The Politics of Big Data
A Three-Level Analysis

By
Alon Peled
Department of Political Science
& Public Administration
The Hebrew University of Jerusalem
Israel
Email: Alon.Peled@post.harvard.edu

The Politics of Big Data
A Three-Level Analysis

Abstract

What role does politics play in the emerging Big Data domain? The paper argues that Big Data political power struggles surface at three distinct levels of analysis: the social sciences, the information state, and bureaucratic politics. At the social sciences level of analysis, Big Data threatens to divide social scientists into antagonistic methodological camps as it does not conform to traditional research techniques. At the information state level of analysis, a handful of powerful agencies and corporations created around data generation are consolidating their competitive advantage and are unlikely to support important data access and privacy protections. The one brighter spot for Big Data is found inside governmental bureaucracy. Here, trends such as "governance by numbers" at the sub-national level and mutually profitable data exchanges at the national level suggest that Big Data may propel agencies to share information better. The article concludes with a proposal to view Dr. John Snow and his work to stop the cholera epidemic in central London in 1854 as an early harbinger of the Big Data movement. Snow's work displays redeeming features that may mitigate less desirable effects of Big Data projects across the three levels of analysis. These features are a sense of purpose, ingenuity, clever data collection design, collaboration, humility and humanity.

Keywords: Big Data, Social Sciences, Bureaucracy, Bureaucratic Politics, the Information State, Data Divide, Privacy, Data Collection, Data Analysis, E-Government, Information Sharing

Last year, Marissa Mayer, Yahoo's CEO, explained that companies like hers single mindedly pursue Big Data analysis and succeed because "data is apolitical" (Schrage 2012). Politics is about who gets what, when, and why and political studies investigate how individuals and organizations compete to acquire legitimate ownership of scarce resources (Lasswell 1958; Weber 2004). So, was Mayer correct in her statement? Have we finally discovered a domain of human activities where scarcity, power struggles, and legitimate ownership are less important than in other domains? If Mayer is wrong, how does politics manifest itself in Big Data?

This paper argues that politics play an important role in the Big Data domain at three distinct levels of analysis: the social sciences, the information state, and bureaucratic politics. At each of these levels of analysis, Big Data re-structures power relationships and re-defines the distribution of information. The politics of Big Data re-draws the fault lines among camps, re-divides ownership of the most important informational resources of the state, and creates new winners and losers. There may be no scarcity of digital data; however, the competition to

interpret and exploit Big Data, and the question of whether to set ethical limits on the use of Big Data are political in nature.

The paper raises pertinent questions regarding the political impact of Big Data on our lives. It concludes by presenting the work of Dr. John Snow to stop the cholera epidemic in Central London in 1854. His work displays features that could potentially nudge social scientists, bureaucrats, and citizens to work together in pursuit of important Big Data projects for the benefit of society at large. These features are a sense of purpose, ingenuity, clever data collection design, humility, and humanity.

Definition and Characteristics of Big Data

The business-definition of Big Data highlights how agencies and corporations collect, store, and analyze large quantities of data and then extract new revenue from data insights (Gantz and Reinsel 2011). The plummeting cost of storage, computing, and network bandwidth and the conversion of everything to digital data facilitate the efforts to collect and analyze Big Data. For example, the U.S. Department of Agriculture uses hundreds of terabytes of satellite data to estimate the overall food supplies of the USA (Anderson 2008).

Big data practitioners emphasize that data is an ideal "raw material" because it is never consumed like other materials, can be put to work in parallel to support multiple purposes, and is always available for re-use. The value of data is often dormant and increases over time as businesses effectively use it, often to support secondary uses unanticipated at the time when the data was first collected. Even banal or incorrect data (such as data with typing mistakes) can be harvested to generate profits. Google developed powerful commercial spell-checking and voice recognition software by learning from incorrect data about the common mistakes that people make when they type or speak. Data can also be re-combined with other data to generate additional value (Mayer-Schonberger and Cukier 2013, 98-122). An overwhelming number of

corporations (73%) reported to leverage Big Data to increase revenue (Avanade 2012). With Big Data, the old information technology (IT) departments, traditionally deemed as cost centers, have become new corporate revenue generators. Big Data analysis also saves lives including identifying epidemics more quickly and helping find terrorists (as in the case of helping to find the Boston Marathon bombers) (Harris 2013).

The most acclaimed raw source of Big Data is known as "the digital exhaust." Organizations often collect data about people's digital traces including registering streams of computer mouse-clicks, logs reporting where we drive or walk, and our financial transactions. Big Data scientists mine the digital exhaust to discover what clients want. In this way, Big Data becomes the competitive advantage of corporations such as Amazon; these corporations appear at times to know ahead of us what we will want to buy next (Mayer-Schonberger and Cukier 2013, 111-115). If data is the new "coin of the realm" than Big Data is the foundation of the new data-centered economy (Economist 2010, 7). Experts estimate that, since 2005, corporations increased their investment in the digital universe by 50%, to four trillion dollars (Gantz and Reinsel 2011, 1)**.**

Scholars provide a more nuanced understanding of Big Data that focuses on how to work with it, what Big Data can tell us, and what new skills we need to acquire. This definition emphasizes that we must carefully design our data collection efforts to ensure that we collect *all* data about a given challenge. Like Google, scholars must embrace 'messy data' because it too contains potential valuable insights (Mayer-Schonberger and Cukier 2013, 32-48). With Big Data, scholars can easily discover non-linear patterns that statistical methods do not reveal. Often, Big Data enables scholars to drill into outlier cases that traditional statistics omit. Big Data allows scholars to rely less on their intuition and more on mathematical and correlational

data-analysis techniques (Mayer-Schonberger and Cukier 2013, 50-72). Big Data empowers scholars to make broad stroke predictions that, despite a certain lack of accuracy, can point out important future trends. Interestingly, scholars do not consider the sheer 'bigness' of Big Data to be an important defining quality of Big Data (Mayer-Schonberger and Cukier 2013, 192).

Big Data is a buzzword that generates excitement. Practitioners suggest that the possibilities inherent in Big Data are "endless" (Manovich 2011, 13)**.** Politicians such as President Obama attempt to ride on the Big Data wave (Barton 2012). Readers are taught new words to describe the size of Big Data including Petabyte (1,000 terabytes), Exabyte (1,000 Petabytes), Zettabyte (1,000 Exabytes), and quintillion (10^{18} – used to describe how many computer files the world will have in 2020: 25 quintillion files) (Gantz and Reinsel 2010). Other social scientists remain unimpressed by these new words and on guard against some of the less desirable attributes of Big Data. Scholars are alarmed at how Big Data projects are invading privacy and concerned about the hubris of Big Data proponents that appear to have excessive confidence in large datasets as a superior form of knowledge (Boyd and Crawford 2012, 663).

The Social Sciences Level of Analysis

At the social sciences level of analysis we ask: Is Big Data analysis merely another 'research tool'? The paper proposes that, on the contrary, Big Data analysis is not just another research technique. It argues that, in the coming years, social scientists may experience new methodological battles that stem from Big Data and could divide the social sciences community.

Humanities scholars, who adopted Big Data before social scientists, are already experiencing such battles. "Digital Humanities" scholars are using Business Intelligence (BI) tools to investigate large digitized corpuses of literary works. However, other scholars have raised difficult questions about what happens when scholars begin viewing everything through digital glasses. These questions include: Must all intellectual work become a "software study"?

Should the digital really become the essence of cultural studies? What epistemic changes are scholars experiencing when they begin viewing culture as computer code? How does discrete digitization impact the continuous flow of a narrative (Berry 2011)?

Social scientists discovered Big Data more recently; they are using smaller datasets than used in Digital Humanities projects (Manovich 2011). Excitement is growing over the new "computational social science" domain. Like corporate analysts, scholars working in this domain analyze the digital exhaust of citizens. Golder invited scholars to imagine the "exciting" future where large datasets could be linked using programming code such as Hadoop (Golder 2010).

Yet Big Data is not necessarily all good news for the social sciences. It may create new political fault lines within the social science community along three axes: data-access, data-analysis, and the ethics of using Big Data. Firstly, data access is a divisive issue. Freely accessible Big Data such as Twitter and Facebook data is the low hanging fruit on the new Big Data tree. Scholars have already discovered that Twitter data often represent the activities of like-minded individuals influenced by a small elite group of gatekeepers with "celebrities following celebrities, media following media, and bloggers following bloggers" (Wu et al. 2011, 9-10).[1] Agencies and corporations own the largest, most interesting, and, potentially, most valuable data repositories. These agencies and corporations have no responsibility to share data with social scientists. Data-giants such as Google protect their valuable transactional datasets and agree to release the less valuable summarized data. So, in the future, will affluent universities buy expensive data for their researchers thus widening the gap with other social scientists? Will new types of social science consortiums composed of the richest research universities emerge to

---

[1] It might therefore not be surprising that one of the first social science Big Data articles that appeared in the prestigious *Science* journal and was based on a breakthrough methodological analysis of billions of Twitter messages reached the somewhat dull conclusion that people are happier when they wake up in the morning than when they are on the job and that people tend to sleep two hours more during the weekend (Golder and Macy 2011).

ensure better or exclusive access to datasets for their researchers (Lazer et al. 2009, 721)? The history of electronic access to scientific journals suggests that this danger is real.

Peter Norvig, Google's Research Director suggested that with Big Data, even the worst algorithm performs far better than what can be achieved with a smaller dataset (Naone 2011, 2). Does this observation imply that, in the future, mediocre social scientists who develop poor algorithms but can run them against high value datasets (purchased by their rich universities) will acquire an edge over more competent social scientists who do not have access to such data?

Second, the social science community could divide from within over the challenge of how to analyze Big Data. Practitioners have argued that Big Data analysis techniques are different than other techniques because the former techniques are focused on discovering random correlations rather than affirming hypotheses (Anderson 2008). In response, other social scientists proposed that interpretation is the very essence of what social scientists do. These scholars also argued that social scientists have a new responsibility to explain the creation process, and the weaknesses of Big Data repositories (Boyd and Crawford 2012, 668).

Finally, the social science community could also split from within in an ethical debate how to use Big Data. Big Data repositories are constructed from linked, smaller datasets. The creators of these smaller datasets and people whose information is contained in them have no way to anticipate how the data will be used once it lands inside a Big Data repository. There is often no incentive to develop mechanisms to inform the creators or data-subjects of small datasets that the data will now be used to support a different cause. The key question is how accountable must social scientists be to the owners and subjects of these datasets? In this debate, we may find on one side social scientists who demand stricter accountability even if it means not exploiting Big Data to its full potential. On the other side, we may discover social scientists who

are willing to sacrifice accountability to advance social science faster.

The Information State

New information resources alter the balance of power among countries and between the public, private, and not-for-profit sectors within the state. In the 19th century, the British built centers of calculation, such as the Royal Botanical Gardens at the Kew, where colonial botanical materials were analyzed. The knowledge gained was distributed throughout the empire resulting in the creation or destruction of local economies (Parry 2004). In what ways might Big Data disrupt established political power relationships among countries and within the state?

Regrettably, the Big Data story at the information state level of analysis is not encouraging. The two Big Data topics that attract attention at the level of the information state are the divide between the data-rich and the data-poor and privacy protection. Creative ideas on improving data access and privacy protection exist. Alas, not much is happening in terms of converting these ideas into action. At the level of the information state, the Big Data winners have already surfaced. These winners are more concerned about consolidating their competitive advantage than about improving data access and privacy protection.

Corporations created around data such as Google and Amazon are the winners in the information state (Schrage 2012). Even within this small group of winners an early Big Data start provides a lasting advantage. Google Maps advantage over Apple Maps is mainly due to the fact that Google entered the geo-spatial data collection domain first, in 2005. Google is seeking to establish a similar early mover advantage in the spell checking and speech recognition domains (Einstein 2012). Likewise, when data giants mesh their resources together they quickly consolidate the field. Recently the investment arms of Google and the CIA created a new company under the apt name of "Recorded Future" to analyze social media content more accurately than any other existing company. Recorded Future will have access to the most

valuable transactional data in the world to "predict the curve" by examining invisible links between billions of web pages (Shachtman 2010). Pollsters and politicians (including President Obama) are striving to generate the same edge over their competitors by investing in Big Data (Anderson 2008; Honan 2012). Sadly, rather than level the playing field of the information state, Big Data appears to be deepening the divides between the data-haves and the data-have-nots.

The same sad conclusion holds true in the domain of privacy protection. Old privacy protections crumble. Citizens cannot provide informed consent to be included in a dataset because no one can anticipate how the data will be used in the future. Citizens cannot remain anonymous in these datasets because, with Big Data, it is easy to re-identify them. It is not easy to opt out from a dataset and the act of opting out might identify a person. Agencies are instructed to collect only the data that they need to fulfill their missions. But corporations have a reverse incentive—to link more data together to generate higher revenue (Bollier 2010).

Good ideas exist to protect privacy more effectively including the creation of anonymized data spaces, empowering citizens to sell personal data, and even using Big Data itself to identify privacy violations. Alas, these ideas, if implemented, will cost corporations such as Google billions of dollars. It is therefore not surprising to hear corporate analysts describe excitedly how Big Data generates better samples for follow-up studies, empowers more eyes to view the data and discover errors, and catch terrorists in near real time. These analysts prefer not to talk about the topics of data-divide and privacy. They also prefer not to discuss what might happen when less desirable characters, such as terrorists, learn to harvest the value of Big Data.

Bureaucratic Politics
The state's bureaucracy has long been the biggest generator, collector and user of data (Economist 2010, 10). The public sector today is the collection of institutions whose administrative staff maintains a monopoly over the legitimate processes of producing, updating

and disseminating the most extensive and authoritative information in the state (Peled 2011).

Scholars at the University of Irvine, California developed more than four decades ago the power

politics approach to explain when, why, and how agencies use their computers in bureaucratic

political fights. The Irvine scholars argued that computers do not revolutionize public sector

organizations; rather, they are instruments used to support a political agenda that these

organizations developed before the arrival of computers (Kling 1980, 60; Kling and Iacono 1984,

1219; Kraemer and King 1986, 494). The Irvine scholars also explained that computers reinforce

the existing organizational status quo because they provide the elite with opportunities to decide

how much to invest in computing, how to control computer access, and which priorities to

promote while developing new systems (Kraemer and King 1986, 492).

Some evidence exists to support the claim that agencies would use Big Data in the same

way that they have used other computer technologies—as weapons to fight over funds, influence,

and autonomy. For example, the American Government Accountability Office (GAO) recently

argued that the continuous refusal of agencies to share their Big Data is one reason why USA

exports are not as competitive as they can be (GAO 2013). In the private sector too, scholars

noted how antagonistic departments harness Big Data as a new weapon in old political struggles.

In one European telecom corporation, analysts used Big Data to reveal that network outages and

the perception by customers that the corporation has made false advertising claims created a

negative corporate social media image. Rather than cooperate to correct these problems, the

marketing and network groups used these insights to blame each other for the negative image

(Bughin; Livingston, and Marwaha 2011, 7).

Yet, there is interesting evidence to support a contrarian claim that Big Data is changing

the rules of the old bureaucratic politics game at the subnational level. Since 2009, Mike Flowers

and his New York City (NYC) data analytics team had been "crunching the numbers" to

discover insights to help politicians govern the city better (Mayer-Schonberger and Cukier 2013,

185-189)**.** Baltimore is doing the same nowadays. City departments are required to surrender

their transactional datasets to these business analytics team as a precondition for receiving

budgets. For the first time ever, elected politicians, aided by small data teams, have the means to

crunch the numbers to run their cities better. Scholars must pay closer attention to this change.

At the national level, agencies largely resist orders to surrender their transactional

datasets (Peled 2011). However, also at the national level, Big Data appears to be nudging

agencies to engage each other in a new "win-win" information sharing arrangement. National

agencies appear to be developing sophisticated new ways to profit from their data and, in the

process, improve data access. The American National Technical Information Service (NTIS) was

created in 1950 as the national clearinghouse to collect and distribute scientific and technical

information (Braman 2006, 303). The NTIS can sell various information products to the public

using flexible pricing schemes. About sixty agencies are taking advantage of the NTIS to sell

information products to the public. For example, the Social Security Agency (SSA) sells its

limited Death Master File (DMF) to the NTIS for $210,000 annual fee. The NTIS, in turn, sells

different access programs to individuals and organizations to access the DMF data. Article B7 of

the NTIS-SSA contract states that "Upon request by SSA, NTIS will provide an analysis of the

DMF customer base in an appropriate format agreed upon by the parties" (NTIS and SSA 2013)**.**

So, data analytics teams at the sub-national level and mutually profitable data-exchanges

at the national level may indicate that Big Data is the first technology to penetrate government

and defy the arguments of the Irvine school of thought. If true, the hidden power of the Big Data

revolution resides not in corporations such as Google but rather in information sharing

improvements inside government.

John Snow, the London Cholera Epidemic of 1854, and the Future of Big Data
It might be possible to harness ideas, people, and resources from across the three levels of

analysis to meet some of the challenges identified above. I will illustrate this idea through the

analysis of a project that may be the earliest harbinger of the Big Data movement going back to

the mid 19th century—the story of Dr. Snow and his cholera dot map.

On August 31, 1854, cholera broke out in the Broad Street area of central London. In

1854, thirty years before Koch discovered how the *vibrio cholera* bacterium spreads through

water contaminated with sewage, dominant theories attributed cholera to the hand of God or to

foul air. Back then, cholera mortality rates were above 50%. Cholera-related death was painful

and took place within 48 hours from infection. The fast spread of the 1854 cholera disease within

the densely populated London area resembled the spread of the Great Plague 200 years earlier

(1665-1666). Panic spread quickly. By September 1854, 616 people had died from cholera, most

of them during the first week of the epidemic. Three quarters of the population in the infected

area fled turning Soho into a ghost town. On September 7, 1854, Snow presented a dot map

displayed in **Figure 1** below to the Board of Guardians of the St. James parish. On this map

Snow plotted the exact location of each one of the first 83 cholera-related deaths as well as the

location of 13 public water pumps. The Guardians immediately saw the concentration of deaths

around the Broad Street pump and ordered the removal of the pump's handle so people could not

drink from this well. The handle was removed in the early morning of September 8 and the

epidemic subsided. Snow and his famous dot map are celebrated for saving thousands of lives.

Figure 1.0 – John Snow's Dot Map (1854)

Long before the invention of the computer, Snow possessed a Big Data state of mind. His work displayed the hallmarks of a true Big Data project. His data addressed the most urgent life-and-death problem of his time. On September 3, 1954, He collected *all* the data that existed about cholera-related deaths from the General Register Office. So the size of his data was N=All=83 cases (Mayer-Schonberger and Cukier 2013, 26). Not Big Data in modern terms, yet the dataset that Snow assembled was the most complete one he could have collected and completeness is more important than "bigness" in defining Big Data. Snow acknowledged some

weaknesses in his data. For example, he failed to find street address numbers for a handful of death cases. However, in a true Big Data spirit that embraces messy data, Snow wrote:

> "If the locality of the few additional cases could be ascertained, they would probably be distributed over the district of the outbreak in the same proportion as the large number which are known" (Tufte 1997, 34).

Additional Big Data features are apparent in Snow's work. He linked geo-spatial and mortality data by recasting the original Registry's death data from its one-dimensional temporal order into a two-dimensional spatial comparison (Tufte 1997, 30). He used data that he collected himself as well as the 19th century version of Open Data that the City of London agreed to release (Johnson 2013). He neither tried to explain the relationship between water and cholera nor to summarize the data. Instead, Snow simply plotted the data points thus enabling the data to reveal the story. Using the available scientific knowledge of his time, Snow could not see the cholera bacterium. Therefore, using correlation and visualization techniques he explored the most complete dataset possible to infer the existence of such a bacterium based on data patterns (Johnson 2013).

Snow's work displays other less glamorous, but equally important data-collection, data-analysis, and ethical features of a true Big Data project. Long before the 1854 cholera outbreak Snow designed a data-collection system to support a clear purpose—investigate potential water-cholera correlations. His 1854 map was merely the "marketing vehicle" that he developed to promote this Big Data project. Moreover, in an interesting interdisciplinary cooperation, John Snow, the scientist, had to convince and then collaborate with Reverend Henry Whitehead to collect the data and, later on, convince the Guardians to remove the pump (Johnson 2013).

By 1854, Snow and Whitehead had completed data collection efforts about four similar cholera outbreaks. In 1849, Snow published the essay *On the Mode of Communication of Cholera* in which he described his ideas about the relationship between water and cholera. Snow understood that quantitative data analysis might fail to capture what the data purports to quantify

13

(Mayer-Schonberger and Cukier 2013, 165-166). He therefore considered his dot map to be merely the beginning of his work. On this map, Snow identified interesting outliers. For example, **Figure 2** below highlights the locations of a brewery and a Work House (e.g., a correctional facility) where few people died from cholera. Snow visited these locations and discovered that the brewery employees and Work House inmates did not drink water from the Broad Street pump.
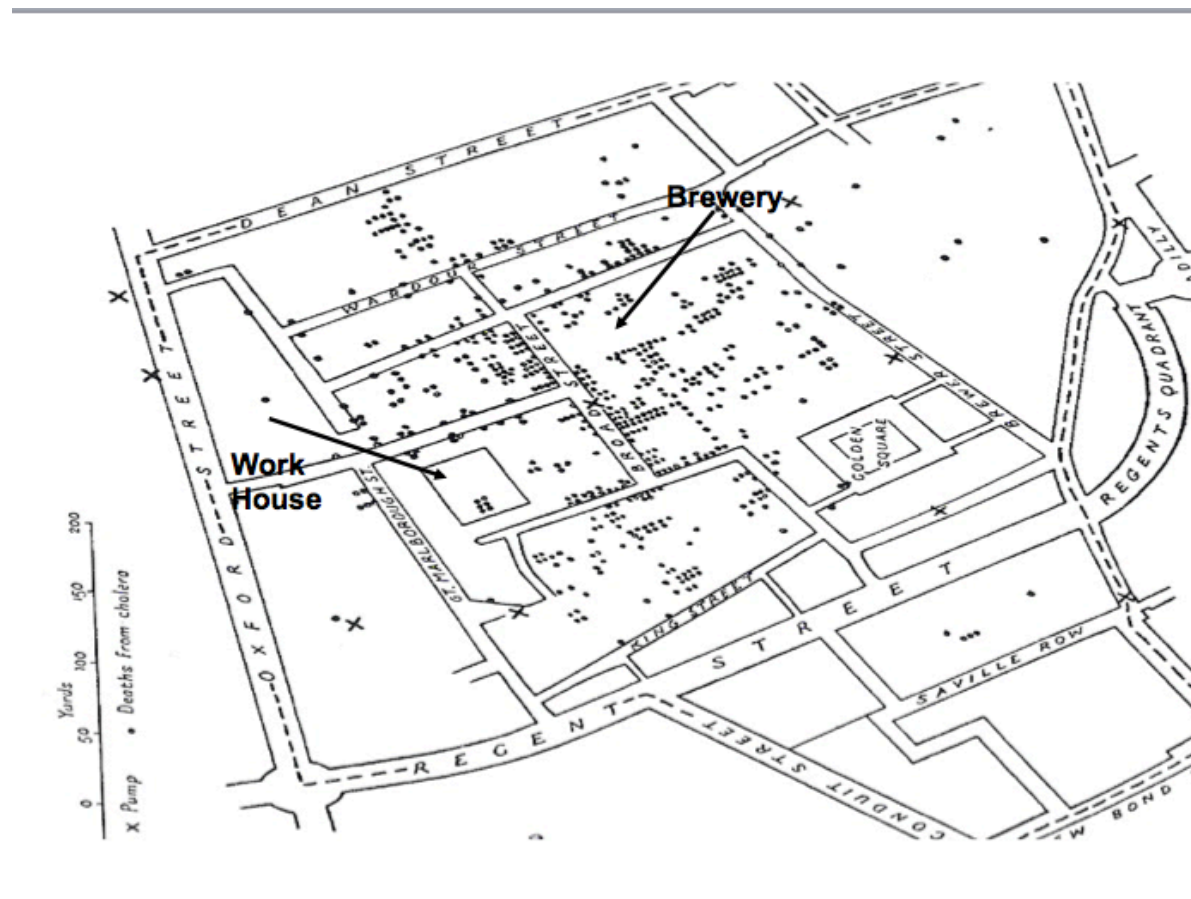

Figure 2 – The Brewery and Work House on John Snow's Dot Map
(Figure from: (Tufte 1997, 30-31)).

Snow could also have treated himself as an outlier, avoiding the contaminated water and remaining healthy while visiting sick houses. He was a relentless data collector. He continued to investigate the water-cholera correlation long after 1854 and assembled additional data to show that the incidence of cholera was ten-fold higher in households supplied by one water company

(the Vauxhall and Southward) as compared to those supplied by another (the Lambeth), the

water extraction point of the former being close to a major sewer (The Vauxhall Society 2012).

Snow was also a Big Data scientist in his analysis method. He was creative, defied the

accepted scientific wisdom of his time, and invited other scholars and laymen to contribute

insights and data to his project. Throughout the effort he remained the architect of the project. He

worked diligently including visiting each of the houses of the first 83 cholera victims to confirm

that they did indeed drink water from the alleged contaminated pump. Continuously, he searched

for opportunities to deepen his analysis. He travelled to interview relatives of more distant

cholera victims and discovered that these victims too drank water from the Broad Street pump

shortly before becoming ill (Tufte 1997).

Today, theoreticians suggest to treat data subjects as individuals and pay close attention

to how these individuals learn and change (Latour 2009). Snow did just that in 1854. He is

considered the father of modern epidemiology because he devised a questionnaire and collected

data from human subjects in a systematic and humane way (John Snow Archive 2013). His

mission could not have been simple. He knocked on doors hours after the death of a beloved

family member to interview grieving family members about the water drinking habits of the

deceased. Nonetheless, he managed to visit every single house and collect the data he needed.

Snow's early Big Data project resulted in several important insights including about the need to

separate completely the drinking water and waste systems of cities (Johnson 2013).

Snow had to overcome political, religious, scientific, and media resistance. The

Guardians were not thrilled to remove access to a water source during an epidemic. A mid 19th

century version of a political lobby composed of water delivery societies acted clandestinely to

foil his work. The Board of Health conducted a large study to demonstrate that cholera

transmitted by air rather than water. This Board continued to reject Snow's findings even after 1854 and only accepted them years later after the Germans adopted Snow's ideas (UCLA Department of Epidemiology 2005). Journalists dismissed his published work. Still, Snow remained undeterred, undistracted, and focused on his water-cholera thesis.

Snow's relentless pursuit of evidence to support a *specific* correlation is seemingly the one feature of his work that does not appear to be in line with the definition of a Big Data project. Are not Big Data projects supposed to scan huge amounts of data in pursuit of insights? How could a project so tightly focused on a *single* correlation be the harbinger of Big Data projects to come? I propose that this seemingly out-of-line feature of Snow's 19th century work is an important missing component in the definition of what a Big Data project must be in the 21st century. Snow designed a data collection system to support a single idea but he also understood that information is power and power corrupts. He therefore deliberately created opportunities for the data to falsify his theory. He understood what quantitative data cannot tell us and how we must supplement it with information from other sources. He worked hard and was humble to the point of undercutting his own achievement; he explained how the flight of the population from Central London might have been the real reason for the demise of the cholera outbreak. He generously credited his colleagues and data subjects for their ideas and data. Snow never believed in the omniscient power of quantitative data to reveal the truth. His sense of purpose, ingenuity, clever data design, collaboration, humility, and humanity are the kind of qualities that must characterize Big Data scientists. These qualities might be the redeeming qualities that will help social scientists avoid internal methodological wars and, instead, harness their minds and work to improve life in the information state.

Peled, Alon. "The Politics of Big Data – a Three-Level Analysis." Paper presentation at the European Consortium of Political Research (ECPR) General Conference, Bordeaux, France (4-7 September 2013).

References

Anderson, Chris. 2008. *The end of theory, will the data deluge make the scientific method obsolete?* http://www.edge.org/3rd_culture/anderson08/anderson08_index.html (accessed August 25, 2013).

Avanade. 2012. Global Survey: Is Big Data Producing Big Returns? http://www.avanade.com/Documents/Research and Insights/avanade-big-data-executive-summary-2012.pdf (accessed August 25, 2013).

Barton, Mike. 2012. Webcast (Thursday): Obama Goes Big on Big Data. http://www.wired.com/insights/2012/03/obama-big-data/ (accessed August 25, 2013).

Berry, David M. 2011. The Computational Turn: Thinking About the Digital Humanities. *Culture Machine* 12:1-22.

Bollier, David. 2010. The Promise and Peril of Big Data. http://www.aspeninstitute.org/publications/promise-peril-big-data (accessed August 25, 2013).

Boyd, Danah, and Kate Crawford. 2012. Critical Questions for Big Data. *Information, Communication & Society* 15 (5):662–679.

Braman, Sandra. 2006. *Change of State - Information, Policy, and Power*. Cambridge, MA: The MIT Press.

Bughin, Jacques, John Livingston, and Sam Marwaha. 2011. Seizing the potential of 'big data'. *McKinsey Quarterly*, Oct., 2011, 1-8.

Economist. 2010. The Data Deluge. *The Economist*:1-15.

Einstein, Dave. 2012. Google vs. Apple Maps: Big-Data Battle, Cloudy Clash. http://www.forbes.com/sites/netapp/2012/10/03/google-apple-maps-big-data-cloud/print/ (accessed August 25, 2013).

Gantz, John, and David Reinsel. 2010. The Digital Universe Decade - Are You Ready? http://www.emc.com/collateral/analyst-reports/idc-digital-universe-are-you-ready.pdf (accessed August 25, 2013).

-----. 2011. Extracting Value from Chaos. http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf (accessed August 25, 2013).

Golder, Scott A. 2010. Scaling Social Science with Apache Hadoop. http://blog.cloudera.com/blog/2010/04/scaling-social-science-with-hadoop/ (accessed August 25, 2013).

Golder, Scott A., and Michael W. Macy. 2011. Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures. *Science* 333:1878-1881.

Harris, Derrick. 2013. On big data, the Boston Marathon and civil liberties. http://gigaom.com/2013/04/17/on-big-data-the-boston-marathon-and-civil-liberties/ (accessed August 25, 2013).

Honan, Mat. 2012. Twitter's New Political Index Proves Big Data Knows What You're Thinking. *Wired*, 01.08.2012, 9.

Johnson, Steven. 2013. Happy birthday John Snow, father of modern epidemiology: A Q&A with Steven Johnson. http://blog.ted.com/2013/03/15/happy-birthday-john-snow-father-of-modern-epidemiology-a-qa-with-steven-johnson/ (accessed August 25, 2013).

Kling, Rob. 1980. Social Analyses of Computing: Theoretical Orientations in Recent Empirical Research. *Computing Surveys* 2 (4):61-110.

Kling, Rob, and Suzanne Iacono. 1984. The Control of Information Systems Development After Implementation. *Communications of the ACM* 27 (12):1218-1226.

Kraemer, Kenneth L., and John L. King. 1986. Computing in Public Organizations. *Public Administration Review* 46:488-496.

Peled, Alon. "The Politics of Big Data – a Three-Level Analysis." Paper presentation at the European Consortium of Political Research (ECPR) General Conference, Bordeaux, France (4-7 September 2013).

Lasswell, Harold Dwight. 1958. *Politics: Who Gets What, When, How*. New York: New Word.

Latour, Bruno, ed. 2009. *Tarde's idea of quantification*. Edited by M. Candea, *The Social after Gabriel Tarde: Debates and Assessments*. London: Routledge.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Laszlo Barbasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. Computational Social Science. *Science* 323:721-723.

Manovich, Lev. 2011. Trending: The Promises and the Challenges of Big Social Data. *Debates in the Digital Humanities*, 1-17.

Mayer-Schonberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston, MA: Houghton Mifflin Harcourt.

Naone, Erica. 2011. With Big Data Comes Big Responsibilities. http://www.technologyreview.com/news/425648/with-big-data-comes-big-responsibilities/ (accessed August 25, 2013).

National Technical Information Service (NTIS), and Social Security Administration (SSA). 2013. Reimbursable Agreement Between The U.S. Department of Commerce National Technical Information Service (NTIS) and the Social Security Administration (SSA) NTIS-1469.

Parry, Bronwyn. 2004. *Trading the genome: investigating the commodification of bio-information*. New York: Columbia University Press.

Peled, Alon. 2011. When Transparency and Collaboration Collide:The USA Open Data Program. *Journal Of The American Society For Information Science And Technology [JASIST]* 62 (11):2085-2094.

Schrage, Michael. 2012. What Executives Don't Understand About Big Data. http://blogs.hbr.org/schrage/2012/09/what-executives-dont-understan.html (accessed August 25, 2013).

Shachtman, Noah. 2010. Exclusive: Google, CIA Invest in 'Future' of Web Monitoring. *Wired.com*, http://www.wired.com/dangerroom/2010/07/exclusive-google-cia (accessed August 25, 2013).

The John Snow Archive and Research Companion. 2013. *About John Snow*. http://johnsnow.matrix.msu.edu/aboutjohn.php (accessed August 25, 2013).

The Vauxhall Society. 2012. *Cholera*. http://www.vauxhallcivicsociety.org.uk/history/cholera/ (accessed August 25, 2013).

Tufte, Edward R. 1997. *Visual Explanations - Images and Quantities, Evidence and Narrative*. Cheshire, Connecticut: Graphics Press.

U.S. Government Accountability Office. 2013. Export Promotion: Better Information Needed about Federal Resources. http://www.gao.gov/products/GAO-13-644 (accessed August 25, 2013).

UCLA Department of Epidemiology - School of Public Health. 2005. *Father of Modern Epidemiology*. http://www.ph.ucla.edu/epi/snow/fatherofepidemiology.html (accessed August 25, 2013).

Weber, Max. 2004. *The Vocation Lectures ("Science as a Vocation," "Politics as a Vocation")*. Edited by D. Owen and T. B. Strong. Indianapolis, Indiana: Hackett Publishing Company.

Wu, Shaomei, Winter A. Mason, Jake M. Hofman, and Duncan J. Watts. 2011. Who Says What to Whom on Twitter. Paper read at International World Wide Web Conference, at Hyderabad, India.