2-28-2018

# A New Methodological Frontier in Entrepreneurship Research: Big Data Studies

Andreas Schwab
*Iowa State University*, aschwab@iastate.edu

Zhu Zhang
*Iowa State University*, zhuzhang@iastate.edu

# A New Methodological Frontier in Entrepreneurship Research: Big Data Studies

**Abstract**
The emergence of 'big data' and related analytic techniques are creating opportunities to advance empirical entrepreneurship theory and practice. This editorial focuses on the implications for the design and execution of empirical studies. It offers guidance on how to navigate related methodological challenges and outlines what editors, professional associations, research-method teachers and administrators can do to enable high-quality big data research.

**Keywords**
Research Methods, Quantitative Methodologies, Big Data

**Disciplines**
Business Analytics | Entrepreneurial and Small Business Operations | Management Information Systems | Management Sciences and Quantitative Methods

**Comments**
This accepted article is published as Schwab, A. & Zhang, Z. (2018). A New Methodological Frontier in Entrepreneurship Research: Big Data Studies. *Entrepreneurship, Theory & Practice*. First published online: February 28, 2018; DOI: 10.1177/1042258718760841. Posted with permission.

**A New Methodological Frontier in Entrepreneurship Research:**
**Big Data Studies**

**Editorial**

Andreas Schwab
Department of Management
Ivy College of Business
Iowa State University
Ames, IA 50011
aschwab@iastate.edu

Zhu Zhang
Department of Supply Chain and Information Systems
Ivy College of Business
Iowa State University
Ames, IA.  50011
zhuzhang@iastate.edu

Published in
*Entrepreneurship Theory & Practice*

Reference:
Schwab, A. & Zhang, Z. (2018).  A New Methodological Frontier in Entrepreneurship Research: Big Data Studies.  Entrepreneurship, Theory & Practice. First published online: February 28, 2018; DOI: 10.1177/1042258718760841.

We thank Bill Starbuck, Jim Wetherbe, Jackie Runyan, one anonymous reviewer and the senior editors of ETP for their helpful comments on earlier drafts of this editorial.

**A New Methodological Frontier in Entrepreneurship Research: Big Data Studies**

**Abstract**

The emergence of 'big data' and related analytic techniques are creating opportunities to advance empirical entrepreneurship theory and practice. This editorial focuses on the implications for the design and execution of empirical studies. It offers guidance on how to navigate related methodological challenges and outlines what editors, professional associations, research-method teachers and administrators can do to enable high-quality big data research.

**Keywords**

Research Methods, Quantitative Methodologies, Big Data

**A New Methodological Frontier in Entrepreneurship Research: Big Data Studies**

Advancements in information technologies continue to reshape economic and social conditions. The ability to capture, store and process vast amounts of information has enabled the creation of what has been labeled "big data" -- massive data sets that contain fine-grained information of micro events and activities. IBM, Google, Microsoft and many other information technology companies are investing heavily in advancing related hardware, software, and services.

Big data are creating new opportunities that governments and organizations around the world have started to exploit (Sivarajah, Kamal, Irani, & Weerakkody, 2017). For entrepreneurs and managers big data provide novel approaches to observe human behavior and to discover and exploit business opportunities (George, Haas & Pentland, 2014; Wetherbe & Zhang, 2015). This editorial, however, focuses on a different topic: the opportunities of big data for academic research and the implications for the design and execution of empirical studies (Guzzo et al., 2015, Tonidandel, King & Cortina, 2016).

In our opinion, big data embody opportunities to make substantial contributions to entrepreneurship research. This editorial encourages entrepreneurship scholars to consider big data research and offers guidance on how to navigate related methodological challenges. It also discusses how editors, professional associations, research-method teachers and administrators can and should support big data research.

## EMERGENCE AND IMPACT OF BIG DATA

The increasing use of smart electronic devices, digital communication networks and the Internet is at everincreasing rates creating and storing data of human behavior and its outcomes. Hal Varian, the chief economist at Google, famously captured the explosive growth of digital data in every sphere of society and business by pointing out that in 2012 more data were produced every two days than had been produced in all of the years prior to 2003 combined (Smolan & Erwitt, 2012: 4). Only two years later, it was estimated that all the data generated in the world between the beginning of time and the year 2000 represented about the same amount of data that were being generated every minute (Kitchin, 2014b). This explosive growth has continued and is supported by continuing advances in digital technologies, such as extensive communication networks, widespread use of smart electronic devices, embedded sensors and digital recording capabilities of machines and products, as well as availability of inexpensive cloud-based massive data storage.

Today's data collection by modern cities nicely illustrates how digital systems are increasingly capturing human activities, often in real time. These cities, for example, continuously digitally record in visual, audio, text and other files the activities and events that occur in their domains. Public area surveillance systems capture the movements of individuals as well as cars, trucks and other vehicles. Activities and communications of some city employees, such as emergency responders, are recorded and archived nonstop. The hand phones, tablets, computers, vehicles and other equipment used by city employees provide a wealth of additional information -- some submitted by employees and some recorded automatically. Finally, city councils, committees, task forces and other institutional units create additional records and reports of their deliberations, decisions and actions, which today are often captured, communicated and retained in digital form. This available information allows for production of more detailed and comprehensive depictions, models and predictions of urban processes and their outcomes. Beyond their use in technocratic approaches focused on improved management of a city, such data also promise rich opportunities for empirical academic research focused on developing better theories of social behavior (Allwinkle & Cruickshank, 2011).

Modern business organizations are also generating massive data by digitally capturing activities. Google with its Android operating system, for example, can track the location information of every user. The resulting mobility profiles capture visits to stores, offices, homes, churches, or other public or private places anywhere in the world. Mobility profiles are incredibly valuable for analyzing human behavior and activities. These profiles become even more valuable if they are linked with information of an individual's use of search engines, social media, mapping software and electronic payment systems. Today even organizations outside the information technology sector have access a wealth of information from employees' smart phones, personal computers, and other electronic devices. Organizations also use CCTV more extensively to visually record what happens in their facilities. Organizations have used this information to create big data to better identify high-potential employees, to improve team productivity (Pentland, 2012), to reduce turnover (Goldberg, Srivastava, Manian, Monroe, & Potts, 2016) and to increase team creativity and success (Tripathi & Burleson, 2012; Olguín & Pentland, 2010).

Finally, more public big data are emerging. Governments and international associations have collected and shared collected data, such as economic census data or patent databases, for decades. Digitalization, however, has made such data more detailed, more accessible, and easier to combine with additional information. In addition, new big data sets are becoming available as public agencies capture more of their own micro activities and observations and make some of the resulting data available to the public. The various US, EU or UN agencies offer multiple illustrative examples. The US Federal Energy Regulation Commission, for example, made public all of the email records (header and text information) of 151 senior Enron mangers that it obtained in the process of investigating the Enron fiasco. These nearly 500,000 emails and text files have been used subsequently for multiple behavioral and organizational studies (Diesner, Frantz & Carley, 2005). Sports leagues have long collected and shared game statistics. Digital technology, however, has enabled them to collect an unprecedented wealth of information, such as that obtained through the continuous recordings of player movements and activities on the field. Again, scientists have used the resulting data sets to study leadership, team processes and organizational learning, among other topics.

The extensive use of smart phones, social media and other Internet applications by individuals and organizations has also created new opportunities for researchers to create big data sets by downloading and archiving information posted on the Internet, such as reports, transactions, comments and announcements. The corresponding downloading of data are today further facilitated by available webscraping technology.[1] In general, the era of big data has created unprecedented to obtain rich empirical data.

[1] See Braun, Kuljanin and DeShonl (2017) for available software packages to scrape web data.

## HOW IS BIG DATA DIFFERENT?

Large data sets are not a novel phenomenon. Entrepreneurship research has long exploited large data sets, such as the U.S. patent or US Census data. Hence, one might consider entrepreneurship researchers wellprepared to exploit the new opportunities of big data by extending established current methodologies developed for the analysis of traditional large data sets. For several reasons, this is not the case.

A primary reason is that big data tend to differ from traditional large data in several substantive ways. These differences imply not only new opportunities, but also new methodological challenges. The label "big" highlights the large *volume* and sheer size of the associated datasets. Big data, however, tend to also differ from traditional large data sets in other respects as big data tend to exhibit high *variety* and *velocity* -- often combined with low *veracity* (McAfee & Brynjolfsson, 2012). While volume represents a rather straightforward characteristic, the other differences are less obvious, but still deserve careful consideration for their implications for empirical research.

*Variety* refers to the heterogeneity of data sources and types of data captured. In the past, large data sets typically contained structured quantitative information stored in spreadsheets and relational databases. A substantial portion, if not the majority, of today's big data is unstructured or semi-structured. It, for example, may include natural language text, audio recordings, videos and photos. The ease and low cost of recording and archiving enables the indiscriminate continuous accumulation of micro-event information, often with time, device and location stamps. The resulting big data tend to contain highly heterogeneous structured and unstructured information capturing microevents from a variety of data sources.

*Velocity* refers to the speed at which data is generated, diffused and acted upon (Gandomi & Haider, 2015). Continuous capturing of micro events can create a constant stream of new observations. At the same time, the captured information is often responded to by the actors involved in the micro events. Posted social media messages, for example, tend to create dynamic streams of sequential communication. In this context, posted messages even have the potential to go viral and trigger subsequent social media postings on a massive scale.

Low *veracity* refers to the messiness, unreliability or biases that can characterize some of the raw data that go into big data sets. The pervasiveness of user-generated content and opportunistic data collection often reduces opportunities for researchers to ensure the quality and accuracy of initial data inputs. Hence, missing data, coding errors and incomplete documentation require close attention.

**HOW TO CONDUCT BIG DATA STUDIES**

Obviously, each data set is unique. As already outlined, big data tend to differ from traditional large data sets with regard to volume, variety, velocity and veracity. These differences create some fundamentally different methodological challenges (George et al., 2014; Tonidandelet al., 2016). Consequently, exploiting big-data opportunities often depends on the methodological capabilities to address these challenges. The required capabilities fall into two categories: first, data-management capabilities to efficiently clean, restructure, integrate, and combine the massive data, and second, data-analysis capabilities tailored to extract meaningful empirical information from big data. Hence, the era of big data is also associated with the emergence of novel investigative tools and approaches.

**Data Management**

Big data can provide a wealth of information about phenomena of interest. Data on a community's emergency response activities, for example, may contain all received 911 calls, as well as communications among dispatchers, police officers, medical emergency teams, fire fighters, and SWAT units. Additional temporospatial and other information is obtainable from any of the deployed vehicles, phones, tablets, and other electronic communication devices. Potentially, video recordings from body and vehicle cameras, plus public space CCTV footage can provide data. Emergency responders will also create records of their activities as they file reports, make schedules and order resources, which today frequently are entered directly into electronic devices. In addition, relevant information may be available from sources outside of the emergency response system. Journalists and others may report about events. Individuals may share photos and comments via social media. Later, insurance, police and court investigations may provide additional information. By combining information from these various sources, researchers can create a rich big data set to investigate entrepreneurial behavior in the context of responding to emergency situations, without necessarily conducting any primary data collection.

This big data set will contain information that was not collected to answer a specific academic research question. Instead, observations in big data are often collected for operational reasons. Audio and video recordings of emergency responders, for example, are recorded for monitoring, training and legal purposes. Digital devices used by emergency responders may automatically record sensor readings, data entries and data processing activities to allow for maintenance and troubleshooting by device producers. Other information, for example, location information of devices, such as phones, vehicles and tablets, might be recorded with only very general potential uses in mind.

These data-collection features, however, tend to create challenges for academic researchers interested in utilizing the recorded data for their empirical investigations. Researchers, for example, often have to integrate data stored in multiple different file formats. Different recording devices may use different recording protocols and recording protocols may change over time for operational reasons. In addition, recording protocols and their changes may not be well documented. In general, the real-time observation data at the heart of many big datasets are usually not clean and well-structured, tending to contain substantial amounts of missing or miscoded data.

Data cleanup and integration are also challenges associated with traditional large data sets, but in big data these challenges tend to be different in quality as well as in scale. Consequently, they often require fundamentally different cleanup approaches. Researchers, for example, often must spend substantial effort to make sense of what has and what has not been recorded and what the recorded information actually tells them. These efforts may include deep consistency investigations to confirm that coding patterns in the data correspond with recording protocols. Given the volume of observations, manual recoding is typically no longer feasible. Instead, researchers need to write programs for automated data re-coding and cleanup or hire people to create such programs, as off-the-shelf tools are often not available. Consequently, research teams need to be adept in using software tools and diagnosing data in order to identify and address erroneous and missing information (Braun et al., 2017).

A second major challenge is integrating data from various sources. If the same micro-event is captured by several different devices, this information needs to be linked exploiting time, location, actor and other information. Again, researchers often face similar integration challenges in traditional data sets, but with big data, the sheer volume of observations and the level of data heterogeneity and messiness tends to substantially raise the bar with regard to required integration efforts. In the case of the emergency response system outlined earlier, data source, time and location information must be carefully examined to establish linked-action sequences that correctly identify the recorded micro-actions of various emergency responders associated with the response to a specific emergency, such as a specific house fire. These data integration efforts are complicated considerably by missing data, noise and lack of overall systematic structure in the raw data.

Another relevant feature of big data is that they often contain information in the form of written messages, audio and video recordings. Obviously, these types of data imply rich opportunities for qualitative and quantitative investigations. For any quantitative investigations, however, the relatively unstructured data must be transformed into more structured data. In the past, coders read, listened and watched the source data and then categorized observations, using a researcher-provided coding scheme. The volume and velocity of big data often render such approaches unfeasible. Advancements in computer-aided text, audio and video analyses, however, have created alternatives on how to extract meaningful categories from such unstructured data. Sentiment analysis, for example, enables the automated detection of opinions and evaluations from unstructured text data (Zhang, Li & Chen, 2012; Zhang, Guo & Goes, 2013). Video analytics can provide automated facial recognition software that helps retailers to identify their customers in-store movement patterns, age, ethnicity and gender composition from in-store CCTV.[2] Community detection and social influence analyses based on phone communication records can quantify social networking patterns, which can enable the identification of terrorist and criminal networks, for example. Software to analyze text, audio and video data have been available for years. Recent rapid advances and emerging new software tools enable researchers to extract key information from unstructured data in far more efficient and reliable ways. These tools, however, remain far from perfect, and their application typically requires careful adjustments based on a deep understanding of both the raw data and the software employed.

Finally, big data can also create new challenges for human subject protection. In the past, anonymization has been an effective tool to prevent unanticipated harm for study participants. The wealth of information in big data, however, often allows "easy, cheap, [and] powerful" reidentification (Ohm, 2010: 1706). Hence, researchers should consider additional data privacy strategies, such as sharding or key coding (Guzzo et al., 2015). The informed consent and debriefing of participants often required by institutional review boards is frequently infeasible in the context of big data. The American Psychology Association (2010), however, has introduced some useful guidelines on how to address these challenges in its code of ethics (e.g., sections 8.02, 8.03 and 8.05).

In summary, the skills and effort required to prepare big data sets for data analysis are substantial. According to a 2016 CrowdFlower survey[3], big data scientists spend around 80% of

[2] https://www.cnbc.com/2017/11/23/facial-recognition-is-tracking-customers-as-they-shop-in-stores-tech-company-says.html

[3] http://visit.crowdflower.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf

their time on preparing and managing data, and, somewhat ironically, 76% of data scientists view data preparation as the least enjoyable part of their job.

**Data Analysis**

Once researchers have generated a cleaned big data set, they face the challenges associated with high volume, variety, velocity and low veracity for statistical analyses. The currently frequently used statistical significance tests, for example, suffer from sample size sensitivities which render them meaningless when dealing with thousands or millions of observations. Incidental and complex endogeneity issues often present in observational data not specifically collected for causal inference tend to create severe challenges for the application of standard regression techniques. Finally, extremely large sample sizes do not imply representative samples. Big data are not immune to problems of non-random sampling -- quite to the contrary. Hence, researchers need to remain vigilant with regard to sampling-related biases.

In response to these and other challenges, researchers in computer science, statistics and other fields are rapidly developing analytical techniques specifically geared for analyzing big data. IT companies such as Google and IBM are investing heavily in the development and promotion of related hardware and software solutions. The list of available techniques is already long and includes supervised machine learning (e.g., classification, nonparametric regression), unsupervised machine learning (e.g., clustering), spatial-temporal and other forms of data mining and new ways to visualize data, such as heat maps (Jin et al., 2015; Oswald & Putka, 2015). These emergent techniques try to not only address the analytical challenges big data imply, but also exploit benefits related to the often more fine-grained, highly-interrelated and continuous nature of big data (Gandomi & Haider, 2015). Considering the research efforts already under way, researchers should expect continuing rapid advances in big-data analytic techniques.

**Theory Development and Research Design**

Currently published quantitative entrepreneurship research is dominated by deductive hypothesis tests, and researchers tend to prefer qualitative designs for exploratory investigations. Big data, however, provide new opportunities for quantitative exploratory studies. Some software developers, consultants and their lead clients have successfully implemented big data projects focused on discovering and forecasting customer, employee and competitor behavior without much regard for testing theories or applying causal models contained in the academic literature. They have also investigated behavior, micro events and dynamic processes for which models simply did not exist. The emergence of big data has clearly stimulated renewed interest in quantitative inductive studies (e.g., Waller & Fawcett, 2013; Putka & Oswald, 2015).

Dichotomizing studies into inductive and deductive, however, might miss another opportunity here: combining inductive and deductive approaches (Kitchin, 2014a). Big data can enable iterative sequences of inductive, abductive and deductive investigations. Researchers can, for example, first employ "guided knowledge discovery techniques to identify research questions (hypotheses) worthy of further examination and testing" (Kitchin, 2014a: 6). They can use subsets of the data for these exploratory investigations. Subsequently, researchers can validate and refine these exploratory findings through deductive tests, using different subsets of the data or newly recorded data.

Exploration, however, does not have to be the starting point. Instead, researchers can start by investigating the predictive power of established causal models before engaging in inductive investigations to refine them. Theory can also guide the design of inductive investigations. Theory can, for example, inform the creation of initial dictionaries for text mining. Machine-learning algorithms can subsequently improve these initial theory-based or prior research-based dictionaries by adding and dropping terms. Hence, researchers should consider combining and alternating between exploratory investigations and deductive tests in order to iteratively develop more powerful, accurate and predictive causal models and theories (Kitchin, 2013; Shmueli, 2010). Researchers can also use predictive models to estimate counterfactuals for causal models (Varian, 2016). Some methodologists, including Tuckey (1997) and Platt (1964), have long argued for the potential of such approaches and provided related general guidance. Several initiatives have started to develop guidelines and research methodologies to support such data-intensive scientific research (Hey, Tansley, & Tolle, 2009).

## HOW TO EXPLOIT BIG DATA FOR ENTREPRENEURSHIP RESEARCH

Academic entrepreneurship research is a human endeavor. Exploiting big data opportunities requires scholars who are motivated and able to execute such studies. Hence, it requires scholars who are curious, risk-taking and interested in the next "big thing".

### What individual researchers can do

Despite broad recognition of the research opportunities associated with big data, entrepreneurship scholars have been hesitant to design and execute big data studies. In our opinion, this is no surprise, for several reasons. Big data studies require some fundamentally different research skills, skills that few entrepreneurship scholars currently possess. Hence, a fundamental "retooling" on behalf of the researchers is needed.

So, what are feasible and efficient ways for interested scholars to acquire these skills? Beyond learning related skills in workshops and seminars or from methods published in the literature, researchers should consider collaborations with researchers who have expertise in big data investigations. This may require interdisciplinary studies with researchers from fields that have made more headway in exploiting big-data-related opportunities, such as supply chain management and marketing, more recently, finance (George et al., 2014). In marketing, for example, scholars have developed expertise in analysis of big data from social media communication to investigate the attitudes and behaviors of customers. In supply chain management, researchers use big data to understand and optimize logistic systems and entire production ecosystems (Waller & Fawcett, 2013). Entrepreneurship researchers could create or join interdisciplinary big data projects and initiatives at their respective universities.

In addition, big data initiatives in various fields of science are currently under way, some of them with the explicit objective of bringing researchers from various disciplines and various universities together. In 2012, for example, the US government, through the National Science Foundation, launched the Big Data Research and Development Initiative (NITRD & NCO, 2016), which has brought together six different federal government agencies, academia, corporations and non-profit organizations to develop infrastructures and methods needed to accelerate related scientific progress (Jin et al., 2015). Hence, entrepreneurship scholars have opportunities to get involved in and benefit from big data initiatives on the local, national and international level.

## What methodology teachers and doctoral students can do

As outlined earlier, big data studies require a substantially different set of method skills and capabilities, acquisition of which requires time and systematic effort. Both tenure-track and tenured professors face severe demands on their time to satisfy their research, teaching and service obligations. Consequently, learning fundamentally new research skills constitutes a substantial challenge. In contrast, doctoral students are in a stage of their academic career that is explicitly devoted to learning research method skills. Therefore, one of the most promising approaches to developing big-data research skills in the entrepreneurship field is to integrate the learning of these skills into the curriculum of Ph.D. programs. These skills include familiarity with software tools geared for big data management and analysis (e.g., Python, R and SQL) and with various data analysis techniques, including machine learning and neural networks. These learning efforts may again require an interdisciplinary approach, with students taking method seminars in disciplines and departments with expertise in big-data analytics. Another promising approach is exposing doctoral students to online courses on MOOC platforms such as Coursera, edX, and Udacity, or other online resources (see Table 1). Senior entrepreneurship and methodology teachers will play a crucial role in motivating and guiding doctoral students in obtaining big-data research skills.

*Table 1 various online resources for data science*

| Provider | Description | URL |
|---|---|---|
| IBM Big Data University | Various data science and cognitive computing courses | https://cognitiveclass.ai/ |
| Microsoft Professional Program | Data science courses hosted on edX | https://academy.microsoft.com/en-us/professional-program/tracks/data-science/ |
| Data Science Central | Online resource collection for big data practitioners | https://www.datasciencecentral.com/ |
| KDnuggets | Online resources for big data researchers | https://www.kdnuggets.com/education/index.html |
| Sage Campus | Online data science courses for social scientists | https://campus.sagepub.com |

**What editors, associations and administrators can do**

Finally, editors and administrators are also crucial to enabling big-data research. Editors can motivate researchers to conduct big data studies by indicating interest in publishing these studies. In a "publish-or-perish" academic environment, such encouragement might prove crucial to triggering big-data research and efforts to acquire related skills. As outlined earlier, big data inherently require substantially more time and effort to clean, integrate and analyze. Hence, researchers focused on publishing quickly and frequently are discouraged from engaging in big-data research. In addition, researchers may be concerned that the use of less established research methodologies and of the same data for multiple investigations may reduce their publication chances in the top journals. Hence, editors signaling their interest to publish big data studies are likely very important to encouraging related research activities.

Similarly, academic associations should send signals of interest for big-data conference submissions and can offer workshops to teach big-data research skills. They can also support knowledge transfer and community building among big data researchers.

Finally, administrators at the university, college and department levels also play crucial roles in enabling big data research. Multiple universities have started serious disciplinary and interdisciplinary big-data initiatives (Table 2). These institutions have recognized the enormous potential of these activities for their long-term contribution to the advancement of science and to their reputations as ground-breaking research universities. More business schools and entrepreneurship departments are likely to start or join similar big-data initiatives in the future.

***Table 2 big data initiatives in universities***

| **Institution** | **Initiative** | **URL** |
| --- | --- | --- |
| U. of Michigan | Michigan Institute for Data Science | http://midas.umich.edu/ |
| UC Berkeley | Berkeley Division of Data Sciences | https://data.berkeley.edu/ |
| Johns Hopkins U. | JHU Data Science Lab | http://jhudatascience.org/ |
| UC Irvine | UCI Data Science Initiative | http://datascience.uci.edu/ |
| Iowa State University | Data Driven Science | https://www.datadrivenscience.iastate.edu/ |

Administrators are crucial not only for bringing together scholars to conduct big data studies, but also for providing the necessary resources. In the case of big data, the necessary computer and data processing capabilities go far beyond what is typically provided to business school faculty members. Creating and maintaining the needed infrastructure is expensive, because big data researchers often need expensive hardware and software. At the authors' home university, for example, it costs $30,000 upfront to rent a capable "compute node" in the university's high-performance computing infrastructure to store and process a big data set, plus other possible maintenance costs. An alternative is to run a virtual server in Amazon EC2, which costs $200 to $800 dollars every month. Big data researchers may also need funds to hire help to cleanup and integrate the data. Hence, supportive and visionary administrators are crucial to enabling big data research.

## CONCLUSION

The topic of big data is very timely for both academics and practitioners. Even after discounting some of the recent hype created by information technology firms with vested commercial interests in selling their products and services (Gandomi & Haider, 2015), the immense potential opportunities of big data for empirical entrepreneurship research are hard to deny. The earlier outlined example of big data capturing the activities of emergency responders, for example, presents fascinating new opportunities to investigate entrepreneurial learning behavior, such as experimentation, bricolage and pivoting in response to a variety of routine or non-routine events. It allows dynamic modeling and process studies of these phenomena. Incubators, research parks, software development companies, and other existing organizations and communities represent settings where the capturing and mining of micro-activities and events has created unprecedented opportunities for investigating entrepreneurial behavior. This new 'big data' world presents both opportunities and challenges. Exploiting related opportunities requires new research skills and capabilities. As things stand, entrepreneurship researchers have to develop expertise in big data analysis. In addition, universities need to provide the computational infrastructure needed for conducting big data studies. These adjustments are necessary to catch up with a rapidly digitalized world and to exploit related research opportunities. In other words, the entrepreneurship research community would benefit from a healthy dose of related scholarly entrepreneurial spirit and initiative.

## REFERENCES


Allwinkle, S., & Cruickshank, P. 2011. Creating smarter cities: An overview. ***Journal of Urban Technology***, 18(2): 1-16.

American Psychological Association. 2010. ***Publication manual of the American Psychological Association*** (6th ed.). Washington, DC: American Psychological Association.

Braun, M. T., Kuljanin, G., & DeShon, R. P. 2017. Special considerations for the acquisition and wrangling of Big Data. ***Organizational Research Methods***. Published online February 28, 2017; DOI: 1094428117690235.

Diesner, J., Frantz, T. L., & Carley, K. M. 2005. Communication networks from the Enron email corpus "it's always about the people. Enron is no different". ***Computational and Mathematical Organization Theory***, 11(3), 201-228.

Gandomi, A., & Haider, M. 2015. Beyond the hype: Big data concepts, methods, and analytics. ***International Journal of Information Management***, 35(2): 137-144.

George, G., Haas, M. R., & Pentland, A. 2014. Big data and management. ***Academy of Management Journal***, 57(2): 321-326.

Goldberg, A., Srivastava, S. B., Manian, V. G., Monroe, W., & Potts, C. 2016. Fitting in or standing out? The tradeoffs of structural and cultural embeddedness. ***American Sociological Review***, 81(6): 1190-1222.

Guzzo, R. A., Fink, A. A., King, E., Tonidandel, S., & Landis, R. S. 2015. Big data recommendations for industrial–organizational psychology. ***Industrial and Organizational Psychology***, 8(4): 491-508.

Hey, T., Tansley, S., & Tolle, K. M. 2009. ***The fourth paradigm: Data-intensive scientific discovery***. Redmond, WA: Microsoft Research.

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. 2015. Significance and challenges of big data research. ***Big Data Research***, 2(2): 59-64.

Kitchin, R. 2013. Big data and human geography: Opportunities, challenges and risks. ***Dialogues in Human Geography***, 3(3): 262-267.

Kitchin, R. 2014a. Big Data, new epistemologies and paradigm shifts. ***Big Data & Society***, 1(1): 1-12.

Kitchin, R. 2014b. The real-time city? Big data and smart urbanism. ***GeoJournal***, 79(1): 1-14.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. ***Harvard Business Review***, 90(10): 60-66.

NITRD, & NCO. 2016. The Federal Big Data Research and Development Strategic Plan: Accessed April 5, 2017 at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/bigdatardstrategicplan-nitrd_final-051916.pdf.

Ohm, P. 2010. Broken promises of privacy: Responding to the surprising failure of anonymization. ***UCLA Law Review***, 57: 1701-1711.

Olguín, D. O., & Pentland, A. 2010. Assessing group performance from collective behavior. ***Paper presented at CSCW-2010 Workshop on Collective Intelligence in Organizations: Towards a Research Agenda***. Savannah, GA. Retrieved April 23, 2017 at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.481.3448&rep=rep1&type=pdf.

Oswald, F. L., & Putka, D. J. 2015. Statistical methods for Big Data: A scenic tour. In S. Tonidandel, E. King & J. M. Cortina (Eds.), ***Big data at work: Data science revolution and organizational psychology*** (pp. 43-63). New York: Routledge.

Pentland, A. 2012. The new science of building great teams. ***Harvard Business Review***, 90(4): 60-69.

Putka, D. J., & Oswald, F. L. 2015. Implications of the big data movement for the advancement of IO science and practice. S. Tonidandel, E. King, and J. M. Cortina (eds), ***Big data at work: The data science revolution and organizational psychology*** (pp. 181-212). New York: Routledge.

Platt, J. 1964. Strong Inference. **Science**, 146(3642): 347-353.

Shmueli, G. 2010. To explain or to predict? ***Statistical Science***, 25(3): 289-310.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. 2017. Critical analysis of Big Data challenges and analytical methods. ***Journal of Business Research***, 70: 263-286.

Smolan, R., & Erwitt, J. 2012. ***The human face of big data.*** Sausalito, CA: Against All Odds Production.

Tonidandel, S., King, E. B., & Cortina, J. M. 2016. Big Data methods leveraging modern data analytic techniques to build organizational science. ***Organizational Research Methods***, Published online Nov 16, 2016, DOI: 1094428116677299.

Tripathi, P., & Burleson, W. 2012. ***Predicting creativity in the wild: Experience sample and sociometric modeling of teams***. Paper presented at the ACM 2012 conference on Computer Supported Cooperative Work.

Tukey, J. 1997. More honest foundations for data analysis. **Journal of Statistical Planning and Inference**, 57(1): 21-28.

Varian, H. R. 2016. Causal inference in economics and marketing. ***Proceedings of the National Academy of Sciences***, 113(27): 7310-7315.

Waller, M. A., & Fawcett, S. E. 2013. Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. **Journal of Business Logistics**, 34(2): 77-84.

Wetherbe, J., & Zhang, Z. 2015. Data analytics help entrepreneurs decide where to boldly go. ***Entrepreneur & Innovation Exchange***. Published online January 2015, https://eiexchange.com/content/64-data-analytics-help-entrepreneurs-decide-where-to-boldly-go.

Zhang, Z., Guo, C., & Góes, P.B. 2013. Product comparison networks for competitive analysis of online word-of-mouth. ***ACM Transactions on Management Information Systems***, 3(20):1-20:22.

Zhang, Z., Li, X., & Chen, Y. 2012. Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews. ***ACM Transactions on Management Information Systems***, 3(1): 5.

**Author Biographies**

**Andreas Schwab** is an Associate Professor and Dean's Fellow in the Ivy College of Business at Iowa State University and a visiting senior Fulbright scholar in the School of Business and Management at the Institute of Technology Bandung. He studies organizational learning processes in dynamic environments, such as project ventures and temporary organizations. He is also engaged in several initiatives to advance methodological practices in the management field by promoting alternatives to statistical significance tests and introducing scholars to Bayesian statistics. His research has been published in the *Academy of Management Journal*, *Organization Science*, *Strategic Organization*, *Entrepreneurship Theory & Practice*, *Academy of Management Learning and Education*, *Industrial and Corporate Change*, *Industrial Relations*, *Family Business Review*, *Group and Organization Management* and other outlets. He serves on the editorial boards of *Entrepreneurship Theory & Practice*, *Organization Research Methods*, *Organization Science*, *Group and Organization Management* and the *Strategic Entrepreneurship Journal*. He received his Ph.D. in management and organization theory from the University of Wisconsin - Madison.

**Zhu Zhang** is an Associate Professor of Information Systems and Kingland Business Analytics Fellow in the Ivy College of Business at Iowa State University. He is a passionate researcher, teacher, and practitioner in the areas of big data analytics and artificial intelligence. His intellectual aspiration is to discover valuable knowledge and intelligence from massive amount of data available in various structures, genres and streams. In particular, his research focuses on computational models for textual data and their applications in various domains (e.g., business and healthcare). His research has been published in *Management Information Systems Quarterly*, *Journal of Management Information Systems*, *ACM Transactions on Management Information Systems*, *IEEE Transactions on Knowledge and Data Engineering*, and other outlets. He received his Ph.D. in information and computer science from the University of Michigan.