LSE Research Online

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Ioanna Constantiou and Jannis Kallinikos

New games, new rules: big data and the changing context of strategy

Article (Accepted version) (Refereed)

Original citation: Constantiou, Ioanna D and Kallinikos, Jannis (2015) *New games, new rules: big data and the changing context of strategy.* Journal of Information Technology, 30 (1). pp. 44-57. ISSN 0268-3962

DOI: 10.1057/jit.2014.17

© 2015 JIT Palgrave Macmillan.

This version available at: http://eprints.lse.ac.uk/63017/
Available in LSE Research Online: August 2015

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (http://eprints.lse.ac.uk) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.lse.ac.uk

This is a nearly identical version to the article published in JIT

http://www.palgrave-journals.com/jit/journal/vaop/ncurrent/abs/jit201417a.html#close

New Games, New Rules: Big Data and the Changing Context of Strategy

Ioanna Constantiou
Department of IT Management, Copenhagen Business School

Jannis Kallinikos
Department of Management, London School of Economics

Abstract

Big data and the mechanisms by which it is produced and disseminated introduce important changes in the ways information is generated and made relevant for organizations. Big data often represents miscellaneous records of the whereabouts of large and shifting online crowds. It is frequently agnostic, in the sense of being produced for generic purposes or purposes different from those sought by big data crunching. It is based on varying formats and modes of communication (e.g. texts, image and sound) raising severe problems of semiotic translation and meaning compatibility. Crucially, the usefulness of big data rests on their steady updatability, a condition that reduces the time span within which this data is useful or relevant. Jointly, these attributes challenge established rules of strategy making as these are manifested in the canons of procuring structured information of lasting value that addresses specific and long-term organizational objectives. The developments underlying big data thus seem to carry important implications for strategy making, and the data and information practices with which strategy has been associated. We conclude by placing the understanding of these changes within the wider social and institutional context of longstanding data practices and the significance they carry for management and organizations.

Keywords

Big data, business environment, data practices, management, organization, social data, strategy making, time, updatability

Introduction

Big data is becoming a highlighted topic in IS, management and social science research. As the label itself indicates, big data is commonly deployed to refer to large data volumes generated and made available on the internet and the current digital media ecosystems. But data volume alone would have never sufficed to encapsulate the novelty of the phenomenon (boyd an Crawford, 2012). Closely associated with large data volumes are the diversity of such data, the frequency by which it is updated and, more generally, the speed by which it grows (O'Reilly, 2012; Davenport, 2014). Taken together, these attributes are meant to give rise to large, diffuse and shifting data structures that challenge traditional techniques (e.g. database and server technologies) and practices (e.g. statistics, accounting, professional classifications) with which large data volumes have hitherto been dealt. A heated and often cross-disciplinary debate has emerged over the last few years on the opportunities and challenges these developments posit for organizations, communities and individuals.

Part of this debate concerns the exact nature of this data, its origin, private or public, and its individual or social implications (e.g. Bollier, 2010; Ekbia *et al.*, 2014). But much research has as well been directed to the implications big data carry for businesses. LaValle *et al.* (2011), for instance, viewed big data as an extension of business data analytics. They recommended its inclusion to existing strategy practices to extend traditional information sources about the business environment and enable the organization to remain competitive by responding promptly and continuously to market disruption or change. McAfee and Brynjolfsson (2012) underlined the opportunities big data creates with respect to improving specific activities such as customer relationship management, while Davenport *et al.* (2012) and Oestreicher-Singer and Zalmanson (2013) described opportunities for new business models predicated on data analytics and social data (user data on social media platforms) respectively. Varian (2010) offered examples of economic transactions in fields like transportation, advertising or contracting which benefit from the information provided by big data and highlighted its importance in making short term economic predictions (Choi and Varian, 2012). In all these accounts, big data is claimed to have a significant impact on the ways firms and organizations assess, analyze and address environmental trends.

More recently, Brynjolfsson and McAfee (2014) described big data as a new and advanced stage of digitization. Combined with existing layers and data techniques, big data provides the premises for combining information across sources and applications, and developing a range of innovative services (see also Varian, 2010). On this view, big data enables the development of new intangible assets, and contributes to the competitive strategy of an organization by enabling the development of innovative services (e.g., new services based on social or transaction data offered by companies such as Facebook, Google or Amazon). A similar claim is put forward by Davenport (2014) suggesting that organizations may derive value from combining big data with other types of predominantly internal data, often used in the context of business analytics. While highlighting various impacts that big data may have on organizations, the issues these authors raise indicate that big data reshapes the means and operations through which information becomes available for decision makers in organizations. In this regard, big data is linked to the context of organizational intelligence and strategy, and how organizations perceive, assess and act upon their internal and external environments. We focus on this changing context of strategy and investigate how it might be reshaped in the presence of big data.

Traditional models and tools in strategy require the collection and use of predefined data. Strategy is based on the use of information derived from data collected through systematic and purposeful processes that address specific information needs of the decision-makers and improve accuracy of prediction of future events. Standard strategy texts describe well-defined measures and indicators that once calculated are supposed to provide input to specific financial or accounting models. In turn, these last offer estimates of specific performance measures of other financial indicators, depending on which strategic decisions are made. Recently, the evidence-based management movement (Rousseau, 2006) reinforced this approach by emphasizing the need to draw on systematic evidence in strategy and management, and collect data driven from theoretical models' requirements. This approach is meant to test the validity of propositions in specific contexts, which in turn is expected to improve organizational performance and raise competitiveness. Thus, existing strategy literature assumes that structured (often quantitative) data is intentionally collected to inform specific models and provide predefined input to the decision making process (Pfeffer and Sutton, 2006). A number of systematic and relevant data collection methods (e.g., market research tools and statistical models of inference, classification or management accounting systems) are commonly used to support strategy in modern corporations.

Such standard, prescriptive models of strategy are emblematic of traditional, mainstream assumptions concerning organizational intelligence and how organizations address their environments (for the recent debate in the field see Priem *et al.*, 2013). For these reasons, we deal only in passing with other theories of strategy with less prescriptive undertones such as strategy as practice (Jarzabkowski *et al.*, 2007; Carter *et al.*, 2008) or emergent strategy (Mintzberg, 1991). At any rate, ours is not an alternative theory of strategy. We use standard approaches to strategy as a means of showcasing how the developments associated with big data erode the very ground on which widely diffused models of decision making associated with strategy as prescriptive game rest. The central place of strategy in our paper is motivated by the fact that big data is commonly presented as a context within which business opportunities or other sources of change in the environment, such as new social trends, can be identified (George *et al.*, 2014), and these are traditional and standard concerns of prescriptive approaches to strategy. Contrary to what is commonly advocated in the current burgeoning literature on big data, the developments associated with big data challenge many of the canons of standard, prescriptive approaches to management and strategy. Much of what comes under the heading of big data is not collected intentionally (Varian, 2010), it is haphazard, hugely heterogeneous, and, not infrequently, trivial, messy and agnostic (Anderson, 2008), as it happens with much user generated content and data logs of various kinds.

Big data can be made business relevant subsequent to its collection but such relevance may not straightforwardly derive from the original data records. It is important therefore to make clear that such data cannot often be used as a direct input to strategy making. Its by and large unstructured or miscellaneous nature fits uneasily with the existing strategic tools and methods, and the maxims derived from them. This raises a number of challenges in the contexts in which these tools and methods are drawn upon to support business objectives or other purposeful interventions (public or third sector organizations). In order to address and exploit big data, existing models and tools based on predefined data collection will need significant modification, if not replacement. For example, the constant updatability big data affords enables real-time responses (known as nowcasting) that involve sophisticated algorithms dealing with dynamic datasets and requiring the developments of new indicators as well as new forms of graphical or visual representation that support sense making under conditions of rapid and shifting environmental change. In other cases, such as those occasioned by social media platforms, the generation of large amounts of social data reconfigures the relationship between an organization and its surrounding constituencies (Oestreicher-Singer and Zalmanson, 2013). In such contexts, social-platform users become producers of information and consumers of services based on the information they themselves produce, as shown by recommender systems and personalization services.

These observations indicate that the issues big data raises for strategy are part and parcel of a wider change that transcends the nature and business usage of big data and concerns organizations as social entities. Whatever the attributes of big data, these are closely associated with the mechanisms or processes by which big data is generated and made available. It is therefore important to emphasize that in a great deal of circumstances big data is produced by a hugely heterogeneous user base, actively through user generated content or indirectly through various kinds of data logs and tracking devices. These mechanisms of big data generation and sharing challenge the technical, expertise based models by means of which data practices and organizational intelligence have developed over long time. Big data is different data, thanks to it being generated by social and organizational arrangements that vary substantially from the technical models and expertise structures that have traditionally been deployed in organizations. Internet and social media users do not often belong to the organizations or networks to which they contribute data. They are not members of these organizations (at least not in the ordinary sense) nor are they commonly utilizing meticulously worked out data models or practices (marketing or accounting procedures, professional classifications) for data generation, as Facebook, Flickr, YouTube or other social networks show. Big data owes much of its distinctiveness to the mechanisms by which it is generated and the messy or trivial everydayness these mechanisms help install at the heart of the processes of data generation and use. In this regard, strategy making is part and parcel of a wider context of social relationships that are shaped in ways that tend to redefine organizational membership and boundary maintenance, and ultimately modes of producing data and consuming data-based services.

The paper is organized as follows. Next section presents an overview of research in strategy making in order to assess the role information is assumed to play in standard models of strategy making and describes the properties of information used. This is followed by an account of the distinctive profile of big data and the social arrangements or mechanisms that make big data different from canonical sources of information, used in purposeful action in general and strategy in particular. We subsequently move on discussing how the developments associated with big data challenge some of the fundamental assumptions on which data and information have traditionally been generated and made relevant for strategic purposes. We identify important implications for strategy making, highlighting specific changes in the context of strategy. In the final section, we assemble the key claims of our paper and outline some wider implications of big data for management and organizations that to a significant degree run counter to the prescriptive role accorded to big data by the burgeoning literature on the subject.

The role of information in strategy making: An overview

External, internal information sources and competitive strategy

Industry competition has been the focus of research for a large number of strategy scholars. Schumpeter (1934) described the dynamic market process through which firms compete in a "perennial gale of creative destruction," and depicted a disequilibrium leading to shifts in market status quo. Since then, researchers have investigated competitive actions taken by leaders and challengers in the market, focusing on different aspects of the process such as competitive dynamics (Ferrier *et al.,* 1999) or hypercompetition (D'Aveni with Gunther, 1994). These studies underline the importance of sensing the environment, by collecting information on market dynamics and competitors' activities, in order to maintain or improve the market position of a firm.

As a means to building and sustaining competitive advantage, traditional strategic planning techniques and tools focus on the analysis of the external environment and the identification of the firm's strengths and weaknesses in relation to it. Well-known models for analysis of the business environments have originated from the field of industrial organization (Caves, 1964), which takes a positioning view of strategy, such as Porter's five forces model (1980; 1985). The positioning school has had strong influence on strategy making since it provides its philosophical outlook, logical justification and the guidelines for data collection and use (Gavetti and Rivkin, 2007).

A number of strategic tools, still used frequently and offering more detailed analysis of the industry, have been developed. Examples include competitor analysis, which provides information to the firm on its main competitors and allows it to predict their behavior; strategic group analysis, which provides performance comparisons for firms with similar business models or similar combinations of strategies (Newman, 1978; Porter, 1980; 1985; McGee and Thomas, 1986) also SWOT analysis (Learned *et al.*, 1965). These tools provide a "snapshot" of the industry environment under consideration, offering a static picture at the moment data is collected. The data is each time collected based on the requirements of specific quantitative measures, defined by theory as representative indicators of different environmental characteristics (e.g., market share, Herfindahl index for market concentration). These indicators entail a simplified representation of the business environment and provide information then used in strategic decisions. The fundamental assumptions behind these strategic tools derive from models of industrial organization and the neat, simplified picture such models convey about the conduct of firms and economic agents. For example, it is assumed that the firm's main objective is to maximize profits, with the key competition dimensions involving cost minimization and product differentiation (Porter, 1980; 1985).

The strategic fit to the external environment is also the main focus of strategy researchers who view the firm as a collection of resources, bundled together by a specific governance structure (Penrose, 1966; Wernerfelt, 1984; Peteraf, 1993). Ownership of a superior resource is considered a critical source of competitive advantage (Barney, 1986). In this respect, the internal view of the firm is relational. While predominantly focusing on the internal relations of the firm, this view still draws on key assumptions of theories of industrial organization, highlighting the importance of the firm owning heterogeneous resources vis-à-vis its competitors. Know-how and other types of idiosyncratic tacit knowledge or long formed IT capabilities (Lim *et al.*, 2011), are typical examples of resources that can create and sustain competitive advantage in this context. Information technologies can be a valuable resource when combined with co-specialized assets or used in business partnerships (Melville *et al.*, 2004).

The internal view underlines the need to collect data on business activities and performance indicators in order to identify reliable signals of efficient processes and mechanisms. Internal data is combined with external data from the environment, enabling the firm to identify new business opportunities and exploit its superior resources. Many firms have historically created business intelligence units to collect data and calculate departmental performance based on balanced scorecards and use of key performance indicators that are expected to reveal through reliable information sources specific organizational capabilities or identify weaknesses. Recently, a number of researchers have underlined the need to use evidence-based management, and especially the development of indicators derived from theory, to increase the validity of the measures and the reliability of the data (Pfeffer and Sutton, 2006).

The focus on industry dynamics and firm's unique resources have been central in the analysis of strategic information systems impact, especially in the context of a network economy characterized by competition in volatile environments, high speed technological change, and uncertainty (Castells, 2001; Merali *et al.*, 2012). A number of sophisticated methods and tools have been developed, enabled by the ever increasing processing capacity of information technologies, to support strategic planning (e.g., decision support systems, enterprise resource planning, business intelligence tools). These solutions address specific information needs; for example they reduce transactions costs (Malone *et al.*, 1987; Ray *et al.*, 2013) by improving coordination of transactions with other business partners (Lajili and Mahoney, 2006) as well as supporting the governance of transactions within organizations (Ciborra, 1993; Cordella, 2006). In particular, scholars highlighted the importance of information technologies in relation to opportunism and operation risks as well as coordination risks (e.g., Clemons, 1993; Cordella, 2006).

Despite the undeniably important developments of information technologies providing opportunities for much improved information provisioning, some major challenges still remain, especially in the domain of business intelligence. First, the uncertainty surrounding strategic planning activities is still prominent since exogenous change of discontinuous status cannot be predicted. Most of the forecasting models use a backward looking approach and are based on historical data in the form of time-series that enable the calculation of specific theory-derived and validated measures and act as indicators of an environmental trend. Data is collected in discrete time intervals, e.g., annually, monthly, weekly. Second, the use of business intelligence output in strategic decisions is sometimes problematic because of the unreliability of data sources. Despite the theoretically validated nature of the measures or quantitative indicators, the lack of reliable data may counterbalance the benefits from their use in strategic decisions. As one example, the majority of investment decisions seem to be made based on underestimated costs and overestimated expected benefits from the project (Shollo and Constantiou, 2013). Third, existing measures are not always the appropriate indicators of emerging environmental characteristics or trends (e.g., value co-creation) or intangible parameters (e.g., intangible benefits such as improving the corporate image) which need to be included in the assessment of an investment decision (Shollo and Constantiou, 2013).

These theories from the positioning school and the internal view of the firm underline the need for several sources of information necessary to strategy making. Among these sources figure market share and concentration indicators, competitor information, consumer market information and information derived from financial and accounting statements. As a rule, information of these kinds is static, periodically updated and by and large specific and structured. None of these theories focuses however on the detailed processes by means of which these sources of information are used, combined and interpreted to become relevant to strategic pursuits. For this we have to turn to other, or more recent, theories of strategy.

Routines and dynamic capabilities as mediating structures

Firms are not just placeholders of resources but social units as well, marked by several characteristics that filter and mediate information about the internal and external environment. Evolutionary economics originally formulated by Nelson and Winter, puts specific emphasis on organizational routines conceived as crystallizations of a firm's experience (1982; 2002). Routines are treated as building blocks of organizational capabilities (Dosi *et al.*, 2001). On this view, the firm's search for business opportunities is mediated by the routines that represent the codification of learning from experience directing attention to local or known events (Nelson and Winter, 1982). In this respect, routines can be viewed as institutional mechanisms that store experience and govern attention to internal and external stimuli. Observed environmental cues and information, collected to provide input to strategic choices, are thus closely associated with the firm's routines.

Building on the internal view and the notion of strategic fit to the environment, researchers have investigated growth strategies through the concept of dynamic capabilities (Teece and Pisano, 1994; Teece *et al.*, 1997). Dynamic capabilities include cognitive skills and organizational processes which are very difficult to imitate. Recently, Teece (2007) described dynamic capabilities as the capacity '(1) to sense and shape opportunities and threats, (2) to seize opportunities, and (3) to maintain competitiveness through enhancing, combining, protecting, and, when necessary, reconfiguring the business enterprise's intangible and tangible assets' (1319). Information flows from both internal and external sources play an important role in dynamic capabilities and their strategic use. The three types of dynamic capabilities described by Teece underlie strong intentionality, as the firm selects from the business environment opportunities compatible with its capabilities, in an attempt to achieve 'best fit'. This approach suggests that the scanning of the environment is based on collecting data with the purpose of finding the optimal fit between the firm's capabilities or routines and the characteristics of the environment.

Teece (2007) identified specific processes of sensing market and technological opportunities in the business environment, for example, processes to tap suppliers and complementors' innovations, or processes to identify customer related opportunities. These processes are influenced by the organizational structure, as well as the members' capabilities and knowledge. According to Teece (2007), discovery and creation are grounded on the organizational (collective) processes but also depend on individual capabilities and personal extant knowledge. The governance structure supports the collection and use of information in strategic management; organizational processes are in place to collect and filter information, while hypothesis development and testing in relation to the meaning of information should be done by top management (Teece, 2007). Next, dynamic capabilities for seizing opportunities are required for the firm to sustain its competitive advantage. These involve investment decisions. Literature on investment decisions and finance offer a number of models and rules that provide guidance for project selection (e.g., discounted cash flows, payback periods and risk assessment models). These models are populated by data derived from systematic information processes that feature expert knowledge and rules, and professional classifications.

Some researchers view the business environment as an ecosystem in which the firm should collect and use information from all partners, when sensing the environment for new business opportunities and seeking input to its competitive strategy (Iansiti and Levien, 2004). As the unit of analysis shifts from the industry to the business ecosystem, the need for research to identify business opportunities transcends local conditions and spans over the core and the periphery of ecosystems. The underlying complexity of the ecosystem requires more sophisticated approaches to discovering and exploiting business opportunities. This raises a fundamental problem in the use of existing techniques and financial tools that cannot capture the complexity of ecosystem relations and the value of co-created future actions of the firm with a range of suppliers or customer groups (Pitelis and Teece, 2009). Some of the most popular tools, such as net present value (NPV) and return on investment (ROI) are based on simple measurements that provide information on specific dimensions of a project, like return on investment or the present value of future cash flows. Another challenge is that the data required from existing financial tools to provide an accurate result are not readily available, (because it involves aggregate data using different sources – an example is when measuring returns on co-created activities by more than one firm, compared to the single source used previously).

The current research agenda on strategy highlights the role of managerial cognition in interpreting the environment and shaping strategy (e.g., Gavetti and Levinthal, 2000). An organization's sensors select information from the environment and channel it to the managers. The information only becomes meaningful after interpretations through the lenses of specific mental representations and values of the managers (Gavetti and Rivkin, 2007). According to Gavetti (2012), managerial cognition plays an important role in identifying cognitively distant business opportunities, which should be matched with organizational capabilities, and are then expected to lead to superior performance of the firm relative to its competitors who may not be able to see these opportunities because of lack of the necessary cognitive skills, e.g., analogical reasoning. Eggers and Kaplan (2013) have underlined the pivotal role of routines as the main building blocks of capabilities that are linked through managerial interpretations to specific strategic choices.

Overall, these theories do not entail or prescribe new sources of information, even though they occasionally underline the need for extending traditional sources of information, as is for instance the case with business ecosystems (Drnevich and Croson, 2013). They view the firm as the main actor, either adapting to or shaping the business environment based on its capabilities and the managerial cognition to sense and seize opportunities. Most of research in this area underlines the prescriptive conclusions that firms need to develop more efficient routines, and build on their capabilities to exploit new business opportunities by combining internal information with external information of the competitive environment. This, in turn, can be achieved by efficient application of theoretical models and tools and expert knowledge that by and large feature top-down processes of strategy making.

Anatomy of Big Data

Data practices and mechanisms of data generation

Do the trends associated with big data make any difference to the problematics of strategy outlined above? As indicated in the introduction, much of the literature on the subject assumes that big data can be recycled into the usual apparatus of strategy making and used to enhance and occasionally expand traditional techniques of designing new products and services, and planning, measuring and controlling organizational outcomes. Such a view assumes that big data is similar (more of the same) to data that firms and organizations have traditionally drawn upon to formulate and implement strategies. To some degree, such an assumption is warranted. Part of big data derives from traditional sources that have been rendered interoperable. Even where big data is different in the sense of being unstructured, it is assumed that it can be made relevant to firms and organizations through various operations of tidying and systematizing (O'Reilly, 2012).

We argue, however, that the developments with which big data is associated establish a new and distinctive context for data generation and use. In many cases, big data represents no more than haphazard user-generated content or default recordings of people's whereabouts in social media, internet sites or other contexts of social life that are digitally recorded (e.g., Facebook "likes", blogs, YouTube video uploadings, RFID sensors, CCTV). Masses of data of this sort flow daily to social networks, corporate databases and other data repositories (IDC, 2011; Schmarzo, 2013). In providing details about user behavior, part of this data can certainly be used to monitor contracts and develop personalized services, as Varian (2010) suggests. It can also be drawn upon to chart the flow and availability of people and resources within the bounds of a system (e.g., firm, platform or network, city, health care system) in ways that allow meaningful interventions. A close look, however, at the mechanisms by which big data is generated and shared suggests a qualitative change of the context in which data collection and use intended to serve strategic goals have traditionally made sense. Such a change is not immediately evident and demands considerable effort to unravel.

In the traditional context of strategy (and organizational governance), data collection and use have, as a rule, taken place along well-trodden paths and by means of established and expertise-based practices that served particular corporate purposes such as control and accounting, finance or marketing. But even within the wider context of society, data have been collected along deliberate and premeditated paths that addressed specific information needs of the state, professions, trade associations and other economic agents (Porter, 1995; Desrosières, 1998). In most of these cases, data collection and use occurred via the deployment of established systems (e.g., accounting and auditing, statistical records), classifications and categories, and topical (theme based) or alphabetic data organization schemes (Bowker and Star, 1999; Rosenfeld and Morville, 2002). In other words, data generation and use have always been bound with one or another form of expertise through which selected facets of reality have been singled out for measurement or monitoring. Organizations continue to find it important to uphold these purposeful and systematic techniques of data gathering, in which data has always been part of some wider cognitive structure provided by categories, classifications or other cognitive placeholders such as alphabetic and thematic data schemes.

Technical, computer-mediated models of data gathering and management (e.g. databases) introduced important innovations to these established and expertise-based categories and schemes (Rosenfeld and Morville, 2002), some of which we deal in some more detail below. Suffices here to say that within the normal space of organizations, computed-based models of data collection, generation and use continued, by and large, this systematic tradition by encoding carefully engineered data fields (along which data was entered, stored and processed) that responded to the functional demands of the contexts within which they were deployed. Some of the trends big data signifies may be compatible with these established and expertise-based methods and techniques of data generation and use. Yet, others such as user generated content, tagging and other forms of social data are probably not. It is not evident how masses of haphazardly generated data that reflect trivial and passing concerns of large and diffuse user crowds can be made relevant for business purposes and, if so, by what processes of data reduction and aggregation.

Obviously, in the current context dominated by computer technologies, even the most haphazard data collection needs an underlying technical infrastructure of data fields, data structures and architectures through which data are captured and stored. However, it makes a difference, and a rather important one, what types of data such a system admits. By the same token, it makes a great deal of difference whether data is gathered through a carefully laid out cognitive (semantic) architecture or, by contrast, is captured and stored without such a plan and on the assumption that it may be variously used a posteriori. This is a fundamental difference that Weinberger (2007) captures by the terms “sorting in the way in” versus “sorting in the way out”. “Sorting in the way in” implies a clear structure and information architecture of data collection whereby data is ordered and its location fixed once-and-for-all at the moment it enters a system or data infrastructure. “Sorting in the way out” entails the capacity to meaningfully categorize and assemble unstructured and miscellaneous data and information that have been gathered or generated on loose premises. It thus seems to us worthwhile to have a closer look at the changes which big data introduces to the prevailing traditions of data collection, generation and utilization supposed to address organizational needs in terms of strategic intent.

The profile of big data

If big data were just a matter of data volume and diversity as well as speed of data growth, it would have most probably continued a long tradition of data collection, generation and use, and the statistical methods and techniques by which large and diverse data volumes have been dealt with (Beniger 1986, Porter 1995). There is, however, something qualitatively different in the developments associated with big data that tends to be obscured by the ways the relevant issues are framed in the mushrooming literature on the matter.

To begin with, much data produced and stored in current digital ecosystems represent a hugely *heterogeneous* spectrum of transactions or behaviours. Such data is the result of haphazard behaviour of user crowds (much user generated content) or is motivated by very generic purposes (such as logs or CCTV records, Facebook "likes") on the assumption that it may become useful or relevant at a later point, and usually at an aggregate level. It seems important to us to disclose how aggregation, by which big data is intrinsically connected, works. As the term indicates, aggregation is concerned with big lumps of data. It therefore implies a transition from particular data records to whatever patterns or relations can be extracted out of large data pools that accrue by means of lumping up individual or function specific data tokens (Chen *et al.*, 2012). In such a context, data of this sort (e.g., a Facebook user's "like" or Google logs) often has little value as individual or specific data pieces.

It comes therefore as no surprise that aggregate records deliberately displace or otherwise sketchily represent the specific conditions or context within which what is recorded makes sense. Context simplification is certainly an attribute of all standardized and aggregate representation (Desrosières, 1998; Ekbia and Evans, 2009). Yet, such an attribute is strongly reinforced by the massive, vastly heterogeneous and, crucially, agnostic nature of default recordings. The same holds generally true for user-generated content whose specific or contingent character is abstracted away in the huge records in which content is registered and aggregated. The ways by which such data is simplified make it scarcely relevant for the ordinary contexts in which it occurs and where the value of information is strongly tied to its specificity (Ekbia and Evans, 2009). It can be recycled to such contexts but only on the basis of some similarity or other generic (decontextualized) attributes that emerge as the outcome of aggregation and the comparison and clustering aggregation affords. In other words, aggregation tends to trade context specificity for generality, an issue that becomes increasingly trickier the greater the heterogeneity and context embedment of the original data are.

Second, and as a consequence of the above, data thus amassed *escape the systematic nature of professional classifications* and other methods of recording, measuring and assessing human efforts, such as those entailed in accounting or auditing systems (Bhimani and Willcocks, 2014). Formal systems of ordering represent an approach to record keeping, classification or measurement characterized by high observational selectivity that is supposed to serve particular projects or purposes. In bookkeeping or statistical records, for instance, data entry takes place along well-trodden paths and according to standard rules and procedures. Such carefully engineered data records have been the longstanding outcomes of professional definitions and longstanding cumulative learning based on practical problem solving. In an analogous way, classifications are made of categories that are the outcomes of professional definitions (e.g., medical or legal records) resting on longstanding empirical observations, scientific developments and pragmatic compromises (Desrosières, 1998; Bokwer and Star, 1999; Bowker, 2005). No matter how much or how often formal systems of ordering and calculation can be bent, worked around or otherwise manipulated to serve specific interests, they exemplify a very different approach to data gathering than the current trends we have been outlining.

These techniques have been abundantly used in traditional marketing strategies that more than any other managerial or economic field has ventured to deal with the diffuse landscape of daily habits, occurring in the contexts of domestic or public life. Traditional techniques of market segmentation have relied on standard criteria for category making such as age, income or life style. The loose criteria by which data is generated and stored in the internet and the current digital ecosystems have occasionally been designated by the term 'folksonomy' as a contrast to the taxonomic nature of classifications (Morveille, 2005; Weinberger, 2007). As opposed to the tidy nature of the latter, folksonomies entail the production of data out of the daily concerns, trivial matters and contingent pursuits of online user crowds, as is the case for Flickr photo or YouTube video tagging. These developments signify a friction between expert-based versus lay cultures and, in this regard, introduce significant modifications in longstanding traditions whereby data has been generated and used.

Third, contemporary data repositories often entail data of various formats. Such formats often cross the border of *alphanumerical* systems that have populated organizations and shaped their management, and include varying cultural artefacts cast in the media of *text, image and sound*. This is far from trivial a development. What is often called unstructured data refers to data tokens other than alphanumerical. Traditional systems of data collection have been made possible through carefully crafted data fields, systematic data entry, enumeration, clustering, aggregation and processing. None of these techniques are directly transferable to big data, specifically big data *qua* social data. In several cases such as those represented by data logs, RFID records or aggregates of structured data (e.g. credit card transactions), some of these older techniques still apply. Yet these techniques become less suitable for capturing the nature of events that populate social media platforms such as Facebook, Instagram, Twitter or Google+. How is that daily updatable text mass (and mess) to be reduced, made sense of and used for specific purposes?

The problem gets really intractable in the case of digital images that exhibit a different anatomy or semiotic composition and whose significance is bound to rise in the convergence space current technologies of computing and communication construct. Even though digital images can ultimately be reduced to bits and bites and dealt with computationally (Borgmann, 1999; Eco 2000; Manovich, 2001), it is important to recognize that their meaning cannot be thus controlled and manipulated (Kallinikos and Mariategui, 2011; Mingers and Willcocks, 2014). An aggregate analysis of images and visual records in general (e.g., YouTube video content) posits formidable problems of relevance and meaning about which the debate on big data is virtually silent. Most of this data is still accessed and treated on the basis of brief verbal descriptions known as metadata, e.g., a tag that describes an image. Yet the image and the tag (text) are very different cognitive and cultural artifacts (Kallinikos and Mariategui, 2011). An aggregation of titles, topics or any other attribute of visual records may offer an inspection or visibility of whatever attributes are thus measured. This inspection or visibility that literary theorist Franco Moretti (2005) calls *distant reading* may be variously useful. But it can scarcely say anything substantial about the content (that is, *close reading*) of individual or less inclusive aggregates of visual records.

Fourth, much of what comes under the heading of big data entails data repositories or platforms whose value is intimately related to their constant *renewal* or *updating*. The quest for continuous data generation and constantly updatable data sources is immanent in the internet, its real time orientation and the instant or quasi-instant communication habits that current digital ecosystems promote. Yet, what thus becomes updatable inevitably moves along a short time horizon the privileges the present at the expense of past and future. Such a strong focus to the present inevitably reframes the established time divisions (past, present, future) and the significance they have had on the planning and management of organizations (Kallinikos 2006, 2009).

Taken together, the attributes of big data and the mechanisms by which it is generated and used bring important changes to entrenched paradigms through which data has traditionally been collected to serve the strategic objectives of firms and organizations. Big data is not just data of any sort. The developments with which it is associated signal important shifts in the mechanisms of data generation and the nature or quality of data that such mechanisms generate. These developments, we suggest, have far-reaching implications for management and strategy making. It should make a rather stark difference whether data is gathered on the basis of a well-trodden cognitive architecture that serves preconceived organizational purposes and objectives or whether it is the data itself that is used to distill a posteriori an organizational plan, a purpose, a human intention, a service or a set of services. Let us now turn to the consideration these critical issues have for strategy.

Big Data and the Changing Context of Strategy

The ideas presented in the preceding section suggest that big data tends to be heterogeneous, unstructured or semi-structured and agnostic. It is also trans-semiotic, entailing different combinations of data formats and communication modes that extend beyond alphanumeric systems to include tokens cast into the media of text, sound and image. We also associated big data with a pronounced focus on real time events, which undermines standard premises of strategy making and challenges long entrenched cognitive and behavioural habits based on a linear conception of time and long-term commitments (Bhimani and Willcocks, 2014).

These attributes of big data and the social arrangements or mechanisms by which big data is produced suggest a more cautious approach as regards the relevance of big data for strategy and more widely the management of organizations. There is no doubt that big data may be variously drawn upon to assemble the big picture of phenomena or dissect the behavior or attributes of large segments of population, such as spending habits of people in designated dates, hours or urban spaces, vehicle traffic patterns in cities or motorways. Such data have not been easy to compile before and can be a significant input to strategy making, for example offering opportunities to develop new services or prepare policy interventions. To illustrate, business applications, which offer traffic monitoring services, build on location-based technologies and use data on actual traffic patterns (Brynjolfsson and McAfee, 2014). Similarly, new business opportunities based on data mining of consumption patterns in relation to particular places or events are being exploited in mobile business ecosystems (Clemons, 2009; Chen *et al.*, 2012). In other contexts such as those represented by the management of cities, critical policy interventions can be derived from the mapping of traffic and people across metropolitan spaces made possible by the developments that are associated with big data.

In yet other cases that stand close to standard techniques of data generation, such as those represented by profiling methodologies and recommender systems, big data may make important contributions to marketing strategies (Davenport, 2014). Recommender engines are currently ubiquitous and can address personal needs by recommending potential consumption items, services, habits, experiences or persons (e.g., dating recommender systems) on the basis of an individual's affinities to other users worked out algorithmically on the basis of massively available and constantly updatable data pools. Overall, such systems offer data on the basis of which standard segmentation strategies can be radically reframed, improved or even abandoned as audience behaviour can be tracked through logs and other actual, as opposed to assumed, choices (Aaltonen and Tempini 2014). Online retailers (e.g., Amazon) successfully use big data for several marketing purposes (Clemons, 2009; Davenport, 2014). In yet other cases, data generated internally by means of routine operations can alone, or in combination with externally sourced data, be crunched through automated techniques of data clustering, aggregation, comparison and association, data mining and profiling, and used to suggest important patterns and trends in the external and internal environment of the firm or organization (e.g., Davenport *et al.*, 2005; boyd and Crawford, 2012; Davenport, 2014). Insurance companies draw on a variety of data to better design premium policies and enforce contracts (Ayers, 2007; Varian, 2010). Successful examples of this type of use of big data by businesses have been repeatedly described in practitioners' reports (e.g., McKinsey or IDC).

However, the attributes of big data summarized above and the radically different social arrangements (internet, social media platforms) through which big data are generated suggest a more complex picture that challenges key premises of strategy as predictive game.

Top-down versus bottom-up strategic decision process

First, the top-down, deductive approach to data gathering and utilization, a cornerstone of standard and predominantly prescriptive ways of depicting sensible strategy making, is challenged. Many of the methods and techniques of data crunching are strongly associated with bottom-up procedures of data processing that are supposed to (re)discover patterns in huge data masses. The overall scheme, which these methods epitomize, can roughly be summarized as follows: first data then search for any possible uses of what is already available as data (Anderson, 2008; Lee *et al.*, 2014). An *ad hoc*, inductivist way of strategy making seems to be emerging as the outcome of the trends we are describing. This seems to undermine the foundations of predictive models of strategy making based on the purposeful, deductive and ex-ante character of organizational intelligence and information collection and use.

As briefly indicated in our introduction, the top-down, deductive approach to strategy has been severely criticized by what currently goes under the label 'strategy-as-practice' (Jarzabkowski *et al.*, 2007; Carter *et al.*, 2008). However, the challenge to strategy stemming from the developments we associate with big data is of an entirely different dimension and scale, occurring at aggregate levels that transcend the "nitty-gritty" practices by which situated decision makers make sense of locally embedded phenomena, tweak and fix data, descriptions and objectives. There is little doubt that specific organizations may need to work out the details, skills, practices and mindsets for capitalizing on big data. Yet, the overall picture cannot adequately be grasped by such context-sensitive approaches. The developments we pinpoint are of a larger scale, framing attention and sampling events before reaching local contexts and they cannot therefore be dealt with the local, embedded processes of meaning making valorized by strategy-as-practice. Nor can they be accommodated by the behavioural models of managerial cognition (Gavetti, 2012; March, 1999), the cognitive habits or skills of managers, and the routines to which these are tied.

This pragmatism, as it were, of strategy cannot address the shifting business parameters introduced by big data. For big data epitomizes the move to de-contextualization *par excellence*. If anything, the trends we describe transcend local contexts of strategy making, inserting at the heart of the process of strategy making agnostic models of data gathering, featuring decontexualized, standardized, technology-driven ways of sense making. We expect these shifts to have a profound change in those old and recent theories and models proposed in the strategic management literature that have focused on managerial cognition, understood either in rationalistic or behavioural terms (e.g., Gavetti, 2012) and that have overemphasized the role of executive managers in interpreting and using environmental information for making strategic choices. The new way information is generated, aggregated and presented to managers will significantly alter the importance of their mental representations and more generally reframe the role individuals have assumed in the process of strategy making.

The challenge big data posits to the top-down, deductive approach to strategy is closely associated with the mechanisms of big data generation. It might be useful to distinguish data generation from data use around which most of the current debate on the subject evolves. Certainly, the attributes of big data we summarized earlier (unstructured, heterogeneous, agnostic, trans-semiotic) do not fit well with the structured and relatively tidy sources of data used by traditional, prescriptive models of strategy. It is also important to recognize that these attributes condition what can be done by big data. But these, no doubt, crucial attributes are closely associated with the mechanisms by which big data is produced. It is diffuse and distributed lay actors *en masse* rather than experts or other kind of dedicated personnel that lie at the heart of big data generation. These different arrangements and mechanisms lends big data many of its distinctive and also fascinating qualities. The circumstances of big data production are, in most cases, not controlled by organizations, nor are they subject to the widespread principles of expert rule on which data and expert knowledge have commonly relied.

An interesting and dramatic reminiscence of these observations is provided by the ever-increasing volume of data in the form of images and videos that massive user involvement in social media platforms generate. Existing strategy models, tools and indicators cannot accommodate these data formats and genera. They simply have no use for data of this sort that traditional views still bound to the realm of amusement. Researchers in the field of strategy and management have not investigated how to analyze and use this data in a systematic way in strategy making. Yet, the richness of this data may offer new perspectives on social life and specific behaviours and thus enable in-depth understanding of market or social trends if analyzed effectively. For example, video recordings of people performing organizational activities are shown to be highly valuable for researchers in the field of managerial cognition, as such records are able to capture behavioural reactions and use alternative means of analyzing strategic decision-making (Connors *et al.,* 2014). However, current knowledge on how to use video data and images are far from developing robust models and tools for capturing the rich and dense information provided by this data.

Time horizon: Long term versus short term decision spans

Secondly, as briefly mentioned above, data produced by means of the processes and operations we associate with big data results in the steady updatability of the relevant data sources (Kallinikos, 2006; 2009). User generated content changes every moment as users add *en masse* perpetually new content. Logs and similar records are modified continually as user crowds enter and exit websites and social media platforms. Under these conditions it is difficult to forecast future trends. To put it succinctly: the present is only fugitive while the past does not provide solid evidence for what is to come. Strategic tools geared to make long-term predictions based on standard statistical ways of inference do not work very well in these environments. For example, "likes" in social media platforms provide a live dataset that is difficult to use in traditional marketing tools, not least because it is challenging to represent such "likes' in meaningful forms of sense making (e.g. visualization) of enduring value. The same holds true for the patterns constructed by recommender engines, as new user choices *en masse* (added transactions or purchases in a day) continually reconfigure the relations between older data items or persons.

Updatability is, in this respect, a double-edged sword (Kallinikos, 2009). The adaptation to emerging conditions that updatable data enables reduces the purchase of older data descriptions and, not infrequently, speeds up the rate of their obsolescence. These short life spans of data generated under the conditions with which big data is associated challenge another cornerstone of strategy as depicted by standard wisdom: its long term commitment and capacity to make predictions that transcend the immediate present of real time and nowcasting. Real time orientation (nowcasting) may be particularly valuable for location-based applications or businesses where it is important to adjust constantly to the current market trends and habits of customers (e.g., fashion or music industry). But the focus on real time also undermines long-term planning, and reframes the trade-offs between short-term and long-term decisions (Bhimani and Willcocks, 2014; Choi and Varian, 2012). The distinction between operational (tactical) and strategic decisions relates to both subject importance and time length. Strategic decisions are based on data and information that supposedly describe enduring conditions of the internal and external environment of the organization. Strategic decisions are also assumed to hold for a considerable period of time or, certainly, beyond the immediate present. The trends we describe and the attributes of big data seem to us to provide enough of an incentive for rethinking these venerable distinctions. Big data both makes what is going on in the present variously relevant to strategic decisions and undermines the significance forecasting has assumed in the traditional context of strategy.

Shifting premises of strategy making

A number of other important observations can be made in relation to strategy making. First, the underlying assumption that organizational routines and capabilities built on cognitive patterns solidify the firm's ability to adjust to the environment (Gavetti and Rivkin, 2007) becomes problematic. For, the firm's survival in the new environment is characterized by volatility conferred by the developments associated with big data. Second, in this volatile environment, where long term forecasts may be less relevant, the firm's search mechanisms cannot be based on the application of a rational (optimization) approach as suggested by the positioning school in strategy (Porter, 1996; Gavetti and Rivkin, 2007). Now, information ambiguity is bound to be pervasive. This also challenges most traditional strategy tools. At the same time, the evolutionary approach (Nelson and Winter, 2002) which highlights the importance of local search (see, Cyert and March, 1963; Gavetti and Rivkin, 2007) is also challenged. Why? Because the dynamic nature of the environment does not allow for choices exclusively based on the success and reality test of past solutions, that is, learning from experience.

Recent research on business implications of big data is caught in an enthusiasm of better understanding the business environment through the means big data is supposed to offer (Davenport, 2014). However, existing strategy tools are not adequate for the interpretation of big data trends, especially when looking for new business opportunities. In a recent editorial on big data of one of the most prominent academic outlets in the management field, the authors praised the importance of data volume as well as the value of outliers possible to extract from such data volumes (George *et al.,* 2014). On the view of the editorial, outliers can help identify the behaviours of people who drastically deviate from the average trends observed in big data. Given the data volume of big data, these outliers may include sizeable numbers and hence represent a source of a social or economic change with long-term effects on the business environment. Outliers then may form the basis for identifying new business opportunities in areas such as product innovation or new service development. However, the concept of 'outlier' is to a significant degree a technical (statistical) construct that makes sense against the background of a relatively homogeneous or standardized measurement that allows the distance from the average to be perceived, calculated and rendered socially relevant. The heterogeneous and unstructured nature of big data tends to make the concept superfluous or, in the best case, difficult to interpret. In other words, the outliers of big data aggregations result from miscellaneous data sources that involve a much wider range of abstraction and context simplification of the original data sources than is the rule otherwise. Outliers thus identified may be fabrications or not behave as they usually do in specific statistical distributions made of reasonably standardized observations. In practical terms, and since it is not possible to trace or understand the specific context from where the data is extracted, any interpretation of outliers observed in big data aggregations might ultimately turn out misleading.

Table 1. The nature of data and information: standard context of strategy versus big data

Standard Strategy Context	*Big Data-Digital Ecosystem*
Relatively Homogeneous	Heterogeneous
Structured	Unstructured or Semi-structured
Purposeful, Theory driven	Agnostic, Haphazard
Mono-semiotic, Alpha-Numerical	Trans-semiotic (Text, Image Sound)
Deductive, Top-down	Inductive, Bottom-up
Long-Term Horizon	Short-Term Horizon
Forecasting	Nowcasting

Table 1 summarizes in a simplified form the changes we have been describing in this paper. Standard strategy is still by and large predicated on an understanding of organizations that derives from the entrenched yet receding industrial paradigm of mass production (Fordism), clear and stable markets and identifiable client or consumer groups separated from these organizations by fixed and inflexible boundaries (Priem *et al.*, 2013). Organizations such as Facebook, Wikipedia or Google do not simply address users. In some important respects, they are the users by which they are populated. In similar contexts, users and user crowds cannot adequately be understood on the standard model of consumers or clients. They are consumers or clients and network participants at the same time. In some important ways, social networks are constituted by the user crowds that populate them and the data these online crowds generate. This takes us to some wider big data issues we need to consider.

Postscript on Big Data, Human Behaviour and Management

Data practices and organization Forms

The changes we describe and schematically depict in Table 1 are symptomatic of the contemporary digital ecosystem and the ways it allows daily and trivial concerns of users to enter institutional worlds via the Web, and do so on a massive scale. To some degree these developments can be viewed as signalling a transformation of contemporary economy and society. They are part and parcel of a much wider shift that makes e*verydayness* qua data imprints an intrinsic component of organizational and institutional life (Kallinikos 2006; 2009; Kallinikos *et al.*, 2013; Yoo, 2010), and also a primary target for a range of commercialization strategies. The importance of these developments emerges against the standard organizational templates, whereby external and trivial concerns have traditionally been kept outside the operative bounds of organizations, via a range of solutions from physical architectures to rigorous admission procedures (Luhmann, 1982; 1993; Thompson, 1967; Zuboff and Maxmin, 2003). Organizations have been bounded sites in which products and services have been produced via strict membership control but also sites of calculation in which data and information have been produced (and owned) by a range of specialized and expert-based operations from accounting to marketing (Bhimani and Willcocks, 2014).

Nowhere else are these trends better manifested than in the context of social media platforms and the huge amounts of trivial data they generate every day. As noted above, there is a large co-extensive territory between social data (i.e., data derived from social media platforms) and big data. The significance of traditional and increasingly interoperable sources of information notwithstanding, the distinctive nature of big data is closely associated with huge online crowd trafficking on social medial platforms and the Web more generally. In fact the value of many internet-based businesses and social media platforms is contingent on the size and, critically, the intensity of the activity of their user base. Though a closer look at these developments is beyond the scope of this paper, they do indicate that the changing context of strategy we describe is indicative of wider social and economic transformations. Big data as a label may stay or go, yet the processes we pinpoint are much more profound and enduring. They signal the blurring of long established social and institutional divisions, whereby the distinction between organizations and their environments, characteristic of standard ways of understanding and running firms and organizations, is radically reframed.

The changes we refer to in this paper therefore carry wider implications that transcend the context of strategy. If we are right then these changes concern the very nature of firms and organizations and their relations to individuals *qua* users, customers and clients, or citizens in the case of governments. In so doing, they recast management (including public management) as a field and social practice in a new context whose exact outlines still remain unclear (Yoo, 2013). Placed in such broader purview, strategy is never an isolated activity. It is rooted in organizational structures, established routines, labour divisions, industry relations and institutional solutions. It is also closely associated with longstanding conventions and methods mentioned above, whereby the division between internal and external environment of firms and organizations has been reproduced and reinforced.

In questioning standard maxims of strategy and the philosophy on which they are predicated, we offer an indirect critique of the rosy and often naively optimistic and ultimately rationalistic picture of the business role and functions of big data. We have deliberately steered away from this rather simplistic construal of the relevant issues, characteristic of a great deal of the current literature on the subject (e.g., Davenport, 2014; Mayer-Schonberger and Cukier, 2013; Schmarzo, 2013). At the same time, we have sought to remain open to the wider significance of the phenomenon and its IS and organizational implications. Big data raises many challenges for organizations, communities and individuals. As a way of demonstrating some of these challenges, we have problematized the context of strategy making. Viewed as an emblematic activity of standard organizational forms predicated on boundary maintenance and long-term trends and relationships, strategy is becoming one of the battlefields through which key trends associated with big data manifest. Making long-term choices becomes a tricky issue in a world dominated by the short-term concerns (Kallinikos, 2009) and saturated by the impulses, trivialities and notorious shifts of everydayness.

As far as we know, ours is the first systematic effort to produce an anatomy of the attributes of big data that ties the phenomenon to longstanding traditions of data collection and use for commercial and other institutional purposes. Data practices or techniques, organizational activities and forms have historically been linked to one another (Beniger, 1986; Kallinikos, 2006; Yates, 1989; Zuboff, 1988). Central organizational operations and structural attributes have been closely associated with established practices and techniques of information collection, processing and use. The recognition of this fundamental fact prompts, we suggest, the careful examination of the ways in which the distinctive profile of big data may reconfigure the link between data practices, economic activities and organization forms.

Unsurprisingly, some of the properties of big data we discuss in this paper have earlier been noted in the literature. However, the assembly of the heterogeneous, unstructured, agnostic, trans-semiotic nature of big data marked by short time value into a distinctive profile that carries important organizational implications is, as far we know, a novel contribution. The significance of that synthetic profile emerges more clearly against the historic background of data practices and cognitive methods and habits by which corporations, states and state agencies have sought to describe and control economic actions and outcomes. There is no way to appreciate the business, organizational and institutional impact of big data in depth, without meticulous attention to the centralized principles of data organization that have been characteristic of modern organizations, and the practices such principles have fostered (Beniger, 1986; Porter, 1995; Zuboff, 1988). This requires an open and interrogative attitude towards the possible novelties introduced by big data and the range of trade-offs that may emerge as the phenomenon diffuses throughout economy and society. We do not wish to dismiss other studies that focus on the immediate gains that big data could bring to people or organizations. Yet the limitations of such studies emerge clearly against the complexity of data and information practices and their social and institutional correlates.

The reality of big data

There are several other serious issues concerning the diffusion of big data that are worth revisiting as a way of concluding our paper. One pivotal issue concerns the relationship between descriptions of reality that are based on aggregate data and context sensitivity, a fundamental issue in IS research but also in social science at large (Ekbia *et al.*, 2014; Floridi, 2012).

Big data does not simply enable capturing events, states or conditions that would have otherwise evaded attention. There is no way to be all embracing and comprehensive without compromising variety, richness or complexity. There is no magic in what big data accomplishes. Aggregate descriptions become possible only through abstraction from the messiness of life and contextual detail. The higher the degree of aggregation is, the wider tends to become the richness or variety that is thereby compromised. This is, of course, a problem intrinsic to all statistical, or otherwise standardized, aggregate descriptions of reality (Desrosières, 1998). Yet big data, we suggest, accentuates that ever-present problem and reawakens the ghost of abstract or generic descriptions that may carry dubious social relevance (Porter, 1995; Desrosières, 1998). This is a fundamental issue that needs, at the very least, to remain open to critical scrutiny and cross-disciplinary debate. One gets a glimpse of the gravity of the problem by linking big data to the 2008 mortgage-originated financial crisis. While that crisis constitutes a complex and multidimensional issue, it can to a certain degree be attributed to the vague and, in many respects, misleading descriptions of reality mediated by the aggregate data, at the heart of the entire institutional apparatus of creating and trading financial instruments (Kallinikos, 2013).

A further extension of this problematic takes us at another thorny problem; namely, the relationship between the standardized or aggregate descriptions of reality afforded by big data and long established habits and forms of social interaction and communication. Social media platforms have often been seen as important means of sociality and community building. To a certain degree this is true. However, it is also important to recognize that the ways by which humans interact on social media platforms and the Web are heavily mediated by large amounts of user data aggregated and computed automatically with the aim of recommending links to persons and, in commercial sites, to products and services. This computed identity and sociality has by and large remained outside the purview of critical research and, certainly, beyond the picture of big data commonly celebrated for its immense business opportunities. As nicely described by Van Dijck (2013:12) " 'making the web social' in reality means 'making sociality technical' ". Some of these issues have occasionally been subsumed under the label of algorithmic power (e.g., Gillespie, 2014). Yet algorithms alone cannot capture the complex edifice of technologies and operations in which big data is embedded. Though certainly important, algorithms are the last step in a complex chain of data operations, data structures and architectures that harvest and make data available for aggregation and computation. Algorithms without data are just a mathematical fiction.

Social data versus big data

Do we equate big data with social data? There certainly are other ways of capturing and disseminating large amounts of data than those social media platforms enable to which we have made occasional references in this paper. These include sensor-generated data, log records, RFID systems, camcorders and CCTV; they also entail more elaborate information-based solutions of careful and deliberate aggregation of various registers or records into common data pools. Examples are again provided by data pools derived from different financial transactions; or credit card records, shopping records or medical records of various kinds. There is no doubt that these trends are part and parcel of the picture commonly subsumed under the label of big data (Aaltonen and Tempini, 2014; Bhimani and Willcocks, 2014; Ekbia *et al.*, 2014; Kallinikos, 2013).

And yet our observations suggest that much of the distinctive mark of big data derives from the social arrangements and mechanisms through which social media platforms afford the production and distribution of large amounts of data via massive, heterogeneous and dispersed user activity. User generated content, tracking of user whereabouts, modes of online socializing and the data they produce provide much of the distinctive mark of the age of big data. Short of these ways of capturing, sharing and using data, big data loses much of its characteristic flavour. It tends to become predominantly an issue of volume (boyd and Crawford, 2012; Van Dijck, 2013) possible to deal with by recourse to standard statistical ways of processing large and structured records. Social media platforms and social data introduce a qualitative break with these old and venerable methods of resourcing and processing large amounts of data. It is the massive online capturing of everydayness, the tweaks and trivial concerns of online crowds and the different social and institutional habits they embody that set social data apart from other modes of recording social events. It is equally the methods and techniques through which unstructured, ephemeral and trans-semiotic data call for analysis, relevance and meaning making. It is in big data *qua* social data that much of the novelty and reforming potency of the trends subsumed under the label of big data reside.

Does the stress we put on aggregation, standardization and abstraction deny the personalization, individualization and re-contextualization possibilities big data offers? In commercial activities, technologies that ride on large records made of aggregated individual choices (such as recommender systems and CRM systems) enable the extraction of insights on individual behaviour and the development of personalized services. Similarly, affinities between individuals, worked out through aggregation and computing capacity, enable the development and enactment of local or contextualized social relationships that would have otherwise not been possible (Van Dijck, 2013). Other examples may be provided by medical facts possible to extract out of large data pools through comparisons of records of individuals, communities or regions and apply them to particular cases. In all these instances, aggregation is just the ascending step in a far more elaborate sequence of steps by means of which knowledge and insight produced by data comparison (aggregation) are recycled to individuals and communities (Borgmann, 1999; Kallinikos, 2006; 2013; Zuboff, 1988) in the form of particularizing descriptions and solutions.

If achieved, these social goals and objectives no doubt show the potency of big data and the great contribution it can make to social life. But these benign prospects do not eliminate the risks of aggregation described above and the cognitive and social habits underlying the lumping up of different facets of reality. These debates on the pros and cons of aggregation, standardization and abstraction have a long precedent in social sciences, as the monumental book of Alain Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning*, we frequently referred to in this paper demonstrates. We want to keep them afresh and on the forefront, all the more so against the background of the utter simplicity with which the debate on big data is often conducted. Big data raises new issues but also reframes and occasionally reasserts the significance of old and rather vexed human questions. In a recent paper, Ekbia *et al.* (2014) nicely capture the way we should look at the broader picture of issues and questions related to big data. Framing them as dilemmas, they go on to note that:

> *'Some of these dilemmas are novel, arising from the unique phenomena and issues that have recently emerged due to the deepening penetration of the social fabric by digital information, but many of them are the reincarnation of some of the erstwhile questions and quandaries in scientific methodology, epistemology, aesthetics, ethics, political economy and elsewhere.'*

As a way of concluding, we restate that big data is part and parcel of wider developments that concern contemporary patterns of living on and off the Web. These in turn are closely associated with the ways by which information is generated, made available or shared, and information-based services are produced and consumed. To a certain degree, individuals, communities and organizations will be able to draw on this data to improve their choices and operations, including strategy making and strategy implementation as much of the mushrooming literature on big data claims. In yet other, important and fundamental ways that we sought to pinpoint in this paper, these developments keep on redefining the relationships between organizations and their environments (including individuals *qua* customers, consumers or users) that have been variously supported by older, established and usually centralized data and information practices. These issues need systematic research beyond the limelight of the hype.

References

Aaltonen, A. and Tempini, N. (2014). Everything Counts in Large Amounts: A Critical Realist Case Study on Data-based Production, *Journal of Information Technology* **29**(1): 97-110.

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, *Wired Magazine*, July, http://www.wired.com/science/discoveries/magazine/16-07/pb_theory (accessed 11 May 2014).

Arthur, B. W. (2011). The Second Economy, *McKinsey Quarterly*, October, http://www.mckinsey.com/insights/strategy/the_second_economy (accessed 11 May 2014).

Ayres, I. (2007). *Super Crunches: How Anything Can Be Predicted*, London: John Murray.

Barney, J. (1986). Strategic Factor Markets: Expectations, Luck and Business Strategy, *Management Science* **17**(1): 99–120.

Beniger, J. R. (1986). *The Control Revolution: Technological and Economic Origins of the Information Society*, Cambridge, MA: Harvard University Press.

Bhimani, A. and Willcocks, L. (2014). Digitisation, 'Big Data' and the transformation of accounting information, *Accounting and Business Research* **44**(4):1-22.

Bollier, D. (2010). The Promise and Perils of Big Data,The Aspen Institute, http://www.aspeninstitute.org/publications/promise-peril-big-data (accessed 11 May 2014).

Borgmann, A. (1999). *Holding on to Reality: The Nature of Information at the Turn of the Millennium*, Chicago: The University of Chicago Press.

Bowker, G. C. (2005). *Memory Practices in Sciences*, Cambridge, MA: The MIT Press.

Bowker, G. C. and Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*, *Inside Technology*, Cambridge, MA: The MIT Press.

boyd, d. and Crawford, K. (2012). Critical Questions for Big Data, *Information, Communication & Society* **15**(5): 662-679.

Brynjolfsson, E. and McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, New York: WW Norton & Company.

Carter, C., Clegg, S. R. and Kornberger, M. (2008). *A Very Short, Fairly Interesting and Reasonably Cheap Book about Studying Strategy*, London: SAGE Publications Limited.

Castells, M. (2001). *The Internet Galaxy: Reflections on the Internet, Business, and Society*, New York: Oxford University Press.

Caves, R. E. (1964). *American Industry: Structure, Conduct, Performance*, New Jersey: Prentice-Hall.

Chen, D. Q., Mocker, M., Preston, D. S. and Teubner, A. (2010). Information Systems Strategy: Reconceptualization, Measurement, and Implications, *MIS Quarterly* **34**(2): 233-259.

Chen, H., Chiang, R. H. and Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact, *MIS Quarterly* **36**(4):1165-1188.

Choi, H. and Varian, H. (2012). Predicting the Present with Google Trends, *Economic Record* **88**(s1): 2-9.

Clemons, E. K., Reddi, S. P. and Row, M. C. (1993). The Impact of Information Technology on the Organization of Economic Activity: The" Move to the Middle" Hypothesis, *Journal of Management Information Systems* **10**(2): 9-35.

Clemons, E. K. (2009). Business Models for Monetizing Internet Applications and Web Sites: Experience, Theory, and Predictions, *Journal of Management Information Systems* **26**(2): 15-41.

Ciborra, C. U. (1993). *Teams, Markets and Systems*, Oxford: Oxford University Press.

Connors, B. L., Rende, R. and Colton, T. J. (2014). Inter-rater reliability for movement pattern analysis (MPA): measuring patterning of behaviors versus discrete behavior counts as indicators of decision-making style, *Personality and Social Psychology*, **5**:605.

Cordella, A. (2006). Transaction Costs and Information Systems: Does IT Add Up? *Journal of Information Technology* **21**(3): 195-202.

Cyert, R. and March, J. G. (1963). *The Behavioral Theory of the Firm*, Englewood Cliffs: Prentice Hall.

D'aveni, R. A. with Gunther, R. E. (1994). *Hypercompetition: Managing the Dynamics of Strategic Maneuvering*, New York: The Free Press, Simon and Schuster.

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

Davenport, T. H., Barth, P. and Bean, R. (2012). How 'Big Data' is Different, *MIT Sloan Management Review* **54**(1): 22-24.

Davenport, T. (2014). *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*, Boston, MA: Harvard Business Review Press.

Desrosières, A. (1998). *The Politics of Large Number: A History of Statistical Reasoning*, Cambridge, MA: Harvard University Press.

Drnevich P. and Croson D. (2013). Information Technology And Business-Level Strategy: Toward An Integrated Theoretical Perspective, *MIS Quarterly* **37**(2):483-509.

Dosi, G., Nelson, R.R. and Winter, S.G. (2001). *The Nature and Dynamics of Organizational Capabilities*, New York: Oxford University Press.

Eco, U. (2000). *Kant and the Platypus: On Language and Cognition*. London: Vintage.

Eggers, J. P. and Kaplan, S. (2013). Cognition and Capabilities: A Multi-Level Perspective, *The Academy of Management Annals* **7**(1): 293-338.

Ekbia, H. and Evans, T. (2009). Regimes of Information: Land Use, Management, and Policy, *The Information Society* **25**(5): 328–343.

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V.R., Tsou, A., Weingart, S. and Sugimoto, C.R. (2014). Big Data, Bigger Dilemmas: A Critical Review, *Journal of the American Society for Information Science and Technology* (forthcoming).

Ferrier, W. J., Smith, K. G. and Grimm, C. M. (1999). The Role of Competitive Action in Market Share Erosion and Industry Dethronement: A Study of Industry Leaders and Challengers, *Academy of Management Journal* **42**(4): 372-388.

Floridi, L. (2012). Big Data and their Epistemological Challenge, *Philosophy and Technology* **48**(2), 103-121.

Gavetti, G. and Levinthal, D. (2000). Looking Forward and Looking Backward: Cognitive and Experiential Search, *Administrative Science Quarterly* **45**(1): 113-137.

Gavetti, G. and Rivkin, J. W. (2007). On the Origin of Strategy: Action and Cognition Over Time, *Organization Science* **18**(3): 420-439.

Gavetti, G. (2012). PERSPECTIVE—Toward a Behavioral Theory of Strategy, *Organization Science* **23**(1): 267-285.

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

Gillespie, T. (2014). The Relevance of Algorithms, in T. Gillespie, P. Boczkowski and K. Foot (eds.) *Media Technologies*, Cambridge, MA: MIT Press, pp: 167-193.

Iansiti, M. and Levien, R. (2004). *The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability*, Boston, MA: Harvard Business School Press.

IDC (2011). Digital Universe Study: Extracting Value from Chaos, http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf (accessed 11 May 2014).

Jarzabkowski, P., Balogun, J. and Seidl, D. (2007). Strategizing: The Challenges of a Practice Perspective, *Human Relations* **60**(1): 5-27.

Kallinikos, J. (2006). *The Consequences of Information: Institutional Implications of Technological Change*, Cheltenham: Elgar.

Kallinikos, J. (2009). The Making of Ephemeria: On the Shortening Life Spans of Information, *The International Journal of Interdisciplinary Social Sciences* **4**(3): 227-236.

Kallinikos, J., Aaltonen, A. and Marton, A. (2013). The Ambivalent Ontology of Digital Artifacts, *MIS Quarterly* **37**(2): 357-370.

Kallinikos, J. and Mariategui, J.-C. (2011). Video as Digital Object: Production and Distribution of Video Content in the Internet Media Ecosystem, *The Information Society* **27**(5): 281-294.

Kallinikos, J. (2013). The Allure of Big Data, *Mercury Magazine* **2**(3): 40-43.

Lajili, K. and Mahoney, J. T. (2006). Revisiting Agency and Transaction Costs Theory Predictions on Vertical Financial Ownership and Contracting: Electronic Integrations as an Organizational Form Choice, *Managerial and Decision Economics* **27**(7): 573-586.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S. and Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review* **52**(2): 21-31.

Learned, E.P., Christensen, C.R., Andrews, K.E. and Guth, W.D. (1965). *Business Policy: Text and Cases*, Homewood, IL: RD Irwin.

Lee, Y., Madnick, S., Wang, R., Wang, F. and Zhang, H. (2014). A Cubic Framework for the Chief Data Officer: Succeeding in a World of Big Data, *MIS Quarterly Executive* **13**(1):1-13.

Lim, J. H., Stratopoulos, T. C. and Wirjanto, T. S. (2011). Path Dependence of Dynamic Information Technology Capability: An Empirical Investigation, *Journal of Management Information Systems* **28**(3): 45-84.

Luhmann, N. (1982). *The Differentiation of Society*, New York: Columbia University Press.

Luhmann, N. (1993). *Risk: A Sociological Theory*, New York: A. de Gruyter.

Malone, T. W., Yates, J. and Benjamin, R. I. (1987). Electronic Markets and Electronic Hierarchies, *Communications of the ACM* **30**(6): 484-497.

Manovich, L. (2001). *The Language of New Media*, Cambridge, MA: The MIT Press.

March, J.G. (1999). *The Pursuit of Organizational Intelligence: Decisions and Learning*, Cambridge, MA: Blackwell.

Mayer-Schönberger, V. and Cukier, K. (2013). *Big Data: A Revolution that Will Transform how we Live, Work, and Think*, London: John Murray.

McAfee, A. and Brynjolfsson, E. (2012). Big Data: The Management Revolution, *Harvard Business Review* October: 1-9.

McGee, J. and Thomas, H. (1986). Strategic Groups: Theory, Research and Taxonomy, *Strategic Management Journal* **7**(2): 141-160.

Melville, N., Kraemer, K. and Gurbaxani, V. (2004). Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value, *MIS Quarterly* **28**(2): 283-322.

Merali, Y., Papadopoulos, T. and Nadkarni, T. (2012). Information Systems Strategy: Past, Present, Future? *Journal of Strategic Information Systems* **21**(2): 125–153.

Mingers, J. and Willcocks, L. (2014). An Integrative Semiotic Framework for Information Systems: The Social, Personal and Material Worlds, *Information and Organization* **24**: 48-70.

Mintzberg, H. (1991). The Five Ps for Strategy, in H. Mintzberg, and J.B. Quinn (eds.) *The strategy process: concepts, contexts, cases,* Englewood Cliffs, NJ: Prentice-Hall, pp.12-19.

Moretti, F. (2005). *Graphs, Maps, Trees: Abstract Models of Literary History*. London: Verso.

Morville, P. (2005). *Ambient Findability,* O'Reilly Media Inc.

Nelson, R. R. and Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*, Cambridge MA: Belknap.

Nelson, R.R. and Winter, S.G. (2002). Evolutionary Theorizing in Economics, *Journal of Economic Perspectives* **16**(2): 23–46.

Newman, H. H. (1978). Strategic Groups and the Structure-Performance Relationship, *The Review of Economics and Statistics* **60**(3): 417-427.

Oestreicher-Singer, G. and Zalmanson, L. (2013). Content or Community? A Digital Business Strategy for Content Providers in the Social Age, *MIS Quarterly* **37**(2): 591-616.

O'Reilly (2012). *Big Data Now: 2012 Edition,* O'Reilly Media Inc.

Penrose, E. T. (1966). *The Theory of the Growth of the Firm*, Oxford: Basil Blackwell.

Peteraf, M. A. (1993). The Cornerstones of Competitive Advantage: A Resource Based View, *Strategic Management Journal* **14**(3): 479–488.

Pfeffer, J. and Sutton, R. I. (2006). *Hard Facts, Dangerous Halftruths, and Total Nonsense: Profiting from Evidence-Based Management*, Cambridge, MA: Harvard Business School Press.

Pitelis, C. N. and Teece, D. J. (2009). The (New) Nature and Essence of the Firm, *European Management Review* **6**(1): 5-15.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*, New York: The Free Press.

Porter, M. E. (1985). *Competitive Advantage: Creating and Sustaining Superior Performance*, New York: The Free Press.

Porter, M. E. (1996). What Is Strategy? *Harvard Business Review* **74**(6): 61–78.

Porter, T. M. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*, New Jersey: Princeton University Press.

Priem, R. L., Butler, J. E. and Li, S. (2013). Toward Reimagining Strategy Research: Retrospection and Prospection on the 2011 AMR Decade Award Article, *Academy of Management Review* **38**(4): 471-489.

Ray, G., Xue, L. and Barney, J. B. (2013). Impact of Information Technology Capital on Firm Scope and Performance: The Role of Asset Characteristics, *Academy of Management Journal* **56**(4):1125-1147.

Rosenfeld, l. and Morville, P. (2002). *Information Architecture for the World Wide Web*, O'Reilly Media Inc.

Rousseau, D. M. (2006). Is There Such a Thing as Evidence Based Management? *Academy of Management Review* **31**(2): 256–269.

Schmarzo, B. (2013). *Big Data: Understanding how Data Powers Big Business,* Indianapolis, IN: John Wiley & Sons.

Schumpeter, J. A. (1934). *The Theory of Economic Development*, Cambridge MA: Harvard University Press.

Shollo, A. and Constantiou, I. (2013). Self-Reinforcing Mechanisms and Organizational Decision Making: The Case of Project Prioritization in a Financial Institution, in J. Sydow and G. Schreyögg (eds.) *Self-Reinforcing Processes in and among Organizations*, London: Palgrave Macmillan, pp. 104-124.

Teece, D.J. and Pisano, G. (1994). The Dynamic Capabilities of Enterprises: An Introduction, *Industrial and Corporate Change* **3**(3): 537–556.

Teece, D. J., Pisano, G. and Shuen, A. (1997). Dynamic Capabilities and Strategic Management, *Strategic Management Journal* **18**(7): 509–33.

Teece, D.J. (2007). Explicating Dynamic Capabilities: The Nature and Microfoundations of (Sustainable) Enterprise Performance, *Strategic Management Journal* **28**(13):1319–1350.

Thompson, J. D. (1967). *Organizations in Action*, New York: McGraw & Hill.

Yates, J. (1989). *Control through Communication: The Rising of System in American Management*, Baltimore: John Hopkins University Press.

Yoo, Y. (2010). Computing in Everyday Life: A Call for Research on Experiential Computing, *MIS Quarterly* **34**(2): 213–231.

Yoo, Y. (2013). The Tables Have Turned: How Can the Information Systems Field Contribute to Technology and Innovation Management Research, *Journal of the Association of Information Systems* **14**: 227-236.

Yoo, Y., Henfridsson, O. and Lyytinen, K. (2010). The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research, *Information Systems Research* **21**(4): 724–735.

Van Dijck, J. (2013). *The Culture of Connectivity: A Critical History of Social Media*, Oxford: Oxford University Press.

Varian, H. R. (2010). Computer Mediated Transactions, *American Economic Review* **100**(2): 1-10.

Weinberger, D. (2007). *Everything Is Miscellaneous: The Power of the New Digital Disorder*, New York: Times Books.

Wernerfelt, B. (1984). A Resource-Based View of the Firm, *Strategic Management Journal* **5**(2): 272–280.

Zuboff, S. (1988). *In the Age of the Smart Machine: The Future of Work and Power*, Oxford: Heinemann Professional.

Zuboff, S. and Maxmin, J. (2003). *The Support Economy: How Corporations Fail Individuals and the Next Episode of Capitalism*, London: Allen Lane.