Technological Forecasting & Social Change xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

Governance of big data collaborations: How to balance regulatory compliance and disruptive innovation

Tijs van den Broek[a,*], Anne Fleur van Veenstra[b]

[a] University of Twente, NIKOS, Hallenweg 17, 7522 NH Enschede, The Netherlands
[b] TNO, Anna van Buerenplein 1, 2595 DA The Hague, The Netherlands

ARTICLE INFO

Keywords:
Disruptive innovation
Data protection regulation
Big data
Governance
Inter-organizational collaboration

ABSTRACT

Big data is an important driver of disruptive innovation that may increase organizations' competitive advantage. To create innovative data combinations and decrease investments, big data is often shared among organizations, crossing organizational boundaries. However, these big data collaborations need to balance disruptive innovation and compliance to a strict data protection regime in the EU. This paper investigates how inter-organizational big data collaborations arrange and govern their activities in the context of this dilemma. We conceptualize big data as inter-organizational systems and build on IS and Organization Theory literature to develop four archetypical governance arrangements: Market, Hierarchy, Bazaar and Network. Subsequently, these arrangements are investigated in four big data collaboration use cases. The contributions of this study to literature are threefold. First, we conceptualize the organization behind big data collaborations as IOS governance. Second, we show that the choice for an inter-organizational governance arrangement highly depends on the institutional pressure from regulation and the type of data that is shared. In this way, we contribute to the limited body of research on the antecedents of IOS governance. Last, we highlight with four use cases how the principles of big data, specifically data maximization, clash with the principles of EU data protection regulation. Practically, our study provides guidelines for IT and innovation managers how to arrange and govern the sharing of data among multiple organizations.

1. Introduction

Organizations increasingly collect, store, process, analyze and extract value from vast amounts of data. Big data refers to the unconventional infrastructures, such as hardware, software, and skills, to extract value out of the increasing variety, volume, velocity, variability and complexity of data available to organizations (Chen et al., 2012; Fosso Wamba et al., 2015; Wang et al., 2016). Big data may improve operational efficiency and the effectiveness of organizations as well as lead to the development of new products, services and business models. Big data technologies are considered disruptive (Brown et al., 2011; Frey and Osborne, 2017; Loebbecke and Picot, 2015). Disruptive technologies make established technologies obsolete and diminish the value of investments made in legacy technologies (Christensen, 1997; Danneels, 2004). Similarly, applications of big data substitute services previously provided by professionals, ranging from advertising to medical diagnosis (Wang et al., 2016). In this way, big data may contribute to social innovation, as public value (e.g. in medicine or education) is created in a different way by private organizations.

Consequently, organizations are keen to invest in big data infrastructures and highly skilled personnel (Chen et al., 2012; Fosso Wamba et al., 2015). Furthermore, the value of big data for innovation increases when datasets are combined across organizations and sectors. Hence, organizations form big data collaborations to share investments and increase the number of datasets. These data collaborations are arrangements between three or more organizations that jointly establish data protocols, data exchange and reporting mechanisms (Bertot and Choi, 2013). However, besides being a disruptive innovation, big data raises ethical concerns: large-scale collection, storing and processing of personal data may cause privacy infringements (Cecere et al., 2015). Hence, the European Commission (2016) is implementing the General Data Protection Regulation (GDPR) to clearly outline rules for organizations how to deal with data and to minimize privacy risks by guaranteeing citizens more control over their data. It remains, however, unclear how this regulation impacts the way organizations share data to jointly innovate. Academically, we conceptualize big data collaborations as form of Inter-organizational Systems (IOS) that require a governance arrangement by the network members employing the IOS

* Corresponding author.
E-mail addresses: t.a.vandenbroek@utwente.nl (T. van den Broek), annefleur.vanveenstra@tno.nl (A.F. van Veenstra).

https://doi.org/10.1016/j.techfore.2017.09.040
Received 22 November 2016; Accepted 17 September 2017

(Chatterjee and Ravichandran, 2013; Robey et al., 2008). In literature, it remains unclear why organizational networks choose for a certain governance arrangement, as well as the directionality between organizational and technical governance elements (Chatterjee and Ravichandran, 2013). We focus on the interaction of coercive institutional pressures from regulation (Krell et al., 2016; Lyytinen and Damsgaard, 2011; Teo et al., 2003) and data characteristics (de Corbiere and Rowe, 2013; Kumar and Van Dissel, 1996) to explain how IOS governance is arranged.

Organizations may arrange a decentralized governance structure to foster innovation. However, big data collaborations in the EU are required to carefully design and implement arrangements that formulate, communicate and assess policies and procedures to control the collection, storing and processing of personal data. Therefore, this paper explores this apparent tension between control (as result of regulatory compliance) and innovation in big data collaborations. The research question is: "*How do big data collaborations balance the need for control and innovation in their governance arrangement?*"

The theoretical contributions are threefold. First, we conceptualize the organization behind big data collaborations as IOS governance. Second, we show that the choice for an inter-organizational governance arrangement highly depends on the institutional pressure from regulation (i.e. GDPR) and the type of data that is shared. In this way, we contribute to the limited body of research on the antecedents of IOS governance. Last, we highlight with four use cases how the principles of big data, specifically data maximization, clash with the principles of EU data protection regulation. Practically, our study provides guidelines for IT and innovation managers how to arrange and govern their sharing of data among multiple organizations.

2. Theory

2.1. Big data as a disruptive technology

An importance focus in IS research remains how and why organizations develop and maintain joint information systems (Chatterjee and Ravichandran, 2013; Robey et al., 2008). Inter-organizational Systems (IOS) are "automated systems shared by two or more organizations, and designed to link business processes" (Robey et al., 2008, p. 498). Organizations develop and adopt IOS to improve operational their efficiency, share data and knowledge with external partners, enter or open up markets, and develop new products and services (Robey et al., 2008). Scholars stress that IOS are not only an instrument to minimize transaction costs between organizations, such as traditionally in Electronic Data Interchange systems, but are increasingly employed as digital platforms that provide a business ecosystem (Barrett et al., 2015; Yoo et al., 2012). In our research, we conceptualize (big) data collaborations as a specific type of IOS.

Big data covers unconventional infrastructures, such as processing power and storage capacity, methods, knowledge and skills, to collect, store and analyze vast quantities of both structured and unstructured data to inform organizational decisions and public policies (McAfee and Brynjolfsson, 2012; Wang et al., 2016). Data may be collected from humans, machines and increasingly from sensors, and may include personal data. Big data is more than an array of infrastructures: Boyd and Crawford (2012) also consider big data as a cultural and scholarly concept and provide a more holistic definition that includes technology, analysis, and mythology (referring to the visionary promise of big data to society). Furthermore, Fosso Wamba et al. (2015) add the importance of skills (e.g. of data scientists) that are required to extract value from big data. The ability to extract value out of this data, e.g. with machine learning and data mining, may enhance organizations' competitiveness by increasing efficiency and effectiveness, improving decision-making and fostering innovation (Brown et al., 2011; McAfee and Brynjolfsson, 2012). Innovation refers to the "production or adoption, assimilation, and exploitation of a value-added novelty in economic and social spheres; renewal and enlargement of products, services, and markets; development of new methods of production; and establishment of new management systems" (Crossan and Apaydin, 2010). Following this definition, big data analytics may spur innovations in the social and economic sphere. We use the term social innovation to refer to the development of products, processes, services aimed at social purposes that are mediated by technologies or closely linked to technological innovations (Edwards-Schachter and Wallace, 2017).

Big data analytics are increasingly acknowledged by innovation and IS scholars as a disruptive technology (Baesens et al., 2014; Brown et al., 2011; Frey and Osborne, 2017; Loebbecke and Picot, 2015). Disruptive technologies make established technologies employed by incumbent firms obsolete and diminish the value of investments made in those legacy technologies (Christensen, 1997; Danneels, 2004). The underlying mechanism is that an emerging technology may lower the entry barriers for newcomers in markets. Similarly, the 'datafication' of markets and the lower costs to store, process, and analyze the growing amount of data have lowered the barriers for challengers (e.g. digital startups) to disrupt incumbent firms' technologies (Loebbecke and Picot, 2015). Furthermore, big data analytics enable artificial intelligence and subsequent autonomous decision-making that increasingly replaces skilled labor among incumbent organizations, ranging from administrative work to medical diagnosis (Loebbecke and Picot, 2015; McAfee and Brynjolfsson, 2012). Examples of disruptive big data driven innovations become more and more apparent in society. First, the availability of data from electronic learning environments stimulated new startups that provide learning analytics, which is the optimization and personalization of learning processes based on data from learners (Anderson and McGreal, 2012). Incumbent players, such as traditional educational institutions, are pressured to invest in learning analytics to keep up with newcomers. A second example of disruption from big data driven innovation is the emergence of the Fintech (a combination of Finance and Technology) industry. Startups, often without a background in the financial industry, develop smart, personalized and low-cost financial services (e.g. lending, sharing bills or stock investment) that may challenge the financial services of incumbent firms in the coming years (Zavolokina et al., 2016).

2.2. The role of compliance in big data innovations

An important discussion among innovation scholars is the impact of regulation on innovation. Traditionally, economists consider regulation to be a coercive force that requires firms to reduce negative externalities, such as privacy infringements in the case of digital services. Reducing these negative externalities is considered to limit the firms' profitability. However, Porter and Van der Linde (1995) posed the rival hypothesis that environmental regulation does not hinder businesses' competitive advantage against competitors when they are well-designed. Regulation may stimulate firms to improve their technology, may increase corporate awareness of the negative externalities, reduces regulatory uncertainties of technological investments by providing a level playing field, and may even trigger innovation (e.g. in the case of environmental regulation). For example, Chakraborty and Chatterjee (2017) found that R&D expenditure in the Indian dye-producing industry increased within a range of 11–61% as a result of environmental regulation. However, results on Porter's hypothesis do not find a consistent relationship between regulation and innovation (Chakraborty and Chatterjee, 2017). Hence, research following up on Porter's hypothesis provides several contingencies for the influence of regulation on innovation, for example flexibility in how the regulation is met, a focus on technological improvement, firm size, ownership and reducing regulatory uncertainty (Chakraborty and Chatterjee, 2017; Johnstone et al., 2010). Similar results are found by Pelkmans and Renda (2014) on EU regulation: prescriptive regulation that requires strict compliance hampers innovation activity in an industry, whereas flexible regulation