

Big and Open Linked Data (BOLD) in research, policy, and practice

Marijn Janssen & George Kuk

To cite this article: Marijn Janssen & George Kuk (2016) Big and Open Linked Data (BOLD) in research, policy, and practice, Journal of Organizational Computing and Electronic Commerce, 26:1-2, 3-13, DOI: [10.1080/10919392.2015.1124005](https://doi.org/10.1080/10919392.2015.1124005)

To link to this article: <https://doi.org/10.1080/10919392.2015.1124005>



© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 07 Mar 2016.



Submit your article to this journal [↗](#)



Article views: 2924



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 22 View citing articles [↗](#)

Big and Open Linked Data (BOLD) in research, policy, and practice

Marijn Janssen^a and George Kuk^b

^aFaculty of Technology, Policy, and Management, Delft University of Technology, Delft, the Netherlands;

^bNottingham Business School, Nottingham Trent University, Nottingham, United Kingdom

ABSTRACT

The value of data as a new economic asset class is seldom realized on its own. With less reliance on self-administered survey, it offers new insights into behaviors and patterns. Yet, it involves a huge undertaking of bringing together multiple actors from different disciplines and diverse practices to examine the underexplored relationships between types of data. There are different inquiry systems and research cycles to make sense out of big and open data (BOLD). We argue that deploying theories from diverse disciplines, and considering using different inquiry systems and research cycles, offers a more disciplined and robust methodological approach. This allows us to break through the limits of backward induction from the evidence by moving back and forward in exploring the unknown through BOLD. As such, we call for developing a variety of rigorous approaches to counter-balance the current theory-free practice in the analysis and use of BOLD.

KEYWORDS

Big data; BOLD; dataset; e-government; inquiry systems; internet of things; open data; policy-making; practice; theory

1. Introduction

Despite the hype of big data as a new economic asset class, the research opportunities are less understood in academic scholarship and discussion. In particular, the relationships between Big and Open Linked Data (BOLD) for use by policymakers and researchers are little explored. Big *data* is typically tantamount to amassing a large (consolidated) dataset (Wigan and Clarke 2013) and, yet, it can remain proprietary. Whereas *open data* enables access for everybody to data without any predefined restrictions or conditions of use, *big data* is about large volumes of data from a variety of sources (Janssen, Matheus, and Zuiderwijk 2015) and linked data is about connecting structured and machine-readable data that can be semantically queried (Bizer, Heath, and Berners-Lee 2009). Research in open data has shown that quality, rather than the quantity, of data matters for service and digital innovation (Kuk and Davies 2011). The exploration of linking big and open data in value creation requires combinations of data from different data sources (Janssen, Estevez, and Janowski 2014). Although there are anecdotal examples, we need further research to examine why and how organizations can generate value from big and open data and which approaches should be followed.

The use of BOLD is often to harness current services and systems. For example, BOLD can be used to improve fraud detection by customs and tax organizations (Klievink and Zomer 2015). The unprecedented increase in the availability and specificity of BOLD can enable the creation of new insights and applications. Yet, with the continuous shift toward data exploitation, we need to rebalance this with a more data-exploration strategy in a new field of data science and research, aiming to create a transdisciplinary research approach and perspective. Transdisciplinary refers to the collaboration of people from different disciplines to create value out of BOLD. There is a need to for frameworks to stimulate thinking about nature, role, and future of data in relation to business analytics (Holsapple, Lee-Post, and Pakath 2014).

CONTACT Marijn Janssen  m.f.w.h.janssen@tudelft.nl,  Faculty of Technology, Policy and Management, Delft University of Technology, Building 31, Jaffalaan 5, 2628 BX Delft, the Netherlands.

© 2016 The Author(s). Published by Taylor & Francis

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Rather than incrementally refining and integrating the new insights in what we already know and do, BOLD can offer some unusual inflection points for new insights and understanding. An illustration is that combining data sources might result in the discovery of previously hidden patterns and the revelation of new insights. For example, the Dutch Tax organization found that persons who were divorcing have a bigger chance to make mistakes in their tax applications. This has resulted in implementing remedial measures to mitigate the occurrence of mistakes (<https://decorrespondent.nl/2720/Baas-Belastingdienst-over-Big-Data-Mijn-missie-is-gedragsvrandering/83656320-f6e78aaf>). Policymakers typically want to have insights about such behavior to improve their policies. Also, for researchers, such insights are of interest as they can advance our understanding of human behavior.

BOLD refers to a diversity of data that needs to be combined to generate new insights. BOLD provides policymakers and researchers with direct measurements and more factual data in comparison to self-administrated surveys. BOLD can be used to explore new applications and transform current practices and processes in various fields, including policymaking, service provisioning, inspection, and enforcement. Yet, the availability of data can result in neglecting theories and become data-led (Anderson 2008). We argue that the role of theorizing is crucial, in addition to the fact that BOLD allows us to focus on exceptions and gives us better insight. Nevertheless, BOLD enables a combined use of inductive, abductive, and deductive approaches in our research enquiries. In the next section, we investigate the characteristics of data and how data are used to make inferences and draw conclusions about reality. This is followed by a discussion about the impact of BOLD on the way we make sense of the world. Finally, we provide an overview of inquiry systems and research cycles that can offer researchers alternative ways of theorizing. We argue that rigorous and solid research approaches should be followed when dealing with BOLD, approaches in which the limitations of BOLD are recognized.

2. BOLD is diverse

All too often “data” is viewed as a homogenous concept; however, BOLD may take various forms and can be collected from many sources. As a result, BOLD comes with distinct known and unknown qualities. The data variety suggests that research should harness the specificity of certain types of data in the contexts of its provenance and potential limits. Although much literature at the early stages focused on the generic and general use of open data, specific focus should be given toward the nature and the intrinsic facets of data characteristics that are underexplored. [Figure 1](#) depicts four main data characteristics between degree of structuredness and openness in the data. The current distinction between types of data is implicitly assumed. For instance, closed data might be opened and open data might be combined with proprietary data in algorithmic processing. Also, initially unstructured data can become structured through meta-data tagging at some stage. The linking of data sources plays a crucial role in combining datasets for creating more insight.

Most of the data policy of democratic governments is through the route of open data. Recently, we have witnessed an increase in a similar practice in some of the businesses through a combination of a semi-open and proprietary route toward data. Some of the current accounting and information-based systems require data to be structured to enable processing the data. For example, by tagging faces in a picture (unstructured), the tags represent the identities of the people and form the metadata (structured). Furthermore, the semantic linking of datasets makes it easier to discover the value of datasets. These two developments result in having access to more data and processing the data automatically.

Data can be collected and used by a variety of organizations. These organizations also determine how and with whom data are shared. Most governments have a policy to open raw data whenever possible to ensure reuse. Nevertheless, a lot of data are still being locked away, as the process of cleaning up the data is tedious and resource-intensive. Some companies have followed the open data route, aiming to attract others through bringing different users groups together and to build

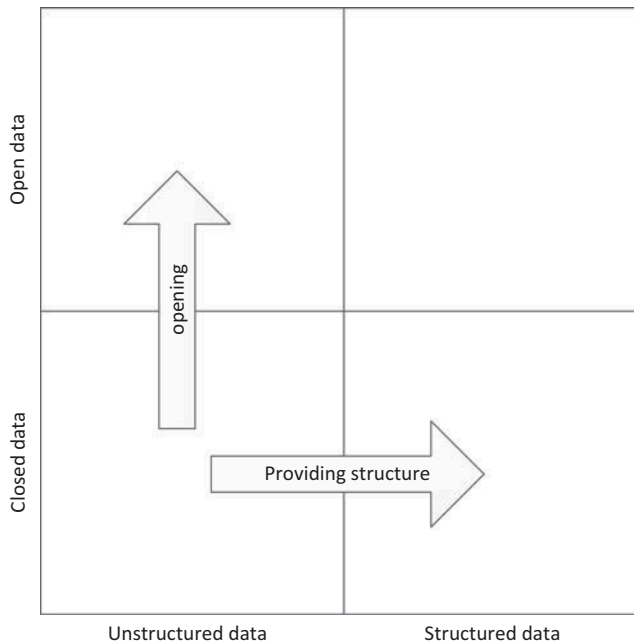


Figure 1. Overview of BOLD areas.

complementary services. For example, Nike has used application programming interfaces (APIs) to make their data available, including detailed information about the source of data quality scores to decrease the environmental impact of data (<http://www.smartdatacollective.com/daniel-castro/137216/how-nike-using-data-help-save-planet>). The opening of data provides an opportunity for creating new business models, which is addressed by the paper “A Value-Centric Business Model Framework for Managing Open Data Applications” included in this special issue.

Data are not always opened for all. Some organizations do not share their data or only share their data as a club good within a small community. The latter is well known in the scientific world, as data are only shared with other researchers in the community who have to agree on a code of conduct to prevent data misuse. The data might be privacy-sensitive or interpretation might be difficult, in that only partial data are made available. This may be problematic, as the partial data may not be randomly sampled and subject to sampling bias. Consequentially, the use of BOLD might result in bias. As such it is necessary to have insight into how the data are collected to understand limitations and to determine what value can be generated from the data.

3. BOLD can provide better and more factual data

Increasingly, various areas of our daily activities are being digitally observed and stored, which potentially offer a basis for deeper insight and understanding of human behavior. With traditional research approaches, actual behaviors are seldom known. Many researchers have to adopt a positivism approach through self-reporting instruments, such as a survey, to collect perceived intention and action. These approaches have to deal with respondent bias, which originates from the inability or unwillingness of the respondents to provide accurate or honest answers. Overly enthusiastic respondents might guess their answers in order to fill in questionnaires. Others simply have no knowledge of the topics or are unaware of what they are doing. An enthusiastic respondent is willing to provide a guess as answers in order to “help” the study, whereas others are simply unaware of their ignorance on the survey’s topic. Respondent bias created by the unwillingness to

provide honest answers stems from the participant's natural desire to provide socially acceptable answers in order to avoid embarrassment or to please the organization conducting the study. This phenomenon is widely known as social desirability bias.

Whereas survey questions are subject to limits of attentional and cognitive resources of the respondents, automatic data collection does not suffer from the problem. The availability of data makes it possible to focus on actual behaviors. People often share their location data with their friends, publicly expressing their sentiments and photographs on social media and, with user consent and the use of data obfuscation, the publicly available data can make sampling somewhat redundant. As such, this reduces various biases, including nonresponse bias. However, data coverage is practically challenging; to collect data of our entire population raises feasibility questions, but also may be unethical without due care in obfuscating the data to ensure privacy.

Data bias can result in the inability to replicate studies and compromising the generalizability. The level of abstraction of the results obtained determines the generalizability. A too-high level of abstraction may be at the expense of real insights and practical value, and might result in a bigger gap between scientific rigor and practical relevance. The availability of data might influence theorizing. We need theories to have plausible explanations for these behaviors. The availability of more data from multiple sources will also enable the development of richer models and advance our understanding. Whereas traditional means of data collection are subject to limits of the instruments, the data-rich model can provide a deeper level of insight into actual behavior, which was not feasible in prior research. Each model is a reduction of reality and the modeler needs to make choices in light of limited resources for data collection and modeling. When more data are available models can become more complex and too detailed to interpret.

The use of BOLD is often tied to evidence-based policymaking (see for example Daniell, Morton, and Ríos Insua 2015; Ferro et al. 2013). With less dependent on opinions and subjective assessment, it enhances evidence-based policymaking through subjecting policies and models to rigorous testing of their underlying assumptions and predictions. Two papers published in this special issue address this topic (“The Role of e-Participation and Open Data in Evidence-Based Policy Decision Making in Local Government” and “Big Data in the Policy Cycle: Policy Decision Making in the Digital Era”).

BOLD can provide insight into real behavior instead of perceived behavior. The digital and contextual footprint is used, rather than perception and intended behaviors. They can reveal insights into the differences between reality and perception. Although BOLD might provide more accurate and factual data, the data still need to be interpreted and processed.

4. Data as an abstraction of reality

As governments, organizations, and persons are making their private and personal data public, other forms of data are also collected through sensors, such as the Internet of Things (IoT). Sensors and actuators are used for the ubiquitous sensing enabling ability to measure (Gubbi et al. 2013). Data should be placed in the context to enable interpretation; only then do data becomes information (Ackoff 1989). Data can be defined as symbols, the products of observation (Ackoff 1989; Rowley 2007). In contrast, data need to be processed to create useful information. Information provides answers to who, what, where, and when questions (Ackoff 1989). Only when data are used do they become information.

A major risk is that data become reality. For example, inspectors might look only at the data to inspect a good and, if the data do not show any anomalies, the good will not be inspected. Nevertheless, there might be data misuse, deliberate bias, data omission, and data manipulation, which can only be detected when looking at the reality. As such, it is essential to validate the results obtained with the data in reality. Figure 2 shows the cycle in which data are collected from reality, then published, analyzed, interpreted, and used to make inferences about reality. All of these steps should be related to reality. Analyses should be validated by looking, not only at the data, but also at reality. Interpretation of data requires deep domain expertise about the situation that is interpreted. Finally, the conclusions result in

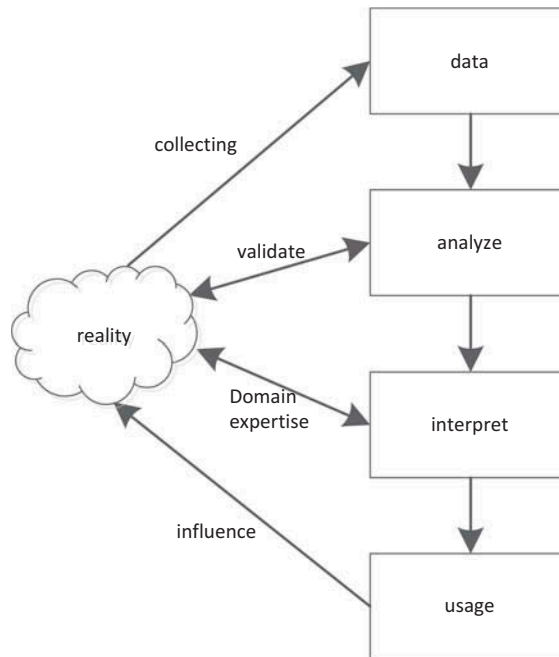


Figure 2. Relating data and reality.

actions to influence reality and even might result in changes. Data may have various information qualities, which can easily blur the reality. Also, data can give only an incomplete picture, as not everything is measured that might be of relevance. The context in which data are collected might prevent its use in another context (Janssen, Estevez, and Janowski 2014).

Data become valuable when it is combined with investments in data verticals including algorithms, infrastructures, multimodal devices, and services (Yoo et al. 2010), but also when investments are made in structuring and linking datasets. The use of open data can be viewed as a social construction process in which many actors interact with each other. A transdisciplinary approach is necessary in which domain experts, data scientists, social scientists, and so on work together. There are many statistical methods and analytic methods to analyze data. Often, several methods are employed to analyze data, different data sets might be used and other data left out, and so on before new insights are gained. This process requires interpretation of the data and understanding the limits of what can (not) be inferred. As models are abstractions of reality, in a similar vein, data are abstractions based on measured observations of reality. The vast amount of data bring reality and the world, as captured in models and data, closer together. Having better data about reality enables us to improve our theorizing.

The digital and contextual footprint requires a different set of social algorithms that size up, evaluate what we want, and provide customized experiences (Lazer 2015). This shifts the focus from what data are collected and used to how data are processed. For example, the US travel platform Trip Advisor has changed not only the ways we book and travel but also the ways hoteliers operate (Orlikowski and Scott 2014). However, although user-contributed reviews are public, its relationship with the ranking algorithm is not known. Public organizations can explore this data relationship with algorithms, seeking to make not only data open but also the existing algorithms transparent. With initiatives such as “do not track W3 C initiative,” we have witnessed a growth of technologies that allow users to verify results by social algorithms (e.g., Google’s AdChoices system) and, notably, how organizations involve users to co-curate their experiences through social algorithms.

Business communities and practices will definitely need our theoretical inputs to better inform their inductive approach (data exploration strategy) and, likewise, there is much that we can learn from them. So, we envisage that practices in academia and governments can both use BOLD to learn from each other and develop a more transdisciplinary research agenda.

5. Just finding patterns is not sufficient

In the article “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete,” Anderson (2008) argued that the vast amount of data and processing capacity availability will bypass the “hypothesize, model, test” model in science, because scientific theorizing simply cannot cope with the deluge of data. His argument was that inquiries can result in useful models that cannot be expressed analytically. Searching for patterns and exceptions in large datasets is a new and increasingly useful application in BOLD. Nevertheless, there has been much criticism on the thought written in this article. Data without interpretation are useless. Without any doubt, BOLD will contribute to our observation and the data will be used to derive patterns and bypass the making hypothesis; however, these new patterns can result in the formulating of hypotheses that can be tested with other data. Models can be derived from BOLD to gain insight and make progress based on what we have learned; only then, is understanding in the relationship between data gained. Using data analysis, the Dutch tax organization discovered the hidden knowledge (recall the earlier example). But, without models, no interpretation would be possible, no causes could be found, and no actions could be taken to improve the quality of tax filing.

Harford (2014) provided the example of flu prediction to make a plea for theories for explaining cause and effects. Google predicted flu outbreaks for several winters based on the terms people used to search. However, in one season, the prediction was not right anymore. This is because there is only a correlation between search terms and flu outbreak, and there is no causality. This shows that a theory-free analysis and looking merely at correlations eventually might not be correct. Furthermore, such a practice does not improve our understanding of underlying factors that result in the flu outbreak.

There is a difference between policymaking and science in this respect. Both are about finding patterns and using them for predictive purposes. Policymaking is focused on using the data for the situation at hand, whereas science focuses on the explanation, which involves some kind of causality, and on generalization of the findings to other situations. Gregor (2006) argued that every theory contains the following elements

- *Generalization*: Abstraction and generalizations from one situation to another situations are a key aspect of any theory
- *Causality*: Causality is the relationship between cause and event and necessary for
- *Explanation and Prediction*: Explanation is closely linked to human understanding; predictions allow the theory to be both tested and used to guide action.

Just mining data and looking for patterns is not be sufficient to label something as science. The selection, measurement error, and other sources of bias should be taken into account, which might block the use of readily available data for these purposes. A key aspect remains the design of the research or policy. As such, data might be less suitable for theory testing, but might be suitable for theorizing.

6. BOLD can provide the counterfactuals to what we do and know in practice

Data analytics inquiries are often used for predictive purposes, whereas the exceptions are often neglected. Model building ignores and minimizes the effect of events that are outside model, and

data cleansing efforts are often focused on removing these type of data. People have a tendency to extrapolate what they already know and to neglect signals that might result in a complete change. BOLD enables discovery of those signals that are hidden and not easy to see. This is like the black swan theory, which looks at unpredictable events that have a major impact (Taleb 2007). Black swans were viewed as impossible until, in 1697, a Dutch discoverer found them in Australia.

Instead of seeking to confirm theory using deductive approaches, BOLD provides the opportunity to look for counterfactuals, rather than gathering a representative sample of data to confirm our theory. Rather than looking for patterns and confirming our predictions, BOLD can be used to seek the unknown. This follows the Bayesian thinking of ascertaining prior and likelihoods of false positives and negatives of prior distribution and likelihoods. In *Bayes Theorem*, the probability of an event occurring is based on conditions related to the event. The Bayes theorem can move away from the limits of backward induction from the evidence to a more iterative approach of moving back and forward in exploring the unknown. In this way, a more fine-grained and detailed estimation can be gained. BOLD offers a way to not only test our beliefs and presuppositions and biases but also recalibrate our beliefs based on the data.

The common use of Bayes rules in business analytics is to update the prior distribution to calculate the posterior distribution of our predictive models, which involves an incremental update of the coefficients of our models. For instance, a public organization can filter emails based on some prior beliefs that the thresholds of probabilities of certain keywords will signify spam mails, and this will form the prior distribution. Together with users' input, those probabilities will be updated using the Bayes formula to recalculate the posterior probabilities for those keywords. The spam filter and rules become more accurate over time. However, the involvement of users to co-curate the underlying algorithm of SPAM filtering will require other theoretical models to better understand how data can transform our understanding of the intricate relationship between data and technology and the social practices. This provides a fruitful avenue for research on materials and materiality in Science and Technology Studies, Information Systems, and Software Studies.

7. Inquiry systems

The previous sections suggest that analyzing data and drawing conclusions from BOLD requires a sound inquiry approach. A variety of inquiry systems and research approaches to make sense out of data exist. An inquiry system is a process that is aimed at solving a problem and creating knowledge. No single best approach exists and the approach taken is dependent on such elements as the problem at hand, research objectives, type and quality of data, and people involved. Churchman (1971) identified five archetypal types of inquiring systems. Policymakers and researchers can use them to understand their actions when pursuing their research. Churchman describes these archetypes as:

- A *Leibnizian* inquiring system is a closed system with a set of built-in elementary axioms that are used along with formal logics to generate more general fact nets or tautologies.
- *Lockean* reasoning is experimental and consensual. Empirical information, gathered from external observations, is used inductively to build a representation of the world.
- A *Kantian* system is a mixture of the Leibnizian and Lockian approaches and generates hypothesis based on tacit and explicit knowledge. This system contains both theoretical and empirical components in which the theoretical component allows an input to be subjected to alternative interpretations.
- A *Hegelian* inquiry system's inquire function is based on dialectic process wherein knowledge is created from conflicting ideas.
- A *Singerian* system is based on disagreements and gradually expanding and adapting knowledge to create agreement. When models fail to explain a phenomenon, new variables and laws are "swept in" to provide guidance and overcome inconsistencies.

These inquiry system provides an overview of distinct ways for gathering evidence and building models to represent a view of the world (Mason and Mitroff 1973). The co-existence of different inquiry systems suggests that different inquiry systems can be used to analyze BOLD.

8. Understanding data research cycles

Even in the big data and IoT era, not all data will be collected automatically. Most data are collected in the context of having a certain objective in mind and cannot be (easily) used for other purposes. For example, if people think that data are not purposeful, it is likely that the data will not be collected. If we want to improve a supply chain, the systems containing such data will be investigated. If we want to know about the adoption of a new information system, we search for this data. But, whereas the first type of data might be collected by supply chain management systems operating the supply chain, the second still requires a research design to observe and see what is happening.

BOLD sets can be approached from the three principal modes of inference: deduction, induction, and abduction as shown in Table 1. A researcher having a certain theory might want to use the data to test hypotheses resulting predictions in certain situations. These predictions can then be tested using the data. This is the deductive research cycle. At the same time, data can be used as deduction from an abstraction from empirical situations and deductively hypotheses can be formulated resulting in theory development. This theory can be subsequently tested and evaluated. Such a strategy is often based on the Leibnizian inquiring system based on a set of elementary axioms that are used along with formal logic to generate more general fact nets or tautologies (Churchman 1971).

Inductive research combines theory and practice and adopts existing problems by an inductive-hypothetic research strategy (Churchman 1971; Sol 1982). Such a strategy is based on a Singerian inquiry system by expanding and adapting knowledge (Churchman 1971). The hypotheses result in predictions for certain situations. These predictions can then be tested using the data.

Looking for patterns is abduction, which is sometimes viewed as a type of guessing. In abduction, premises do not guarantee the conclusion. Abduction is about a logical “guess” of a concept theory based on a surprising observation or an unusual phenomenon. Often these observations are not compatible with our present theories. Abduction results in informed guesses, after which, deduction can explicate the guesses and induction can be used to evaluate these.

BOLD does not automatically mean deductive, inductive, or abductive. The start can be a theory that is used to analyze data and a deductive cycle can be followed. Using a deductive method should provide careful attention to the way the data are collected. The data might not be a sample and the way it is measured might create already a certain view on reality, which can be blurred.

BOLD can be viewed as empirical measures from reality. As such, it can be argued that data-driven is derived from observations and consequently inductive. Yet reality is not observed directly, but it is already abstracted in some data. The data might not have been collected for this purpose. This means that observation is made on some derivative of reality.

By looking at exceptions in the data, an abductive research cycle might be followed. Guesses and theories might be created and additional analysis might be conducted. It might be necessary to analyze the exception into details and go back to the reality in which the data were collected. Furthermore, analysis is necessary to determine whether the exception cannot be attributed to such factors as measurement mistakes.

Table 1. Data driven research cycles.

Deductive cycle	Inductive cycle	Abductive cycle
1. Existing theories from literature	1. Observation	1. Finding anomalies, exceptions and patterns
2. Hypothesis	2. Induction	2. Hypothesis formulation
3. Observation	3. Theory development	3. Theory testing
4. Confirmation/rejection	4. Theory testing	4. Evaluation
	5. Evaluation	

Other typical challenges in data-driven research approaches include being aware that different datasets should be used for theory formulation and theory testing. Furthermore, as data are abstractions of reality, it is imperative to have knowledge about the reality, and about how data are collected, under which circumstances, and from which population. Data might simply not be suitable for theory testing, but might be suitable for abduction and theorizing. For theory testing, experiments might need to be designed and the setting of conditions for data collecting might be necessary. We argue that data-driven research is not new in this sense and can follow the existing ways of doing research, but we plea for careful consideration of which approach is followed, being aware of the needs that such an approach requires and the limits. The integration of insights from the data analytics and science field can help to understand the limitations of BOLD.

9. Article overview

The seven articles in the special issue illustrate the diversity of aspects in the big and open linked data field, show the possible impacts of open data, and highlight some of the main development in this field.

BOLD is a relatively new field of research that has been given attention only recently. The number of articles in this field has increased significantly over the past few years. The first article of this special issue is titled “State of the Art Review of Open Data Research: Insights from Existing Literature and a Research Agenda.” In this article Hossain, Dwivedi, and Rana review the literature and create a research agenda for the open data field. The theories used and models developed are evaluated, and the most productive journals, authors, and institutions are analyzed. The authors plea for more research in the domains of open data behavioral models, the value of open data, the misuse of open data, and legal and ethical implications.

The second article, “A Taxonomy of Open Government Data Research Areas and Topics” co-authored by Charalabidis, Alexopoulos, and Loukis, aims to understand the emerging field of open data by developing a taxonomy of research areas and corresponding research topics. The authors base their insights on policy documents, expert input, and the research literature. The resulting taxonomy is created as open data and can be accessed using <http://mind42.com/public/f2a7c2f6-63ec-475f-a848-7ed5abe6c5a4>. The authors introduce a life cycle consisting of nine stages (create, preprocess, curate, store/obtain, publish, retrieve/acquire, process, use and collaborate). The taxonomy provides insight into the inherent complexity of the opening of government data and the creation of value from it. The mapping of literature on this taxonomy shows that in some areas hardly any research is conducted.

The availability of BOLD affects decision-making and participation in decision-making. Sivarajah, Weerakkody Waller, Lee, Irani, Choi, Morgan, and Glikman address this topic in their article “The Role of e-Participation and Open Data in Evidence-Based Policy Decision Making in Local Government.” An e-participation platform utilizing open data is investigated and the authors find that the use of open data for policymaking is a complex and challenging undertaking. Nevertheless, the benefits can be high as the use of open data might result in more evidence-based and transparent decision-making. Data visualization is found to be a key condition for enhancing engagement between civil society and local government authorities.

Yu focusses on the development of suitable open data business models in his article titled “A Value-Centric Business Model Framework for Managing Open Data Applications.” Value creation using open data-based business models is not well understood in practice and theory. In this article, an open data value-centric business model framework is developed. This framework can guide the development and evaluation of new operational business models and covers value identification, proposition, creation, and assessment.

In the fourth article “Improving the Speed and Ease of Open Data Use through Metadata, Interaction Mechanisms and Quality Indicators” Zuiderwijk, Janssen, and Sussha employ a design science approach to develop an open data infrastructure. They argue that such an open data

infrastructure should support searching, analyzing, visualizing, discussing, providing feedback on, and assessing the quality of open data. Using quasi-experiments, the authors show that metadata, interaction mechanisms, and data quality indicators contribute to making open graph data use easier and faster, and enhance the user experience.

The article “Big Data in the Policy Cycle: Policy Decision Making in the Digital Era,” authored by Höchtl, Parycek, and Schoellhammer, focuses on the interdependency between technological and political change. The authors take the policy cycle as a starting point and identify opportunities and challenges associated with big data analytics in government. They find that technological advances can reduce the time frame and increase the evidence base for policy decisions.

The final article, co-authored by Sandoval-Almazan and Gil-Garcia, titled “Towards an Integrative Assessment of Open Government: Proposing Conceptual Lenses and Practical Components,” focuses on measuring elements of open government to provide guidance for further improvement. The authors introduce an integrative model capturing both practically relevant and theoretically supported variables. Open government is an evolving area and as such measurements need to be continuously adapted to evolve with the developments.

The articles in this special issue give a substantial overview of the state of the art in this domain and, at the same time, contribute to advancing our knowledge in this emergent field. In particular, we emphasize the need for theory development and changing the methods for performing theory-driven quantitative studies in this field.

10. Conclusions

More and more data are created that can be used for a variety of purposes ranging from surveillance to government policymaking. Furthermore, the availability, opening, and linking of data provides ample opportunities for policymakers and researchers. The vast amount of data brings reality and the world, as captured in models and data, closer together. However, BOLD are outcomes of measuring reality and should not be confused with reality. The use of data requires insight into how the data are collected and its qualities. Always, the limitations should be considered when inferring.

The use of BOLD can result in a need to rely less on subjective, often survey-administrated, data. The memory of people might be biased and BOLD provides factual data instead of depending on the memory of people and how they want to be seen. Data directly observing behavior can be collected instead of having to rely on self-administrated survey data. Using the Bayes theorem, we can move away from the limits of backward induction from the evidence to gain fine-grained and detailed estimations. Furthermore, new algorithms can be developed to make use of the data. BOLD will change the way governments operate and the ways policymakers do research.

Systematic approaches are necessary to deal with BOLD, which should take into account the context in which the data are collected. Theories for understanding the relationships among data are key ingredients for conducting research in this area. Having better data about reality enables us to improve our theorizing and to deepen our insight. Once the data become available, and can be combined, a main challenge will be to follow a sound inquiry approach. Without having any knowledge about the context in which the data are collected, the limitations of what can be done with the data are not clear. Collaboration of people from different disciplines to understand the data is often necessary. Focusing on data only, without any theorizing, might result in correlations that are mixed up with cause-effect. Different inquiry systems can be followed and inductive, abductive, or deductive methods may be followed. At the same time, practice needs to be informed by theory to avoid making mistakes and drawing wrong conclusions. We plea for rigor in addressing BOLD and for taking the appropriate approach that best fits the problem under investigation and, at the same time, recognizing the limitation originating from the data and the approach taken.

References

- Ackoff, R. L. 1989. From data to wisdom. *Journal of Applied Systems Analysis* 16(1):3–9.
- Anderson, C. 2008. The end of theory: The data deluge makes the scientific method obsolete. 2014. <http://www.wired.com/2008/06/pb-theory/>.
- Bizer, C., T. Heath, and T. Berners-Lee. 2009. Linked data—The story so far. *International Journal on Semantic Web* 5(3):1–22.
- Churchman, C. W. 1971. *The design of inquiring systems: Basic concepts of systems and organizations*. New York: Basic Books.
- Daniell, K. A., A. Morton, and D. Ríos Insua. 2015. Policy analysis and policy analytics. *Annals of Operations Research*. doi:10.1007/s10479-015-1902-9.
- Ferro, E., E. N. Loukis, Y. Charalabidis, and M. Osella. 2013. Policy making 2.0: From theory to practice. *Government Information Quarterly* 30(4):359–368. doi:10.1016/j.giq.2013.05.018.
- Gregor, S. 2006. The nature of theory in information systems. *MIS Quarterly* 30(3):611–42.
- Gubbi, J., R. Buyya, S. Marusic, and M. Palaniswami. 2013. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems* 29(7):1645–1660. doi:10.1016/j.future.2013.01.010.
- Harford, T. 2014. Big data: A big mistake? *Significance* 11(5):14–19. doi:10.1111/j.1740-9713.2014.00778.x.
- Holsapple, C., A. Lee-Post, and R. Pakath. 2014. A unified foundation for business analytics. *Decision Support Systems* 64:130–41. doi:10.1016/j.dss.2014.05.013.
- Janssen, M., E. Estevez, and T. Janowski. 2014. Interoperability in big, open, and linked data—Organizational maturity, capabilities, and data portfolios. *Computer* 47(10):44–49. doi:10.1109/MC.2014.290.
- Janssen, M., R. Matheus, and A. Zuiderwijk. 2015. Big and Open Linked Data (BOLD) to create smart cities and citizens: Insights from smart energy and mobility cases. In *Electronic Government*, eds. E. Tambouris, M. Janssen, H. J. Scholl, M. A. Wimmer, K. Tarabanis, M. Gascó, B. Klievink, I. Lindgren, and P. Parycek, Vol. 9248, 79–90. Cham, Switzerland: Springer International Publishing.
- Klievink, B., and G. Zomer. 2015. IT-enabled resilient, seamless and secure global supply chains: Introduction, overview and research topics. In *Open and big data management and innovation*, eds. M. Janssen, M. Mäntymäki, J. Hidders, B. Klievink, W. Lamersdorf, B. Van Loenen, and A. Zuiderwijk, Vol. 9373, 443–453. Cham, Switzerland: Springer International Publishing.
- Kuk, G., and T. Davies. 2011. *The roles of agency and artifacts in assembling open data complementarities*. Paper presented at the Thirty Second International Conference on Information System, Shanghai, China.
- Lazer, D. 2015. The rise of the social algorithm. *Science* 348(6239):1090–1091. doi:10.1126/science.aab1422.
- Mason, R. O., and I. I. Mitroff. 1973. A program for research on management information systems. *Management Science* 19(5):475–487. doi:10.1287/mnsc.19.5.475.
- Orlikowski, W., and S. Scott. 2014. What happens when evaluation goes online? Exploring apparatuses of valuation in the travel sector. *Organization Science* 25(3):868–891.
- Rowley, J. E. 2007. The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science* 33:163–80. doi:10.1177/0165551506070706.
- Sol, H. G. 1982. Simulation in information systems development. Doctoral thesis, University of Groningen, Groningen, The Netherlands.
- Taleb, N. N. 2007. *The black swan—The impact of the highly improbable*. London: Penguin.
- Wigan, M. R., and R. Clarke. 2013. Big data's big unintended consequences. *Computer* 46(6):46–53. doi:10.1109/MC.2013.195.
- Yoo, Y., O. Henfridsson, and K. Lyytinen. 2010. The new organizing logic of digital innovation: An agenda for information systems research. *Information Systems Research* 21(4):724–735.

Notes on contributors

Marijn Janssen is a full Professor in ICT & Governance and head of the Information and Communication Technology section of the Technology, Policy and Management Faculty of Delft University of Technology. He is Co-Editor-in-Chief of *Government Information Quarterly*, Associate Editor of the *International Journal of Electronic Business Research* (IJEBR), *Electronic Journal of eGovernment* (EJEG), *International Journal of E-Government Research* (IJEGR), *Decision Support Systems* (DSS) and *Information Systems Frontiers* (ISF). He was ranked as one of the leading e-government researchers in a survey in 2009 and 2014 and has published over 350 refereed publications. More information: www.tbm.tudelft.nl/marijn.j.

George Kuk is a full professor of innovation and entrepreneurship at the Nottingham Business School. His research focuses on open innovation and strategy in software, data, design, and platform within the creative industries. He examines how companies can attract creative resources for digital and service innovation.