons learned from new data-oriented partnerships among government, universities, businesses, and civil society organizations, including the data partnerships created under the auspices of President Obama's Big Data initiative?[1]
2) *The new data economy*: What are the practices and consequences of the new data economy, including the mechanisms for monetizing and selling data and the economic and public value created by data-oriented initiatives?
3) *Implications for academia*: What are Big Data's implications for universities, academic disciplines, and research practices and policies?

Three key research priorities related to *individuals and everyday life* were discussed at the workshop:

1) *Privacy and privacy protection*: What are the privacy and privacy protection issues associated with new data sources and technologies?
2) *Data amateurs*: What are the activities, careers, and contributions of individuals who use Big Data and data science tools for career development, for personal fulfillment, or even for illicit purposes?
3) *Benefits and burdens*: What are Big Data's personal benefits, for example, better decision-making and precision medicine, and its personal burdens, for example, new forms of "social sorting" and discrimination?

The third domain in which workshop participants discussed research priorities was that of *organizations and work*. Participants noted that far more attention has been paid to

the implications of Big Data for the science and civic domains than for the domain of corporations and their employees. Yet this domain is highly significant for innovation, employment, and national productivity and competitiveness. Three research priorities in this domain were identified:

1) *Corporate data scientists*: What are the professional competence and ethical conduct requirements for corporate data scientists?
2) *Corporate data practices*: How do organizations govern and manage data and algorithms?
3) *The changing world of work*: What changes might Big Data bring for jobs and workers' knowledge and expertise?

These three contexts (science, everyday life, and work) differ considerably in their dynamics and challenges. In science, Big Data projects are often one-time studies, typically published in academic journals after peer review. An example is research to determine whether a change in medical practice guidelines leads to improvements in patient outcomes. By contrast, commercial Big Data projects frequently lead to recurrent automated decision-making processes (e.g., automated mortgage lending, automated pricing and delivery of online advertisements, and automated electricity load balancing and pricing) powered by algorithms that may be non-transparent both to their developers and to those to whom the decisions apply. These different contexts have specific research issues that may require tailored research strategies. At the same time, several important issues cut across the three domains. Three such themes are discussed in this report: 1) *value conflicts and ethical dilemmas*, 2) *data, algorithm, and decision quality*, and 3) *the balance between human expertise and computer-based automation*.

Big Data implications research is a cost-effective way to support national interests in innovation, economic development, and productivity. Big Data implications research is *policy-aware*, that is, it aims to provide business and public policy-makers with scientific evidence that can inform their decision-making on important practical questions. As such, Big Data implications research must be *value-sensitive*, that is, explicitly attentive to the values, needs, incentives, and goals of multiple stakeholder groups. It should also be *design-oriented*, that is, focused on suggesting and evaluating alternative strategies for realizing opportunities and overcoming challenges. The implications of Big Data are many, varied, and not fully knowable before they unfold. These characteristics suggest that Big Data implications research is at its best when it is *transdisciplinary* (that is, bringing multiple academic perspectives to bear in order to address complex challenges), *sociotechnical* (that is, examining both social and technical factors and their interrelationships), *anticipatory* (that is, future-oriented and precautionary) as well as focused on current situations and their historical roots, and, ideally, *participatory* (that is, involving multiple stakeholder groups, including practitioners).

Introduction

This report summarizes the conclusions of a National Science Foundation funded workshop to prepare a research agenda for the social, economic and workforce implications of "Big Data." Invited academics from several fields (including computer science, data analytics, economics, ethics, information systems, organizational behavior, law, public policy, and sociology) met with members of the business community and representatives of research funding agencies and foundations in Washington, DC, in January, 2014, to discuss research topics and priorities.

The focus of the workshop was research on the *implications and consequences* of Big Data, rather than on research that builds or uses large datasets, complex analysis techniques, or specialized technology infrastructures. The remainder of this chapter summarizes workshop participants' discussions and conclusions about what is meant by "Big Data," the kinds of implications that Big Data might have, and the socioeconomic significance of Big Data consequences.

About The NSF Workshop (1/2)

In Summer 2013, the US National Science Foundation awarded Bentley University (M. Lynne Markus, Principal Investigator; Kevin Crowston, NSF Program Officer) a grant (#1348929) for holding a research agenda-setting workshop on the social, economic, and workforce implications of Big Data analytics and decision-making. The workshop was held on January 30-31, 2014. Adopting an interdisciplinary approach, the workshop and brought together almost fifty scholars and policy makers from academia, industry, and various government agencies to develop a research agenda for and stimulate research in this important area. The scholars represented fields as diverse as computer science, information systems, law, health care, economics, public policy, and philosophy.

The participants were asked to prepare brief position papers before the workshop, exploring their perspectives on the researchable consequences of big data analytics, the importance of these consequences, the most important policy questions raised by big data analytics, and the key challenges faced by researchers working in this area. These position papers were shared among participants before the workshop, and they formed a solid foundation for the workshop conversations.

About The NSF Workshop (2/2)

The first day of the workshop opened with a plenary address by Jules Polonetsky (Future of Privacy Forum). The two core sessions of the day addressed the topics "Big Data, decision-making, and knowledge" and "Big Data and sociotechnical infrastructure," respectively. In the panel portion of each session, a number of invited workshop participants first presented their perspectives on the topic of the session. After this, all participants discussed the topic and related questions in small groups and in a summarizing plenary conversation.

During the second day of the workshop, academic participants of the workshop continued their deliberations by identifying the most important research questions and cross-cutting research themes, building on the work of the first day. The final session of the workshop covered challenges affecting research on the implications of Big Data analytics and decision-making, specifically focusing on questions related to infrastructure, coordination, access to data, methods, education, and dissemination of results.

A list of the participants is available in Appendix A. M. Lynne Markus (PI) chaired the workshop steering committee, and its other members were Carol Ann Boughrum (local arrangements), Fred Ledley, Kevin Mentzer, Sue Newell, and Heikki Topi (all representing Bentley University).

What is Big Data?

The term Big Data has many definitions and connotations. Big Data is often defined in terms of ***data characteristics*** such as volume, variety, and velocity.[2] For example, "Big data is too voluminous, too varied, and too fast moving to be handled via conventional tools and techniques" or "Big Data is data too big to be moved around."

These definitions pose a moving target. What is considered large today will be seen as small tomorrow. They are also unnecessarily restrictive, since many of the societal consequences attributed to Big Data can emerge with relatively small data sets. In addition, workshop participants concluded that typical definitions of Big Data are too limited because they obscure organizational practices of data collection, analysis, and use.

Defining Big Data as a ***particular kind or source of data***—such as social media data (e.g., Twitter feeds or online product reviews),[3] sensor data, or open government data—is also limiting, because it directs attention to particular kinds of implications while obscuring others. Social media and government data, for example, focus attention primarily on concerns about individual information privacy and secondarily on data security and unlawful or unacceptable discrimination.[4] The President's 2014 90-day review of Big Data[5] focused almost entirely on these three issues. But social media and government data also raise other important, if less obvious, issues *for universities* (e.g., how to fund Big Data research infrastructure; how to avoid losing research talent to social media companies) and *for academic disciplines* (e.g., how to evaluate Big Data research contributions in relation to "little data" methods; how to ensure research validity when data are proprietary and cannot be made open for review).

Workshop participants discussed a wide range of Big Data sources and implications and concluded that a broader view of Big Data is called for. Participants adopted a sociotechnical perspective, viewing Big Data as a ***cluster or assemblage of data-related ideas, resources, and practices***. Big Data *ideas* include "Data is a monetizable asset" and "Decisions should be data-driven, not based on human intuition." Big Data's *resources* include data stores, tools like Hadoop or NoSQL data storage, and infrastructures of equipment and skilled personnel. Big Data also involves *techniques and practices* such as data collection and retention, predictive analytics, or automated decision-making.

In other words, Big Data is not just *data*, but also ***what is or could be done with data*** and the goals and values that motivate that use. Workshop participants described a "data supply chain"[6] comprising both "upstream" and "downstream" activities. Upstream activities include the collection, retention, analysis, and sale of data, along with related activities such as categorization, standardization, curation, management, and packaging, to name a few.[7] The downstream side of the supply chain involves purchasing data, conducting analysis, interpreting analyses, making decisions based on data analysis, and achieving value from data-driven changes in such activities as marketing, operations, benefits eligibility determination, hiring and promotion, investing, prescribing, and so forth.[8] The downstream side can also involve onward sales of data.[9]

Workshop participants observed that downstream users of Big Data often do not know where the data they use comes from. Consequently, they may have a limited understanding of data representativeness, quality, and biases. At the same time, the people who supply or collect data are often ignorant of the downstream and future uses and consequences of data. For example, medical records sometimes contain errors when doctors code illnesses in ways that help their patients obtain insurance reimbursement. This type of data error does not affect the patient's quality of care, but it can lead to erroneous conclusions later when researchers mine medical records for evidence-based treatment guidelines.

A particularly important aspect of the data downstream is the decisions that are made based on Big Data analytics. Much discussion of data-driven decision-making assumes that decisions are made by subject-matter experts, who are, or who work closely with, data collectors or analysts. These assumptions are appropriate for much academic Big Data research and for science and engineering activity in corporations. However, they do not always hold *in commercial and industrial settings.* One reason is the lack of skilled data scientists and of knowledgeable managers who can guide data science projects. Another reason is that the purpose of analytics in many commercial and industrial settings is *not* to support human experts, but rather ***to replace people***, either by employing less-skilled workers instead of scarce or expensive experts or by entirely automating the decision-making process.[10]

When human experts are replaced by either partially or totally ***automated decision-making***, there can be profound changes in the flexibility and reliability of organizational processes, the nature of workers' subject matter knowledge and expertise, organizational memory and learning, job characteristics, and employment opportunities.[11] Some of these changes are highly beneficial to organizations by reducing costs, reducing decision biases, and increasing decision quality. However, these changes may also involve risks of various kinds for some stakeholder groups. For example, almost everyone has been inconvenienced when service workers (in government offices, banks, or retail establishments) have been unable to make warranted changes in organizational policies "because the computer wouldn't allow it."[12] People also risk loss of jobs from automation or loss of autonomy in making job-related decisions. For instance, transportation safety is often said to demand total automation to prevent human tampering and errors. There is no easy answer to the question of whether and when humans should or should not be involved in the operation and control of machines,[13] but the trend is toward ever-greater automation.[14] This Big Data implication is so important that it deserves systematic academic research to support public discussion and debate.

Automated decision-making is already far advanced.[15] One example involves the sale (actually, the auctioning) of advertising on websites.[16] Other examples show that decision automation is making inroads into high-skill, knowledge-based occupations such as engineers and technical workers, bankers and insurance underwriters, and lawyers and doctors:

- The physical testing of automobile safety via crash-testing has largely been replaced by simulation[17]
- In the U.S. home mortgage industry, the underwriting of mortgages (even subprime) was almost completely automated before the housing crisis[18]
- Insurance underwriting is also being automated[19]
- Legal automation, such as e-discovery, has developed rapidly to the point where it can replace much legal labor[20]

- Sixty percent or more of securities trading has already been automated.[21]

In the case of medicine, for example, technologies such as IBM's Watson[22] have the potential to rival the diagnostic capabilities of expert physicians. If the pattern observed elsewhere repeats itself in medicine, paraprofessionals will take over more and more of the physicians' role, perhaps relegating physicians to a "quality control" capacity. While this development could leave more time for physicians to interact with patients, automation could also affect this activity: one workshop participant described clinical information systems that were purposely designed to take over some aspects of physician–patient communication. This pattern is not confined to medicine, but is also affecting science, engineering, and many other fields.

These examples show that Big Data has implications not just for academic data scientists, government agencies, marketing professionals, and citizens who use social media. Big Data is not just important for the "invisible" technical workers[23] who develop and support Big Data and algorithms in business organizations. Big Data also has important implications for organizations, for knowledge workers like engineers, auditors, and human resources specialists, and for consumers at large.

In short, Big Data is much bigger than just "data."

What kinds of implications might Big Data have?

Big Data has implications ***for science and universities***. It is changing the methods and tools that scientists use in fields as diverse as astronomy, biology, history, and sociology. Big Data is providing scientists with new strategies for solving problems that previously could not be solved. For example, Big Data provides predictions about the best ways to treat cancer in the absence of good causal theories. But additional resources are required to develop the technical infrastructure to support Big Data research, and scientists need to develop new skills to evaluate Big Data research quality. In fields that use Big Data about persons, new procedures must be devised to protect personal information privacy and to protect the interests of the organizations that collect, manage, and sell that data.

Big Data also has implications ***for science–technology policy***. Big Data's *potential* to improve the efficiency and effectiveness of existing enterprises and to create innovative new products, services, jobs, business models, and industries is well known. But there are frequent reports of massive shortfalls in the number of data scientists and related technical workers needed to capitalize on this potential. There are also frequent calls for government investment in research that uses Big Data tools and techniques to answer scientific and engineering questions and for building the infrastructure to support that research.

Big Data also has implications ***for every government agency*** with data resources

that could be "opened" for use by citizens, universities, social service providers, or businesses. These data resources are seen as sources of government transparency, new scientific knowledge, better ways to improve citizen welfare, and economic development.[24] However, existing open data repositories lack the organization and integration that would unleash their full potential, and the resources and incentives to do this work are missing. In addition, government officials are naturally concerned about how to protect the privacy of personal information, about how to ensure data security, and about how to validate and replicate scientific research that makes use of government data.

In addition, there is widespread awareness of Big Data's possible benefits and concerns ***for individuals as users of Big Data*** in the form of social media, e-commerce website sites, and "willpower-support apps" (e.g., fitness monitors). For example, the Internet of Things (widespread interconnection of very large numbers of computing and sensing devices) is expected to increase consumers' product knowledge and convenience, reduce waste in the supply chain and the home, and make target marketing more accurate. At the same time, some people worry about threats to personal privacy and the possibility of financial or health identity theft.

These are among the most frequently discussed areas of Big Data implications. But as discussed above, they are not the only ones. Big Data is currently being used in many, if not all, social and economic spheres. It is used in engineering, accounting, financial services, energy, agriculture, health care, manufacturing, the humanities, social entrepreneurships, corporate boardrooms, and government tax auditing. It is used to detect and prevent frauds, to insure crops, to design industrial and consumer products, to set prices, for hobby genealogy, for archaeological research, and for citizen science.

Thus, Big Data is not just about science, science policy, open government data, academic research, marketing, and social media. Big Data is everywhere, and its implications are, too[25]. Among the important, but less discussed, issues and potential implications of Big Data are:

- Changes in academic practices and values, such as the role of theory and prediction, statistical sampling, and hypothesis testing, the effectiveness of peer review and open data policies, and collective understanding of the concept of "knowing"[26]
- Changes in resource requirements and funding sources for university-based research, and their consequences for the employment of scientists, research productivity, and the quality of research
- Changes in the character of civil society[27] in an era of ubiquitous monitoring through object-based sensors and body cameras[28]
- How data-sharing by some individuals (e.g., social media data; genetic data by hobby genealogists) affects choices and outcomes for others (e.g., friends and relatives who would prefer to opt-out of data sharing)
- The ethical choices involved in building Big Data applications and algorithms and how those choices are made[29]

- Changes in employment opportunities and job quality for all knowledge workers (not just data scientists)[30]
- How sale of data affects the economics of product and service delivery and the ability of citizens to access data about them
- The quality, safety, and oversight of automated decision-making systems, and redress in the case of inaccurate or unfair decisions.[31]

How significant might Big Data's implications be?

Big Data is sometimes portrayed as a ***revolution***[32] in science, business, and professional work. More skeptical observers dismiss that talk as hype and describe Big Data as a transitory management ***fad or fashion***. Others argue that Big Data is ***nothing new***: Workshop participants noted that Big Data can be seen as the extension of the "avalanche of printed numbers" that began in the early 1800s.[33] That view argues for Big Data's staying power (under changing labels), but does not address the question how important its ***future*** consequences might be. Could Big Data transform society and the economy by displacing older patterns of activity and by creating new products or services, job types and opportunities, organizational forms or business models, industries, etc.? In other words, do the changes promised (and ultimately delivered) by Big Data have the potential to rival earlier transformative innovations, like internal combustion, electric power, information technology, and biotechnology?

In general, workshop participants shared the view that Big Data has the ***potential*** *for* ***transformative socioeconomic consequences.*** That is, for purposes of public- and business-policy making, the prudent course of action is to regard Big Data as an innovation that could, like automobiles, nuclear power, or pesticides, bring both enormous benefits and considerable side effects. Taking this stance toward Big Data would favor investing in Big Data applications and infrastructure, while also investing in ***implications research*** that could improve Big Data's benefits, foster its acceptance, understand its implications, and reduce its risks.

The emergence of transformative innovations is difficult or impossible to predict before the fact. However, after-the-fact research has identified three defining characteristics of innovation clusters that have resulted in socioeconomic transformations.[34] First, transformative innovations garner ***substantial support and resources*** from powerful opinion-leaders, technology product and service firms, consultants, entrepreneurs, investors, and customers or users.[35] Second, transformative innovations are ***multipurpose***, that is, widely adopted in multiple sectors or economic niches.[36] Third, transformative innovations have ***unintended side effects.***[37]

Some of the side effects of transformative innovations are beneficial, while others are not. Even initially negative side effects can be beneficial, by spurring subsequent innovation. But some negative side effects of transformative innovations (e.g., air pollution

or personal injuries caused by automobiles) have proved challenging to eliminate or control. In addition, only some of the side effects of transformative innovations can be accurately foreseen. Some anticipated outcomes never occur, and sometimes the actual consequences are the opposite of expectations.

The difficulty of anticipating an innovation's consequences is occasionally cited as a reason for not even trying to anticipate the outcomes: What happens if we get it wrong, take actions to reduce risks, and inadvertently block progress? Why not just wait and see what happens, and then fix any problems that arise? To the contrary, there are powerful reasons ***to try*** to anticipate the implications and consequences of potentially transformative innovations, regardless of foresight accuracy. One important reason is that potentially transformative innovations can fizzle, failing to bring the hoped-for benefits, or be long delayed.[38] Some earlier transformative innovation clusters, notably electric power, took decades to diffuse and to produce substantial benefits, because of barriers such as traditional mindsets and unsuitable factory layouts. Similar gaps between early investments and realized value have been seen in information systems[39] and technologies of all kinds.[40] Systematic attempts to foresee barriers and consequences and to study them as they emerge can identify conditions that limit the acceptance of an innovation and suggest strategies for overcoming them. (Other reasons to try to anticipate a potentially transformative innovation's consequences, discussed in the next chapter, include: 1) to increase human proactivity, resilience, and ability to respond appropriately, if and when negative outcomes occur and 2) to improve technology designs and economic benefits.)

Observers have already begun to enumerate barriers to Big Data's uptake, and to blame those factors for limited evidence of Big Data's benefits to date. Among those barriers are:

- Lack of trained data scientists,[41] especially ones who combine technical skills, consulting skills, domain knowledge, and awareness of legal and ethical issues
- Inadequate tools and technical infrastructures[42]
- Insufficient or poor quality data in some areas (or a superabundance of potentially useful data in others)[43]
- Organizational and business leaders who do not understand Big Data's benefits or potential areas of application[44]
- Inappropriate rules and regulation,[45] and
- Public concerns about data privacy and security risks and unintended consequences.[46]

Research that studies these barriers systematically is an important way to unblock innovation and to realize Big Data's positive potential.

Big Data is starting to show the characteristics of a transformative innovation cluster. First, Big Data is garnering lots of support. The Obama Administration launched a

$200 million Big Data initiative in 2012.[47] The Big Data business is currently said to be worth more than one hundred billion dollars.[48] And a flood of publications attests to widespread business interest. Second, Big Data is being applied in many sectors, including natural science and biomedical research, engineering, social science research, humanities research, government open data initiatives, social entrepreneurship, marketing, financial services, agriculture, transportation, accounting, health care, and insurance. Third, Big Data is driving important changes in government, business, and societal practices whose implications can only weakly be foreseen. Some of those changes include:

- Enterprises of all kinds are attempting to monetize their data assets or create data businesses[49]
- Accountants are starting to provide valuations of corporate data assets,[50] and insurers are offering coverage for data leaks and thefts
- There is an active debate in the artificial intelligence, law, and ethics communities about the extent to which *algorithms* should be considered morally responsible and legally liable for automation outcomes[51]
- Legal contracts governing data sale and terms of use have evolved rapidly
- Regulators are demanding the use of data-based models in compliance reporting
- Individuals are sharing their genetic or health data online on sites like www.ancestry.com or www.23andme.com
- Social media data are being mined to enable home invasion, cyberbullying, and identity theft
- Algorithms are being used in employment, pay, and benefit decisions.

In short, Big Data is neither a fad nor a revolution—it is a ***potentially transformative innovation*** that could bring great benefits or be blocked by barriers such as public concerns about potential harms including loss of personal privacy, discrimination, and pervasive surveillance. Comprehensive study of Big Data's social, economic, and workforce implications can help ensure that Big Data's hoped-for benefits actually come about.

Summary

In summary, workshop participants concluded that:

"Big Data" should be understood to include not just *data* but also the *models, algorithms, processes, and human actions* involved in the analysis, interpretation, and use of data. Models are already commonly used to inform expert decision-making in science and engineering, business, and health care. In addition, models and data-driven analyses are increasingly important in compliance with government regulations (e.g., in financial services and transportation). Those models *support* human decision-makers, but in other cases, algorithms are used to *automate* decision-making. Automated decision-making, in which data-driven decisions or activities are executed with minimal human intervention, is

extensively used in financial services (e.g., fraud detection and the auditing of tax returns) and has the potential to be applied in many more domains. Automated decision-making should be a particular area of focus of Big Data implications research.

Big Data has important implications for nearly every sector of social and economic activity. Although some people associate Big Data with particular activities, e.g., academic research or consumer marketing, and with particular data sources (especially social media), Big Data's implications extend far beyond any single domain or data source. Specific domains (such as health or marketing) have unique issues (e.g., related to the ownership or privacy of personally identifiable information or the possibility of discrimination) that require focused study. However, the pervasiveness of Big Data also suggests the need for research on implications that cross many domains. Examples of such cross-cutting implications include 1) the inevitability of value conflicts over the ends and means of Big Data, which requires informed ethical decision-making by clients, developers, and users, 2) the importance of quality in data, algorithms, and data-driven decisions, and 3) changes in the relationships between humans and algorithms in knowledge-based activities.

Big Data has the potential to be socioeconomically transformative, that is, to displace older activities and to create entirely new opportunities. This means that Big Data may bring tremendous benefits, including those that are unintended and currently unforeseen. It also means that Big Data has the potential for big risks. Concerns about those risks could create public resistance and initiate government regulation that might stifle beneficial innovation. But, as discussed in the next chapter, ignoring or belittling those concerns is counterproductive for two reasons. First, it can actually *increase* public resistance. Second, it can ironically promote a design mindset in which important technical and economic opportunities are missed. Better scientific knowledge, deeper understanding, and public awareness about Big Data's implications can *promote innovation* by 1) identifying or evaluating alternative technical and nontechnical strategies for *reducing Big Data's actual or potential harms* and for *reducing barriers to the achievement of Big Data's benefits,* 2) generating or testing new approaches for *increasing Big Data's economic and public value*, such as through better tools and techniques, improved services and business models, enhanced work practices, and augmented educational programs, and 3) *allaying unfounded fears* through careful analysis of Big Data's realized or plausible costs, benefits, and risks and their distributions across stakeholder groups.

What is Big Data?

- *An innovation cluster that brings together a set of related:*
  - *Ideas, such as "data is an asset that should be monetized"*
  - *Resources, such as "data lakes," Hadoop, and NoSQL data storage*
  - *Techniques and practices, such as data collection, data brokerage, predictive analytics, and automated decision-making*
- *With the potential to be socioeconomically transformative, that is:*
  - *Displacing previously important ideas or practices, such as certain job categories or scientific methods*
  - *Creating sizable new opportunities, for example, new products, services, business models, occupations, organizational forms, and industries*
- *If actually transformative in the future, Big Data will be seen after the fact to have:*
  - *Garnered sustained support and resources from powerful actors in government, business, academia, the media, and the public*
  - *Been widely adopted in multiple sectors or areas of socioeconomic activity, for example, health care, higher education, finance, public services, and civil society*
  - *Had both positive and negative outcomes, including many that were unintended and some that were unforeseen*

The Case for Big Data Implications Research

Big Data is a big subject. Not surprisingly, there are many avenues of research related to Big Data. Some Big Data research *applies* the tools and techniques of Big Data to scientific and industrial problems in health care, marketing, or natural science. Some Big Data research *develops* tools, techniques, and infrastructures for use by other researchers. The focus of this report is research that inquires about the societal *implications* and *consequences* of Big Data's use by academics, governments, businesses, and individuals.

All three types of Big Data research are needed and deserving of public and private support. The reason this report focuses on Big Data *implications* research is that this type of research rarely gets adequate attention compared to research on applications and infrastructure. Implications research is often seen as unnecessary or even as harmful to the interests of Big Data businesses and promoters. To the contrary, workshop participants believe that implications research is necessary, societally beneficial, and highly cost-effective.

This chapter makes the case that Big Data implications research promotes public and private interests in Big Data in several ways:

- It develops knowledge about ways to eliminate barriers to Big Data adoption and value realization and to mitigate realistic risks of harm from Big Data applications
- It reveals opportunities for breakthrough innovations and benefits that might otherwise be missed, and
- It provides evidence to allay unfounded fears about Big Data risks.

Categories of research about Big Data

Implications research is only one of several categories of important and valuable research pertaining to Big Data. A great deal of research is already underway, with substantial government and foundation support, in the categories of Big Data application research and Big Data infrastructure research.

Big Data application research

Big Data tools and techniques are used in academic and non-academic research in the natural and life sciences, engineering, the social and behavioral sciences (including business disciplines), and in the humanities. Big Data ***application*** research aims to answer research questions concerning topics such as:

- The evolution of the universe
- The properties of materials
- The causes of diseases and their most effective treatments

- The consequences of educational reforms, macroeconomic policies, or medical guidelines
- The design of more efficient and effective transportation routes
- Patterns of social media use, and
- The structure and meaning of ancient texts.

Big Data infrastructure research

In addition, Big Data tools and techniques are themselves the objects of research in fields such as computer and information science, mathematics, statistics, information systems, finance, sociology, bioinformatics, physics, etc. Big Data ***infrastructure*** research builds and evaluates tools and techniques, addressing issues such as:

- Data storage and curation
- Data analysis and search algorithms
- Data standards and semantics
- Data visualization and interpretation
- Social media and communication technology features
- Privacy- and security-enhancing technologies, such as encryption or differential privacy[52]
- Identity authentication systems, and so forth.

Big Data implications research

This report focuses on a third type of Big Data research, that is, research on what are sometimes called Social, Economic, and Workforce (SEW) implications and sometimes called Ethical, Legal, and Social Implications (ELSI).[53] Big Data ***implications*** research is conducted by sociologists, economists, and anthropologists, by computer and information scientists and information systems specialists, and by lawyers, public policy scholars, ethicists, and other humanists. It aims to answer questions about topics like:

- Changes in the skills needed by data scientists and how those skills are best taught
- The realized economic and public value of Big Data investments and applications, for example, in the area of open government data
- The uses and consequences of Big Data in everyday life (e.g., the Internet of Things, ubiquitous body cameras, and "willpower support" applications)
- The effects of Big Data on individual privacy, data security, unlawful discrimination, and new forms of social sorting and labeling
- Changes in professional and business practices in specific domains such as finance, health care, education, and science
- The values reflected in the design of the tools data scientists use and how these tools can be improved
- Automated decision-making (e.g., in work scheduling, benefit verification, or loan approvals) and what it means for employment opportunities, quality of work, and the errors citizens experience and the ease of getting them fixed.

These are just a few of the topics that fall under the umbrella of Big Data implications research. A key contribution of implications research is to provide evidence and insights to inform research in the other two categories of Big Data research.

The categories are summarized in Table 1.

Table 1. Categories of Big Data Research

Category	*Typically Conducted By*	*Example Research Areas*
Big Data Application Research	Academic and non-academic scientists and scholars in almost every field	• Ancient texts • Astronomy • Economics • Financial or education policy • Genomics and medicine • Social media use • Transportation and engineering
Big Data Infrastructure Research	Researchers in data-related technical fields such as computer science, mathematics, information science and information systems, bioinformatics	• Data storage • Algorithms • Visualization • Standards • Privacy, security and identification technologies
Big Data Implications Research —the focus of this report	Computer and information scientists, information systems scholars, sociologists, economists, public policy scholars, lawyers, ethicists, other humanists	• Data scientist skill requirements and educational needs • The limitations and improvement possibilities of data science research methods • Economic and public value of Big Data investments • Ethical, legal, social implications of Big Data use in academic research and commercial applications • Social strategies to complement technologies for promoting privacy, security, non-discrimination, etc. • Better practices for Big Data investment, deployment, and adoption • Standards, methods, and role responsibilities for auditing big data algorithms and decisions

Costs and controversies of Big Data implications research

Big Data implications research differs from Big Data applications and infrastructure research in terms of the disciplines of those who do the research and the research questions they ask. It also differs in cost and perceptions of its value.

Costs

Big Data applications and infrastructure research is often quite expensive because of the technical resources needed to conduct it. For example, Big Data application researchers may need to pay private data brokers or aggregators for access to data. They may also find it costly to buy or even to rent from web services providers the computational resources needed for their projects.

By comparison, Big Data implications research can be relatively inexpensive because the labor of researchers and their students is often the major resource required. This means that Big Data implications research can be adequately supported by a small portion of the total funding provided for Big Data research. For instance, genetic engineering and nanotechnology are two prior instances in which significant public support was allocated to implications research. In those cases, there was broad consensus that implications research would be appropriately and adequately served by about 3-5% of the total research funding. Studies have shown that, even with a low level of funding relative to other sciences, implications research can make significant contributions to public policy, economic prosperity, and public understanding.[54]

Controversies about Big Data implications research

At the same time, implications research is more controversial than application or infrastructure research. One important reason is that implications research involves questions of, and potential conflicts over, human values. "What is the distribution of costs and benefits across stakeholder groups?"[55] "Is there a better way to accomplish the same social or economic objectives?" "Where should we invest?" These are among the most fundamental questions of implications research. *Any* answer to these questions faces challenges about the legitimacy of assumptions, the appropriateness of evaluation criteria, the stakeholder groups considered, the validity of the evidence, and so forth.

Some people oppose all implications research, simply because it is vulnerable to such challenges. Other people are willing to support implications research for specific technologies when risks are obvious, as with nuclear power. For example, the scientists involved in genetic engineering and nanotechnology research recognized that their fields have major ethical, legal, and social implications: They advocated for research to illuminate those implications and to evaluate alternative courses of policy and action.

In the case of Big Data, unfortunately, many of those who conduct Big Data *application* and *infrastructure* research do not yet see the need to support the systematic study of Big Data *implications*. Rhetoric about Big Data implications has become quite heated. "Data enthusiasts" accuse those who raise concerns about risks of being "neo-Luddites"[56] or "technophobes", whereas the latter label the former as "data evangelists," "data fundamentalists," or even as "solutionists."[57] Although many people share the goal of "maximizing the benefits of data-driven innovation," they differ sharply on preferred ways of achieving that goal. Those concerned about Big Data risks propose efforts to anticipate risks and evaluate alternative approaches for addressing them. Those concerned about negative public opinion and increased regulation refuse to acknowledge the possibility of problems that are not yet obvious.

Different beliefs or fears could motivate unwillingness to think and talk about the potential risks associated with Big Data. Among them are:

- Belief that Big Data poses no risks or that risks are minimal and far outweighed by benefits
- Belief that any Big Data problems that may eventually arise can be easily dealt with by technology enhancements when the time comes[58]
- Belief that any eventual Big Data problems are the fault of Big Data's *users* and thus not the responsibility or the concern of those who *develop* or commercialize Big Data products and services
- Fear that open discussion of concerns about Big Data, whether well founded or not, will turn the public against Big Data
- Fear that open discussion of concerns about Big Data will lead to pre-emptive or ill-conceived government regulation that will block innovation.[59]

Some of these beliefs and fears are better founded than others. For example, the Big Data educational initiative InBloom was terminated in the wake of public opposition.[60] "Taxi app" company Uber, which uses its considerable Big Data resources as a bargaining chip in its dealings with local governments,[61] is facing regulatory scrutiny and constraint. But other beliefs are more questionable: 1) whether risks can adequately be dealt by technical enhancements after-the-fact, 2) whether the threats to "innovation" are increased by open discussion and research about potential risks, and 3) whether technologists benefit from ignoring the unintended consequences of their innovations.

Without attempting to treat these issues comprehensively, the following points bear mention. First, even some strong proponents of Big Data acknowledge that Big Data has great potential risks that may be difficult to address after-the-fact.[62] Today's increasingly wired and interconnected economy makes it challenging to limit harms,[63] as the increasing frequency and intensity of cyberattacks and data breaches shows.[64] Deanonymization may be getting easier[65]: datasets that are now secure may become insecure later. Furthermore, the history of research on technological innovation indicates that additional technical

development alone is rarely sufficient to address societal challenges. To manage risk and to realize benefits, technological innovations must be complemented by "social" innovations in education, use practices, and shared governance.[66]

Second, failure to acknowledge the possibility of risks may be more likely than open public discussion to lead to rejection or unwanted regulation. The InBloom initiative did not fail because of early and open debate about what the initiative implied for student privacy; it failed because parents were not consulted during the initiative's design. Uber avoided working with regulators until regulators started to take action. Lack of open discussion of potential risks and consequences can be more harmful to innovation than public debate, because dismissing or belittling people's concerns convinces them that there is indeed something to hide. As noted by a workshop participant whose employer sells Big Data products and services: "If Big Data turns out to be nothing more than Big Business, Big Government, and Big Brother, we're all in trouble." (Another participant observed that many people now believe that Big Data is "by the people, of the government, for the private sector.")

Some experts further claim that public discussion that anticipates "worst case scenarios" is socially beneficial even when predictions of future consequences are flawed. Risk assessment encourages people to take actions that prevent risks from materializing, and it increases people's resilience when bad things actually do happen.[67]

Third, technology developers should be concerned about the unintended consequences of their innovations,[68] because *it is in their interest to do so*. For one thing, developers "know the technology and hence are in the best position to create norms that will maximize the benefits and minimize the risks of using such systems."[69] For another thing, reluctance to think through potential unintended consequences of technology may reflect a flawed design theory. Educators in the field of user-experience design report that failing to consider the potential unintended consequences of high-tech innovations leads to missed opportunities, whereas explicit technology foresight generates new insights, innovative designs, and better economic outcomes.

Better design theories can produce better results for both innovators and technology users. The insight of modern design theories, which go under names like "value-sensitive design," "privacy-by-design," and "responsible innovation,"[70] is that processes that attempt to satisfy seemingly conflicting goals or values often lead to breakthrough innovations in a way that straightforward functional engineering cannot[71]. By considering value-focused non-functional requirements (such as quality, manufacturability, biodegradability, and users' concerns for privacy and security) in addition to functional requirements, designers can generate ideas that lead to radical innovation rather than incremental improvements. The value of this design orientation is captured in the book title "Quality is Free"—a book that overturned 1970s-era conventional wisdom that

producing high-quality products and services always costs more.[72] In this light, open discussion and systematic research on Big Data's social, ethical, economic, and workforce implications is both necessary and beneficial for promoting innovation. Transparency and ethical business are "good business."[73]

Workshop participants acknowledged that legitimate controversies exist about the value of Big Data implications research and the ways in which it is conducted. However, they concluded that implications research has tremendous potential to improve the applications and innovations of Big Data and their economic benefits. As one participant put it: "We need data humanists, not just data scientists." They further agreed that public and private investment in implications research is a highly cost effective way of achieving those desirable outcomes.

Summary

Public and private investment in Big Data implications research is needed and valuable because it can:

- Develop knowledge about ways to eliminate barriers to Big Data adoption and value realization and to mitigate risks of harm from Big Data applications
- Reveal opportunities for breakthrough innovations and benefits that might otherwise be missed, and
- Provide evidence to allay unfounded public fears about Big Data risks.

Ignoring or dismissing the possible risks of Big Data is counterproductive, because it can actually *increase* the chances of public backlash or restrictive regulation. By contrast, implications research is a cost effective way to increase Big Data's public and private benefits. The level of investment required to achieve these objectives is modest compared to current and likely future investments in Big Data application and infrastructure research—on the order of 3-5% of total Big Data research funding.

The number of research topics and questions subsumed under the label of Big Data implications research is vast. Research priorities will undoubtedly change as experience with Big Data grows, together with the body of research findings. The next section presents workshop participants' current views of the major priorities for Big Data implications research.

Big Data Implications Research Priorities

Workshop participants discussed three domains of high-priority Big Data implications research: 1) science and science–technology policy, 2) individuals and everyday life, and 3) organizations and work. Several themes cut across all three domains. This section describes three such themes: 1) values and ethics, 2) data, algorithm, and decision quality, and 3) knowledge and skill.

Science and science–technology policy

Big Data has important implications for science and science–technology policy. Workshop participants identified three research priorities in this domain: 1) research on the success of new data partnerships, 2) the collection and analysis of data about the new data economy, and 3) the study of emergent changes in universities, academic disciplines, and journal publishing.

New data partnerships

The Big Data era is challenging neat divisions of scientific labor across government, universities, and businesses. Government is not just a funder of scientific research; it is also a major data broker. Universities not only house scholars who collect and analyze research data; universities also partner with businesses, purchase data and analysis facilities from businesses, and spin-off research-based businesses. Businesses not only consume and produce research; they also buy and sell data that can be used for research by other businesses, governments, and universities.

Workshop participants raised many questions and a few concerns about these shifting activity patterns. For example, are partnerships between universities (or individual scientists) and private-sector businesses an efficient and effective way to provide researchers with access to data and analytic resources? If so, how can universities ensure the rights of researchers to publish their findings? (This is important because open dissemination of research is a fundamental academic mission.)

In a related development, government agencies have been charged to make their records available for use by individuals, businesses, and universities.[74] How can governments ensure the protection of the personal information in government databases (or the personal information that can be created by linking official records with other data sources)?[75] Today, protection of government data relies mainly on bringing researchers to sites where access to and use of data can be physically controlled. Are there ways (e.g., encryption, differential privacy, or analysis behind firewalls) to achieve adequate protection while allowing remote data access? How can government agencies vet the quality of the analyses and conclusions based on government data?

Questions like these naturally draw attention to a growing number of public–private partnerships focused on open government data. Many such partnerships are forming under the auspices of President Obama's $200 million Big Data initiative.[76] Evidence on the success of these partnerships will provide useful information to guide future investments. Also useful will be careful documentation and wide dissemination of the "lessons learned" employed in successful partnerships. That is, future data initiatives can benefit from knowledge about the social and technical strategies that successful partnerships devise for the protection of data and personal information privacy and for the validation, dissemination, and realization of value from analysis results.

Data metadata

Workshop participants bemoaned the fact that so little data is available about Big Data. Which companies buy data? Which companies sell data? How much data is bought and sold? What are the business models of companies that sell data?[77] What terms of use govern data sales, and what are the implications of these terms? How many hands does data pass through in the "data supply chain" from initial collection to their eventual reuse in science, engineering, marketing, or consumer-oriented apps? In short, participants identified a lack of what might be called "data metadata." (Metadata is data about data.)

If the new data economy turns out to be as socioeconomically transformative as many observers claim, there will be great societal benefit in research on the processes and outcomes of transformation. How many new industries, companies, and jobs does Big Data create? Do metropolitan "data parks" spur economic development and growth? How does the buying and selling of data change the economics of non-data products and services? What legal frameworks and regulatory regimes promote or inhibit the new data economy, and what opportunities are there to achieve better outcomes?

Answers to questions like these will be beneficial in two ways. First, they provide the evidentiary basis for business and government policy decisions to improve Big Data investments, uptake, and outcomes. Second, through comparisons with prior transformations, knowledge about the changes brought by the Big Data innovation cluster can be useful for policy-making in the future.

Evolution of the academy

A third research priority in the domain of science and science-technology policy concerns emergent changes in universities, academic disciplines, research publishing, and other aspects of "knowledge infrastructure."[78] Some observers have claimed that Big Data will change the nature of scientific research, reducing the roles of theory, statistical sampling, and hypothesis testing. In their place, it is said, will be the analysis of "N=All" using massive computational resources.[79]

There is certainly evidence suggesting movement in that direction. In business- and policy-related disciplines, Big Data is a popular theme of conferences and journal special issues. Many new Big Data educational programs are being launched, and academic position announcements calling for Big Data researchers are numerous. So the implications of this development are worth considering.

Universities are justifiably concerned about the resources required to acquire, store, and analyze massive datasets. Universities that fail to acquire the needed resources may lose top research talent. Will universities experience an exodus of researchers to social media companies and other data brokers? Partnerships with industry are one way to acquire the needed resources, but those relationships sometimes come with strings attached. Will partnerships affect 1) the kinds of research questions asked, 2) ability to publish research findings, and 3) ability of other scientists to replicate and validate research findings? How do or should partnerships between researchers and corporations affect universities' human subjects protection responsibilities?[80] Research on these questions and strategies for maintaining scientific integrity in research partnerships are clearly needed.

Changes are also likely to be felt within and across academic disciplines. Disciplines already vary considerably in the kinds and availability of data used by researchers. They also differ in norms about data sharing[81] and the recognition accorded for developing and supporting data resources that other scholars can use.[82] Similar divisions can also occur *within* disciplines, raising concerns that qualitative, survey, and experimental research methods may lose prestige and resources in comparison to Big Data analytics (or may, in extreme cases, be entirely rejected as allowable research approaches in certain fields).

The rise of Big Data research in traditionally "small data" disciplines poses questions about the need for new types of *research and publication reviews* and incentives.[83] For example, social media data are often claimed to be "public data," implying that privacy protection is not the researcher's concern. However, by the terms and agreements of social media websites, much so-called public data is actually the *private property* of social media companies. This suggests that universities ought to review and monitor Big Data research proposals for potential violations of intellectual property rights. Yet universities' existing policies and procedures may not be adequate for this kind of review. For instance, university Institutional Review Boards (IRBs, required by government funding agencies to protect the rights, including the privacy rights, of human research subjects) are expressly prohibited from considering issues other than human subjects protection. Consequently, IRBs cannot oversee intellectual property rights or other rights of corporate research subjects.

Publication review policies may also need Big Data-related enhancements in some fields. In the natural sciences, journals often now require "open data" as a condition of

article publication,[84] and this policy has been proposed for researchers in business disciplines as well. But, when research data are proprietary (owned by a corporation), providing open access to the data is not an option. When, in addition, a dataset includes personally identifiable (or re-identifiable) information, the legality and ethics of open data access are questionable. On the other hand, access to other researchers' data and analytic code seems one of the most promising ways to ensure proper validation and replication of research results. Academic fields and journal editors need to devise and share strategies for addressing this dilemma.

In short, workshop participants called for a program of Big Data research focusing on implications for universities, university-based researchers, and academic journal publishing. This research would also explore emergent "better practices" that deserve to be widely disseminated. For example, today, individual universities and the editorial teams assigned to submitted articles conduct reviews of research ethics. Could these reviews be better handled by academic associations or even through crowdsourcing?

Individuals and everyday life

Three key research priorities related to individuals and everyday life were discussed at the workshop: 1) new ways of thinking about privacy and privacy protection, 2) the careers, activities, and contributions of "data amateurs," and 3) the personal benefits and burdens of Big Data applications and use.

Privacy reimagined

Workshop participants raised the topic of personal information privacy more frequently than any other topic during the workshop. They acknowledged the benefits of Big Data but observed that there has not yet been sufficient attention to how to achieve those benefits without incurring individual privacy harms. (Participants also insisted that "Big Data is not *just* about privacy." That is, privacy harms are not the only important concerns that Big Data raises for individuals. Big Data also has implications for discrimination, surveillance, inequality, and human dignity.)

Old ways of thinking about privacy do not fit today's realities.[85] First, data brokers use Big Data matching techniques to compile detailed profiles on individuals from data sources that may not contain personally identifiable information like name, address, and social security numbers, thereby defeating traditional protection approaches (e.g., anonymization).[86] The locus of privacy threats has shifted to sources of information that are not under regulatory protection,[87] such as social media, location-based tracking, and body cameras. (Alternative data sources are now being used instead of official records for purposes like identifying people for clinical drug trials.[88])

Second, the harms associated with disclosure of personal information are poorly understood by many. Not limited to the annoyances of spam and cross-platform online behavioral advertising, personal information disclosure (whether voluntary, as on social media sites, or involuntary, as in the case of sensor tracking and credit card data mining) exposes individuals to risks such as not getting a job or getting fired from one, not getting insurance or getting much more expensive insurance, theft of financial, health, and social identity, online harassment or targeting for crime, etc. When data collected in one context is used in another, the opportunities for unfairness multiply. Out-of-context privacy breaches can occur through automation without any single individual viewing private data—a situation that reduces transparency and the ease of correction when errors and unfairness occur.

Third, it has become harder for people, especially young people, to protect the privacy of their personal information.[89] In part this stems from desire to reap the immediate benefits of "free" apps and social network sites. In part it occurs because commercial privacy policies[90] and the privacy control features of today's technologies are complex, opaque, and frequently changed.

Finally, any individual's attempts at privacy protection can be rendered irrelevant by decisions of their friends and family members, such as decisions to use a particular email service, to employ a particular device for texting, or to post photos and other content on a particular social media site. When friends and family members opt in, individuals cannot opt out. In brief, ways of thinking about and regulating privacy have not kept up with technology, commercial business practices, or changing social norms.

The conclusion of workshop participants was that privacy should remain an important focus of Big Data implications research, and that priority should be given to attempts to ***understand*** and ***protect*** privacy in new ways. For instance, should the starting point for thinking about privacy be the assumption that privacy is "dead" and that monitoring is ubiquitous (through security and body cameras and the Internet of Things)?[91] Can the case for public ownership of private patient data be made?[92] Should privacy be thought of as "the freedom to be forgotten?" Should efforts to regulate data privacy shift from focusing on data collection and sharing to focusing on data use? Could innovative identity and access management techniques that are intended to ensure data *security* (such as those under development by the National Institute for Standards and Technology[93]) also have application in *privacy* protection?

Privacy protection research should *not*, participants argued, focus solely on *technological* solutions. High-tech solutions that people ignore or circumvent will invariably fail. Therefore, low- or no-tech strategies—including education, training, human oversight, and sanctions—are needed to complement privacy-enhancing technology.[94] Designing and evaluating effective *combinations* of technologies and social practices is a

clear priority for Big Data implications research.

Data amateurs

A second priority for Big Data implications research related to everyday life concerns the behavior of data amateurs, that is, people who are not (or not yet) acting on behalf of an organization and who may have no formal training in data science. It is important to understand amateur data scientists for several reasons. One reason is that informal learning is both a complement to, and an alternative point-of-entry into, occupations requiring strong technical skills, such as programming and data science. In other words, given the projected shortfall in the number of data scientists, it is important to explore alternative ways to fill those needs.

Amateur data scientists include, for example, the would-be social or business entrepreneurs who participate in data "hackathons." Casual conversations with hackathon conveners suggest that it is rare for these events to be formally evaluated. As a result, little is currently known about the opportunities that these events provide, outside of formal educational programs, for citizens to learn data science skills and gain access to data and tools. A related unanswered question is the social and economic value created by citizen data scientists when they establish social or business enterprises.

Data amateurs also include individuals who participate in the movement called "the quantified self." A growing number of people use devices and apps to log every aspect of their daily lives, including their physiological and psychological states. Intentionally or unintentionally (through the features and policies of websites and apps), they share this data with others. Research is needed to understand the consequences of these behaviors, not only for personal health and privacy, but also for society.

Another group of data amateurs is the growing number of "crowdworkers" who volunteer their efforts in Big Data science projects and industrial innovation contests or who work for very low wages at tasks that are assigned and coordinated by crowd-sourcing "platforms." Some observers claim that these data amateurs are already making an appreciable contribution to the economy, possibly without adequate recompense—two hypotheses worthy of implications research.

Unfortunately, the category of data amateurs also includes people intent on online harassment (including cyber-bullying and hate crimes) and would-be thieves. Understanding how these individuals pursue their dark-side data science careers might reveal the keys to better deterrence and detection. Other amateur uses of Big Data may not be illegal, but may nonetheless raise social and ethical concerns. For example, people have been known to use Big Data capabilities to identify closeted gays or the parents of unacknowledged children. Research on these Big Data uses and their societal consequences could contribute to public understanding and the promotion of social norms.

Benefits and burdens

A third priority for citizen-oriented implications research is documenting both the benefits and the burdens that individuals experience from Big Data. As noted earlier in this report, proponents of Big Data anticipate many health, financial, and other personal benefits. But actual benefits may be less, more, or different than expected. Learning about the outcomes actually achieved can help business and public policy makers and technology developers make better Big Data investment and design decisions.

For example, under the category of *benefits*, it would be useful to know whether and to what extent individuals make better health decisions when they use fitness monitors or "willpower-support" apps on their smart watches or when they acquire online access to their electronic medical records. Similarly, does better public information about dodgy investment advisors or the real costs of payday loans actually reduce financial victimization? Does information about one's driving habits change driving behavior (with or without the added incentives of savings on insurance or fuel)?[95] Do consumers really benefit from data-driven price and/or feature comparisons among products and services like health insurance or telecommunication plans?[96]

The *burden* side of the equation is equally important to understand. For example, the Justice Department annually publishes estimates of the dollar impacts of *financial* identity theft. But identity theft can also involve *health* data[97] and *social identity,*[98] where costs are not so easily measured in financial terms. What are the *personal* costs in time and stress, as well as money, of these data-oriented crimes?[99] What are their aggregate implications for national wellbeing and productivity?

Among the other burdens that individuals may bear is not having access to their own data. For example, telecommunications operators maintain detailed consumption data, but do not provide it to consumers in a form in which it can be analyzed easily to compare subscription plans. Medical device manufacturers have not always been willing to share detailed personal medical data with individuals and their physicians. What are the consequences of being denied access to one's own information or to the economic benefits of one's data because of data ownership or licensing agreements?[100]

Big Data also harbors the potential for unfair treatment. What informational and institutional resources are available for redress when citizens think they may have suffered from corporate or government decisions based on inaccurate data or faulty algorithms?[101] What happens to the graduates who cannot get jobs because of their social media footprints that the systems never forget? How is individual identity and life experience shaped by "algorithmic regulation" and new forms of "social sorting" and discrimination?[102] How and how much is civic discourse shaped by "filter bubbles" in which algorithms match us with people, news, and views that they think we want to hear?[103] Does Big Data enable perfect price discrimination, eliminating consumer

surplus?[104] These are just a few of the research questions related to the burdens individuals may experience from the Big Data revolution.

Organizations and work

The third domain in which workshop participants discussed research priorities was that of organizations and work. Participants noted that far more research attention has been paid to the implications of Big Data for the science and civic domains than for the domain of corporations and their employees. Yet this domain is highly significant for national employment, productivity, innovation, and competitiveness. Three research priorities in this domain were identified: 1) the skill requirements and professional conduct of corporate data scientists, 2) the ways in which corporations manage data and algorithms and make decisions about decision automation, and 3) emergent changes in jobs, knowledge, and expertise brought about by Big Data.

Corporate data scientists

Huge projected shortfalls in the number of corporate data scientists are a much-discussed barrier to achieving Big Data benefits.[105] In response, university departments of mathematics, statistics, computer science, information systems, and management science (and some combinations) across the country have set up programs to train data scientists. Observing that some of these programs seem to consist entirely of courses on different data analysis techniques, workshop participants discussed what *else* corporate data scientists might need to know to perform effectively in their roles. A surprising number of issues surfaced in workshop discussions that might never be mentioned in courses devoted to the data science techniques of classification tree analysis, machine learning, genetic algorithms, sentiment analysis, or social network analysis. Among those other issues are:

- ***Biases in available data, errors of interpretation, and strategies for dealing with biases and error.*** For example, the millions of interactions captured on social media websites do not actually constitute "N=All" (because some people still opt-out), and hence may not yield representative findings. Data scientists studying housing trends need to know that county deed records capture both housing sales and mortgage foreclosures and that the trending topic of "coke" may not mean "Coca-Cola." Another area of potential concern is the substitution of correlational thinking for causal analysis. Data scientists should be well schooled in practices for validating analyses and interpretations.
- ***Data quality problems and strategies for overcoming them.*** Conversations with data scientists reveal that high quality data are key to data science successes. But corporate data is rarely of sufficient quality. The projects needed to improve data quality can be politically challenging, time-consuming, and costly.[106] Whether or not data scientists perform this work themselves, they need to know when it needs to be done and how to convince others to do it.

- ***Challenges arising in the attempt to move from data-driven insight to action.*** The valuable insights generated by data science need to be put into organizational practice. This is the responsibility of managers, not of data scientists. However, data scientists have essential roles in persuading and educating managers and preparing them for the ongoing operation and maintenance of data streams and algorithms. Having the ability to analyze organizational authority structures and process flows can help data scientists contribute more effectively to constructive organizational change.
- ***The potential "reactivity" of predictive analytics.*** Humans are subjects, not passive objects. They have the ability to react to the decisions and actions taken on the basis of Big Data analyses. They can learn to "game" analytic predictions in ways that defeat intended purposes. They can also react in ways that create "self-fulfilling prophecies." This reactivity has important implications for data scientists and their clients. For instance, predictive models must be repeatedly reassessed for the validity of their underlying assumptions and for the fit between their predictions and reality. To be effective, data scientists need the knowledge and skills to do this (with the help of their clients).
- ***Awareness of potential legal and ethical concerns and potential unintended consequences of Big Data projects, and ways to deal with them.*** Data scientists need to know about relevant privacy and security laws and regulations and how they apply to Big Data projects. They also need to be able to recognize and find acceptable resolutions to ethical concerns and value conflicts.[107] For example, customers or other stakeholders may find a company's *legal* uses of Big Data to be *unacceptable* as well as unavoidable. Data scientists should be aware of the potential value conflicts and ethical dilemmas and should be skilled in discussing them with clients and in resolving them through approaches like "value sensitive design."[108]

How well do today's educational programs prepare future data scientists in these areas? The answer is not very well. The majority of current programs focus almost exclusively on either the building of data science tools and infrastructure or on the use of data science techniques (such as machine learning) to solve research or practical problems. But a few programs offer limited opportunities for students to learn about data protection law, ethical dilemmas in big data use, and how to confront ethical concerns constructively.

The ethics of computing is an established research area within computer science. A number of computer and information science programs offer courses on ethical theory and its application[109] or approaches for building positive values like privacy into the design of systems and devices.[110] Despite these efforts, educators observe "an 'ethics gap'" in technical education, noting that technical professionals are rarely prepared to deal well with the many ethical challenges in contemporary professional practice.[111] Among those challenges are pressures from clients and superiors to do something that a professional

believes is wrong. (Clients are problem owners such as business managers, who commission data scientists to work on particular projects.)

Data science is also taught in the business school departments of management science and information systems. Business analytics program descriptions typically make little or no mention of data privacy and security,[112] and few programs dedicate a full course to these topics. Of non-technical subjects, data science programs in business schools emphasize "consulting skills" to help future data scientists converse with clients and understand business requirements. The business schools in which these programs are housed often *do* offer courses in business ethics, but the content of those courses does not usually deal with Big Data-related topics like data protection.

One possible explanation for "the ethics gap" of technical professionals is that dealing with data quality issues, post-project action, privacy, security, value conflicts, and ethical dilemmas is seen as not belonging to the data scientist's role,[113] but is instead viewed as the responsibility of clients. However, it is not clear that clients today have the knowledge and skills to perform well in this role. Ethics educators in business schools point out that managers-in-training generally lack the training and skills they need to speak up when they believe something is wrong.[114] Managers in the workplace may also be ill prepared to handle the issues raised by big data. A workshop participant described a data science project in which the human resources (HR) professionals were so ill equipped to use analysis software that data scientists ended up making key HR decisions. Along similar lines, the business press has cited lack of business knowledge or confidence about Big Data as one of the barriers to its uptake,[115] and experts are urging managers to take on a bigger role in Big Data decision-making.[116]

Workshop participants did not know of any courses (e.g., "Data science for non-data scientists") that might prepare future managers and other knowledge workers for taking on a Big Data decision-making role, outside of executive education. Such courses are clearly needed and are likely to emerge over time. But such courses will not obviate the need to also educate *data scientists* about value conflicts and ethical dilemmas and how to deal with them. Responsibility for dealing with such issues is inevitably a shared one, and *all parties*—data science professionals, data science clients, users, and people affected by Big Data use—must be, but generally are not yet, educated to deal with them.

Attention to data protection and ethical issues is also limited in the certification programs offered by industry groups. INFORMS,[117] a professional society with origins in management science, offers a certification program (CAP) for analytics professionals. The CAP Code of Ethics/Conduct[118] admonishes professionals to act professionally, to follow applicable laws, to resist pressures to produce analyses biased toward a particular result, and to avoid unauthorized or illegal use of intellectual property. The single mention of privacy occurs in the context of human subjects protection, which is familiar in academic

and health care *research* contexts but much less known in the general corporate world. Privacy does get a few mentions in the materials provided to help people prepare for the CAP certification exam, but the sample test questions do not cover data protection law or the ethical dilemmas related to big data.

The Code of Conduct of the Data Science Association[119] is more specific than that of INFORMS about data protection issues. While not actually using the word privacy, it offers a detailed definition of confidential information that clearly includes the personal data of employees and customers. Among other things, the Code enjoins professionals to "do no harm," that is, avoid to negative side effects. (The Code of Ethics of ACM,[120] a computer science professional association, makes a similar exhortation.) The Data Science Association Code also states that the professional is to abide by the client's objectives and acknowledges the dilemmas that doing so may entail. However, it stops short of clarifying the data scientists' obligations in cases where the client wants the data scientist to do something that is legal, but ethically questionable:

> "A data scientist shall not counsel a client to engage, or assist a client, in conduct that the data scientist knows is criminal or fraudulent, but a data scientist *may* discuss the consequences of any proposed course of conduct with a client and *may* counsel or assist a client to make a good faith effort to determine the validity, scope, meaning or application of the data science provided." (emphasis added)

In short, neither budding data scientists or managers-in-training appear well prepared to address Big Data's ethical grey areas, which continue to evolve with technology development and social change. Data scientists and clients need education in this area. Although it would be ideal to devote an entire course to ethical issues (as is done now in some computer science programs), most programs could benefit from packaged course modules that could be included in existing courses on consulting skills or requirements analysis. Most instructors of such courses do not have the time or expertise to develop such modules on their own, so a high priority for Big Data implications research funding is to support the development of appropriate educational materials.

The literature on ethics education concludes that abstract discussion of ethical theories is only one component of effective education. Students also need the opportunity to discuss and role-play cases that are representative of the kinds of situations they will encounter on the job.[121] Producing course materials that provide opportunities for experiential learning about the legal and ethical issues associated with Big Data is a top priority of Big Data implications research.

Corporate data practices

A second important priority for Big Data implications research in the organizations and work domain concerns how organizations do and should govern Big Data and make

decisions about its use. It has frequently been observed that data brokers and social media companies are not very transparent about their policies and practices of data collection, use, and reuse (including sale to other organizations), despite (or possibly because of) their lengthy privacy policies and terms and conditions of use. On the other hand, complete external transparency about Big Data uses and policies is not a current requirement for businesses, and there is no social consensus that it should be.[122] Furthermore, companies may fear that greater voluntary disclosure could expose them to legal liability or put them at a disadvantage compared to companies that do not disclose.

In this situation, implications researchers can provide real societal benefits by disseminating information about actual corporate Big Data practices that (possibly anonymous) organizations provide to researchers voluntarily. To the extent that implications research identifies "better practices," research findings can help allay public fears about the downsides of Big Data, and it can also promote the adoption of good practices. To the extent that implications research identifies unresolved challenges or problems, research can stimulate the design of innovative solutions.

What aspects of corporate Big Data practices might benefit from implications research? Possibilities include (but are not limited to) the following:

- Innovative and effective strategies for dealing with information privacy and data security threats
- Corporate values about data use and stakeholder protection;[123] how organizations monitor and control the external business partners that access and use the organizations' data
- How organizations decide which Big Data projects to pursue and which not to pursue; whether these decisions are well informed by model-builders' knowledge of technical limitations[124]
- How organizations decide which decisions should be fully or partially automated; how they validate, maintain, and ensure oversight over algorithms and the people who use them, for example, through peer reviews of code[125]
- How organizations cope with algorithm opacity; for example, by triangulating multiple algorithms types, through visualization techniques, through explanation facilities, or by building "self-documenting" models
- What the critical threats to data quality are; whether there are questions for which poor data quality does *not* affect decision quality; the social and technical strategies that organizations use to assess and improve data quality and data "fitness for use"[126]
- How organizations "monetize" their data assets; whether and how data sales affect the economics of non-data products and services
- Whether (and how) the terms and conditions governing use of purchased data limit the ability of data-buying organizations to commercialize innovative new products and services

- What outcomes organizations achieve from Big Data projects; how organizations bridge the gap between "data and insights" on the one hand and "action" on the other hand; how frequently Big Data uses lead to the development of innovative new products, services, production processes, etc.
- How organizations build and maintain data scientists' *business* knowledge and the *data-related skills* of other professionals and managers; how organizations enhance the knowledge and skills of both data scientists and their clients in dealing about the legal and ethical implications of Big Data projects
- How organizations can leverage the activities of auditors and certifying organizations to ensure that progress toward corporate *social* responsibility includes corporate *data* responsibility.

A particularly important area for research on corporate practices is whether, how, and to what effect corporations embrace initiatives, such as the ***New Deal on Data***,[127] that aim to accommodate citizens' values and needs along with those of corporations. In short, implications research that focuses on corporate decisions and practices related to Big Data can help allay unfounded fears, disseminate practices that reduce risks, and promote innovation in areas needing improvement.

Knowledge and expertise

A third research priority concerns the implications of Big Data for knowledge work and human expertise. As noted, Big Data has created big demand for data scientists and for a whole range of other data-related technical jobs. Less clear are Big Data's implications for the employment opportunities, job skills, and work quality of *other* workers,[128] especially *high-skill* knowledge workers, including engineers, technicians, doctors, lawyers, and college professors.

Since the industrial revolution, there have been many waves of anxiety about the potential impact of automation on jobs, skills, and wages.[129] Wave after wave of past anxiety has faded. Although automation has eliminated numerous jobs and some craft skills, many observers believe that automation creates more jobs than it destroys, because of the innovation it unleashes.[130] These new jobs, often entirely new occupations, are thought to be of higher skill than the ones eliminated and therefore able to command higher wages. However, it is rarely the case that the new jobs can be filled by displaced workers.

Big Data is again provoking automation anxiety,[131] and there is renewed academic interest in the effect of automation on jobs and employment.[132] Although many scholars remain highly optimistic, a few observers argue that this time may indeed be different.[133] One recent analysis concluded that 47% of all US jobs are vulnerable to automation;[134] those researchers concluded, however, that high-skill and high-wage professional jobs were not at risk (yet[135]). Other scholars disagree, noting that automation (specifically,

artificial intelligence and robotics) has already progressed past the midpoint of the human occupational skill spectrum.[136] Systems today are rivaling or exceeding human performance in an increasing array of tasks formerly thought to be the sole province of humans, from the ability to read facial expressions and medical images to the ability to decode street signs. As a result, traditional high-skill and high-wage jobs may also be at risk.

Although it may be a long time before artificial intelligence becomes general enough to replace the *entirety* of the high-skill jobs that people perform today,[137] automation is steadily encroaching on knowledge occupations. A typical (though not universal) strategy is to divide up a highly skilled job, once part of that job can be automated.[138] The task of using the automated system is assigned to less-skilled workers, while experts are retained to oversee their work and handle especially difficult situations. This process is usually considered a win-win scenario: Lower-skilled workers now do work that is important, prestigious, and better paid; experts are relieved of routine, boring jobs and become more productive. At the same time, the number of job opportunities available to experts may decrease.

This scenario raises additional issues that require systematic study, as automation continues its advance.[139] One issue is the possible loss of expert knowledge and skill (a process called "deskilling"). Knowledge and skill in *using automation* is not the same as knowledge or skill in *performing the task* to which automation is applied.[140] Therefore, lower-skilled workers using "smart machines" that make them more productive may not actually have deep knowledge of the work they perform. They may not be able to do the work at all if automated systems become unavailable. And they may not have the knowledge (or authority) to deviate from system-generated scripts when deviations are called for. In addition, knowledge and skill erode when a task is infrequently performed, as happens when experts are only called in to handle difficult problems. Therefore, even experts may find themselves less expert over time and thus less able to provide knowledge inputs for system maintenance and improvement. There are also questions about where future generations of experts will come from, if entry-level paths to expertise are closed off through automation.[141] Finally, the tendency to assume that machines, unlike people, make unbiased and correct decisions, may lead to a reduction in human decision autonomy in jobs and, consequently, to a reduction in job quality.

In another automation scenario, traditionally trained professionals may not have the skills or inclination to make data-driven decisions, leading to their substitution by "quants," who are more skilled with technology and numbers than with the subject matter. In other cases, the need for speed (e.g., in automated securities trading or vehicle operation) may exceed human capabilities to monitor, control, or enhance technology, with the result that humans are largely removed from the automation control loop.[142]

High-Priority Domains for Big Data Implications Research

1. ***Science and science-technology policy***
   a. *The characteristics and success factors of new data partnerships*
   b. *The collection and analysis of data about the new data economy, and*
   c. *Emergent changes in universities, academic disciplines and journal publishing*
2. ***Individuals and everyday life***
   a. *New ways of thinking about privacy and privacy protection*
   b. *Careers, activities, and contributions of "data amateurs," and*
   c. *Personal benefits and burdens of Big Data applications and use*
3. ***Organizations and work***
   a. *Skill requirements and professional conduct of corporate data scientists*
   b. *Management of data and algorithms and decision making regarding decision automation within corporations and government units, and*
   c. *Emergent changes in jobs, knowledge, and expertise brought about by Big Data*

Big Data presents an additional challenge to work through intensive employee surveillance. Even in its more benign form, the measurement of every aspect of work coupled with transparent reporting—sometimes called "the quantified organization"—can lead to workplace stress and alienation.[143] But the powerful tools Big Data provides to organizations to monitor employee's actions and communications also raise employee privacy concerns. In today's environment of heightened regulatory scrutiny[144], organizations are clearly motivated to use employee surveillance tools, as a recent headline shows: "JPMorgan Algorithm Knows You're a Rogue Employee Before You Do."[145] The question is whether, in the attempt to protect *corporate* assets and *customer* data, workplace surveillance infringes on *employees'* rights and undermines the quality of work.

The point of this discussion is *not* that the cumulative effects of knowledge automation are certain to be negative. The opposite may, in fact, be the case. The point is

that the implications of Big Data for knowledge work and expertise are *potentially so significant socioeconomically* that they demand systematic study. There are no easy answers to questions about what the role of humans should be, and the capabilities of machines are a moving target in any case. Nevertheless, without sound evidence about how companies make decisions about automating jobs, and what the outcomes of those choices are, policy-makers and the public will not have a say in one of the most important current developments affecting jobs, work, employment, and productivity. Instead, the future world of work will be largely shaped by the unnoticed day-to-day decisions of people whom workshop participants called "invisible technical workers."[146]

Cross-cutting themes

The sections above presented Big Data implications research priorities grouped into the three broad domains of science, civic life, and organizations. These domains are already much broader than is typical; it is more usual to parse Big Data implications into specialized areas such as health, finance, and education. At the same time, even the broader categories used in this report exhibit shared themes. Here we consider three: values and ethics; data, algorithm, and decision quality; and knowledge and skill.

Values and ethics

Value conflicts pervade Big Data. Many groups have a stake in what Big Data can bring: universities and academic researchers, research funding agencies, government agencies, consumers, citizen scientists, fitness buffs, data brokers, social media companies, other organizations like health providers and consumer products companies, data scientists, and many other kinds of knowledge workers. Some value conflicts reflect differences in people's incentives, since Big Data's costs, benefits, and risks are not evenly distributed across stakeholder groups. Some value conflicts have their origins in cultural differences. But given the number and diversity of Big Data stakeholders, there will always be conflicts about the goals to be achieved and the means to get there.[147] This means that any person or organization that commissions, develops, commercializes, or uses Big Data applications confronts ethical dilemmas about ends and means.

A common tendency is to avoid value conflicts, in the belief that efforts to balance conflicting objectives involve tradeoffs that reduce each party's benefits. Proponents of value-sensitive design approaches believe, however, that confronting value conflicts squarely promotes innovation that can improve outcomes for all.[148] Workshop participants advocate for Big Data implications research that acknowledges potential value conflicts and seeks their innovative resolution.

Data, algorithm, and decision quality

Big Data is not always better data.[149] For some purposes, small data is sufficient or even superior.[150] Big Data can be bad data, when data is inaccurate, misanalysed or misinterpreted, or acted on inappropriately. This observation puts a premium on the quality of data, analyses, and decisions in every domain, whether science, everyday life, or organizational behavior. Transparency,[151] validation and verification, and corrigibility are essential characteristics of good academic research, government policy-making, organizational processes like eligibility decision-making, and knowledge-based work such as engineering design and medical diagnosis. Workshop participants advocate for Big Data implications research that identifies barriers to, and enablers of, quality in data, algorithms, and decisions. They also call for research that develops or evaluates innovative social and technical ways to ensure data, algorithm, and decision quality.

Knowledge and skill

Big Data has the potential both to enhance human expertise and to replace it (whether through automation or through depreciation of formerly valued human capabilities). This dual potential can be seen in science,[152] in everyday life, and in the world of enterprises and work. How best to combine the capabilities of humans and machines has probably been an issue since humans first used tools. It is, however, a particularly important question today as automation encroaches ever faster into the most highly skilled and rewarded human activities. Workshop participants advocate for Big Data implications research that provides valid evidence to inform public choice about the appropriate and acceptable uses of automation.

Cross-Cutting Themes in Big Data Implications Research

1. ***Values and ethics***
   a. *Multiple stakeholders with different incentives and backgrounds affected by Big Data applications*
   b. *Differences in ends and means leading to value conflicts and ethical dilemmas*
   c. *Need for value-sensitive design: confronting value conflicts openly promotes innovation that benefits everybody*
2. ***Data, algorithm, and decision quality***
   a. *Big Data as bad data because of inaccuracies, misanalysis, misinterpretation, or inappropriate actions*
   b. *Importance of transparency, validation and verification, and corrigibility*
   c. *Need for research on barriers to and enablers of quality in data, algorithms and decisions*
3. ***Knowledge and skill***
   a. *Two roles of Big Data: enhancing and potentially replacing human expertise*
   b. *Need for research on understanding the best ways to integrate the capabilities of humans and computational tools and the appropriate and acceptable uses of automation*

Summary

This section outlined an agenda for research on Big Data's social, economic, ethical, and workforce implications.

Big Data is creating major changes in how science gets done, inside and outside university settings. New partnerships among universities, governments, and businesses need new governance models to generate sufficient resources and participation, to protect

personal data and intellectual property, and to ensure that they deliver public as well as private value. Research is needed to provide evidence on what works and what doesn't work.

Big Data is already seen as a significant and potentially transformative area of economic activity. But little is known about how the business of Big Data works. Research is needed to map the territory, document its business models and practices, and understand its contributions to the U.S. economy as a whole.

Big Data alters the resource requirements for research and the practices of researchers, portending major changes for universities, academics, and academic publishing. Research is needed to identify the challenges and emerging strategies for dealing with them successfully.

Big Data poses risks of harm to individuals from personal information privacy breaches, illegal discrimination, and illegitimate or unfair characterization of identity (i.e., algorithmic social sorting.) These risks are not adequately addressed by existing legal and technological protection mechanisms. The threat of harm has led to great public anxiety, growing distrust of governments, businesses, and employers, and the blockage of some potentially valuable innovations. Research is needed to assess the evolving threats and to develop new approaches for combating them, so that individuals will be willing to contribute their data for use toward socially desirable ends.

Big data is used by individuals in everyday life, not just by academics, governments, and businesses. Data amateurs are important because they can augment the ranks of professional data scientists, launch valuable social and business entrepreneurships, contribute personal data and insights to a variety of initiatives, and create harm through illegal, unethical, or socially unacceptable uses of data. Research is needed to understand Big Data in the realm of amateurs, not just professionals.

Big Data promises tremendous benefits—financial, health, social—to *all* individuals. Evidence about the nature and magnitude of these benefits could help allay unfounded fears about Big Data's downsides. Research is needed to understand whether and how people take advantage of Big Data's potential: whether, for example, they make better decisions when given access to their data about their health or consumption. Research is also needed to quantify actual harms and to provide evidence for designing better technologies, social practices, and procedures for due process and redress.

Big Data is threatened by a shortage of skilled people, including data scientists, technical support personnel, and managers who are willing and able to make sound investment, deployment, and use decisions about Big Data. There is little consensus about what knowledge and skills these professionals need, but current educational programs appear to underprepare their graduates, particularly in the social implications, the legal

issues, and ethical dilemmas raised by Big Data. Research is needed to identify knowledge and skill requirements and to develop the kinds of educational materials that will close "the ethics gap."

Big Data is largely an organizational activity, conducted in businesses, not-for-profits, and governments. Little is currently known about how organizations make decisions about Big Data investments, how Big Data work is organized, governed, and managed, and how Big Data changes what organizations do and the value they create. Research is needed on organizations' Big Data practices at the Board level, during data science projects and technical development, and after technologies and decisions are deployed.

Big Data changes work, not only of the employees who "do" Big Data, but also of many others. The insights from Big Data analyses are used to redesign and sometimes to automate organizational activities. New jobs may be created, but some old jobs are threatened. In addition, jobs may change in ways that reduce human knowledge and skill, discretion, and pay. Research is needed to provide evidence about Big Data-driven changes in work and about whether (and what kind of) interventions may be needed.

The Big Data phenomenon is usually analyzed by sector, e.g., health care, marketing, or financial services. This research agenda takes a different approach and is organized by the domains of science, everyday life, and work, because many issues (e.g., personal information privacy and deskilling of work) are common to several sectors. At the same time, the domains discussed here also share some common themes. This report highlighted three shared themes: values and ethics, quality (of data, algorithms, and decisions based on data and algorithms), and human knowledge and skill. Big Data implications research is also needed on these themes, regardless of sector or domain.

Guidelines for Big Data Implications Research

Before presenting a summary of our recommendations, we close this report with a brief discussion of the desirable characteristics of Big Data implications research.

As noted above, scientists have supported implications research for other potentially transformative innovation clusters such as genetic engineering and nanotechnology. Experience with those initiatives identified opportunities for improvement[153] that should be reflected in Big Data implications research. The primary purpose of implications research is to provide evidence that can inform public and business policy-makers. This can only happen when implications research is ***policy-aware***, that is, when it identifies or investigates questions with important implications for policy and practical action. One suggestion for increasing the policy relevance of implications research is to embed implications researchers in teams of scientists and engineers, as the teams confront design decisions, value choices, and ethical dilemmas.[154] Additional strategies include observing and interacting with users (and affected stakeholders who may not be users) while analyzing the sociotechnical features and affordances of Big Data initiatives. Close engagement[155] with the people who do, use, or are affected by Big Data is a hallmark of good Big Data implications research.[156] So are **value-sensitivity**[157] and **design-orientation**,[158] where "design" refers to both social practices and technical features.[159]

The core ideas underlying value-sensitivity and value sensitive design are particularly relevant for Big Data implications research. Value sensitive design "accounts for human values in a principled and comprehensive manner throughout the design process."[160] Conducting technology design processes in a way that truly accounts for human values encourages designers to consider the implications of the technology before the design is completed and to take steps to mitigate potential negative consequences.

Another desirable characteristic of Big Data implications research is ***transdisciplinarity***. Big Data's implications are many and varied. If Big Data proves to be like earlier transformative innovation clusters (such as the telephone and the automobile), it will have myriad consequences. In addition to intended benefits, Big Data is likely to have unintended benefits and harms. It may have "dual effects,"[161] that is, opposite effects occurring at nearly the same time and place. For instance, the telephone promoted both the building of skyscrapers and suburban sprawl.[162] Individual academic disciplines tend to emphasize particular kinds of consequences (e.g., economic or social outcomes but not both). The diversity of Big Data's potential consequences strongly suggests that transdisciplinary collaborations will generate more comprehensive analyses and deeper general understanding.

All research projects are necessarily bounded in scope for manageability. For instance, a manageable study of data-driven clinical decision-making in dermatology would

probably not also cover the marketing of treatments. But restrictions on the outcomes of interest within a bounded investigation are more a matter of disciplinary orientation than of pragmatic necessity. Within a transdisciplinary team, it would be feasible as well as desirable to explore varied conditions and outcomes. In the dermatology example, the team might investigate a suite of outcomes including costs, decision correctness, practitioner reactions, patient acceptability, and quality of care. It is also important that transdisciplinary teams publish integrative findings instead of reporting only those outcomes of interest to narrow disciplinary journals.

The outcomes of Big Data will continue to unfold and cannot be fully known today. But in other policy domains the techniques of technology foresight (e.g., scenario planning and the Delphi method) have proved useful for targeting investments, reducing risks, and increasing human resilience in the face of change.[163] These ***anticipatory*** techniques can be combined with systematic research on current and historical conditions and with stakeholder ***participation*** to provide helpful insights on Big Data policy-oriented questions.

An indispensable characteristic of Big Data implications research is a ***sociotechnical*** perspective. Many Big Data commentators have noted that technology appears to have outstripped society's ability to regulate it. At the same time, workshop participants observed that social conditions such as information-sharing practices appear to be changing even faster than technology. Big Data implications research needs to explore the interplay between technology and social practices, rather than focusing exclusively on either. This requires, again, collaboration between scholars with specialized expertise in a variety of areas with the specific intent of working towards integrated outcomes. In this process, academic disciplines with a specific focus on understanding the integration of technology and organizational practices—such as information systems,[164] information science, and social informatics—can play a particularly important role.

In short, Big Data promises to be a transformative innovation. It is only fitting for this innovation to demand transformative scientific methods—methods that bring together the domains of science, policy, and design; a focus on the social, the technical, and their interplay; and retrospective, current, and prospective investigative approaches.

Desired Characteristics of Big Data Implications Research

1. ***Policy-awareness***
   - *Research that identifies or investigates questions with important implications for policy and practical action*
2. ***Value-sensitivity***
   - *Research that pays close attention to and explicitly addresses the differing value preferences of multiple stakeholder groups*
3. ***Design-orientation***
   - *Research that leads to strategies for creating and evaluating alternative solutions for the problems and opportunities identified in the research*
4. ***Sociotechnical approach***
   - *Research that takes into account both social and technical factors and deepens our understanding of their interconnectedness*
5. ***Transdisciplinarity***
   - *Research that integrates multiple disciplinary perspectives to generate more comprehensive analyses and deeper general understanding*
6. ***Anticipatory stance***
   - *Research that incorporates techniques of technology foresight to imagine future possibilities.*
7. ***Participatory approach***
   - *Research that actively engages all relevant stakeholders in the research process*

Summary of Recommendations

In this final section of the report, we present three recommendations to agencies and foundations that already have or are considering programs of support for Big Data research. These recommendations are based on the assumption that Big Data is a potentially transformative innovation. Prior transformative innovations have exhibited both major benefits and negative consequences. Systematic research on the implications and consequences of Big Data will help maximize the socioeconomic benefits while mitigating negative consequences.

The discussions in the workshop summarized in this report show the breadth of Big Data's potential implications and consequences. Many similar conversations focus exclusively on individual privacy or data security, but, as discussed earlier, there are many other issues. Unlawful discrimination, extensive social sorting and labeling, errors propagated by automated decision-making, and changes in professional practices and human decision authority are just some of the areas that deserve systematic research.

Research on the implications and consequences of Big Data is a highly cost-effective way to increase the societal benefits of this potentially transformational innovation. Therefore, workshop participants make the following recommendations regarding Big Data implications research:

1. **Provide financial support for research on the implications and consequences of Big Data**, *in addition to* the support that is currently provided to research on Big Data applications and infrastructure. The level of support required to make a significant difference in Big Data implications research is modest compared to the resources needed for research on applications and infrastructure. We recommend that at least 3-5% of total Big Data research funding be allocated to implications research (following the practice established in the areas of genetic engineering and nanotechnology). 3-5% of President Obama's $200 million Big Data initiative would be $6-10 million, or about 12-20 small research awards, across multiple agencies.

2. **Focus research support on the three core domains of:**

    1) *Science and science–technology policy*, including data-oriented partnerships, the new data economy, and the implications of Big Data for universities, academic disciplines, and research publishing
    2) *Individuals and everyday life*, including privacy and privacy protection, the behavior of data amateurs, and Big Data's individual benefits and burdens, and
    3) *Organizations and work*, including the knowledge and skill needs of corporate data scientists, organizational practices related to the management of data and algorithms, and the impact of Big Data on jobs, knowledge, and expertise.

In particular**, recognize the importance of three cross-cutting themes and fund research that addresses them, regardless of domain:**

- Better approaches for addressing value conflicts and ethical dilemmas
- Barriers to and enablers of data, algorithm, and decision quality, and
- The tensions between human expertise and computer-based automation and their impact on human knowledge and skill.

3. Give preference to proposals with the following characteristics:

1) *Policy-awareness*. Given that one of the primary purposes of implications research is to provide evidence to inform policy-making, it is important for Big Data implications research to identify and address issues related to policies and practical actions.
2) *Value-sensitivity*. It is essential that Big Data implications and outcomes research pay close attention to, and explicitly address, the differing value preferences of multiple stakeholder groups.
3) *Design-orientation*. Effective implications research should lead to strategies for creating and evaluating alternative solutions for the problems and opportunities identified in the research.
4) *Transdisciplinarity*. Understanding and articulating the complex implications of Big Data requires the integration of multiple disciplinary perspectives.
5) *Sociotechnical thinking*. As demonstrated in this report, Big Data is not solely a technical phenomenon, and its implications cannot be analyzed without an approach that provides deep understanding of both social and technical factors and the ways they are related.
6) *Anticipatory focus*. Research on the implications of Big Data should not be based solely on an analysis of what has happened in the past. Instead, forward-looking techniques are needed to imagine future possibilities.
7) *Participatory approach*. The active engagement of stakeholders in the conduct of Big Data implications research is essential, because it fosters the other desirable characteristics of Big Data implications research.

Our hope is that the findings of the workshop on the social, economic, and workforce implications of Big Data and the recommendations based on them will contribute toward enabling a comprehensive, systematic, and integrative research program on the implications and consequences of Big Data.

Appendix A: Workshop Participants

Participants of the Workshop on Social, Economic and Workforce Implications of Big Data Analytics, January 30-31, 2014.

Mark Adams	Good Start Genetics
Ritu Agarwal	University of Maryland
Diane Bailey	University of Texas at Austin
Carol Ann Boughrum	Bentley University
Geoffrey Bowker	University of California, Irvine
danah boyd	Microsoft Research
Eric Clemons	University of Pennsylvania
Kevin Crowston	NSF
Mary J. Culnan	Future of Privacy Forum
Faisal D'Souza	National Coordination Office
Joseph Dery	EMC
Cheryl Eavey	NSF
Hamid Ekbia	Indiana University
Mike Fisher	AKF Partners
Kenneth R. Fleischmann	University of Texas at Austin
Bill Franks	Teradata
Arnie Greenland	IBM
Danny Goroff	Sloan Foundation
Frances Grodzinsky	Sacred Heart University
Suzanne Iacono	NSF
Daniel Katz	Michigan State University
John Leslie King	University of Michigan
Fred Ledley	Bentley University
Peter Lyster	NIGMS / NIH
Kalle Lyytinen	Case Western Reserve University
Stu Madnick	MIT
Ann Majchrzak	University of Southern California
Stephen Marcus	NIGMS / NIH
M. Lynne Markus	Bentley University
Connie L. McNeely	George Mason University
Kevin Mentzer	Bentley University
Sue Newell	Bentley University
Wendy Nilsen	NIH
Paul Ohm	University of Colorado
Theresa Pardo	University of Albany
Jules Polonetsky	Future of Privacy Forum
David Ribes	Georgetown University
Christine Ries	Georgia Tech
Joshua L. Rosenbloom	NSF

Laurie Schintler	George Mason University
Katie Shilton	University of Maryland
George Strawn	National Coordination Office
Heikki Topi	Bentley University
Anne L. Washington	George Mason University
Wendy Wigen	National Coordination Office
Susan Winter	University of Maryland
Heng Xu	NSF
Fen Zhao	NSF

The authors also acknowledge with gratitude Wenxiu (Vince) Nan's highly valuable technical contributions during the document production process.

Endnotes

[1] Kalil, Tom. "Big Data is a Big Deal." The White House Briefing Room (2012). Blog. <https://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal>. Last accessed September 6, 2015.
[2] *Demystifying Big Data: A Practical Guide to Transforming the Business of Government*. Washington, DC. Foundation TechAmerica. <http://www-304.ibm.com/industries/publicsector/fileserve?contentid=239170>. Last accessed September 13, 2015; Schroeck, Michael, et al. *Analytics: The Real-World Use of Big Data*. Somers, NY. IBM Institute for Business Value (2013). <https://www.ibm.com/smarterplanet/global/files/se__sv_se__intelligence__Analytics_-_The_real-world_use_of_big_data.pdf>. Last accessed September 13, 2015.
[3] Zhou, Xujuan, et al. "The State-of-the-Art in Personalized Recommender Systems for Social Networking." *Artificial Intelligence Review* 37.2 (2012): 119-32.
[4] Romei, Andra, and Salvatore Ruggieri. "Discrimination Data Analysis: A Multi-Disciplinary Bibliography." *The Knowledge Engineering Review* 29.5 (2014): 582-638.
[5] *Big Data: Seizing Opportunities, Preserving Values*. Washington DC. Executive Office of the President (2014). <http://www.whitehouse.gov/issues/technology/big-data-review>. Last accessed May 26, 2014; *Report to the President, Big Data and Privacy: A Technological Perspective*. Washington, DC. Executive Office of the President (2014). <http://www.whitehouse.gov/issues/technology/big-data-review>. Last accessed May 26, 2014.
[6] Martin, Kirsten E. "Ethical Issues in the Big Data Industry." *MIS Quarterly Executive* 14.2 (2015): 68-85; Washington, Anne L. *Can Big Data Be Described as a Supply Chain?* (2014). <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2577989>. Last accessed August 24, 2015.
[7] Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: The MIT Press (2000).
[8] Chui, Michael, et al. *Ten IT-Enabled Business Trends for the Decade Ahead*. McKinsey Global Institute (2013). <http://www.mckinsey.com/~/media/mckinsey/dotcom/insights/high%20tech%20telecoms%20internet/ten%20it-enabled%20business%20trends%20for%20the%20decade%20ahead/mgi_it_enabled_trends_report_may%202013_v2.ashx>. Last accessed September 13, 2015.
[9] Hartzog, Woodow. "Chain-Link Confidentiality." *Georgia Law Review* 46 (2012): 657-75.
[10] Arthur, W. Brian. "The Second Economy." *McKinsey Quarterly*. October (2011): 90-99; Manyika, James, et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute (2011). <http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation>. Last accessed September 13, 2015; Carr, Nicholas. *The Glass Cage: Automation and Us*. New York, NY: W.W. Norton & Company (2014).
[11] Carr, Nicholas. *The Glass Cage: Automation and Us*. New York, NY: W.W. Norton & Company (2014).
[12] Irwin, Neil. "Why Ben Bernanke Can't Refinance his Mortgage." The Upshot (2014). Blog. <http://www.nytimes.com/2014/10/03/upshot/why-ben-bernanke-cant-refinance-his-mortgage.html?_r=0>. Last accessed September 13, 2015.
[13] Ribes, David, et al. "Artifacts that Organize: Delegation in the Distributed Organization." *Information and Organization* 23 (2013): 1-14.

[14] Carr, Nicholas. *The Glass Cage: Automation and Us*. New York, NY: W.W. Norton & Company (2014).
[15] Nof, Shimon Y., ed. *Springer Handbook of Automation*. Berlin, Germany: Springer (2009).
[16] Stewart, Christopher S., and Merissa Marr. "Inside the Effort to Kill a Web Fraud "Botnet"." *The Wall Street Journal.* December 5, 2013.
[17] Leonardi, Paul M. *Car Crashes without Cars: Lessons about Simulation Technology and Organizational Change from Automotive Design*. Cambridge, MA: The MIT Press (2012).
[18] Markus, M. Lynne, et al. "The Computerization Movement in the US Home Mortgage Industry: Automated Underwriting from 1980 to 2004." *Computerization Movements and Technology Diffusion: From Mainframes to Ubiquitous Computing*. Eds. Kraemer, Kenneth L. and Margaret S. Elliott. Medford, NY: Information Today (2008): 115-44.
[19] Batty, Mike, and Alice Kroll. *Automated Life Underwriting: A Survey of Life Insurance Utilization of Automated Underwriting Systems*. Deloitte. Society of Actuaries (2009). <https://www.soa.org/Files/Research/Projects/research-life-auto-underwriting.pdf>. Last accessed September 13, 2015.
[20] Katz, Daniel M. "Quantitative Legal Prediction—Or—How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry." *Emory Law Journal* 62 (2013): 909-66.
[21] *Foresight: The Future of Computer Trading in Financial Markets, Final Project Report*. London, UK. The Government Office for Science (2012). <http://www.cftc.gov/ucm/groups/public/@aboutcftc/documents/file/tacfuturecomputertrading1012.pdf>. Last accessed September 13, 2015.
[22] Ungerleider, Neal. "IBM's Watson Is Ready To See You Now-In Your Dermatologist's Office." *Fast Company* (2014). <http://www.fastcompany.com/3030723/ibms-watson-is-ready-to-see-you-now-in-your-dermatologists-office>. Last accessed May 26, 2014.
[23] Ribes, David, et al. "Artifacts that Organize: Delegation in the Distributed Organization." *Information and Organization* 23 (2013): 1-14.
[24] Manyika, James, et al. *Open Data: Unlocking Innovation and Performance with Liquid Information*. McKinsey & Company (2013). <http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information>. Last accessed September 13, 2015.
[25] Ekbia, Hamid, et al. "Big Data, Bigger Dilemmas: A Critical Review." *Advances in Information Science* 66.8 (2015): 1523–45.
[26] boyd, danah, and Kate Crawford. "Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon." *Information, Communication & Society* 15.5 (2012): 662-79; Newell, Sue. "Managing Knowledge and Managing Knowledge Work: What We Know and What the Future Holds." *Journal of Information Technology* 30.1 (2015): 1-17; Manyika, James, et al. *Help Wanted: The Future of Work in Advanced Economies*. McKinsey Global Institute (2012). <http://www.mckinsey.com/insights/employment_and_growth/future_of_work_in_advanced_economies>. Last accessed September 13, 2015.
[27] Morozov, Evgeny. "The Real Privacy Problem." *MIT Technology Review* 116.6 (2013): 32-43.
[28] Introna, Lucas D., and David Wood. "Picturing Algorithmic Surveillane: The Politics of Facial Recognition Systems." *Surveillance & Society* 2.2/3 (2004): 177-98.
[29] Culnan, Mary J., and Cynthia Clark Williams. "How Ethics Can Enhance Organizational Privacy: Lessons from the ChoicePoint and TJX Data Breaches." *MIS Quarterly* 33.4 (2009):

673-87; Davis, Kord. *Ethics of Big Data: Balancing Risk and Innovation*. Sebastopol, CA: O'Reilly Media (2012); Grodzinsky, Frances S., Keith W. Miller, and Marty J. Wolf. "The Ethics of Designing Artificial Agents." *Ethics and Information Technology* 10 (2008): 115-21.
[30] MacCrory, Frank, et al. "Racing With and Against the Machine: Changes in Occupational Skill Composition in an Era of Rapid Technological Advance." *Proceedings of the 35th International Conference on Information Systems.* Auckland, NZ. (2014): 1-17.
[31] Hoofnagle, Chris Jay. *How the Fair Credit Reporting Act Regulates Big Data*. Future of Privacy Forum Workshop on Big Data and Privacy: Making Ends Meet (2013); Crawford, Kate, and Jason Schultz. "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms." *Boston College Law Review* 55.1 (2014): 93-128; *Unlocking the Value of Personal Data: From Collection to Usage*. World Economic Forum (2013). <http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf>. Last accessed September 13, 2015.
[32] Mayer-Schonberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York, NY: Harcourt Mifflin Harcourt Publishing Company (2013); McAfee, Andrew, and Erik Brynjolfsson. "Big Data: The Management Revolution." *Harvard Business Review* 90.10 (2012): 60-68.
[33] Ambrose, Meg Leta. "From the Avalanche of Numbers to Big Data: A Comparative Historical Perspective on Data Protection in Transition." *Digital Enlightment Yearbook 2014: Social Networks and Social Machines, Surveillance and Empowerment*. Eds. O'Hara, K., M-H.C. Nguyen and P. Haynes. Amsterdam, Netherlands: IOS Press (2014): 27-48.
[34] Arthur, W. Brian. *The Nature of Technology: What It Is and How It Evolves*. New York, NY: Simon and Schuster (2009); Geels, Frank W., and Johan Schot. "Typology of Sociotechnical Transition Pathways." *Research Policy* 36 (2007): 399-417.
[35] Latour, Bruno. *ARAMIS or the Love of Technology*. Cambridge, MA: Harvard University Press (1996).
[36] Arthur, W. Brian. *The Nature of Technology: What It Is and How It Evolves*. New York, NY: Simon and Schuster (2009).
[37] Tenner, Edward. *Why Things Bite Back: Technology and the Revenge of Unintended Consequences*. New York, NY: Vintage (1997); Winston, Lord Robert. *Bad Ideas?: An Arresting History of our Inventions: How our Finest Inventions Nearly Finished Us Off*. New York, NY: Bantam Press (2010); Harrison, Michael I., and Ross Koppel. "Interactive Sociotechnical Analysis: Identifying and Coping with Unintended Consequences of IT Implementation." *Handbook of Research on Advances in Health Informatics and Electronic Healthcare Applications: Global Adoption and Impact of Information Communication Technologies*. Eds. Khoumbati, Khalil, et al. Hershey, PA: IGI Global (2010): 33-51; Pool, Ithiel de Sola. *Forecasting the Telephone: A Retrospective Technology Assessment of the Telephone*. Norwood, NJ: Ablex (1983).
[38] Arthur, W. Brian. *The Nature of Technology: What It Is and How It Evolves*. New York, NY: Simon and Schuster (2009); Geels, Frank W., and Johan Schot. "Typology of Sociotechnical Transition Pathways." *Research Policy* 36 (2007): 399-417.
[39] Brynjolffson, Erik. "The Productivity Paradox of Information Techology." *Communications of the ACM* 36.12 (1993): 66-77
[40] Rosenberg, Nathan. *Exploring the Black Box: Technology, Economics, and History*. Cambridge, UK: Cambridge University Press (1994).

[41] Manyika, James, et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute (2011). <http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation>. Last accessed September 13, 2015.
[42] Dwoskin, Elizabeth. "The Joys and the Hype of Software Called Hadoop." *The Wall Street Journal.* December 17, 2014.
[43] Redman, Thomas C. "Data's Credibility Problem." *Harvard Business Review* 91.12 (2013): 84-88.
[44] Court, David. "Getting Big Impact from Big Data." *McKinsey Quarterly.* January (2015). <http://www.mckinsey.com/insights/business_technology/getting_big_impact_from_big_data?cid=other-eml-alt-mkq-mck-oth-1501>. Last accessed September 13, 2015.
[45] Eqbuono-Davis, Lisa, and Tanisha Carino. "System Transformation Intensifies Demand for Evidence: The Trust Conundrum." Health Affairs Blog (2013). Blog. <http://healthaffairs.org/blog/2013/09/10/system-transformation-intensifies-demand-for-evidence-the-trust-conundrum/>. Last accessed September 13, 2015.
[46] Polonetsky, Jules, and Omer Tene. "Privacy and Big Data: Making Ends Meet." *Stanford Law Review Online* 66 (2013): 25-33.
[47] Kalil, Tom. "Big Data is a Big Deal." The White House Briefing Room (2012). Blog. <https://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal>. Last accessed September 6, 2015.
[48] Press, Gil. "6 Predictions for the $125 Billion Big Data Analytics Market in 2015." *Forbes* (2014). <http://www.forbes.com/sites/gilpress/2014/12/11/6-predictions-for-the-125-billion-big-data-analytics-market-in-2015/>. Last accessed September 4, 2015.
[49] Kesmodel, David. "Monsanto to Buy Climate Corp. in Data-Science Push." *Wall Street Journal.* October 3, 2013.
[50] Monga, Vipal. "The Big Mystery: What's Big Data Really Worth?" *The Wall Street Journal.* October 12, 2014.
[51] Dahiyat, Emad Abdel Rahim. "Intelligent Agents and Liability: Is It a Doctrinal Problen or Merely a Problem of Explanation?" *Artificial Intelligence and Law* 18.1 (2010): 103-21; Dahiyat, Emad Abdel Rahim. "Intelligent Agents and Contracts: Is a Conceptual Rethink Imperative." *Artificial Intelligence and Law* 15 (2007): 275-390; Stahl, Bernd Carsten. "Responsible Computers? A Case for Ascribing Quasi-Responsibility to Computers Independent of Personhod or Agency." *Ethics and Information Technology* 8 (2006): 205-13.
[52] Dwork, Cynthia. "Differential Privacy." *Automata, Languages and Programming*. Eds. Bugliesi, Michele, et al. Vol. 4052. Lecture Notes in Computer Science. Berlin, Germany: Springer (2006): 1-12.
[53] Fisher, Erik. "Lessons Learned from the Ethical, Legal and Social Implications Program (ELSI): Planning Societal Implications Research for the National Nanotechnology Program." *Technology in Society* 27 (2005): 321-28.
[54] Bastow, Simon, Jane Tinkler, and Patrick Dunleavy. *The Impact of the Social Sciences: How Academics and Their Research Make a Difference*. Thousand Oaks, CA: Sage Publications (2014).
[55] Gillespie, Tarleton. "The Politics of Platforms." *New Media & Society* 12.3 (2010): 347-64.
[56] Miller, Ben, and Robert D. Atkinson. *Are Robots Taking Our Jobs, or Making Them?* The Information Technology & Innovation Foundation (2013). <http://www2.itif.org/2013-are-robots-taking-jobs.pdf>. Last accessed September 13, 2015.

[57] Morozov, Evgeny. *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York, NY: PublicAffairs (2013).
[58] Ridley, Matt. "The Scarcity Fallacy." *The Wall Street Journal.* April 26-27, 2014.
[59] Hemerly, Jess. "Public Policy Considerations for Data-Driven Innovation." *Computer*. June (2013): 25-31.
[60] Kharif, Olga. "Privacy Fears Over Student Data Tracking Lead to InBloom's Shutdown." *Bloomberg Business* (2014). <http://www.bloomberg.com/bw/articles/2014-05-01/inbloom-shuts-down-amid-privacy-fears-over-student-data-tracking>. Last accessed September 6, 2015.
[61] MacMillan, Douglas, and Lisa Fleisher. "How Sharp-Elbowed Uber Is Trying to Make Nice." *The Wall Street Journal.* January 29, 2015.
[62] Davis, Kord. *Ethics of Big Data: Balancing Risk and Innovation*. Sebastopol, CA: O'Reilly Media (2012); Biola, Holly, et al. "With Big Data Comes Big Responsibility." *Harvard Business Review* 92.11 (2014): 101-04; Buecher, Jeff. "Artificial Moral Agents: Saviors or Destroyers?" *Ethics of Information Technology* 12 (2010): 363-70.
[63] Sommerville, Ian, et al. "Large-Scale Complex IT Systems." *Communications of the ACM* 55.7 (2012): 71-77.
[64] Fritz, Ben, and Rachel Emma Silverman. "Data Breach Sets Off Upheaval at Sony." *The Wall Street Journal.* December 4, 2014; Matthews, Anna Wilde, and Danny Yadron. "Heath Insurer Anthem Hit by Hackers." *The Wall Street Journal.* Feburary 4, 2015.
[65] Ohm, Paul. "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization." *UCLA Law Review* 57 (2010): 1701-77.
[66] Brynjolfsson, Erik, and Lorin M. Hitt. "Beyond Computation: Information Technology, Organizational Transformation and Business Performance." *The Journal of Economic Perspectives* 14.4 (2000): 23-48.
[67] Clarke, Lee. *Worst Cases: Terror and Catastrophe in the Popular Imagination*. Chicago, Illinois: University of Chicago Press (2006); DeLeo, Rob A. "Anticipatory–Conjectural Policy Problems: A Case Study of Avian Influenza." *Risks, Hazards & Crisis in Public Policy* 1.1 (2010): 147-84.
[68] Grodzinsky, FS, K Miller, and MJ Wolf. "Moral Responsibility for Computing Artifacts: "The Rules" and Issues of Trust." *SIGCAS Computers and Society* 42.2 (2012): 15-25; Grodzinsky, Frances S., Keith W. Miller, and Marty J. Wolf. "The Ethics of Designing Artificial Agents." *Ethics and Information Technology* 10 (2008): 115-21; Johnson, Deborah G, and John M. Mulvey. "Accountability and Computer Decision Systems." *Communications of the ACM* 38.12 (1995): 58-64. See also the ACM *Code of Ethics*. Web. <http://www.acm.org/about/code-of-ethics>. Last accessed February 18, 2015.
[69] Johnson, Deborah G, and John M. Mulvey. "Accountability and Computer Decision Systems." *Communications of the ACM* 38.12 (1995): 58-64.
[70] Friedman, Batya, Peter H. Kahn, and Alan Borning. "Value Sensitive Design and Information Systems." *Human-Computer Interaction and Management Information Systems: Applications*. Eds. Galletta, Dennis and Ping Zhang. Armonk, NY: ME Sharpe Inc. (2006): 348-72; Owen, Richrd, John Bessant, and Maggy Heintz. *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*. Hoboken, NJ: Wiley (2013); van den Hoven, Jeroen. "Value Sensitive Design and Responsible Innovation." *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*. Eds. Owen, Richard, John Bessant and Maggy Heintz. Hoboken, NJ: Wiley (2013): 75-83; Nissenbaum,

Helen. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Redwood, CA: Stanford University Press (2009).
[71] Shilton, Katie. "Value Levers: Building Ethics into Design." *Science, Technology, and Human Values* 38.3 (2012): 374-97.
[72] Crosby, Philip B. *Quality Is Free: The Art of Making Quality Certain: How to Manage Quality - So That It Becomes A Source of Profit for Your Business*. New York, NY: McGraw-Hill Companies (1979).
[73] Culnan, Mary J., and Cynthia Clark Williams. "How Ethics Can Enhance Organizational Privacy: Lessons from the ChoicePoint and TJX Data Breaches." *MIS Quarterly* 33.4 (2009): 673-87.
[74] Weiss, Peter. "Borders in Cyberspace: Conflicting Government Information Policies and Their Economic Impacts." *Open Access and the Public Domain in Digital Data and Information for Science: Proceedings of an International Symposium*. Eds. Esanu, Julie M. and Paul F. Uhli. Washington, DC: The National Academies Press (2004): 592-608; Helbig, Natalie, et al. *The Dynamics of Opening Government Data*. Albany, NY: The Center for Technology in Governement (2012).
[75] *Demystifying Big Data: A Practical Guide to Transforming the Business of Government*. Washington, DC. Foundation TechAmerica. <http://www-304.ibm.com/industries/publicsector/fileserve?contentid=239170>. Last accessed September 13, 2015.
[76] "NSF Advances National Efforts Enabling Data-Driven Discovery." National Science Foundation (2013). Web. <http://www.nsf.gov/news/news_summ.jsp?cntn_id=129244>. Last accessed February 18, 2015.
[77] Sharda, Ramesh, Daniel Adomako Asamoah, and Natraj Ponna. "Business analytics: Research and teaching perspectives." *Proceedings of the 35th International Conference on Information Technology Interfaces (ITI)*. Dubrovnik, Croatia. (2013): 19-27.
[78] Edwards, Paul N., et al. *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Ann Arbor, MI: Deep Blue (2013).
[79] Anderson, Chris. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired Magazine* (2008). <http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory/>. Last accessed September 13, 2015; Mayer-Schonberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York, NY: Harcourt Mifflin Harcourt Publishing Company (2013); Kell, Douglas B., and Stephen G. Oliver. "Here Is the Evidence, Now What Is the Hypothesis? The Complementary Roles of Inductive and Hypothesis-Driven Science in the Post-Genomic Era." *BioEssays* 26.1 (2003): 99-105; Hey, Anthony J. G., D. Stewart W. Tansley, and Kristin M. Tolle, eds. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA: Microsoft Research (2009).
[80] Grimmelmann, James. "The Law and Ethics of Experiments on Social Media Users." *Colorado Technology Law Journal* 13 (2015): 219-72.
[81] Wallis, Jillian C., Elizabeth Rolando, and Christine L. Borgman. "If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology." *PLOS ONE* 8.7 (2013): e67332.
[82] Borgman, Christine L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: The MIT Press (2015).
[83] "Scientific Research: Looks Good on Paper." *The Economist* (2013). <http://www.economist.

com/news/china/21586845-flawed-system-judging-research-leading-academic-fraud-looks-good-paper>. Last accessed September 13, 2015; Dutton, William H., and Paul W. Jeffreys. *World Wide Research: Reshaping the Sciences and Humanities*. Cambridge, MA: The MIT Press (2010).

[84] Crotty, David. "PLOS' Bold Data Policy." The Scholarly Kitchen (2014). Blog. <http://scholarlykitchen.sspnet.org/2014/03/04/plos-bold-data-policy/>. Last accessed September 13, 2015.

[85] Garfinkel, Simson. *Database Nation: The Death of Privacy in the 21st Century*. Sebastopol, CA: O'Reilly Media (2000).

[86] Ohm, Paul. "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization." *UCLA Law Review* 57 (2010): 1701-77.

[87] Rubinstein, Ira S. *Big Data: A Pretty Good Privacy Solution*. Future of Privacy Forum (2013). <http://www.futureofprivacy.org/wp-content/uploads/TECH-Rubinstein-Big-Data-A-Pretty-Good-Privacy-Solution.pdf>. Last accessed February 18, 2015

[88] Walker, Joseph. "Data Mining To Recruit Sick People." *Wall Street Journal.* December 17, 2013.

[89] Clemons, Eric K., and Josh Wilson. *Students' and Parents' Attitudes toward Online Privacy: An International Study*. Philadephia, PA. The Wharton School (2015).

[90] Bruening, Paula L., and Mary J. Culnan. "Through a Glass Darkly: From Privacy Notices to Effective Transparency." *North Carolina Journal of Law and Technology* (Forthcoming, 2015).

[91] Ohm, Paul. "The Fourth Amendment in a World Without Privacy." *Mississippi Law Journal* 81.5 (2013): 1309-55; Rifkin, Jeremy. *The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism*. London, UK: Palgrave Macmillan (2014).

[92] Rodwin, Marc A. "The Case for Public Ownership of Patient Data." *The Journal of the American Medical Association* 302.1 (2009): 86-88.

[93] http://www.nist.gov/nstic/

[94] Heffetz, Ori, and Katrina Ligett. "Privacy and Data-Based Research." *Journal of Economic Perspectives* 28.2 (2014): 75-98.

[95] Morris, Betsy. "Truckers Tire of Government Sleep Rules." *The Wall Street Journal.* November 14, 2013.

[96] Banjo, Shelly, and Sara Germand. "The End of the Impulse Shopper." *The Wall Street Journal.* November 25, 2014.

[97] Gelfand, Alexander. "Privacy and Biomedical Research: Building a Trust Infrastructures; An Exploration of Data-Driven and Process-Driven Approaches to Data Privacy." *Biomedical Computation Review*. January (2012): 22-28.

[98] Seltsikas, Philip, and Robert M. O'Keefe. "Expectations and Outcomes in Electronic Identity Management: The Role of Trust and Public Value." *European Journal of Information Systems* 19.1 (2010): 93-103.

[99] Hirsch, Dennis. *The Glass House Effect: Why Big Data is the New Oil, and What To Do About It*. Future of Privacy Forum. <http://www.futureofprivacy.org/wp-content/uploads/Hirsch-Glass-House-Effect1.pdf>. Last accessed September 13, 2015.

[100] Marcus, Amy Dockser, and Christopher Weaver. "Heart Gadgets Test Limits of Privacy Laws on Health." *Wall Street Journal.* November 29, 2012; MacMillan, Douglas. "Fight Over Flickr's Use of Photos." *The Wall Street Journal.* November 25, 2014.

[101] Hoofnagle, Chris Jay. *How the Fair Credit Reporting Act Regulates Big Data*. Future of Privacy Forum Workshop on Big Data and Privacy: Making Ends Meet (2013).
[102] Pewen, William F. "Protecting our Civil Rights in the Era of Digital Health." *The Atlantic* (2012). <http://www.theatlantic.com/health/archive/2012/08/protecting-our-civil-rights-in-the-era-of-digital-health/260343/>. Last accessed February 18, 2015; Strahilevitz, Lior Jacob. "Toward a Positive Theory of Privacy Law." *Harvard Law Review* 126 (2010): 2010-42.
[103] Pariser, E *The Filter Bubble: What the Internet Is Hiding from you*. New York, NY: The Penguin Press (2011).
[104] Mikians, Jakub, et al. "Detecting Price and Search Discrimination on the Internet." *HotNets-XI Proceedings of the 11th ACM Workshop on Hot Topics in Networks*. Redmond, WA. (2012): 79-84.
[105] Craig, Elizabeth, Charlene Hou, and Brian F. McCarthy. *The Looming Global Analytics Talent Mismatch in Insurance*. Accenture (2012). <http://nstore.accenture.com/IM/FinancialServices/AccentureLibrary/data/pdf/looming-global-analytics-talent-mismatch-in-insurance.pdf>. Last accessed September 13, 2015; Craig, Elizabeth, Charlene Hou, and Brian F. McCarthy. *The Looming Global Analytics Talent Mismatch in Banking*. Accenture (2013). <https://www.accenture.com/us-en/~/media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Technology_6/Accenture-The-Looming-Global-Analytics-Talent-Mismatch-in-Banking.pdf>. Last accessed September 13, 2015.
[106] Redman, Thomas C. "Data's Credibility Problem." *Harvard Business Review* 91.12 (2013): 84-88.
[107] Fleischmann, Kenneth R., and William A. Wallace. "Value Conflicts in Computational Modeling." *Computer* 43.7 (2010): 57-63.
[108] Friedman, Batya, Peter H. Kahn, and Alan Borning. "Value Sensitive Design and Information Systems." *Human-Computer Interaction and Management Information Systems: Applications*. Eds. Galletta, Dennis and Ping Zhang. Armonk, NY: ME Sharpe Inc. (2006): 348-72; van den Hoven, Jeroen. "Value Sensitive Design and Responsible Innovation." *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*. Eds. Owen, Richard, John Bessant and Maggy Heintz. Hoboken, NJ: Wiley (2013): 75-83.
[109] Fleischmann, Kenneth R., William A. Wallace, and Justin M. Grimes. "How Values Can Reduce Conflicts in the Design Process: Results from a Multi-Site Mixed-Method Field Study." *Proceedings of the Annual Meeting of the American Society for Information Science and Technology*. New Orleans, LA. (2011): 1-10.
[110] Friedman, Batya, Peter H. Kahn, and Alan Borning. "Value Sensitive Design and Information Systems." *Human-Computer Interaction and Management Information Systems: Applications*. Eds. Galletta, Dennis and Ping Zhang. Armonk, NY: ME Sharpe Inc. (2006): 348-72.
[111] McGinn, Robert. *The Ethically Responsible Engineer*. Hoboken, NJ: Wiley-IEEE Press (2015).
[112] Pearson, Travis, and Rasmus Wegener. *Big Data: The Organizational Challenge*. Bain & Company (2013). <http://www.bain.com/images/bain_brief_big_data_the_organizational_challenge.pdf>. Last accessed August 23, 2015.
[113] *The Field Guide to Data Science*. Booz Allen Hamilton (2013). <http://www.boozallen.com/insights/2013/11/data-science-field-guide>. Last accessed September 13, 2015.
[114] Gentile, Mary C. *Giving Voice to Values*. New Haven, CT: Yale University Press (2010).

[115] Court, David. "Getting Big Impact from Big Data." *McKinsey Quarterly.* January (2015). <http://www.mckinsey.com/insights/business_technology/getting_big_impact_from_big_data?cid=other-eml-alt-mkq-mck-oth-1501>. Last accessed September 13, 2015.
[116] Davenport, Thomas H. "Keep Up with Your Quants." *Harvard Business Review* 91.7/8 (2013): 120-23.
[117] https://www.informs.org/
[118] Code of Ethics/Conduct. Web. <https://www.informs.org/Sites/Certified-Analytics-Professional-Program/Applicants/CODE-OF-ETHICS>. Last accessed August 24, 2015.
[119] The Data Science Code of Conduct. Web. <http://www.datascienceassn.org/code-of-conduct.html>. Last accessed August 24, 2015.
[120] http://www.acm.org
[121] Gentile, Mary C. *Giving Voice to Values*. New Haven, CT: Yale University Press (2010).
[122] Calo, Ryan. "Consumer Subject Review Boards: A Thought Experiment." *Stanford Law Review Online* 66 (2013): 97-102.
[123] Davis, Kord. *Ethics of Big Data: Balancing Risk and Innovation*. Sebastopol, CA: O'Reilly Media (2012).
[124] Bailey, Diane, Paul M. Leonardi, and S.R. Barley. "The Lure of the Virtual." *Organization Science* 23.5 (2012): 1485-504.
[125] Helbing, Dirk, and Stefano Balietti. "From Social Data Mining to Forecasting Socioeconomic Crisis." *European Physical Journal Special Topics* 195 (2011): 3-68; Reed, Albergotti. "Furor Erupts Over Facebook Study." *The Wall Street Journal.* June 30, 2014.
[126] Helbig, Natalie, et al. *The Dynamics of Opening Government Data*. Albany, NY: The Center for Technology in Governement (2012).
[127] Staff, Harvard Business Review. "With Big Data Comes Big Responsibility: An Interview with Sandy Pentland." *Harvard Business Review* 92.11 (2014): 100-04.
[128] Newell, Sue. "Managing Knowledge and Managing Knowledge Work: What We Know and What the Future Holds." *Journal of Information Technology* 30.1 (2015): 1-17.
[129] Carr, Nicholas. *The Glass Cage: Automation and Us*. New York, NY: W.W. Norton & Company (2014).
[130] Levy, Frank, and Richard J. Murnane. *The New Division of Labor: How Computers Are Creating the Next Job Market*. Princeton, NJ: Princeton University Press (2005); Autor, David H., and David Dorn. "How Technology Wrecks the Middle Class." *The New York Times.* August 24, 2013.
[131] Head, Simon. *Mindless: Why Smarter Machines Are Making Dumber Humans*. New York, NY: Basic Books (2014).
[132] Brynjolffson, Erik, and Andrew McAfee. *Race Against the Machine*. Digital Frontier Press (2011); Brynjolffson, Erik, and Andrew McAfee. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York, NY: W.W. Norton & Company (2014).
[133] Ford, Martin. *The Lights in the Tunnel: Automation, Accelerating Technology and the Economy of the Future*. CreateSpace Independent Publishing Platform (2009); "The Onrushing Wave." *The Economist* (2014). <http://www.economist.com/news/briefing/21594264-previous-technological-innovation-has-always-delivered-more-long-run-employment-not-less>. Last accessed September 13, 2015.

[134] Frey, Carl Benedict, and Michael A. Osborne. *The Future of Employment: How Susceptible are Jobs to Computerization?* Oxford, UK. Oxford Martin Programme on the Impacts of Future Technology (2013). <http://www.futuretech.ox.ac.uk/future-employment-how-susceptible-are-jobs-computerisation-oms-working-paper-dr-carl-benedikt-frey-m>. Last accessed September 13, 2015.
[135] MacCrory, Frank, et al. "Racing With and Against the Machine: Changes in Occupational Skill Composition in an Era of Rapid Technological Advance." P*roceedings of the 35th International Conference on Information Systems*. Auckland, NZ. (2014): 1-17.
[136] Elliott, Stuart W. "Anticipating a Luddite Revival." *Issues in Science and Technology* 30.3 (2014): 27-36.
[137]Davenport, Thomas H. "The Confusing Landscape of 'Cognitive Computing'." CIO Journal (2014). Blog. <http://blogs.wsj.com/cio/2014/12/17/the-confusing-landscape-of-cognitive-computing/>. Last accessed September 13, 2015.
[138] Markus, M. Lynne, et al. "The Computerization Movement in the US Home Mortgage Industry: Automated Underwriting from 1980 to 2004." *Computerization Movements and Technology Diffusion: From Mainframes to Ubiquitous Computing*. Eds. Kraemer, Kenneth L. and Margaret S. Elliott. Medford, NY: Information Today (2008): 115-44.
[139] MacCrory, Frank, et al. "Racing With and Against the Machine: Changes in Occupational Skill Composition in an Era of Rapid Technological Advance." *Proceedings of the 35th International Conference on Information Systems*. Auckland, NZ. (2014): 1-17; Carr, Nicholas. *The Glass Cage: Automation and Us*. New York, NY: W.W. Norton & Company (2014).
[140] Bainbridge, Lisanne. "Ironies of Automation." *Automatica* 19.6 (1983): 775-79.
[141] Davenport, Thomas H., and Jeanne G. Harris. "Automated Decision Making Comes of Age." *MIT Sloan Management Review* 46.4 (2005): 83-89.
[142] *Foresight: The Future of Computer Trading in Financial Markets, Final Project Report*. London, UK. The Government Office for Science (2012). <http://www.cftc.gov/ucm/groups/public/@aboutcftc/documents/file/tacfuturecomputertrading1012.pdf>. Last accessed September 13, 2015.
[143] Kantor, Jodi, and David Streitfeld. "Inside Amazon: Wrestling Big Ideas in a Bruising Workplace." *The New York Times.* August 15, 2015.
[144] *Executive Perspectives on Top Risks for 2015: Key Issues Being Discussed in the Boardroom and C-Suite*. North Carolina State University's Enterprise Risk Management Initiative and protiviti Risk & Business Consulting and Internal Audit (2015). <http://www.protiviti.com/en-US/Documents/Surveys/NC-State-Protiviti-Survey-Top-Risks-2015.pdf>. Last accessed August 24, 2015.
[145] Son, Hugh. "JPMorgan Algorithm Knows You're a Rogue Employee Before You Do." *BloombergBusiness* (2015). <http://www.bloomberg.com/news/articles/2015-04-08/jpmorgan-algorithm-knows-you-re-a-rogue-employee-before-you-do>. Last accessed September 7, 2015.
[146] Ribes, David, et al. "Artifacts that Organize: Delegation in the Distributed Organization." *Information and Organization* 23 (2013): 1-14.
[147] Fleischmann, Kenneth R. *Information and Human Values*. San Rafael, CA: Morgan & Claypool (2014).
[148] Friedman, Batya, and David G. Hendry. "The Envisioning Cards: A Toolkit for Catalyzing Humanistic and Technical Imaginations." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Austin, TX. (2012): 1145-48.

[149] boyd, danah, and Kate Crawford. "Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon." *Information, Communication & Society* 15.5 (2012): 662-79.
[150] Ross, Jeanne W., Cynthia M. Beath, and Anne Quaadgras. "You May Not Need Big Data After All." *Harvard Business Review* 91.12 (2013): 90-98.
[151] Zarsky, Tal Z. "Transparent Predictions." *University of Illinois Law Journal*. 4 (2013): 1504-70.
[152] King, Ross D., et al. "The Robot Scientist Project." *Discovery Science*. Vol. 3735: Lecture Notes in Computer Science (2005): 16-25
[153] Fisher, Erik. "Lessons Learned from the Ethical, Legal and Social Implications Program (ELSI): Planning Societal Implications Research for the National Nanotechnology Program." *Technology in Society* 27 (2005): 321-28.
[154] Fisher, Erik. "Lessons Learned from the Ethical, Legal and Social Implications Program (ELSI): Planning Societal Implications Research for the National Nanotechnology Program." *Technology in Society* 27 (2005): 321-28.
[155] McDowall, Will. "Technology Roadmaps for Transition Management: The Case of Hydrogen Energy." *Technological Forecasting and Social Change* 79 (2012): 530-42; van den Ende, Jan, et al. "Traditional and Modern Technology Assessment: Toward a Toolkit." *Technological Forecasting and Social Change* 58 (1998): 5-21; Cuhls, Kerstin. "From Forecasting to Foresight Processes--New Participative Foresight Activities in Germany." *Journal of Forecasting* 22.3 (2003): 93-111.
[156] Ries, Christine P. "Market Driven Impact of Big Data on Public Policy Analysis and Practice: Public Sector Efficiency, Voter Centric Policies and Public Advocacy Management." Atlanta, GA: School of Economics, Georgia Institute of Technology (2015).
[157] Friedman, Batya, Peter H. Kahn, and Alan Borning. "Value Sensitive Design and Information Systems." *Human-Computer Interaction and Management Information Systems: Applications*. Eds. Galletta, Dennis and Ping Zhang. Armonk, NY: ME Sharpe Inc. (2006): 348-72; van den Hoven, Jeroen. "Value Sensitive Design and Responsible Innovation." *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*. Eds. Owen, Richard, John Bessant and Maggy Heintz. Hoboken, NJ: Wiley (2013): 75-83.
[158] Brown, Tim. "Design Thinking." *Harvard Business Review* 86.6 (2008): 84-92.
[159] Silver, Mark S., and M. Lynne Markus. "Conceptualizing the SocioTechnical (ST) Artifact." *Systems, Signs & Actions* 7.1 (2013): 82-89.
[160] Friedman, Batya, Peter H. Kahn, and Alan Borning. "Value Sensitive Design and Information Systems." *Human-Computer Interaction and Management Information Systems: Applications*. Eds. Galletta, Dennis and Ping Zhang. Armonk, NY: ME Sharpe Inc. (2006): 348-72.
[161] Markus, M. Lynne. "Information Technology and Organizational Structure." *Information Systems and Information Technology, Computing Handbook, Volume 2*. Eds. Topi, Heikki and Allen Tucker. London, UK: Chapman and Hall/CRC Press (2014): 67-1 - 67-22; Pool, Ithiel de Sola. *The Social Impact of the Telephone*. Cambridge, MA: The MIT Press (1978).
[162] Pool, Ithiel de Sola. *Forecasting the Telephone: A Retrospective Technology Assessment of the Telephone*. Norwood, NJ: Ablex (1983).
[163] Cornish, Edward. *Futuring: The Exploration of the Future*. Bethesda, MD: World Future Society (2004); Cuhls, Kerstin. "From Forecasting to Foresight Processes--New Participative Foresight Activities in Germany." *Journal of Forecasting* 22.3 (2003): 93-111.

[164] Topi, Heikki. "The Evolving Discipline of Information Systems." *Information Systems and Information Technology, Computing Handbook, Volume 2*. Eds. Topi, Heikki and Allen Tucker. London, UK: Chapman and Hall/CRC Press (2014): 1-1 - 1-26.