REFRAMING OPEN BIG DATA

Paper presented at the 21st European Conference on Information Systems, June 5-8, Utrecht, Netherlands.
Track 21 “Openness as an IS/IT Strategy - Open Data, Models, Platforms, and Sources”

Marton, Attila, Copenhagen Business School, Frederiksberg, Denmark, am.itm@cbs.dk

Avital, Michel, Copenhagen Business School, Frederiksberg, Denmark, michel@avital.net

Jensen, Tina Blegind, Copenhagen Business School, Frederiksberg, Denmark, blegind@cbs.dk

Abstract

Recent developments in the techniques and technologies of collecting, sharing and analysing data are challenging the field of information systems (IS) research let alone the boundaries of organizations and the established practices of decision-making. Coined ‘open data’ and ‘big data’, these developments introduce an unprecedented level of societal and organizational engagement with the potential of computational data to generate new insights and information. Based on the commonalities shared by open data and big data, we develop a research framework that we refer to as open big data (OBD) by employing the dimensions of ‘order’ and ‘relationality’. We argue that these dimensions offer a viable approach for IS research on open and big data because they address one of the core value propositions of IS; i.e. how to support organizing with computational data. We contrast these dimensions with two other categories that stem from computer science and engineering, namely ‘big/small’ and ‘open/closed’ to address the complex interplay between people and data, social interaction and technological operations. Thus conceived, this paper contributes an alternative approach for the study of open and big data as well as laying the theoretical groundwork for its future empirical research.

Keywords: open big data (OBD), openness, order, relationality, IS research.

1 Introduction

Digital technologies exhibit immense capabilities of casting minute details about virtually any aspect of social interaction into the form of binary data. In their latest incarnation, these capabilities combine a variety of data sources, ranging from governmental administrations, businesses and scientific research to social network sites and smart phone apps, with sophisticated techniques of data analytics, interoperability and raw processing power. They give rise to new possibilities of mixing and mashing-up data in automated and collaborative ways as illustrated by Wikipedia, Facebook and Google. As a result, we are witnessing an unprecedented level of utilizing computational data for a wide range of tasks, broadly described by the two terms of open data and big data. In this theory development paper, we provide a research framework that enables a structured analysis of the emerging capabilities afforded by data and the new technologies that drive these new capabilities.

We observe big data and open data as different, although overlapping, themes, which can be brought together in a form we refer to as 'open big data' (OBD). We employ the term OBD to direct attention to the underlying complementary relationship and commonalities of big data and open data as comparable ways of increasing the potential of data to trigger new insights, which would not have been seen otherwise. We argue that this is accomplished according to the two dimensions of 'order' and 'relationality'. Order refers to certain contemporary techniques and technologies, such as algorithmic search engines or social tagging, allowing for the flexible ordering of immense heaps of data in ways that can be determined after the data has been collected and made available; i.e. in an ex-post fashion. Relationality refers to the degree to which datasets are linkable or, in more general terms, relatable to other datasets. This is the case when users refine open government data (OGD) for further linking with other open data or when a company combines Facebook messages with data from its customer relationship management (CRM) system to get a better idea of its image amongst the customer base.

Order and relationality are two distinct but complementary ways of increasing the generative capacity (Avital and Te'eni 2009) of data; i.e., the capacity of data to trigger new insights induced by ex-post order and links to other data - by making data 'orderable' and 'relatable'. We submit that this approach is closely aligned to IS research and practice and the discipline's core value proposition to support organizing with computational data. Hence, we contrast these two dimensions with the distinctions predominantly used to describe the rise of data-intensive platforms and services – open/closed and big/small data. As we discuss below, these distinctions come from computer science and emphasize properties fundamental to its domains. Although these distinctions are increasingly used with respect to the study of IS, we argue that, because of their roots in computer science, they do not address the complex interplay between people and data, social interaction and technological operation, which IS research and practice is primarily engaged with. We propose to reframe open and big data in order to bring into focus the techniques and technologies employed to increase the potential of data to be informative for individual or collective sense-making.

Thus conceived, this paper contributes to the development of a sensitizing as well as sensitive framework for IS research; i.e., a framework rich enough to grasp the intricacies of this new level of societal and organizational engagement with computational data. In order to accomplish this goal, the paper is structured as follows. In the first section, we give an overview of the themes and issues that dominate the contemporary discussions based on the distinctions of big/small and open/closed data. The second section focuses on the conceptual elaboration of these themes and issues by introducing a set of different distinctions based on the potential of data to be ordered and related to other data. Finally, we develop the conceptual framework based on the two dimensions of order and relationality. We propose this approach as a way to reframe IS research with respect to the growing influence of these new data-intensive techniques and technologies on individual and organizational lives.

2 Themes and Issues

Big and open data are emerging as major themes in IS research. Although big data can clearly be assigned to the area of business intelligence and analytics, it changes most of the technical and strategic practices on how organizations, especially businesses, should employ data in order to understand their performance and to make decisions (Davenport, et al. 2005, Chen, et al. 2012). As we will discuss below, big data is different from having a lot of data. A similar picture can be drawn with respect to open data. The open data movement has a lot of conceptual and ideological commonalities with open source, open innovation and open access (reference withhold) – domains studied in depth by IS researchers. Open data is clearly building on these domains but is nonetheless different from sharing code or accessing data. Different approaches are required in terms of, for instance, legal considerations such as licensing (Miller, et al. 2008) and technical issues such as the format in which data is to be published (Shadbolt, et al. 2006). This section is dedicated to the introduction of open and big data in order to develop a working definition and understanding of these terms, before we combine them into the form of open big data (OBD).

2.1 Open data

Open data is a movement of publishing digital data online in an open format bringing together a variety of societal actors ranging from individual hacktivists, NGOs and NPOs to businesses, governmental administrators and policy-makers (Hogge 2010). The Open Knowledge Foundation (OKF), a leading NGO in promoting open data, gives the following, widely used definition: "A piece of content or data is open if anyone is free to use, reuse, and redistribute it – subject only, at most, to the requirement to attribute and/or share-alike" (opendefinition.org, last access 12th Oct. 2012). Thus conceived, open data shares the intellectual and ideological views of similar movements that advocate the opening of digital technologies and infrastructures. A telling example is the Budapest Open Access Initiative (BOAI, http://www.opensocietyfoundations.org/openaccess), which focuses on making research free and available online. Open data, however, is not only about sharing data or making it accessible. More importantly, the term 'open data' connotes the publishing of data in ways that enable anybody to repurpose the data and to combine it with other datasets to create new, innovative online services.

In order to assess the openness of a dataset, Sir Tim Berners-Lee (2006), a vocal advocate for open data, developed a 5-star rating system according to the degree to which data can be reused and combined with others; (1 star) any content made available online in any format (e.g. image scanned table published as a .pdf file) is considered open data as long as it is published under an open license (Miller, et al. 2008); (2 stars) data is published in a machine readable format (e.g. excel instead of an image scanned table); (3 stars) data is published in a machine readable, non-proprietary format (e.g. csv instead of excel); (4 stars) open standards developed by the WWW Consortium (W3C) are used to uniquely identify the data enabling others to link to; (5 stars) data is linked to other data providing context.

Put into practice, open data presents an unprecedented level of opportunities for anybody with the necessary skills to repurpose, recycle, link and mash-up datasets using the WWW as the infrastructure for interoperability. In particular, open government data (OGD) or public sector information (PSI), as it is referred to by the EU, has become a main arena (European Commission 2011). Taking a leading role, the US and UK governments have published thousands of datasets on their respective data portals data.gov and data.gov.uk followed by other regional, national and supranational governmental organizations (Hogge 2010). By the same token, Wikipedia's DBpedia offers all its articles and related data in an open and linkable format. It is a key part of the Linked Open Data (LOD; www.linkeddatta.org) network – a collection of open and linked datasets and promoter of the Semantic Web (Auer, et al. 2007). Finally, civic hacktivists and businesses are crucial players as they are creating new and innovative online services and apps by combining diverse datasets into mash-ups for

end-users (Hogge 2010). A case in point, MySociety.org develops online services, non-profit and for clients alike, such as mapumental.com. Providing real-time maps for the London metropolitan area, the service helps users to find a new home according to the expected commuting time to work.

Open data combines unrestricted availability with technical interoperability (Tammisto and Lindman 2012). In order to achieve this, data is going through a value chain from (1) raw open data, such as OGD, (2) its refinement into linked open data, such as DBpedia, to (3) the combination, reuse and mash-up of linked open data into applications that are useful for (4) end-users (Latif, et al. 2009). Given these observations, open data is supposed to increase transparency and accountability in terms of public governance and administration, facilitate civic engagement and give rise to new service providers as intermediaries between data and end-users. As a result, open data holds the potential to lead to innovative services and, ultimately, to insights which would not have been gained otherwise.

2.2 Big data

In computer science and industry, so-called 'small data' refers to structured data managed by means of traditional relational databases and stored in data warehouses. By contrast, big data "[…] refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse. This definition is intentionally subjective and incorporates a moving definition of how big a dataset needs to be in order to be considered big data – i.e. we don't define big data in terms of being larger than a certain number of terabytes […]" (Manyika, et al. 2011:1). The size of the datasets is only one condition for big data. Doug Laney (2001) at META Group (now Gartner) introduced the so-called '3VS' framework in 2001 adding variety and velocity to volume, which has become the standard conceptual approach towards big data (Zikopoulos, et al. 2012:5-8).

For over a decade now, the volatile growth in the sheer volume of bits and bytes produced and collected by digital means has been staggering. Pioneered by Peter Lyman and Hal R. Varian (2003), research into the deluge of data has repeatedly shown an exponential growth. The latest of IDC's annual "Digital Universe" studies reports an increase from 1 zettabyte in 2010 to 1.8 zettabytes (1.8 trillion gigabytes) in 2011 (Gantz and Reinsel 2011). Key drivers of this development are, beside low costs of storage and processing capacity, predominantly the availability of data through online services such as social network sites, the proliferation of smart phones but also RFID tags and sensors networked into the internet of things. Employing data from such diverse sources, however, also increases the variety of the data. In addition to traditional structured data organized into columns and rows, big data is including raw data (e.g. sensors), semi-structured data (e.g. web logs) and unstructured data (e.g. tweeted text), which do not fit the traditional paradigm of data schemas and warehouses (Zikopoulos, et al. 2012). Finally, velocity refers to the ephemeral nature of binary data continuously brought up-to-date (Kallinikos 2006), which, in turn, requires data analysis to happen as close to real-time as possible before the value of the data expires. This notion is sometimes referred to as "nowcasting" (Varian 2010:5). Taken together, all these developments have led to a paradigm shift from pre-determined schemas and relational databases – small data – to distributed infrastructures in the ways data is stored and processed (O'Reilly 2012). By the same token, the analysis relies on recent developments in artificial intelligence and machine learning, automated content analysis and visualization providing the techniques and tools to make big data manageable.

Big data is an intricate assemblage of techniques and infrastructures diffusing into the institutional fabric of society. Business and commerce are the obvious social domains in which big data will flourish (Davenport, et al. 2005). Other domains, however, are expected to benefit as well (Manyika, et al. 2011). Health care is expected to increase its effectiveness in terms of patient care and the development of new treatment regimens. On all levels of government, big data is expected to improve the governance of citizens, administration of services and cutting of costs. The natural sciences already gaze at the sub-atomic through the Large Hadron Collider at CERN and stargaze through telescopes, such as the Sloane Digital Sky Survey, each generating petabytes of data on a daily basis (Shiri 2012). The social sciences hope for a new methodology based on behavioural data revealing what people are

actually doing rather than what they say they are doing (Manovich 2011, Phillipe 2012). Whatever values, facts, truths or, generally, information one is looking for, it is supposedly there in the data waiting to be discovered.

2.3 Open and big data

Open and big data share some commonalities but need to be considered as two distinct phenomena nonetheless. Certainly, open data sources, especially those published by the government, can be added to the pool of data already crunched by big data techniques and technologies. By the same token, if a sufficient number of open datasets are linked in sophisticated ways and assembled into a service, the whole assemblage may present characteristics of big data. The differences become clearer when the respective distinctions, introduced by big and open data, are taken into consideration. With respect to big data, the conventional distinction is obviously big/small; with respect to open data, it is open/closed.

	Small data	**Big data**
Closed data	**Traditional Information Systems** Relational databases and data-warehouses; Structured data; Centralized computing and storage; e.g. ERP, DSS	**Big Data Analytics** Non-relational databases; Structured and unstructured data; Distributed computing and storage; e.g. Recommender systems, Search engines
Open data	**Open Architectures** Machine-readable and "linkable" data; Semantic web technologies; e.g. Linked Open Data, Open APIs	**Open Big Data (OBD)** Emerging research on mash-ups; e.g. Apps for Democracy (www.appsfordemocracy.org)

Table 1. Prevalent themes in the current discourse on data in IS research.

While big data is focused on tapping into as much data as possible, structured or unstructured, small data refers to well-structured and curated sets of data to be managed within the confines of the analytical tools developed over a decade ago (Zikopoulos, et al. 2012). Open data is dedicated to the tearing down of technological walls between long established silos of knowledge (Marton 2011). The control over those silos lies with the holder of the data deciding who has access and how the data is to be used, i.e. closed data. By contrast, open data does not merely grant universal accessibility to the data itself but, crucially, enables the creation of new services beyond the control of the holder of the data. Brought together in Table 1, these distinctions reflect the current state of the discourse on big and open data and can be related to well-established themes in IS research and practice. Closed and small data refers to the origins of IS research as the study of **Traditional IS**, for instance, enterprise resource planning (ERP) or decision support systems (DSS). Making data big while maintaining its closed nature is the area of **Big Data Analytics** as predominantly businesses retain control over their data based upon which they offer services such as recommender systems to their customers (Chen, et al. 2012). Opening small data, on the other hand, refers to **Open Architectures** as data is made available through open Application Programming Interfaces (APIs) in machine-readable formats and control is passed on beyond organizational boundaries (Latif, et al. 2009).

This paper focuses on the combination of open and big data, which is an emerging yet hardly understood theme. As described above, the fundamental concepts of **Open Big Data** are technical in nature as they were developed in the fields of computer science and engineering. While big data is about distributed computation and infrastructures, open data is about standards on how to make data machine-readable, and hence linkable. Conceptualizing these in contrast to small and closed data respectively reflects the views and perspectives of computer science, which is helpful and informative but does not fit the fundamental agenda of IS research and practice.

In the next section, we take open big data (OBD) out of its contemporary context, shown in Table 1, and focus on the comparison of open and big data as such. Thus, we reframe open and big data to look at the potential of data to be ordered ex-post (order) and related to other data (relationality). We argue that such a reframing takes into consideration better the views and perspectives of the IS field.

3 Open Big Data (OBD)

Making sense of a phenomenon is the result of the distinctions an observer introduces into the world (Luhmann 2002). However, distinctions enlighten certain aspects of a phenomenon, while hiding others. It is in this sense that we propose a different approach towards what we term open big data (OBD), in order to enlighten crucial aspects that may otherwise be overlooked. Instead of comparing present with past developments (i.e., big with small, open with closed), we compare big data with open data to develop an alternative view on these two emerging phenomena.

3.1 Data and information

As a first step, we return to the basic concepts of data and information. To be clear, it is not our intention to contribute to the ongoing discussions in the IS field with respect to clarifying or even defining information (McKinney and Yoos 2010). It is, however, required to explicitly state our understanding of information with respect to big and open data, as it is the foundation of the sections that follow. To begin with, we refer to data as binary codified data. Thus conceived, data is the outcome of categorizing events according to the two basic categories of 0 and 1 (Borgmann 1999). As a second step, data needs to be processed in order to be potentially informative for somebody. One may think of analytical tools such as visualization, automated pattern seeking and others as means to increase the potential of the data to be informative for individual or collective sense-making (reference withhold).

A given dataset forms the basis for potential information to emerge but it is not the same thing. By adding new data to an existing dataset one increases the potential informativity of the dataset. However, one could also link two separate datasets. The combination of the two datasets does not lead to more data, but it results in more potential information nonetheless, simply because one can compare, correlate, triangulate or combine the sets of data. By linking datasets one may gain new insights, which would not have been gained, if each dataset was analysed separately (Narayanan and Shmatikov 2009). Given this argument, information is closely related to novelty (Bateson 2000, Kallinikos 2006). One is informed, if one learns something new or, in more abstract terms, information occurs if data triggers change (McKinney and Yoos 2010).

The separation of data and information has been proposed since the early days of IS research. Boland (1987), for instance, already referred to information as ‘inward-forming’: as change in a person from an encounter with data. Conflating data with information conjures up the illusion of entifying information, which is inherently illusive and event-like (Kallinikos 2006). It is the illusion that information can be perfect, when enough data is available. The hype of big data, and to a lesser degree open data, reinvigorates this illusion considerably. IS research, we submit, has the opportunity to enrich the discourse on open and big data by calling attention to the distinction between data and information. It is in this sense that the next section will discuss big and open data as two different but complementary ways of increasing the potential of data to inform.

3.2 Potential information

Big and open data share the commonality of using data in new, technology-intensive ways to gain insights. In terms of open data, this goal is to be achieved by linking, combining and mashing-up open datasets. The mashing-up of separate datasets increases the potential of the data to inform, since two or

more datasets combined allow for new insights based on the mixing, repurposing and contextualization of data (Bauer and Kaltenböck 2012:27). The potential information does not primarily rely on the amount of data but on the potential links that can be established between data sources forming a layer for innovative services to be built upon. As the director of the Open Knowledge Foundation, Rufus Pollock, put it in one of his presentations, "[G]oing forward in some fields like software, data is going to be a platform, not a commodity. […] You need to be building on your data" (Pollock 2012:time index 19:38). Thus conceived, it is more appropriate to observe developments in open data in terms of its immense potential to be linked to other data; i.e., the relationality of data (Boyd and Crawford 2011:1). Given these arguments, a shift in the perspective from open/closed to relationality sheds light on the distinguishing characteristic of open data. Rather than simply being the opposite of closed, open data unfolds a layer of potentialities for interoperability as well as innovation without imposing rules on how the data is to be used.

By contrast, big data is insightful by crunching as much data as possible through automated means of computational processing and a new generation of analytical tools. In the words of David Bollier (2010:8), "the data once perceived as 'noise' can now be reconsidered with the rest of the data, leading to new ways to develop theories and ontologies." The potential for information is increased by adding more data in all its variety and velocity, thus filtering less and less data as noise. In other words, previous approaches in business analytics were based on a pre-determined order. If an event fit into the order, it was data. If it did not, it was noise. The potential of the data to inform was based on a highly selective process of collecting what fit into the pre-defined order and thus ended up in a data warehouse for further analysis (Croll 2012:56). The filter was on the way in (Weinberger 2007). With big data, whatever comes in binary format can be conceived of as data. Data can be ordered, analysed and thus made potentially informative in ways that are not pre-defined. The potential of big data to inform is based on ordering as much data as possible through automated computational operations after its collection; i.e. in an ex-post fashion. Since filtering what gets into an information system is of no concern anymore, the filter is now on the way out (Weinberger 2007:102). Those filters take the form of algorithmic calculations and analytical tools, generating patterns expected to be informative. Thus conceived, it is more appropriate to observe the emergence of big data based on order, because it highlights a fundamental shift from filtering noise through an ex-ante order to an ex-post ordering of data.

Big data is a revelatory illustration for the overall shift from pre-defined categorization schema to ex-post ordering as events which used to be discarded as noise, are captured as data. Open data, we argue, is a revelatory illustration for the potential of data to be linked or rather related to other data, i.e., high relationality. However, ex-post ordering and high relationality are not solely a product of or confined to big data or open data respectively. Ex-post ordering can be employed with respect to open data as well. Open data is made publicly available without an inherent order in the *collection* of datasets, sometimes in such a format that it requires additional efforts of 'cleaning' and 'refining' (Kuk and Davies 2011:4). The ordering of the data is accomplished by social actors by means of the services they develop, which brings order to the open data in an ex-post fashion. By the same token, relationality can be employed with respect to big data. As already discussed above, big data is about the crunching of datasets from a variety of sources and in a variety of formats, which requires the data to be relational in order to be brought together, combined and mixed for the purpose of analysis and reporting. It is also in this sense that, as stated above, big data is not about a lot of data but about its capabilities to network data; i.e., to be relatable in all its variety and velocity.

Given these observations, we argue that ex-post ordering and high relationality are two distinct dimensions according to which data is currently employed to gain new insights. Big and open data are only two instantiations of the increasing potential of data to inform. Both introduce ex-post ordering of data based on the potential of data to be related to other data. However, there are differences as well. Big data relies mostly on automated processes and number crunching leading to correlations between datasets and, ultimately, to patterns deemed informative for strategic decision-making. Open data, on the other hand, relies mostly on social actors and human engagement to select and refine appropriate

datasets for the development of a service. In their working paper, Kallinikos et al. (2012) refer to this difference as information generated while based on semantic and agnostic generativity. Semantic generativity is the mixing and aggregation of existing data arising out of cultural and semantic considerations. By contrast, agnostic generativity is the blind, statistically based manipulation of data arising out of algorithmic calculations producing correlations and patterns. The potential of the data to inform is increased in two dimensions; as Kallinikos et al. (2012) put it, horizontally through meaningful mixing and mashing-up as well as vertically through number crunching. In our terminology, potential information is increasingly generated by broadening as well as by deepening the availability of relational data.

3.3 Research framework

Distinguishing between data and information, rather than between big and small or open and closed data, redirects our attention away from issues that concern computer science to issues which are of genuine concern to IS research. In the most general sense, IS research is studying techniques and technologies of manipulating data in order to support organizing. We propose to characterize the developments discussed above in those techniques and technologies as a shift to ex-post ordering and high relationality. Data is, to an increasing degree, made to be 'orderable' and 'relatable'. Order and relationality, we submit, are more appropriate for IS research as these dimensions focus on the potential of data to be insightful for somebody. Thus, they address the relationship between people and data, social interaction and technological operation that IS research and practice is primarily engaged with. From the perspective of IS, it is not about making data bigger and more open; it is about increasing the potential of data to trigger change, new insights and innovation or, in more abstract terms, information (Boland 1987). This is increasingly accomplished by means of ex-post ordering and high relationality giving rise to new data intensive services we refer to as Open Big Data (OBD). The rationale for choosing the term OBD is to convey that 1) open data is not the same as big data, 2) open data is not an aspect of big data and vice versa. Most crucially, we want to convey that 3) both – open data and big data – are ways of making data orderable and relatable.

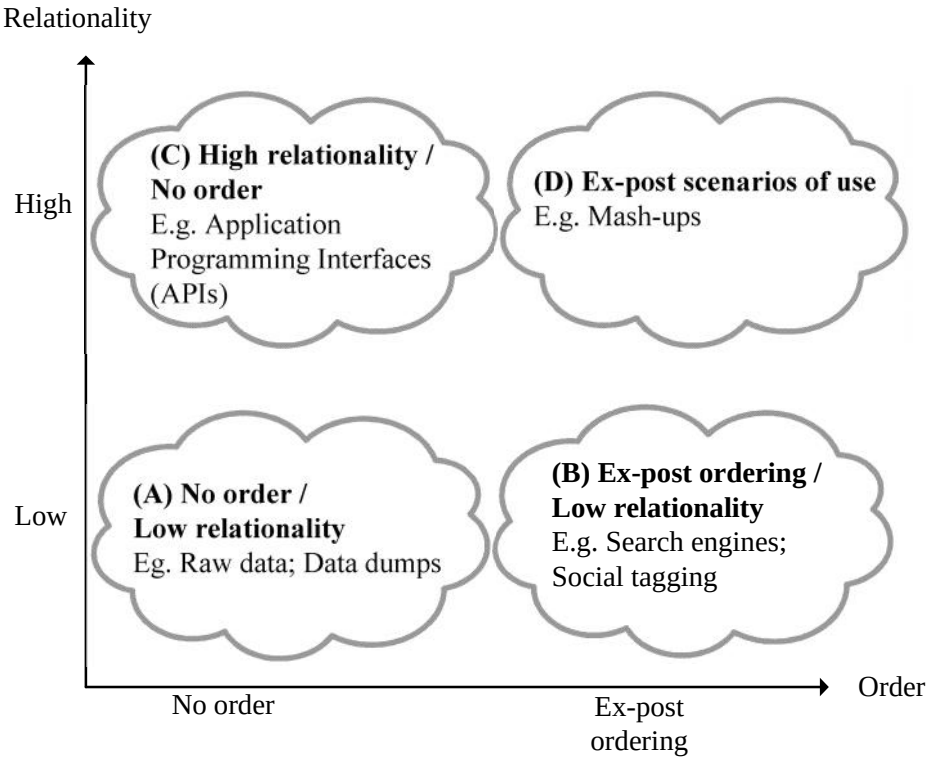

Figure 1. The dimensions of the emerging research framework.

As depicted in Figure 1, we unfolded a research framework for the study of OBD according to the two dimensions of order and relationality. **Quadrant A** is the domain of raw data and 'messy' data dumps without an inherent order to dictate its further use. In this case, data affords a very low degree of relationality, if at all, requiring further efforts of refinement or cleaning. **Quadrant B** is about bringing order into the mess of data in an ex-post fashion without increasing its relationality. This is the world of search engines and social tagging, which increase the potential of data to be informative by making data orderable based on algorithmic calculations of relevance or the adding of descriptive, searchable meta-data by online crowds. **Quadrant C**, on the other hand, is about increasing the potential of data to be related to other data without imposing an order. This is the world of building Application Programming Interfaces (APIs) for datasets allowing software applications to query the datasets from across the WWW. The potential of data to inform is increased by making datasets highly relational with other datasets by means of APIs. Finally, **Quadrant D** combines ex-post ordering and high relationality. We refer to this combination as 'ex-post scenarios of use' as it does not only allow for data to be used as it is but also for data to be repurposed for further reuse by others. One can build one's own scenario of use for the data after its publication and beyond the immediate control of its publisher; i.e., in an ex-post fashion. The archetype is a mash-up, which increases the potential of data to inform by relating and ordering a selection of datasets, thus creating an informative application for end-users.

3.4 Vignette – Wikipedia

We submit that this framework can be applied to analyse the whole range of OBD phenomena – be it the WWW, Facebook or Twitter, to name but a few. For the sake of illustration, Wikipedia is a case in point, since it is an intricate assemblage of data and services with an immense following in terms of users and contributors. To begin with, the actual collection of Wikipedia articles belongs to Quadrant A. Combining structured and unstructured data, the collection as such is like a heap of books; it does not have an inherent, pre-determined order. The articles also display a low degree of relationality as they can only be linked through hyperlinks (**Quadrant A**). In order to increase the potential of the collection to inform, Wikipedia implemented functionalities for full text search and for social tagging. This is clearly a step towards making data orderable (not more relatable) by means of ex-post ordering – be it according to algorithmic relevance ranking or searchable social tags (**Quadrant B**). By the same token, Wikipedia also created DBpedia offering data in a highly relational format through an API. In this case, the potential of the articles to inform is increased by making data relatable rather than orderable (**Quadrant C**). Taken together, Wikipedia allows for its data to be used via Wikipedia's website or, in other words, according to Wikipedia's own scenarios of use (e.g. full-text search for articles, editing content, etc.). Wikipedia also allows for the data to be repurposed in ways that are beyond the immediate control of Wikipedia. It allows for others to create mash-ups; i.e., to create their own scenarios of use for the data (**Quadrant D**). An example among many others, DBpedia Mobile is a mash-up service for smart phones. Combining data from DBpedia, Google Maps, Geonames, Flickr and others, the app allows users to explore the area they are in, highlighting nearby locations of interest. In other words, DBpedia Mobile relates a selection of datasets and orders the data by mapping location data from DBpedia onto Google Maps – a new scenario of use developed by ordering and relating data in an ex-post fashion.

4 Discussion

Our conceptual reframing of open and big data in terms of order and relationality refers to a variety of issues to be taken into consideration by IS research and practice. In terms of IS research, we argue for a conceptualization that acknowledges data-intensive phenomena as intricate assemblages combining order and relationality in various degrees rather than a movement from closed to open *or* small to big data. As we illustrated with Wikipedia, our framework unfolds data-intensive services as a mix of data as well as ways to order (e.g. full text search, social tagging) and relate data (e.g. APIs). Thus

conceived, OBD challenges some core, long-held traditions with respect to the design of information systems and the management of knowledge. Instead of data schema and classification systems, determining the 'informativeness' of data yet to be collected, the focus shifts towards creating and maintaining potentialities for data to be manipulated in ways which are determined in an ex-post fashion (Weinberger 2007).

Still, despite the high degree of granularity that social interaction is recorded with, one must not forget that binary digitization is a categorization. It is a radically selective process that disintegrates complex human activities into a series of a single difference – the difference between the two categories of 0 and 1 (Borgmann 1999). At the most fundamental level, binary data is still what fits the classification system of 0/1 and, thus, is the result of a series of design choices and decisions. Numbers, of course, do not speak for themselves, but require a whole array of practices, routines and rituals of sense-making as well as rhetorical skills in terms of justifying decisions based on the evidence socially constructed into the data (Boyd and Crawford 2012, Kallinikos 2012). Inherently political, open and big data afford considerable risks. For instance, OBD gives rise to a 'data divide' as a new aspect of the digital divide (Gurstein 2011). Having the necessary skills and, more importantly, access to OBD is becoming a critical and divisive issue in a range of domains. A second example is privacy as the immense potential of data to be ordered and related renders the anonymization of data ineffective. Combining diverse datasets allows to triangulate and, thus, to infer on individual identities (Narayanan and Shmatikov 2009).

In terms of IS practice, OBD techniques and technologies afford a considerable leap in the capabilities to employ data for the purposes of innovation and service development. On the one hand, there is a connection between orderable and relatable data and new ways of cooperation that facilitate innovation. OBD leads to new opportunities in terms of designing and implementing information systems that enable the emergence of generative capacities (Avital and Te'eni 2009, Van Osch and Avital 2009, Kallinikos, et al. 2012). On the other hand, OBD is also prone to self-referential phenomena (Marton 2009, Kallinikos, et al. 2010). For instance, the collection, analysis and management of OBD are data-intensive tasks in themselves leading to even more data that may feed back and thus contribute to the deluge of data (Kallinikos 2006). OBD does not only derive from ordered and related data but may also lead to more noise.

Given these observations, OBD unfolds an array of new challenges and domains for IS research and practice. Challenges will include designing appropriate information systems that are integrated with open big data, the sensible employment of data to support decision-making or the critical understanding of the political nature of OBD assemblages and the power-structures they evoke and reinforce.

5 Conclusion

Our analytical comparison of the emerging phenomena of big and open data revealed common denominators, which are indicative of an immense increase in the potential of data to trigger new insights. This potential unfolds along two dimensions we identified as order and relationality. The potential of data to inform is increased by means of ex-post ordering as well as by means of affording datasets to be linked and to interoperate (high relationality). Brought together, ex-post ordering and high relationality allow for ex-post scenarios of use for the data, such as mash-ups, which relate and order a selection of datasets creating an informative application for end-users. Data-intensive ecosystems, such as the WWW but also Wikipedia, Facebook and so forth, unfold along these dimensions of order and relationality forming assemblages, which we referred to as open big data (OBD). Thus, this paper contributed an alternative perspective to the study of contemporary phenomena usually referred to as big data or open data. In contrast to the distinctions of big/small and open/closed, imported from computer science, we developed a conceptual framework that is more appropriate for IS research as it is based on the distinction between data and information. Manipulating data in order to make it informative for individual or collective sense-making is the core

concern of IS research and practice. The two dimensions of order and relationality address this concern with respect to the current developments in the techniques and technologies that support organizing by means of open big data.

References

Auer, S., C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak and Z. Ives (2007). DBpedia: A nucleus for a web of open data. 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, Busan, Korea.

Avital, M. and D. Te'eni (2009). From generative fit to generative capacity: Exploring an emerging dimension of information systems design and task performance. Information Systems Journal, 19(4), 345-367.

Bateson, G. (2000). Steps to an ecology of mind. Chicago, University of Chicago Press.

Bauer, F. and M. Kaltenböck (2012). Linked open data: The essentials: A quick start guide for decision makers. Vienna, Austria, Mono/Monochrom.

Berners-Lee, T. (2006). Linked data. Design Issues http://www.w3.org/DesignIssues/LinkedData.html

Boland, R. (1987). The in-formation of information systems. In Critical issues in information systems research (Boland, R. and Hirschheim, R. A. Eds.). Chichester, Wiley, 363-379.

Bollier, D. (2010). The promise and peril of big data. Washington, D.C., The Aspen Institute.

Borgmann, A. (1999). Holding on to reality: The nature of information at the turn of the millennium. Chicago, University of Chicago Press.

Boyd, D. and K. Crawford (2011). Six provocations for big data. A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, Oxford, UK.

Boyd, D. and K. Crawford (2012). Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication and Society, 15(5), 662-679.

Chen, H., R. H. L. Chiang and V. C. Storey (2012). Business intelligence and analytics: From big data to big impact. MIS Quarterly, 36(4), 1-24.

Croll, A. (2012). Big data is our generation's civil rights issue, and we don't know it. In Big data now: 2012 edition (O'Reilly Ed.), Sebastopol, CA, O'Reilly Media, 55-59.

Davenport, T. H., D. Cohen and A. Jakobson (2005). Competing on analytics. Babson Park, MA, Babson Executive Education.

European Commission (2011). Open data. An engine for innovation, growth and transparent governance. Brussels, European Union

Gantz, J. F. and D. Reinsel (2011). Extracting value from chaos. Framingham, MA, IDC http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf

Gurstein, M. (2011). Open data: Empowering the empowered or effective data use for everyone? First Monday, 16(2). http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3316../2764

Hogge, B. (2010). Open data study, Open Society Foundation

Kallinikos, J. (2006). The consequences of information: Institutional implications of technological change. Northampton, MA, Edward Elgar.

Kallinikos, J. (2012). The allure of big data. ParisTech Review. http://www.paristechreview.com/2012/11/16/allure-big-data/

Kallinikos, J., A. Aaltonen and A. Marton (2010). A theory of digital objects. First Monday, 15(6). http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3033/2564

Kallinikos, J., A. Aaltonen and A. Marton (2012). Information generativity and logics of innovation. Working Paper, London School of Economics and Political Science

Kuk, G. and T. Davies (2011). The roles of agency and artifacts in assembling open data complementaries. 32nd International Conference on Information Systems, Shanghai, China.

Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. Application Delivery Strategies, META Group/Gartner

http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf
Latif, A., P. Höfler, A. Stocker, A. Usswaeed and C. Wagner (2009). The linked data value chain: A lightweight model for business engineers. 5th International Conference on Semantic Systems, Graz, Austria.
Luhmann, N. (2002). Theories of distinction: Redescribing the descriptions of modernity. Stanford, CA, Stanford University Press.
Lyman, P. and H. R. Varian (2003). How much information 2003? Berkeley, CA, School of Information Management and Systems, University of California http://www.sims.berkeley.edu/research/projects/how-much-info-2003/printable_report.pdf
Manovich, L. (2011). Trending: The promises and challenges of big social data http://www.manovich.net/DOCS/Manovich_trending_paper.pdf
Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. Hung Byers (2011). Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute
Marton, A. (2009). Self-referential technology and the growth of information. From techniques to technology to the technology of technology. Soziale Systeme, 15(1), 137-159.
Marton, A. (2011). Social memory and the digital domain: The canonization of digital cultural artefacts. EGOS, Gothenburg, Sweden.
McKinney, E. H. and C. J. Yoos (2010). Information about information: A taxonomy of views. MIS Quarterly, 34(2), 329-344.
Miller, P., R. Styles and T. Heath (2008). Open data commons. A license for open data. 1st Workshop about Linked Data on the Web, Beijing, China.
Narayanan, A. and V. Shmatikov (2009). De-anonymizing social networks. 30th IEEE Symposium on Securty and Privacy, Oakland, CA.
O'Reilly, Ed. (2012). Big data now: 2012 edition. Sebastopol, CA, O'Reilly Media.
Phillipe, O. (2012). Reassembling the data - How to understand opportunities using an interdisciplinary approach. Internet, Politics, Policy, Oxford, UK.
Pollock, R. (2012). Open data: How we got here and where we're going. Presentation at Lift12, Geneva, Switzerland http://videos.liftconference.com/video/4699918/open-data-how-we-got-here-and
Shadbolt, N., T. Berners-Lee and W. Hall (2006). The semantic web revisited. IEEE Intelligent Systems, 21(3), 96-101.
Shiri, A. (2012). Typology and analysis of big data: An information science prospective. Internet, Politics, Policy, Oxford, UK.
Tammisto, Y. and J. Lindman (2012). Definition of open data services in software business. 3rd International Conference on Software Business, Cambridge, MA.
Van Osch, W. and M. Avital (2009). Collective generativity capacity: The seed of IT-induced collective action and mass innovation. Proceedings of the 8th Journal of the Association for Information Systems sponsored Theory Development Workshop, Phoenix, AZ. http://sprouts.aisnet.org/766/2/JAIS-TDW09-013.pdf
Varian, H. R. (2010). Computer mediated transactions. The American Economic Review, 100(2), 1-10.
Weinberger, D. (2007). Everything is miscellaneous: The power of the new digital disorder. New York, Times Books.
Zikopoulos, P. C., C. Eaton, D. deRoos, T. Deutsch and G. Lapis (2012). Understanding big data. Analytics for enterprise class hadoop and streaming data. New York, McGraw-Hill.