

## The Emerging Role of Big Data in Key Development Issues: Opportunities, Challenges, and Concerns

By: [Nir Kshetri](#)

Kshetri, Nir (2014). "The Emerging Role of Big Data in Key Development Issues: Opportunities, Challenges, and Concerns," *Big Data & Society*, July-December, 1-20, DOI: 10.1177/2053951714564227

**Nir Kshetri, The Emerging Role of Big Data in Key Development Issues: Opportunities, Challenges, and Concerns, *Big Data & Society* (vol. 1, no. 2) pp. 1-20. Copyright © 2014 Nir Kshetri. Reprinted by permission of SAGE Publications.**

**Made available courtesy of SAGE Publications:**

<http://dx.doi.org/10.1177/2053951714564227>

### **Abstract:**

This paper presents a review of academic literature, policy documents from government organizations and international agencies, and reports from industries and popular media on the trends in Big Data utilization in key development issues and its worthwhileness, usefulness, and relevance. By looking at Big Data deployment in a number of key economic sectors, it seeks to provide a better understanding of the opportunities and challenges of using it for addressing key issues facing the developing world. It reviews the uses of Big Data in agriculture and farming activities in developing countries to assess the capabilities required at various levels to benefit from Big Data. It also provides insights into how the current digital divide is associated with and facilitated by the pattern of Big Data diffusion and its effective use in key development areas. It also discusses the lessons that developing countries can learn from the utilization of Big Data in big corporations as well as in other activities in industrialized countries.

**Keywords:** Big Data | development | agriculture | healthcare | social media | cellphone | data privacy and security | smallholder farmers

### **Article:**

#### **Introduction**

Big Data (BD) is likely to be of tremendous benefit to developing countries. It is anticipated that geo-locating a rural African farmer working in his farm with the help of an app installed in his cellphone, identifying the soil type and needs of the field, and offering advice regarding appropriate seeds, where they can be purchased, and how they can be planted and harvested is not far in the future (Patel, 2013). As another example, a retrospective analysis of the 2010 cholera outbreak in Haiti showed that mining data from Twitter and online news reports could have given the country's health officials an accurate indication of the disease's spread with a lead time of two weeks (Chunara et al., 2012).

Perhaps the greatest advantage offered by BD in the context of development is that it helps us gain a better understanding of the extent and nature of poverty and devise appropriate policy measures. For instance, mobile data can make it possible to better understand the dynamics of slum residents. The call detail record (CDR) and other information can provide insights into the slum population, which would help forecast the needs for toilets, clean drinking water, and infrastructures (bigdata-startups.com, 2013). To take an example, in Kenya's Nairobi, geo-coded cellphone transaction data is used by the Engineering Social Systems (ESS) project to model slums' growth, which could help the government to optimize resource allocation for infrastructural development and other resources (Bays, 2014). Alternative data collection and analysis techniques such as surveys have a very low degree of usefulness for such purposes, which may take months and even years for getting results and are often out of date.

An encouraging trend is that the tools and expertise that are employed to make decisions and take actions related to behavioral advertising based on consumers' real-time profiling are being used in addressing developmental problems. For instance, data generated by social media such as Twitter is being analyzed in order to detect early signs that can lead to a spike in the price of staple foods, increase in unemployment, and outbreak of diseases such as malaria. Robert Kirkpatrick of the UN Global Pulse team referred to such signs as "digital smoke signals of distress" and noted that they can be detected months before official statistics (Lohr, 2013). The importance of this technique is even more pronounced if we consider the fact that there are no reliable statistics in many developing countries.

Prior research indicates that information infrastructure has social, political, and economic dimensions (Bowker, 1996). This argument can be extended in obvious ways to include BD infrastructures. We argue that different social, political and economic situations and the current state of digital divide may give rise to significantly different rates of diffusion in industrialized and developing countries. Unsurprisingly, there is an enormous gap between the developing and developed worlds in the utilization of BD.

One way to better understand this phenomenon is to consider how the nature of the current digital divide is associated with and facilitated by the pattern of BD diffusion and its effective use. Hilbert (2014) notes that the current inequality of technological capacity represents a more mature and also more persistent stage of the digital divide. Boyd and Crawford (2012) argue that limited access to BD has created new forms of digital divides. They note that the lack of money to afford data is among many factors that may contribute to the digital divide.

The international digital divide may account for the differential rates of BD diffusion and its effective use in industrialized and developing countries. In order to better understand the diffusion and effective use of BD in developing countries, it is thus first important to consider the current stage of the digital divide. Hilbert (2014) has identified three complementary stages associated with the digital divide: access to a technology, its effective usage, and social integration and impact of the technology. The rapidly narrowing access gap between industrialized and developing countries makes the last two stages more relevant. Prior researchers have suggested that factors such as availability of skills and capabilities, social and cultural attitudes towards a technology, the institutional environment, and social reorganization are tightly linked to the ability to use the technology effectively and appropriately (Buente and

Robbin, 2008; DiMaggio and Hargittai, 2001; DiMaggio et al., 2004; Hilbert, 2014; Robinson, 2009). Many of these factors can be considered as key components of a social system, which is tightly linked to the diffusion of an innovation (O'Neil et al., 1998; Rogers, 1995, 2003).

While there are some encouraging trends in the utilization of BD in addressing a number of social and economic problems, developing economies are far from achieving the full transformative potential of BD. It is thus important for researchers and policymakers to have a deeper understanding of social, political, and economic contexts that facilitate and inhibit BD's diffusion and effective utilization in key development areas. This issue also needs to be considered in relation to the broader issue of the digital divide.

In light of the above observations, our goal in this paper is modest and is simply aimed at deepening our understanding of facilitators and inhibitors of diffusion and effective utilization of BD in developing countries. A related goal of this paper is to explore this issue in relation to the current nature of the international digital divide. In order to achieve these goals, we present a review of academic literature, policy documents from government organizations and international agencies, and reports from industries and popular media on the trends in BD utilization in key development issues and its worthwhileness, usefulness, and relevance.

Before getting to the details of this study, it is worthwhile to first clarify some of the key concepts. By developing countries, we mean low-, lower middle-, and lower middle-income countries in the World Bank categorization (The World Bank Group, 2014). In order to define BD for the purpose of this paper, we start with the technology research company Gartner's definition of BD, which are "high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making" (gartner.com, 2013). With regard to volume, Boyd and Crawford (2012: 663) note that BD is a "poor term" and argue that BD "is less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets". In this paper's context, we define BD as data sets that can provide insights into human well-being, which satisfy at least one of the following characteristics compared to data sets that have been traditionally used in developmental issues: (a) are of higher volume, (b) are of wider variety, and (c) enable users to make decisions and act faster. The paper is structured as follows. We proceed by first discussing the methodology. The section following that provides a theoretical framework for assessing the diffusion and effects of BD in developing countries. Next, we analyze the characteristics of BD in the context of development. This is followed by a section on the emerging trends associated with BD utilization in development with a discussion of some present and potential future applications. Then, we investigate the opportunities and challenges of using BD for development with the help of the case of agriculture and farming activities. It is followed by a section on discussion and implications. The final section provides concluding comments.

## **Methodology**

The approach employed in this paper can be described as a positivistic epistemology. Given the minute amount of existing research on BD's use, particularly in developing countries, much initial research in this area needs to be qualitative, concept-, and theory-building in character. This paper thus takes such an approach.

Following the approach often used in positivist research, this paper seeks to identify details associated with BD use in developing countries. We have employed three main sources for the reasoning in the theory-development process: theoretical explanations for “whys” and “hows” of BD diffusion; past empirical findings; and practice or experience (Webster and Watson, 2002). Among these, the logical reasoning is the most important component of our explanations, which provides “the theoretical glue” (Whetten, 1989: 491). We have also included empirical research related to BD uses and from other related areas such as diffusion of cellphones and other innovations. Next, BD-related experiences of organizations (e.g. Land O'Lakes, DuPont, AgriLife), nations (e.g. Kenya, Uganda), regions (e.g. Latin America and the Caribbean (LAC) economies, Africa), and development agencies (e.g. UN Global Pulse team). In addition to academic literature, we reviewed policy documents from government organizations and international agencies and reports from industries and popular media. Materials reviewed also include reports of consultants and *suppliers of BD-related solutions* (e.g. Monsanto, Gartner, Cisco and McKinsey). In order to locate articles, reports, and policy documents related to BD uses in developing countries, we searched using various combinations of keywords such as “data”, “Big Data”, “development”, “developing countries”, “Africa”, “Latin America”, “China”, “India”, “Nigeria”, “Brazil”, “Argentina”, etc. We limited our search to materials published in the English language.

## **A theoretical framework for assessing the diffusion and effects of BD in developing countries**

### Characteristics of the technology and environment affecting BD diffusion

Theoretical and empirical evidence suggests that the diffusion of an innovation is influenced by the environment, and characteristics of the innovation (O'Neil et al., 1998; Rogers, 1995, 2003). Rogers (1995) identified various characteristics of a technology affecting its diffusion pattern. Table 1 presents how BD performs in terms of the various characteristics of an innovation proposed by Rogers: relative advantage, compatibility, complexity, observability, and trialability.

Regarding the environment, prior researchers have noted that an innovation (e.g. BD) is embedded in a social system, which plays an important role in its diffusion. A social system is “a set of interrelated units that are engaged in joint problem solving to accomplish a common goal” (Rogers, 2003: 23). It varies in forms—formal as well as informal.

Bowker (1996) provides further specification and elaboration of the conditions under which new innovations diffuse and thus extends Rogers' (1995, 2003) framework. He argues that, in addition to the economic dimension, information infrastructures also have social (informal) and political (formal) dimensions. Social systems also exist at different levels (individual, organization, network, or national). For instance, the social structure and communication structure affect information flow and other factors that are critical for the adoption of the innovation by the adopting units. Various societal norms also affect the behavioral patterns of the members of a social system.

**Table 1.** Characteristics of a technology influencing its diffusion: the case of BD.

Dimension	Explanation	Situation in developing countries
Relative advantage	<ul style="list-style-type: none"> <li>Perceived benefits of a technology over previous technologies and the extent to which it is better than the idea it supersedes.</li> </ul>	<ul style="list-style-type: none"> <li>Compared to traditional surveys (e.g. household, labor market, living standard) Big Data involves relatively shorter time to collect and analyze.</li> <li>BD initiatives are likely to reduce the costs significantly compared to traditional surveys.</li> <li>SM (e.g. Twitter's API) is more appropriate to detect crisis than traditional techniques such as interviews and surveys (Madsen, 2013).</li> </ul>
Compatibility	<ul style="list-style-type: none"> <li>The degree to which a technology and the tasks it performs are perceived as being consistent with the existing values, beliefs, past experiences, and needs of potential adopters.</li> </ul>	<ul style="list-style-type: none"> <li>The initiatives such as Kenya's ODP are likely to be incompatible with institutions filled with secrecy and corruption.</li> </ul>
Complexity	<ul style="list-style-type: none"> <li>The level of difficulty of installing and using a technology (variety and uncertainty increase complexity).</li> </ul>	<ul style="list-style-type: none"> <li>There is a shortage of skills for data-related manpower. People with the lack of appropriate technical skills and qualifications may find BD techniques more complex.</li> </ul>
Observability	<ul style="list-style-type: none"> <li>The degree to which the features and benefits of a technology are visible, noticeable, and understandable to self/others, the results can be described to non-users.</li> </ul>	<ul style="list-style-type: none"> <li>BD can help document quality standards of agricultural products.</li> </ul>
Trialability	<ul style="list-style-type: none"> <li>The ability to experiment or try (on a limited basis) before formally adopting.</li> </ul>	<ul style="list-style-type: none"> <li>Organizations can try with small data sets to see the appropriateness and effectiveness of BD projects.</li> </ul>

Source: Based on Rogers (1995) and author's research.

Finally, the characteristics of the social system are tightly linked to the real and/or perceived attributes of an innovation (Rogers, 1995). For instance, due to BD's potential role in promoting transparency, decision makers in a social system characterized by secrecy and distrust may view BD's attributes negatively.

#### International heterogeneity in the diffusion rates of innovations

We start this section with a brief description of the international heterogeneity in the diffusion rates of innovations and the key sources of the heterogeneity. Prior research indicates that organizations in a country with a low degree of inter-relatedness with other complementary technologies often find it difficult to obtain the information and skills needed for a new technology (Allen, 1998). Thus, countries with a small base of high technology and innovative capital goods are likely to experience lower rates of diffusion for emerging new technologies (Antonelli, 1986). To put things in context, wider and deeper adoptions of personal computers (PCs), mobile devices, and other information and communications technology (ICT) applications are likely to lead to faster diffusion of BD.

Another source of international variation in BD diffusion lies in the market and infrastructure factors controlling the availability of the technology to potential adopters (Brown et al., 1976). Prior research has suggested that manufacturers of new technological products are more likely to focus their efforts on large distributors often located in developed countries (Gatignon and Robertson, 1985). The environments in developing countries are associated with adverse conditions in terms of markets and infrastructures, which are likely to slow down the diffusion of new technological innovations such as BD. For instance, a large proportion of firms in developing countries may lack readiness to adopt BD. Likewise, under-developed infrastructures such as those related to weather forecasting and satellite-imaging technologies hinder the use of BD in agricultural activities in these countries.

In order to provide insights into the above issues, it is important to discuss the changing nature of the international digital divide. On the basis of a review of academic and practitioner-oriented writings on the digital divide, Hilbert (2014) identifies three complementary stages associated with it: access to a technology, its effective usage, and social integration and impact of the technology. The fact that over 90% of the population in developing economies owned cellphones in 2013 means that the access-related digital divide has significantly narrowed.

Nonetheless, developing economies face a number of challenges that may limit their capability to utilize BD effectively. A low degree of digitization is among the biggest barriers. One way to measure digitization is to consider Booz & Company's Digitization Index (BCDI), which is based on 23 indicators such as access, affordability, reliability, speed and usability of digital services/applications, and users' skills. For instance, the average BCDI score of emerging economies was 27 compared to developed economies' 54. Booz & Company has described most emerging economies in the constrained or emerging stages of digitization, whereas developed economies are in the transitional or advanced stage (Strategy, 2012). In these countries, there are limited hardware, software, and other technology applications to generate and distribute relevant data and knowledge (UN Development Program (UNDP), 2011). Moreover, the lower quality of hardware, software, and infrastructure may lead to a lower amount of information flow in developing countries. For instance, many businesses design websites with features such as "low-graphics" and "text-only" to accommodate the needs of users in developing countries.

It is also reported that most African countries use accounting methods more than five decade old to generate vital statistics such as those related to gross domestic product (GDP) (Bhushan, 2012). The World Bank's study indicated that many countries in the LAC region lack a strategic vision required for the overall management of the statistical system (The World Bank, 2010).

Equality in access is a necessary but not a sufficient condition to lead to digital equality. Factors such as availability of skills and capabilities, social and cultural attitudes towards a technology, the institutional environment, and social reorganization are tightly linked to the ability to use the technology effectively and appropriately (Buente and Robbin, 2008; DiMaggio et al., 2004; Hilbert, 2014; Robinson, 2009). Factors such as the availability of skills and the availability of social support have also been recognized by DiMaggio and Hargittai (2001) as the key dimensions of the digital divide. They have especially emphasized the role of "Internet competence", which is related to the know-how, technical skills and capacity to exploit the Internet's potential by strategically responding to challenges and opportunities (DiMaggio and Hargittai, 2001).

Finally, equality in usage does not necessarily translate to social equality. Some groups of society are in a position to benefit more from ICTs than others. This implies the possibility of dominance by some social groups and may therefore lead to "an increasing social divide" (Hilbert, 2014). For instance, one commentator noted that about 90% of the discussion at the 2013 Internet Governance Forum (IGF) referred to BD as a surveillance tool. At the same time, the debate focusing on developing countries treated BD as a means to "observe" people to fight poverty. The argument provided by IGF participants was that data can help provide access to clean drinking water, healthcare, and other necessities (linnettaylor, 2013).

Some forces to overcome the adverse impacts of markets and infrastructures

While manufacturers of new technological products often find developed countries less attractive, prior researchers have suggested that multinationals exploit technological capabilities internationally by means of activities such as export, movement of production activities abroad, and licensing, which have driven the globalization of technologies (Archibugi and Michie, 1997; Iammarino and Michie, 1998). Such forces have been shown to influence the diffusion of Internet and e-commerce related technologies, especially in developing economies (Kshetri, 2001; Kshetri and Dholakia, 2002). As noted earlier, transnational corporations (TNCs) such as Monsanto and Syngenta are likely to drive the international technology transfer of BD in the agricultural sector.

Prior research has suggested that international institutions influence the global diffusion of ICTs in several ways (Kshetri, 2001). For instance, they introduced the Internet for the first time in many developing countries. The UNDP introduced the Internet in more than 15 countries by connecting them to the global network and deployed the Internet protocol (IP) network in more than 40 countries (UNDP, 2001). By early 2001, the UNDP also trained over 25,000 organizations and created more than 40,000 websites for governments and civil society stakeholders and 3000 national and regional thematic networks (UNDP, 2001). As noted earlier, various BD initiatives have been undertaken by the UN in developing countries.

Another encouraging trend is that there has been an abundance of BD-related entrepreneurship in the developing countries. This is important because, as noted above, Western firms often find developing countries to be less attractive markets. Firms in these countries are rapidly emerging as providers of products, services, software, and solutions related to BD (Kshetri, 2011). For instance, a technology developed by the Brazilian company Cignifi can recognize patterns in consumers' phone-calls, text messages, and data usage, which are used to predict lifestyle and credit risk profile (bigdata-startups.com, 2013). A study conducted by India's National Association of Software and Services Companies (NASSCOM) and CRISIL Global Research & Analytics estimated that the Indian BD industry was US\$ 200 million in 2012, which is expected to increase to US\$ 1 billion in 2015 (Srikanth, 2013). Likewise, the AgriLife platform was developed by Kenya-based IT company MobiPay and was launched in late 2012. Mercy Corps supported the expansion of AgriLife to Uganda and helped build relationships with other service providers and integrate them into the platform, so they can reach rural clients more effectively.

### **Characteristics of BD in the context of development**

Massive amounts of data generated by social media, cellphones, and other digital communication tools, which are being increasingly used in developing countries, are a true form of BD. While such data has not been traditionally used in developmental issues, it is likely to be a useful indicator of human well-being and is thus a relevant BD source for development (Global Pulse, 2012).

It is first important to explain what BD is and how one can say that a data set used for a specific, development-related purpose is BD. As noted earlier, Gartner has defined BD in terms of three Vs: volume, velocity, and variety. The software company, SAS, has added two additional

dimensions: variability and complexity (sas.com, 2013). The following discussion will examine how the various characteristics or dimensions of BD identified by Gartner and SAS are relevant in the context of developmental data (Table 2).

**Table 2.** Relevance of Big Data dimensions in developmental issues.

Characteristic	Explanation	Some examples in the context of development
Volume	Huge amount of data is created from a wide range of sources such as transactions, unstructured streaming from text, images, audio, voice, VoIP, video, TV and other media, sensor, and machine-to-data.	Rapid diffusion of cellphones, social media, and other technologies has led to the creation of huge amount of data.
Velocity (fast data)	Some data is time-sensitive and needs to be collected, stored, processed, analyzed, and acted on quickly. In some cases, speed is more important than volume.	Most decisions are based on data that are near-real-time (e.g. AgriLife).
Variety	Data comes in multiple formats such as structured, numeric data in traditional database, and unstructured text documents; email, video, audio, and financial transactions.	Structured and unstructured data are being used in a number of developmental projects (e.g. Malaria Surveillance & Mapping project in Botswana and "Water Watchers" in South Africa).
Variability	Data flows can vary greatly with periodic peaks and troughs. These are related to social media trends, daily, seasonal and event-triggered peak data loads, and other factors.	Variation of data flow is related to certain developmental indicators (e.g. the correlation of the volume of Tweets about staple foods and increase in the cost).
Complexity	Data come from multiple sources which require linking, matching, cleansing, and transforming across systems.	By matching and linking data from diverse sources such as CDRs, open portals, social media, government, NGOs, and corporations (e.g. prediction of food shortages by combining data related to drought, weather conditions, migration patterns, market prices of staples, seasonal variation in prices and past productions), key insights can be gained regarding a developmental issue.

## Volume

There has been a colossal increase in the digitization rate of developing countries. Of particular importance to the present discussion is the rapid diffusion of cellphones, which are probably the most important source of data in the context of development. Whereas only a little over a quarter of the population in developing countries owns a PC, close to 90% owned cellphones in 2013. One estimate suggested that the mobile data traffic generated by subscribers in emerging markets grew by over 100% in 2013 (cisco.com, 2014).

People with high disposable income in developing economies tend to spend a significant portion on topping up their mobile airtime credit. The monthly airtime expenses can provide background of the household income. This information provides guidance on how to best target appropriate services through advertising. It can be done anonymously. Monitoring airtime expenses for trends and sudden changes provides a measure of the early impact of an economic crisis and the impact of programs designed to improve livelihoods (UN Global Pulse, 2013b).

Mobile phone-related data often provide high quality, valuable information because a mobile phone is often the only interactive technology for most low-income individuals. Moreover, it is easy to link mobile-generated data to individuals, which can help understand their needs and behaviors (World Economic Forum (WEF), 2012). The frequency with which calls are made and received with contacts outside of one's immediate community provides an in-depth understanding of the socio-economic class (UN Global Pulse, 2013a).

Probably the most useful category of data is the CDR, which is automatically generated by mobile network operators for all mobile transactions. Each record contains attributes of the transaction such as the start time and duration of a call. In addition, the operator records cell towers with which the phones of the caller and recipient are connected. This information makes it possible to use CDRs to know the location of both parties (UN Global Pulse, 2013a). CDRs have a number of potential uses. The information about cell towers provides insight into the community's movement patterns such as how people move from home, work, school, markets or clinics. More importantly, such information provides a basis for assessing the potential spread of a disease into the area and the movements of a disaster-affected population (UN Global Pulse, 2013b). This information provides key insight for relief efforts.

Cellphone transactions have been recognized as a major source of data for developmental issues. For instance, the characteristics of data related to microfinance transactions such as the number and characteristics of clients, loan amounts and types, and default rate arguably fall between traditional development data and BD (Global Pulse, 2012). With a more widespread use of mobile and online platforms for microloan transactions, a large amount of microfinance data can be digitized and analyzed in real time.

Activity data generated by social media also constitutes a major data source for developmental issues. For instance, most of Facebook's growth in recent years is coming from emerging markets. Among the 10 countries with the most Facebook users in 2012, six were emerging markets. Five of them, India, Brazil, Indonesia, Turkey, and the Philippines, accounted for 217 million Facebook users in 2012 (Mims, 2012). This growth can be partly attributed to initiatives such as Facebook Zero. Thanks to Facebook's collaboration with mobile operators from a number of emerging economies, users can access 0.facebook.com (Facebook Zero) completely free. Facebook Zero contains the key features of Facebook. A majority of users in developing countries use mobile devices to access Facebook. Most of these phones are feature phones that operate on a pay-as-you-go basis, rather than smart phones with app capabilities. Every Phone app, which runs on around 3000 feature phone models worldwide, has made it possible for these users to access Facebook. As of July 2013, over 100 million people used this app. Some telecom carriers in countries such as India, the Philippines, and Indonesia offer free or discounted data for Facebook Zero users (Byford, 2013).

## Velocity

Velocity is considered as a "competitive differentiator" for businesses using BD (Laney, 2001: 2). In this paper's context, BD provides the possibility for real-time feedback, which can lead to important developmental outcomes. For instance, by monitoring a population in real time, it is possible to understand the areas where developmental policies and programs are failing and make required adjustments (Letouzé, 2012).

A number of initiatives that have been launched to promote a BD ecosystem have focused on velocity of data. In sub-Saharan African (SSA) economies, the use of farm credits is reported to decline due to poor access to financial services, high borrowing costs, and high risks associated with such credits (Oluoch-Kosura, 2010). The creation of high-velocity data has helped address some of these problems. For instance, as of September 2013, the information created by

AgriLife, a cloud-mobile platform in Kenya which provides financial institutions and suppliers “near-real-time information” on farmers’ ability to pay for services (capacity.org, 2013), facilitated over US\$ 2 million in revolving credit lines to about 120,000 small farmers in Kenya and Uganda (G-Analytix, 2013).

As another example, the World Bank’s “Listening to LAC” (L2L) initiative in Latin America deployed mobile technologies to conduct real-time self-administered surveys. The surveys collect life events data on a near real-time basis and generate panel data. The data is expected to inform policymakers on current indicators and help them to respond more quickly and effectively to key trends (The World Bank, 2010). The data collection instrument is also expected to help policymakers assess the impact of their programs in real time and observe coping mechanisms in situations such as migration, school attendance, employment patterns, and nutrition (The World Bank, 2010).

It is especially important to explain the benefits of BD in the context of the lack of availability of data on key developmental indicators. Most traditional development data come from surveys (e.g. household, labor market, living standard) and official statistics. In addition to high costs, key problems of survey data include a relatively longer time to collect and analyze. Painfully, developing nations need to wait for a decade or more to adjust the GDP or estimate poverty indicators (Fengler, 2013). Data that is collected more frequently, which is better organized, can help assess the social and economic conditions faster.

## Variety

It is important to first define structured and unstructured data. Structured data can be organized in an assigned format that can be used by a database management system such as Oracle and Microsoft SQL. Some examples include histories of mobile payment transaction and the date of a Twitter account creation. Such data can be arranged in a list, compared with other data, used to generate new data, and retrieved for decision making. Unstructured data, on the other hand, is unformatted and lacks a predefined standard structure (e.g. cannot be organized in terms of rows and columns). Some examples include email messages, social media posts, pictures, and video. It is also worth noting that some sources involving interactions between people and machines such as web applications or social networks may provide multi-structured data. For instance, web log data includes unstructured data such as text and visual images and structured data such as transactional information (Arthur, 2013).

Structured and unstructured data are being increasingly combined in developmental projects. For instance, the Malaria Surveillance & Mapping project in Botswana was a pilot program launched in 2011, which aims to move away from paper reports towards mobile clouds. Health care workers are equipped with mobile phones to gather and upload malaria-related data to the cloud. The data can also be tagged with structured data such as GPS coordinates and unstructured data such as pictures, video, and audio. If there are signals of an outbreak of malaria, Ministry of Health officials and other health workers in the area receive a real-time notification via text message (mhealthinfo.org, 2011).

Another remarkable example of the utilization of structured and unstructured data is “Water Watchers”, a mobile application developed to report water-related issues in May 2013 by IBM and South Africa’s City of Tshwane. One estimate suggested that 60% of water worldwide is lost due to leaky pipes (Carew, 2013). The app’s users take a picture which shows a water-related problem and answers three questions about the problem. This data is then uploaded in real time on a cloud server. The information generated can be expected to identify a water “leak hot spot” map (Carew, 2013). BD may hold a great promise for finding appropriate steps to prevent or minimize this wastage.

## Variability

The approaches in BD assume that correlations can be considered as pragmatic indications of relations among variables (Mayer-Schönberger and Cukier, 2013). In this regard, one technique which has been of great interest in developing countries is anomaly (or outlier) detection based on the variability over time in the amount of data flow related to a given developmental indicator. Note that anomaly detection involves identifying items or events that fail to conform to an expected pattern. The idea here is that items or events that exhibit an anomalous behavior (e.g. unusually high rate flow of data of certain categories) may be associated with some kind of problem. To take an example, the data formats used in Twitter’s API are such that they provide the dynamic anomaly detection features (Madsen, 2013). The idea here is that the key metadata used in segmenting Twitter data such as hashtags (#) and replies (@) are user-generated and thus are logical targets to follow in order to understand problems and crises that users face. In sum, observing patterns of anomalies inside the data flowing from the Twitter API, one can detect signals of a crisis (Madsen, 2013).

One example to illustrate this point comes from a research project undertaken by the UN Global Pulse, which indicated that analysis of Twitter data can provide information on an increase in food prices. An analysis of a data set containing thousands of Tweets from Indonesia discussing the price of rice indicated that the volume of Tweets about staple foods had a positive correlation with increase in the cost (UN Global Pulse, 2013a).

## Complexity

It would be helpful to first note that the really big difference between variety and complexity concerns multiple data types (variety) versus multiple sources of data (complexity). Matching and linking data from multiple sources such as CDRs, open portals, social media, government sources, non-governmental organizations (NGOs), and corporations can provide a whole picture of the economic and social conditions of the rural population and thus valuable and relevant new insights (bigdata-startups.com, 2013). To take an example, a study of the ESS department of Harvard University (hsph.harvard.edu, 2014) indicated that BD can be employed to predict food shortages by combining variables from a number of sources such as drought, weather conditions, migration patterns, market prices of staples, seasonal variation in prices, and past productions (bigdata-startups.com, 2013). As another example, time-series analyses of CDRs can be combined with random surveys to provide better insights about the dynamics of rural economies and help devise appropriate government responses (bigdata-startups.com, 2013).

## BD in development: Some present uses and potential future applications

These are noteworthy and encouraging trends in BD's utilization in developing economies. BD is playing an increasingly important role in several key development areas such as healthcare, agriculture, biotechnology, education, and environment monitoring.

BD has been effectively used to evaluate and measure the impacts of humanitarian aid and similar interventions. One example is the UN's analysis of social media posts to find out whether its Every Woman Every Child (EWEC) initiative accomplished the goal of effectively delivering the message to the target audience.

The UN trained a team to monitor and recognize relevant Tweets. The EWEC team used the analytical tool Foresight developed by the social media analysis consulting firm Crimson Hexagon to analyze public Tweets from September 2009 to July 2013. The team developed a taxonomy of relevant keywords such as "maternal health", "breastfeeding", and "vaccination of children" to identify messages that are relevant to women's and children's health and searched these keywords (Global Pulse, 2012, 2014). An analysis of millions of social media posts in a two-year period regarding the extent and frequency with which relevant keywords were used indicated a significant shift towards an increased awareness of parental health (Kirkpatrick, 2013).

In the pre-BD era, the principal way to measure the impacts of humanitarian aid programs such as the EWEC would be to conduct household surveys. It is thus important to compare the performance of BD with such surveys. In this regard, Table 3 illustrates how different BD dimensions can be applied in the context of EWEC's analysis of social media posts. Regarding the volume, despite problems such as the relative lack of randomness compared to household surveys, it is possible to inexpensively extract a huge amount of data from social media posts. Sometimes the final results may take years since key processes in household surveys such as visiting households for data collection and entering data into a computer system are extremely time consuming and expensive. Finally, such surveys are not effective in collecting and analyzing unstructured data.

**Table 3.** Relevance of BD dimensions in EWEC's analysis of social media posts.

Characteristic	Explanation in the EWEC context
Volume	The EWEC team analyzed about 14 million Tweets.
Velocity	Utilization of user-generated digital content and Crimson Hexagon's analytical tool led to faster analysis and decision-making.
Variety	The analysis involved structured data such as the dates of Twitter posts and unstructured data such as Twitter texts for the relevant keywords.
Variability	There was a sharp increase in Tweets containing the keywords identified by the EWEC team during key events such as World Aids Day, Mothers Day, International Women's Day and Millennium Development Goals (MDGs) summits.
Complexity	Combining the Tweet data with data from other sources (e.g. Facebook posts) might offer further insights and understanding of the EWEC initiatives and validate the results.

A similar use of BD has been in providing real-time information on important social and economic indicators. For instance, sentiment analysis can provide insights into a community's attitude toward certain key issues. To take an example, the United Nations Children's Fund

(UNICEF) (2013) showed that by analyzing parents' social media posts, it is possible to track attitudes towards immunization.

Another notable use of BD has been in the design of an effective method and system for efficient distribution and management of life-saving medicines that have limited supplies. One example is the World Health Organization's (WHO) pilot program called SMS for Life (<http://www.wksu.org/toolkit/chapter4/section1.html>), which aims to improve the distribution of malaria drugs in Tanzania's rural areas at the health facility level. Demand for antimalarial drugs is highly unpredictable. Such drugs cost as much as US\$ 10 per course, which is prohibitively expensive for most Tanzanians (irinnews.org, 2009).

A creative feature of the process is that front-line workers from every clinic send an SMS with their stock count every week. Based on the figures sent by the clinics, senior coordinating staff can determine an appropriate target restocking level at a given clinic to make sure that no stock-out occurs. The WHO reported highly encouraging results. The proportion of clinics with no stock of at least one antimalarial medicine decreased from 78% to 26%. In one of the three districts, stock-outs were completely eliminated by week 8 of the pilot (Newton, 2012). It is worth noting that the data used in SMS for Life is not as big as that of the EWEC initiative. Nonetheless, by aggregating and cross-referencing data sets collected from a number of clinics and other sources, the program has been able to improve the distribution of malaria drugs. According to Boyd and Crawford's (2012) definition, the data used in SMS for Life can be considered as BD for development.

#### Potential impacts of BD in enhancing transparency

The above discussion indicates BD's potential socioeconomic impacts in a number of domains and sectors. Among the major impacts we may point out the BD-led promotion of transparency and accountability. Note that transparency involves making information about an entity's operations, structures, and other attributes available to the public (Finel and Kristin, 1999; Heald, 2006). Transparency has gained wide support among state decision-making bodies, international organizations, and private companies (Finel and Kristin, 1999).

Governments can increase transparency by making information available to the public. For instance, the US government's launch of the website Data.gov in 2009, which makes statistical information collected by over 50 federal agencies available to the public, is considered as an important transparency measure (Etzioni, 2010). Among developing economies, Kenya is probably the most spectacular example in making data available to the public and facilitating the use of BD. In 2011, Kenya launched an Open Data Portal (ODP) with the help of the World Bank. The project received support at the highest levels of the government. The data in the ODP includes a full digital edition of the 2009 census, government expenditure for 12 years, household income surveys, and data about the location of schools and health facilities.

A popular version of the theory of transparency is that, thanks to their purchasing power, consumers play a key role in controlling the economy by choosing the businesses that are likely to succeed or fail (Sirgy and Su, 2002). In order to exercise their power, consumers like to have information about the attributes of the goods they purchase. Manufacturers thus provide

information such as caloric value, types, and levels of vitamins in the labels of food items in a product. According to the transparency theory, this disclosure enables consumers to make informed choices and reward the businesses that provide the preferred products. This practice puts business disregarding consumers' preferences at a disadvantage (Sirgy and Su, 2002). In this paper's context, developed country-based firms are in a strong position to exercise their power over farmers from developing countries. The use of BD forces farmers to use less pesticide, which is likely to help them enhance the quality of their products and stay competitive. As noted earlier, the availability of digital records of farming activities plays a major role in documenting quality standards of agricultural products.

Apart from the obvious direct economic effects, the use of BD is also associated with a number of non-economic benefits. Strengthening transparency by making information public would help monitor and discipline office-holders and fight corruption (O'Neill, 2006). Bentham (2001: 277) noted that "... the more strictly we are watched, the better we behave". Initiatives such as Kenya's ODP can thus be seen as a key factor in strengthening the performance of government and public administration in developing countries.

Prior research indicates that transparency is likely to be more useful when the level of information costs is lower (Etzioni, 2010). In this regard, the rapidly falling costs of collecting, processing, storing, and transmitting data and information are likely to play major roles in promoting transparency and accountability in the public and the private sectors. This benefit is especially stronger for small- and medium-sized enterprises (SMEs) engaged in exporting products.

### **Opportunities and challenges of using BD: The case of agriculture and farming activities**

One of the most important benefits of BD can be in improving agricultural productivity. This effect is likely to have tremendous developmental benefits as the agricultural sector employs over 60% of the active labor force in SSA economies (Oluoch-Kosura, 2010). According to the UNCTAD's World Investment Report (2009), over 900 million people in the world were undernourished, and 65 countries faced "serious" or "alarming" danger of food shortages and famine. Studies have suggested that ineffective farm operations such as late planting/weeding, the lack of proper land preparation and harvesting techniques, and poor housing and feeding for livestock can reduce smallholder farmers' productivity by up to 40% (Oluoch-Kosura, 2010). BD has a potential to improve this condition. According to Monsanto, the world's biggest seed company, tailoring information and advice to farmers could increase annual world-wide crop production by about US\$ 20 billion (Bunge, 2014).

Some argue that BD is the source of the next revolution in farming (Bunge, 2014). An overview of the deployment of BD in industrialized countries would be helpful for how the condition can be improved in developing countries. On this front, precision agriculture or precision farming has been a key trend in industrialized countries. Data collected on soil conditions, seeding rates, crop yields, and other variables from farmers' tractors, combines, and drones is combined with detailed records on historic weather patterns, topography, and crop performance collected by the providers of prescriptive-planting technology (Bunge, 2014; foxnews.com, 2014). Human experts may need to perform tasks involving decision problems and processes for which no

algorithm exists or the algorithm has not yet been developed. In some cases, due to unknowns no algorithm can solve all instances of the problem. In agriculture, some examples of situations include tasks involving unknown soil types, and extreme weather conditions, which often need to be performed by humans rather than algorithms. The data is thus crunched by algorithms and human experts and turned into customized useful advice and is sent directly to farmers and their machines, instructing them as to the optimum amount of pesticides, herbicides, fertilizer, and other applications.

Many tractors and combines are guided by global positioning system satellites. An article published in *usatoday.com* explained that a corn and soybean farmer in Iowa used a \$30,000 drone to study how the yield in his 900 acre farm is affected by changes in topography and other factors (Doering, 2014). This example is illustrative of a widespread adoption and diffusion of BD in the agriculture sector in industrialized countries. Many farmers who have implemented data-driven prescriptive planting based on the analysis of nutrients in soil and other factors have reported a significant increase in productivity (Bunge, 2014). The point is that even small alterations in planting depth or the distance between rows of crops can lead to a significant increase in agricultural productivity.

The diffusion of BD is associated with and facilitated by measures taken by the providers of prescriptive-planting technology to strengthen their resources and capabilities. In 2013, Monsanto acquired the weather-data-mining firm Climate Corp. Likewise, the agricultural cooperative Land O'Lakes bought satellite-imaging specialist Geosys. In the same vein, in order to provide real-time climate and market information to its data service users, DuPont announced collaboration with the weather-and-market analysis firm DTN/The Progressive Farmer. In 2013, Deere agreed to send data from its tractors, combines, and other machinery to the computer servers of DuPont and Dow (Bunge, 2014).

Developing world-based small farmers face several key challenges. Studies conducted in Sri Lanka and other countries have indicated that farmers are not able to sell harvests due to oversupply or not getting the planned harvest, and the lack of necessary information (Walisadeer et al., 2013). BD can help address this problem.

Nutrient management is another area where BD may be relevant. In Africa, outdated knowledge is pervasive and ubiquitous in recommendations for nutrient management. This often leads to too much fertilizer in relation to potential crop demand and on a uniform basis irrespective of the type of land (Giller et al., 2011). A model-based and data-driven approach is thus likely to reduce the costs of fertilizer and increase productivity.

A further area in which BD might have potential to facilitate agricultural and farming activities in developing countries relates to the availability of near-real-time data and information regarding farmers' needs and capabilities, which can be used by value chain partners to effectively serve the farmers. One example, as noted earlier, is the cloud-based platform AgriLife, which is accessible via mobile phone. It is used for collecting data and analyzing farmers' production capability and history. In order to ensure fast, easy and efficient availability of resources and services to distant, rural farmers, the platform also acts as an integration point for financial institutions, mobile network operators, produce buyers, and their agents (Yeoman,

2013). The data analysis provides a better understanding of small farmers' needs and production capability. Service providers can tailor their offerings such as crop insurance, input payments, and savings accounts based on the data (Yeoman, 2013). Uganda's Farmers Centre (FACE) was an early adopter of AgriLife. FACE started uploading information on its 10,000 farmer clients, who travel long distances to purchase inputs and aggregate their produce at FACE warehouses for processing/sale. Before using AgriLife, FACE collected information by paper-based questionnaires. Small farmers' transaction data would help them build a credit history, which is used by value-chain actors to provide credit and other resources such as seeds, fertilizers, and pest-control chemical agents (Yeoman, 2013).

As mentioned earlier, as of September 2013, AgriLife facilitated over US\$ 2 million in revolving credit lines to about 120,000 small farmers in Kenya and Uganda. The AgriLife platform is also being used in Zimbabwe, Zambia, and Senegal (G-Analytix, 2013).

Prior researchers have recognized that developing world-based farmers face difficulties in meeting the quality and safety standards set by the developed world (Oluoch-Kosura, 2010). In this regard, the availability of easily accessible data that include digital records of farming activities such as the amounts of seeds and pesticides will obviously play a major role in documenting quality standards of agricultural products. It would be interesting to assess the above examples related to the use of data in farming activities in developing countries in terms of BD dimensions. Obviously, higher volume of data on farming activities is available than in the past. For instance, data such as farmers' credit history and the amounts of seeds and pesticides used was not available in the pre BD-environment. As to the data speed, near-real-time data and information on farmers' needs and capabilities are available. This means that financial institutions, produce buyers, and other relevant actors can fulfill farmers' needs more quickly than in the past. Regarding the variety, most data currently used in farming-related activities is structured data. Such data can be combined with unstructured data. For instance, farmers can upload pictures and videos related to a problem they are facing, which can be analyzed by experts to offer customized advice.

In one way, TNCs are likely to be a driving force behind the diffusion of BD in the agricultural sector of developing countries. Large food and biotechnology TNCs such as Monsanto and Syngenta already have a notable presence in developing countries, which is a positive factor from the standpoint of BD-led productivity growth in these countries. During 2005–2007, the share of agriculture in foreign direct investment (FDI) inflows was 15.1% in Cambodia and 12% in Laos (UNCTAD, 2009). Monsanto reportedly has control of over 95% of the Indian cotton seed market (Vidal, 2011). TNCs, which are often producers, processors, or traders of agricultural products or sellers of inputs or machinery, engage in a contracting system in which they assume a variety of responsibilities including providing technical assistance and marketing to developing world-based small farmers (Glover, 1984, 1987). TNCs such as Monsanto and Syngenta, which have become a driving force behind the utilization of BD in the industrialized world, are thus likely to act as a key channel in the international technology transfer of BD.

A related point is that international technology transfer in BD is likely to have differential effects across different categories of crops. For instance, foreign companies are more active in newly emerging export crops, which are integrated into the international supply chain. Traditional cash

crops such as coffee, cotton, tea, and tobacco are thus more likely to realize the need to adopt various aspects of BD (Hoeffler, 2006).

Other potential mechanisms and determinants of BD diffusion among farmers also exist. For instance, Oluoch-Kosura (2010) reported that NGOs, farmers' organizations, and the private sector in Africa are playing important roles in facilitating farmers' education, access to agricultural information, and training. International supply chain structures often tend to exclude smallholder farmers. In Mozambique, farmers, who are engaged in contract farming, pool resources to get technical advice and other services. More than 400,000 smallholders with less than one hectare of land each are reported to benefit from such arrangements (Hoeffler, 2006).

### Critical challenges and issues

Against the backdrop of rapid diffusion of BD among big farmers in industrialized countries, a comparison of their BD ecosystems with those of developing countries would be helpful to understand critical challenges and problems in the effective utilization of BD for developmental issues. First, and perhaps most important, agriculture firms in the industrialized world have a long history of data production and consumption. For instance, DuPont has been making use of farm-level data since the early 2000s (Bunge, 2014). Likewise, it is increasingly common for farmers to monitor the progress of their agricultural activities on iPads and tablets. In industrialized countries, firms in diverse industries such as satellite-imaging, weather-data-mining, and weather-and-market analysis have enabled a rich ecosystem of BD.

While large growers can afford specialized machineries, small farmers are not in a position to do so (Glover, 1987). The conditions that stimulated the growth of BD in the US farming industry such as the widespread adoption of mechanized tractors, genetically modified seeds, computers, and tablets for farming activities are less prevalent in developing countries. Most smallholders in developing countries are not in a position to do so. Smallholder farmers often have no means to access the data and cannot interpret it. A main concern is that BD collection efforts will only benefit big and well-educated farmers (Palmer, 2012).

Accurate and actionable data require considerable technical skills to handle data mining and analysis method and system. The lack of human resources and expertise represents another major barrier to the implementation of BD projects. Even industrialized countries such as those in European Union (EU) economies have reported a huge skills shortage for data-related manpower (Kroes, 2013). Data scientists are both in short supply and expensive to employ in SSA economies (WEF, 2012). Most of the top BD companies are from the industrialized world and developing competitive indigenous companies in the BD area is not an easy task for developing countries. The EU competition commissioner Neelie Kroes noted that of the top 20 global BD companies, 17 are from the US and two from Europe (Kroes, 2013). Another study suggested that of the 15 most powerful BD companies, 14 were US-based and one was Europe-based (Korolov, 2013). It is argued that the highest performance computers are unaffordable even to a member of the EU (Kroes, 2013).

As an upshot of the above discussion, there is a lack of appropriate database systems for agribusiness development, agriculture management, and produce distribution. A BD attempt is

greatly hampered by the lack of reliable infrastructure to collect information. Consider, for instance, climate-related historical data. African countries have limited capacity to develop, generate, disseminate, and effectively use climate data and information (Twomlow et al., 2008). National institutions, leadership, and the civil society are inherently weak and cannot determine the types of climate data and information needed for agriculture and other economic activities. Among the problems faced by policymakers and practitioners to work more effectively to respond to climate changes and other climate related effects concern extremely low number of meteorological stations for climate data collection and the lack of digitization of the data (United Nations Framework Convention on Climate (UNFCCC), 2007).

As another example, the lack of information has been a main barrier to an effective implementation of healthcare systems. For instance, while there is a rising prevalence of diabetes in Indonesia, there is no data available to measure the effects beyond intermediate outcomes such as the number of people trained; percentage of health centers providing education; or development of training material and guidelines (e.g. training's impact on detection rate and outcomes and screening's impact on complications) (Soewondo et al., 2013).

Farmers are concerned about the potential misuse of information at the firm level related to their farming activities and at the industry level. For instance, the trade group American Farm Bureau Federation (AFBF) warned its members that seed companies' prescriptive planting programs have vested interests in higher crop yields associated with BD's use (Bunge, 2014). Big agricultural firms such as Monsanto might influence farmers to buy specific seeds, sprays, and equipment and are likely to profit from the costs of their services and higher seed sales (Seppala, 2014). The gathering of data from sensors on tractors, combines, and other farm equipment by large seed companies is receiving the same level of attention as immigration reform and water regulations (foxnews.com, 2014). Another key concern that farmers have expressed is that their data and information could be used by competitors. For example, other farmers' access to the crop-yield information may create direct and unwanted competition to rent farmland, which may cause a new spike in land values and seed prices (Bunge, 2014). The issue regarding who owns farmers' crop data is also of equal concern (Seppala, 2014).

Another fear is that Wall Street traders could use the data to make bets that could hurt the farmers. For instance, if conditions early in the growing season lead to lower futures contract prices, it may reduce the profits farmers could have made from crops by locking into sell the futures (Bunge, 2014). Likewise, farmers are concerned that hedge funds or big companies might use real-time data at harvest time from a large number of combines to speculate in commodity markets long before official crop-production estimates are available (foxnews.com, 2014). This fear has some foundation as the developments in BD technologies make it possible to do so. For instance, a group at the MIT Media Lab used location data from mobile phones to estimate the number of people in Macy's parking lots on Black Friday. The estimate made it possible to estimate the retailer's sales on that day even before Macy's had recorded those sales. Insights like this are expected to provide competitive advantage to Wall Street analysts and managers (McAfee and Brynjolfsson, 2012).

In the developing world's context, an even bigger question than that of whether agricultural productivity can be improved using BD is who is likely to benefit from the BD-led growth in

productivity. One possibility is that agricultural productivity associated with BD utilization in developing countries may provide benefits primarily to foreign companies. This is because while a number of positive outcomes of agricultural TNCs' operation in developing countries have been recorded, there are also possible negative effects such as the potential abuse of their market power and dominant position. One estimate suggested that foreign investors acquired (or sought) about 15–20 million hectares of farmland in developing countries during 2006–2009 (UNCTAD, 2009). The increasing globalization of agriculture and the food chain means that industrialized world-based agricultural giants may *expand such activities globally*.

Security and privacy issues associated with BD have attained at least some degree of institutionalization in industrialized countries, which is a small comfort for the farmers (Kshetri, forthcoming). Most industrialized countries have more well-developed regulations related to data privacy and security. They also have industry standards, company-specific guidelines, and performance measures. For instance, US-based food and agricultural companies such as Monsanto, DuPont, and other corporations claim that they do not use data for purposes other than providing services requested by farmers, keep the data secure and do not sell it (foxnews.com, 2014). Some companies get consent from customers before sharing their data. The AFBF has put together a “privacy expectation guide” to educate its members. In addition, it has drafted a policy which has emphasized that data should remain the farmer's property (foxnews.com, 2014). Some US farmers are reportedly contemplating a new initiative to aggregate data on their own so that they can decide the type of information to sell and at an appropriate price. Other farmers are teaming up with smaller technology companies in order to challenge the domination of big agricultural giants in the prescriptive-planting business (Bunge, 2014). Many developing countries currently have no regulatory safeguard in place to protect farmers and citizens from possible data misuse. This means that BD-related issues are being considered in a setting of nascent institutionalization. Farmers in developing countries are even more prone to exploitation by big businesses.

## **Discussion and implications**

Each generation has access to more information and experiences more information overload compared to the preceding generation (Blair, 2003). However, a key feature of the recent information revolution is that developing economies are experiencing a rapid explosion of data and information due primarily to a widespread diffusion of cellphones and social media in these economies. If history is any guide, this explosion is likely to change the ways individuals and organizations act and interact. In an analysis of the effects of information overload from the 16th to the 18th century, Blair (2003) found that the availability of more information led to the diffusion and development of various learning aids and tools and also affected the way scholars worked. It is thus reasonable to expect that BD is likely to have a profound impact on development-related activities such as agriculture and healthcare and related decision-making processes.

BD may have different meaning and significance for the purpose of development. For instance, BD applications in developing countries may not necessarily involve petabytes of data as used by TNCs such as eBay and Walmart. They may also be characterized by relatively lesser variety of

data. Likewise, as to the velocity dimension, a data-processing time of several months may be considered fast enough.

BD is being used to understand and respond to important development issues such as water supply, food security, human health, conservation of natural resources, and protection against natural hazards. The above discussion makes it clear that while social media activities have important social aspects and consequences, SM data also has at least some economic importance. There have recently been some encouraging developments in the utilization of BD in improving farmers' livelihood and access to services. The cases of AgriLife and other initiatives indicate that BD has promoted better functioning of the market.

Experts have emphasized the importance of assessing organizations' "information supply chain to identify and prioritize data management issue" (Laney, 2001: 2). This is also relevant to development-related data. Data in different contexts may come from different combinations of sources. The composition and structure of the data may differ across economies. Among developing countries, Indonesia has the most Twitter users (Richter, 2013). More Tweets come from Jakarta than from New York, Tokyo, London, or São Paulo (Florida, 2012). In Kenya, the mobile money transfer is extremely popular and could serve as an important data source (Kaplan, 2013; UN Global Pulse, 2013b). Sixty-eight percent of Kenyan cellphone users regularly use their phones to make or receive payments (pewglobal.org, 2014).

There is likely to be a wide variation across economic activities and industries in the level of the diffusion of BD. Even within an industry, differences in the diffusion of BD are likely to be significant. For instance, in the agricultural industry, cash crops that are integrated in the modern supply chain are likely to see the impacts of BD sooner.

Several important lessons can be drawn from the successful examples. Initiatives in industrialized countries to develop a BD ecosystem should be instructive for gaining an understanding of BD's impact on development. The US Department of Agriculture, for instance, announced the launch of a portal on the Data.gov website, which links to 348 agriculture data sets (Patel, 2013).

Prior researchers have suggested that social and other forms of supports from more experienced users are likely to promote technical competence of new users (DiMaggio and Hargittai, 2001). In this regard, multinational companies and international agencies may act as key forces in providing such supports to facilitate BD diffusion in developing economies.

Political as well as economic factors are of crucial importance in determining the distribution of benefits resulting from the system (Glover, 1984, 1987). Farmers who are able to mobilize and organize themselves (e.g. in cooperatives or other forms of groups) may increase their bargaining power vis-a-vis TNCs based in industrialized countries.

The conditions in developing countries provide limited incentives to encourage the public and private sectors to invest in the creation of relevant database. It is important to develop means to make usable and relevant knowledge available to smallholder farmers in a timely manner. Data accessibility is more challenging in developing countries. In light of the usefulness of climate-

related data as noted above, for instance, more investment into climate observation networks is needed. Historical and projected data on climatic conditions would be of great help to farmers. While the creation of a database that is completely customized to meet the need of every field and every farm may not be feasible in the short run, making information available regarding a basis for even a broad categorization (e.g. soil type) can be of great help. For instance, in Kenya and Zimbabwe, while a wide heterogeneity exists among farms, they can be arguably reduced into three categories in relation to a response to fertilizers: (1) fertile fields unresponsive to fertilizers, which require only maintenance fertilization; (2) intermediate fields highly-responsive to fertilizers, which require managing fertilizers efficiently; and (3) infertile fields unresponsive to fertilizers, which may require complete restoration and rehabilitation (Giller et al., 2011). Providing advice and guidance to farmers based on the responsiveness to fertilizers of their farms is likely to lead to a more appropriate management of fertilizer use.

Regarding the observation that high-performance computers may be unaffordable to utilize BD effectively, it is important to note an encouraging trend: the diffusion of cloud computing in developing economies. These economies are experiencing dramatic and significant cloud-led socio-economic transformations (Kshetri, 2012, 2013). Global cloud vendors have entered in developing markets and some high-profile entrepreneurial firms are emerging in the cloud's supply side in these economies. This means that individuals and businesses in developing countries can effectively utilize BD by renting storage and computing power from the clouds.

Despite the optimistic promises that some consultants and development experts have made regarding the developmental and practical applications of BD (Kirkpatrick, 2013; Laney, 2001; Letouzé, 2012), it should not be viewed as a panacea and the answer to all the many and varied problems facing the developing world. There is a need to fully assess availability, appropriateness, and effectiveness of BD in addressing development challenges. There are only limited types of data that can be found for most developing economies (e.g. related to Tweets in Indonesia and mobile money transfer data in Kenya). Data unavailability thus remains a major challenge, which, according to Boyd and Crawford (2012) has led to a new form of digital divide. While appropriate analysis of BD may provide valuable insights and information for key policy areas, great care must be taken to ensure that data quality standards are satisfied and appropriate methodological steps have been taken. For instance, the use of Twitter API data has been criticized on the grounds that it suffers from questionable quality and serious methodological challenges such as samples of unknown representativeness, a lack of one-to-one correspondence between accounts and users and proliferation of Tweets created by bots (Boyd and Crawford, 2012; Crawford, 2009).

It is important to discuss the above problems in the context of transparency. Whereas transparency is essential to ensure reliability and validity, BD created through the use of social media is often produced by commercial organizations' closed structures (Driscoll and Walker, 2014). For instance, Gillespie (2011), pointing out the fact that Twitter engages in censorship, has argued that users have displayed a "misplaced faith" in Twitter Trends, and has urged them to stop "worshipping algorithms".

There are also important privacy and ethical issues involved around BD. Acquisti and Gross (2009) showed that combination of public databases can lead to serious privacy violations such

as revelation of individuals' social security number and other sensitive information. What is particularly relevant in the context of this paper, as noted earlier, is that some analysts have a tendency to criticize some uses of BD in industrialized countries on the ground that it is a surveillance tool. On the other hand, these analysts' view is that people in developing countries do not need privacy. Some have challenged this view and pointed out that poor people have no less reason than rich people to be worried about surveillance (linnettaylor, 2013). Moreover, in countries characterized by conflict, crisis, and weak law enforcement, the privacy challenge may be a security risk (Letouzé, 2012). In China, a malicious actor reportedly can sell a database containing a specific type of information, for instance, phone numbers, for more than US\$ 1500 on the black market. The illegal companies, in turn, charge their clients between US\$ 1500 and US\$ 150,000 for services such as private investigation, illegal debt collection, asset investigation, and even kidnapping (Yan, 2012).

### Limitations

Several limitations of this paper must be recognized. First, some of the main arguments presented here rely on the reports of companies which are consultants or leading suppliers of BD-related solutions (e.g. Monsanto, Gartner, Cisco, and McKinsey) and developmental agencies promoting BD (e.g. Global Pulse). These companies may have vested interests in promoting the diffusion of BD and thus may overemphasize the positive aspects of BD. An additional limitation of this research is that it covered only materials published in the English language.

### Policy implications

The above discussion suggests some important policy implications which stress the need to emphasize the enrichment of the BD ecosystem and to ensure that appropriate regulations aimed at encouraging organizations' BD adoption in activities with positive social and economic contributions and outcomes are in place. Due to the public goods nature of data, organizations that invest in data collection cannot necessarily reap all the benefits. While some statistics are gathered several times in different ways, others are rarely or never collected (adb.org, 2013). Appropriate incentives are needed to collect relevant data and overcome the fragmentation among the governments and organizations collecting development-related data.

The government is a key actor that can drive the BD ecosystem. Civic organizations, mobile app developers, and media groups are using the data available at Kenya's ODP to improve understanding of population patterns and transparency of public services (WEF, 2012). Governments and international agencies need to support the expansion of initiatives such as Kenya's ODP in areas such as agriculture.

It is especially necessary to introduce policies, procedures, and interventions to ensure the privacy and confidentiality of sensitive data. To return to the Kenya example above, the ODP was put in place without appropriate legal frameworks in key areas such as protections for data reuse (WEF, 2012). While Kenya's constitution guarantees openness, transparency, and participation, which facilitated the ODP's establishment, many authoritarian regimes lack such an environment.

Data consumption and exchange are no less important than data production and analysis. Some argue that the consumption of free data such as time spent on social media and on cellphones may provide a “consumer surplus” not captured in official statistics (Letouzé, 2013). The utilization of BD in key development areas hinges critically upon the availability of manpower with BD competency. It is thus important for national governments and international agencies to direct more efforts towards developing BD manpower.

Guidelines, interventions, supports, and incentives are needed to encourage sharing existing data. In this regard, much of the valuable data that is relevant for the development context is often with the private sector. For instance, networks of mobile phone operators have data related to text messages, digital-cash transactions, and location data. In some cases, the government owns these companies. One way to enrich the BD ecosystem would be to persuade these organizations to make the relevant data available (Lohr, 2013). Universities and research centers also constitute a key source of data and knowledge. It is argued that scientists working in these institutes are against making relevant data accessible due to security, privacy, and other concerns. They often use reasoning against data sharing, such as: “I don’t want to share it”, “it’s mine” or “It’s government property” (Patel, 2013). Analysts have stressed the importance of providing incentives to individuals to share information such as pricing/offers and improved services. It is also important to develop privacy standards and “opt out” ability (WEF, 2012).

Finally, for national governments and international agencies such as the UN or the World Bank, it is critical to engage in new and more comprehensive collaboration with the private sector in order to facilitate the diffusion and effective uses of BD in developing countries. Here we introduce the concept of “Data Philanthropy”, which is likely to play a key role in the diffusion of BD in developing countries (Pawelke and Tatevossian, 2013). This idea emerged at the WEF in Davos in 2011. The basic idea behind Data Philanthropy is simple: it involves a partnership in which businesses share data for public benefit. Data Philanthropy is described as the “next movement in charitable giving and corporate citizenship” (Coren, 2011). Data donated by corporations and governments can be used to track diseases, avoid economic crises, and aid development. While a huge amount of data is held by businesses, they are reluctant to distribute it even anonymously, due primarily to privacy concerns (Coren, 2011). It is thus important for national governments and international agencies to work closely with businesses such as Twitter and Facebook, which have large amounts of data, and encourage them to engage in Data Philanthropy.

### **Concluding comments**

The data being used in a number of developmental purposes can be considered as BD. Preliminary evidence indicates that BD is likely to help better utilize the scarce resources and can help deal with the various sources of inefficiency that have been frequently cited by critics as among the key obstacles for development in developing countries. For instance, BD can help to reduce the waste of inputs such as fertilizer and increase agricultural productivity and control the epidemics of various diseases.

Uses of BD that lead to positive social and economic outcomes and those that benefit socially and economically disadvantaged groups need to be promoted. Responsible uses of BD also require protecting people's dignity and legitimate expectations of privacy and economic interests. For instance, among the key lessons from the overview of risks to farmers over possible misuse of data and information in industrialized countries, these concerns are even more pronounced due to the lack of data protection regulations in developing countries. Moreover, most farmers in developing countries lack the degree of self-awareness and organization that can be observed among some industrialized world-based farmers. For instance, while farming groups in industrialized countries are taking measures to protect against misuse and exploitation of their data, such measures are lacking in developing countries.

BD obviously offers a number of potential benefits and vast possibilities in developed economies. Nevertheless, developing economies are at a nascent stage and far from a full utilization of the great potential of BD. Benefiting from BD requires a drastically different approach. In order to overcome barriers related to BD adoption, policymakers should ensure various enabling conditions for the creation, availability, and use of data. The lack of BD-related skills and competency underscores the importance of moving the focus beyond the numbers of technological devices to the strengthening of national technological capacity to use BD.

BD has not yet been a priority of policy issue for most of the donor agencies nor has this issue been closely tied to national development strategies. While many organizations and data collection process in industrialized countries constitute a rich ecosystem and serve a wide variety of functions needed to achieve various goals, most developing countries lack such ecosystems. It would be critical for the developing world's governments to get support from key stakeholders such as researchers, international agencies, software makers, data intensive sectors, and venture capitalists to create and utilize relevant, development-related data and information. Collaboration and cooperation among these stakeholders are essential to foster a data ecosystem for development. Creation of appropriate databases may stand out as particularly appealing and promising to some entrepreneurial firms. Governments, businesses, and individuals are willing to pay for data when they perceive the value of such data in helping them make better decisions. In the meantime, policymakers, academics, and other stakeholders should make the most of what is available.

### **Acknowledgement**

Constructive comments on earlier versions from co-editor Dr Irina Shklovski and two anonymous *Big Data & Society* reviewers helped to improve the paper substantially.

### **Declaration of conflicting interest**

The author declares that there is no conflict of interest.

### **Funding**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## References

- Acquisti A, Gross R (2009) Predicting social security numbers from public data. *Proceedings of the National Academy of Science* 106(27): 10975–10980.
- adb.org (2013) Big data: Vital statistics for development. Available at: <http://www.adb.org/features/big-data-vital-statistics-development> (accessed 18 August 2014).
- Allen D (1998) New telecommunications services: Network externalities and critical mass. *Telecommunications Policy* 12: 257–271.
- Antonelli C (1986) The international diffusion of new information technologies. *Research Policy* 3: 139–147.
- Archibugi D, Michie J (1997) Technological globalisation or national systems of innovation? *Futures* 29(2): 121–137.
- Arthur L (2013) What is Big Data? Available at: <http://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/> (accessed 18 August 2014).
- Bays J (2014) Harnessing Big Data to address the world's problems. Available at: <http://voices.mckinseyonsociety.com/harnessing-big-data-to-address-the-worlds-problems/> (accessed 18 August 2014).
- Bentham J (2001) Farming defended. In: Quinn M (ed.) *Writings on the Poor Laws*, vol. 1, Oxford: Oxford University Press.
- Bhushan A (2012) Big data. *USAID Frontiers in Development*, 11–13 June.
- bigdata-startups.com (2013) How Big Data can help the developing world beat poverty. Available at: <http://www.bigdata-startups.com/big-data-developing-world-beat-poverty/> (accessed 18 August 2014).
- Blair A (2003) Reading strategies for coping with information overload ca. 1550-1700. *Journal of the History of Ideas* 64(1): 11–28.
- Bowker GC (1996) The history of information infrastructures: The case of the international classification of diseases. *Information Processing & Management* 32(1): 49–61.
- Boyd D, Crawford K (2012) Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication, & Society* 15(5): 662–679.

Brown L, Malecki E, Spector A (1976) Adopter categories in a spatial context: Alternative explanations for an empirical regularity. *Rural Sociology* 41: 99–118.

Buente W, Robbin A (2008) Trends in internet information behavior, 2000–2004. *Journal of the American Society for Information Science and Technology* 59(11): 1743–1760.

Bunge J (2014) Big data comes to the farm, sowing mistrust; seed makers barrel into technology business. *Wall Street Journal (Online)* 26 February.

Byford S (2013) *Facebook's drive into developing nations scores 100 million users*. Available at: <http://www.theverge.com/2013/7/21/4544326/facebook-for-every-phone-has-100-million-users> (accessed 18 August 2014).

capacity.org (2013) *Agrilife platform*. Available at: <http://www.capacity.org/capacity/opencms/en/topics/gender-and-social-inclusion/agrilife-web-platform-kenya.html> (accessed 18 August 2014).

Carew J (2013) World war water. Available at: <http://www.itweb.co.za/?id=64253:World-War-Water> (accessed 18 August 2014).

Chunara R, Andrews J and Brownstein J (2012) Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *American Journal of Tropical Medicine and Hygiene* 86: 39–45. Available at: <http://www.ajtmh.org/content/86/1/39.abstract> (accessed 18 August 2014).

cisco.com (2014) Cisco visual networking index: Global mobile data traffic forecast update, 2013–2018. Available at: [http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white\\_paper\\_c11-520862.html](http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html) (accessed 18 August 2014).

Coren MJ (2011) Data philanthropy: Open data for world-changing solutions. Available at: <http://www.fastcoexist.com/1678963/data-philanthropy-open-data-for-world-changing-solutions> (accessed 18 August 2014).

Crawford K (2009) Following you: Disciplines of listening in social media. *Continuum: Journal of Media & Cultural Studies* 23(4): 532–533.

DiMaggio P and Hargittai E (2001) *From the 'digital divide' to 'digital inequality': Studying Internet use as penetration increases*. Princeton University Center for Arts and Cultural Policy Studies, Working Paper #15. Available at: <http://www.princeton.edu/~artspol/workpap/WP15%20-%20DiMaggio%2BHargittai.pdf> (accessed 18 August 2014).

DiMaggio P, Hargittai E, Celeste C, (2004) From unequal access to differentiated use? A literature review and agenda for research on digital inequality. In: Neckerman K (ed.) *Social Inequality*, New York, NY: Sage, pp. 355–400.

Doering C (2014) Growing use of drones poised to transform agriculture. Available at: <http://www.usatoday.com/story/money/business/2014/03/23/drones-agriculture-growth/6665561/> (accessed 18 August 2014).

Driscoll K, Walker S (2014) Working within a black box: Transparency in the collection and production of Big Twitter Data. *International Journal of Communication* 8: 1745–1764.

Etzioni A (2010) Is transparency the best disinfectant? *Journal of Political Philosophy* 18(4): 389–404.

Fengler W (2013) Big data and development: “The second half of the chess board”. Available at: <https://blogs.worldbank.org/african/big-data-and-development-the-second-half-of-the-chess-board> (accessed 18 August 2014).

Finel BI, Kristin ML (1999) The surprising logic of transparency. *International Studies Quarterly* 43: 315–339.

Florida R (2012) The world cities that tweet the most. Available at: <http://www.theatlanticcities.com/arts-and-lifestyle/2012/08/world-cities-tweet-most/2944/> (accessed 18 August 2014).

foxnews.com (2014) American farmers confront Big Data revolution. Available at: <http://www.foxnews.com/us/2014/03/29/american-farmers-confront-big-data-revolution/> (accessed 18 August 2014).

G-Analytix (2013) Credit information sharing: Unlocking access to affordable credit: Alternative data to drive financial inclusion. In: *The 2nd regional credit information sharing conference*, 24 September, Nairobi, Kenya.

gartner.com (2013) Big data. Available at: <http://www.gartner.com/it-glossary/big-data/> (accessed 18 August 2014).

Gatignon H, Robertson TS (1985) A propositional inventory for new diffusion research. *Journal of Consumer Research* 11: 849–867.

Giller KE, Murwira MS, Dhliwayo DKC, (2011) Communicating complexity: Integrated assessment of trade-offs concerning soil fertility management within African farming systems to support innovation and development. *Agricultural Systems* 104: 191–203.

Gillespie T (2011) Our misplaced faith in Twitter Trends, October 19. Available at: [http://www.salon.com/2011/10/19/our\\_misplaced\\_faith\\_in\\_twitter\\_trends/](http://www.salon.com/2011/10/19/our_misplaced_faith_in_twitter_trends/) (accessed 18 August 2014).

Global Pulse (2012) Big data for development: Challenges & opportunities. Available at: <http://unglobalpulse.org/> (accessed 18 August 2014).

Global Pulse (2014) Advocacy monitoring through social data: Women's and children's health. Available at: <http://www.unglobalpulse.org/EWEC-social-data-analysis> (accessed 18 August 2014).

Glover DJ (1984) Contract farming and smallholder outgrower schemes in less-developed countries. *World Development* 12(11/12): 1143–1157.

Glover DJ (1987) Increasing the benefits to smallholders from contract farming: Problems for farmers' organizations and policy makers. *World Development* 15(4): 441–448.

Heald D (2006) Varieties of transparency. In: Hood C, Heald D (eds) *Transparency: The Key to Better Governance?*, Oxford: Oxford University Press, pp. 23–45.

Hilbert M (2014) Technological information inequality as an incessantly moving target: The redistribution of information and communication capacities between 1986 and 2010. *Journal of the Association for Information Science and Technology* 65(4): 821–835.

Hoeffler H (2006) Promoting the Kenyan potato value chain: Can contract farming help build trust and reduce transaction risks? In: *Proceedings of the 99th EAAE (European Association of Agricultural Economists) seminar, trust and risks in business networks*, Bonn, Germany, 8–10 February. Available at: <http://ageconsearch.umn.edu/bitstream/7726/1/sp06ho02.pdf> (accessed 18 August 2014).

hsph.harvard.edu (2014) Big data for social good. Available at: <http://www.hsph.harvard.edu/ess/bigdata.html> (accessed 18 August 2014).

Iammarino S, Michie J (1998) The scope of technological globalisation. *International Journal of the Economics of Business* 5(3): 335–353.

irinnews.org (2009) Tanzania: Anti-malarial drugs too expensive – minister. Available at: <http://www.irinnews.org/report/82984/tanzania-anti-malarial-drugs-too-expensive-minister> (accessed 18 August 2014).

Kaplan E (2013) Searching for digital signals in Kenya. *UN Global Pulse*, 3 September.

Kirkpatrick R (2013) Use Big Data wisely and it can provide big insights. Available at: <http://www.scidev.net/global/data/opinion/use-big-data-wisely-and-it-can-provide-big-insights.html> (accessed 18 August 2014).

Korolov M (2013) 15 most powerful Big Data companies. Available at: <http://www.networkworld.com/slideshow/114134/15-most-powerful-big-data-companies.html#slide1> (accessed 18 August 2014).

Kroes N (2013) Speech – Big data for Europe, European Commission – SPEECH/13/893. Available at: [http://europa.eu/rapid/press-release\\_SPEECH-13-893\\_en.htm](http://europa.eu/rapid/press-release_SPEECH-13-893_en.htm) (accessed 18 August 2014).

- Kshetri N (2001) Determinants of the locus of global e-commerce. *Electronic Markets* 11(4): 250–257.
- Kshetri N (2011) Cloud computing in the global south: drivers, effects and policy measures. *Third World Quarterly* 32(6): 995–1012.
- Kshetri N (2012) Cloud computing in India. *IEEE IT Professional* 14(5): 5–8.
- Kshetri N (2013) Cloud computing in sub-Saharan Africa. *IEEE IT Professional* 15(6): 64–67.
- Kshetri N (2014) Big data's impact on privacy, security and consumer welfare, *Telecommunications Policy* 38: 1134–1145.
- Kshetri N, Dholakia N (2002) Determinants of the global diffusion of B2B e-commerce. *Electronic Markets* 12(2): 120–129.
- Laney D (2001) 3D data management: Controlling data volume, velocity *and* variety. Available at: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (accessed 18 August 2014).
- Letouzé E (2012) Big data for development: What may determine success or failure? *OECD Technology Foresight*, 22 October.
- Letouzé E (2013) Could Big Data provide alternative measures of poverty and welfare? Available at: <http://www.oecdbetterlifeindex.org/es/blog-es/could-big-data-provide-alternative-measures-of-poverty-and-welfare.htm>(accessed 18 August 2014).
- linnettaylor (2013) Surveil the rich, observe the poor: Big data at the internet governance forum 2013. Available at: <http://linnettaylor.wordpress.com/2013/10/25/surveil-the-rich-observe-the-poor-big-data-at-the-internet-governance-forum-2013/> (accessed 18 August 2014).
- Lohr S (2013) Searching Big Data for 'digital smoke signals'. Available at: [http://www.nytimes.com/2013/08/08/technology/development-groups-tap-big-data-to-direct-humanitarian-aid.html?\\_r=0](http://www.nytimes.com/2013/08/08/technology/development-groups-tap-big-data-to-direct-humanitarian-aid.html?_r=0) (accessed 18 August 2014).
- Madsen AK (2013) *Web-visions: Repurposing digital traces to organize social attention*. The Doctoral School of Organisation and Management Studies, Copenhagen Business School. Available at: [http://openarchive.cbs.dk/bitstream/handle/10398/8746/Anders\\_Koed\\_Madsen.pdf?sequence=1](http://openarchive.cbs.dk/bitstream/handle/10398/8746/Anders_Koed_Madsen.pdf?sequence=1) (accessed 23 September 2014).
- Mayer-Schönberger V, Cukier K (2013) *Big Data: A Revolution that will Transform How We Live, Work and Think*, Boston: Houghton Mifflin Harcourt.

McAfee A and Brynjolfsson E (2012) Big data: The management revolution. Available at: <http://hbr.org/2012/10/big-data-the-management-revolution/ar/1> (accessed 18 August 2014).

mhealthinfo.org (2011) Malaria surveillance & mapping. Available at: <http://www.mhealthinfo.org/project/malaria-surveillance-mapping> (accessed 18 August 2014).

Mims C (2012) Facebook's plan to find its next billion users: Convince them the internet and Facebook are the same. Available at: <http://qz.com/5180/facebooks-plan-to-find-its-next-billion-users-convince-them-the-internet-and-facebook-are-the-same/> (accessed 18 August 2014).

Newton A (2012) Big data for development: Beyond transparency. Available at: <https://blogs.worldbank.org/psd/big-data-for-development-beyond-transparency> (accessed 18 August 2014).

O'Neil, Hugh M, Poudier RW, (1998) Patterns in the diffusion of strategies across organizations: Insights from innovation diffusion literature. *The Academy of Management Review* 23(1): 98–114.

O'Neill O (2006) Transparency and the ethics of communication. In: Hood C, Heald D (eds) *Transparency: The Key to Better Governance*, Oxford: Oxford University Press, pp. 75–90.

Oluoch-Kosura W (2010) Institutional innovations for smallholder farmers' competitiveness in Africa. *African Journal of Agricultural and Resource Economics* 5(1): 227–242.

Palmer N (2012) ICT for data collection and monitoring and evaluation. *Eagriculture*, June.

Patel P (2013) Feeding the world with Big Data. Available at: <http://spectrum.ieee.org/computing/networks/feeding-the-world-with-big-data> (accessed 18 August 2014).

Pawelke A and Tatevossian AR (2013) Data philanthropy: Where are we now? Available at: <http://www.unglobalpulse.org/data-philanthropy-where-are-we-now> (accessed 18 August 2014).

pewglobal.org (2014) Emerging nations embrace internet, mobile technology. Available at: <http://www.pewglobal.org/2014/02/13/emerging-nations-embrace-internet-mobile-technology/> (accessed 18 August 2014).

Richter F (2013) Twitter's top 5 markets account for 50% of active users. Available at: <http://www.statista.com/topics/737/twitter/chart/1642/regional-breakdown-of-twitter-users/> (accessed 18 August 2014).

- Robinson L (2009) A taste for the necessary: A Bourdieuan approach to digital inequality. *Information, Communication & Society* 12(4): 488–507.
- Rogers EM (1995) *Diffusion of Innovations*, 4th ed. New York, NY: Free Press.
- Rogers EM (2003) *Diffusion of Innovations*, 5th ed. New York, NY: Free Press.
- sas.com (2013) Big Data: What it is and why it matters. Available at: [https://www.sas.com/en\\_us/insights/big-data/what-is-big-data.html](https://www.sas.com/en_us/insights/big-data/what-is-big-data.html) (accessed 18 August 2014).
- Seppala TJ (2014) Monsanto pushes Big Data-driven planting but farmers are skeptical. Available at: <http://global.ofweek.com/news/Monsanto-pushes-Big-Data-driven-planting-but-farmers-are-skeptical-8587> (accessed 18 August 2014).
- Sirgy MJ, Su C (2002) The ethics of consumer sovereignty in an age of high tech. *Journal of Business Ethics* 28: 1–14.
- Soewondo P, Ferrario A, Levenus Tahapary D (2013) Challenges in diabetes management in Indonesia: A literature review. *Globalization & Health* 9(1): 1–34.
- Srikanth RP (2013) The Big Data opportunity for Indian IT service providers. *InformationWeek*. Available at: <http://www.informationweek.in/informationweek/news-analysis/277187/opportunity-indian-service-provider>(accessed 18 August 2014).
- Strategy (2012) *Digitization in Emerging Economies: Unleashing Opportunities at the Bottom of the Pyramid*, 1 October. Available at: <http://www.strategyand.pwc.com/global/home/what-we-think/reports-white-papers/article-display/digitization-emerging-economies-unleashing-opportunities> (accessed 18 August 2014).
- The World Bank (2010) Getting the numbers right: Making statistical systems a real plus for results. *The World Bank: IBRD Results*, March.
- The World Bank Group (2014) Country and lending groups. Available at: [http://data.worldbank.org/about/country-and-lending-groups#Low\\_income](http://data.worldbank.org/about/country-and-lending-groups#Low_income) (accessed 18 August 2014).
- Twomlow SJ, Mugabe FT, Mwale M, et al. (2008) Building adaptive capacity to cope with increasing vulnerability due to climatic change in Africa – A new approach. *Physics and Chemistry of Earth* 33: 780–787.
- UN Global Pulse (2013a) Twitter and perceptions of crisis-related stress. *UN Global Pulse*. Available at: <http://www.unglobalpulse.org/projects/twitter-and-perceptions-crisis-related-stress> (accessed 18 August 2014).

UN Global Pulse (2013b) Landscaping study: Digital signals & access to finance in Kenya. *UN Global Pulse*. Available at: <http://www.unglobalpulse.org/Kenyan-access-finance> (accessed 18 August 2014).

UNCTAD (2009) World investment report 2009: Transnational Corporations, Agricultural Production and Development. Geneva: United Nations Conference on Trade and Development.

UNDP (2001) Driving information and communications technology for development: A UNDP agenda for action 2000-2001, United Nations Development Programme. Available at: <http://www.sdn.undp.org/it4dev/ffICTe.pdf> (accessed 18 August 2014).

UNDP (2011) Africa adaptation programme: Capacity building experiences improving access. In: *Understanding and application of climate data and information discussion paper*. Series 2, June. New York: United Nations Development Programme, Environment & Energy Group.

UNFCCC (2007) Climate change: Impacts, vulnerabilities and adaptation in developing countries. Available at: [unfccc.int/resource/docs/publications/impacts.pdf](http://unfccc.int/resource/docs/publications/impacts.pdf) (accessed 18 August 2014).

UNICEF (2013) Tracking anti-vaccination sentiment in eastern European social media networks. New York: UNICEF.

Vidal J (2011) GM crops promote superweeds, food insecurity and pesticides, say NGOs. Available at: [Guardian.co.uk](http://Guardian.co.uk) (accessed 18 August 2014).

Walisadeer AI, Wikramanayake GN and Ginige A (2013) An ontological approach to meet information needs of farmers in Sri Lanka. In: B Murgante et al. (eds). *ICCSA, Part I, LNCS 7971*: 228–240.

Webster J, Watson RT (2002) Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly* 26(2): R13–R23.

WEF (2012) *Big Data, Big Impact: New Possibilities for International Development*. Geneva, Switzerland: World Economic Forum.

Whetten DA (1989) What constitutes a theoretical contribution? *Academy of Management Review* 14(4): 490–495.

Yan Z (2012) Personal data crimes set to be defined. Available at: [www.chinadaily.com.cn/china/2012-07/04/content\\_15546503.htm](http://www.chinadaily.com.cn/china/2012-07/04/content_15546503.htm) (accessed 18 August 2014).

Yeoman K (2013) Mobile phones unleash farmers in Uganda. Available at: <http://www.csmonitor.com/World/Making-a-difference/Change-Agent/2013/0509/Mobile-phones-unleash-farmers-in-Uganda> (accessed 18 August 2014).