CrossMark

# Big Data and Social Indicators: Actual Trends and New Perspectives

Enrico di Bella[1] · Lucia Leporatti[2] · Filomena Maggino[3]

**Abstract** Big Data are a top subject in international research articles and a vast debate is taking place on their actual capability of being used to complement or even substitute official statistics surveys and social indicators in particular. In this paper we analyse the metadata of the Scopus database of academic articles on Big Data and we show that most of the existing and intensively growing literature is focused on software and computational issues whilst articles that are specifically focused on statistical issues and on the procedures to build social indicators from Big Data are a much smaller share of this vast production. Nevertheless the works that focus on these topics show promising results because in developed countries Big Data seem to be a good information base to create reliable proxies of social indicators, whereas in developing countries their use (for instance using satellite images) may be a viable alternative to traditional surveys. However, Big Data based social indicators deeply suffer of a number of open issues that affect their actual use: they do not correspond to any sampling scheme and they are often representative of particular segments of the population; they generally are private process-produced data whose access by national statistical offices is rarely possible although the intrinsic value of the information contained in Big Data has a social importance that should be shared with the whole community; Big Data lack the socio-economic background on which social indicators have been founded and their help to policy makers in their decision process is a fully open point. Therefore Big Data may be a big opportunity for the definition of traditional or new social

✉ Enrico di Bella
  edibella@unige.it

  Filomena Maggino
  filomena.maggino@unifi.it

[1] Department of Economics and Business Studies, University of Genoa, Via Vivaldi 5, 16123 Genoa, Italy

[2] Department of Political Sciences, University of Genoa, Via Vivaldi 5, 16123 Genoa, Italy

[3] Department of Statistics, Computer Sciences and Application "G. Parenti", University of Florence, Florence, Italy

🖄 Springer

indicators but their statistical reliability should be further investigated and their availability and use should be internationally coordinated.

# 1 Introduction

Fifty years passed since the "Social Indicators Movement" of the 1960s–1970s started to describe social phenomena. The work by Land and Michalos (2015) reviews how the international academic and political debate brought to the introduction of concepts and indicators that are nowadays widespread at any level, but that were really a novelty when they were introduced, such as the concepts of quality of life, progress, sustainable development and well-being and their corresponding measures. In their work the authors conclude that the Social Indicators Movement realized most of its agenda, but there is still much to do, not only to improve exiting social indicators, but also to introduce new ones designed on the new societal forms that are far different from the one in which the "Social Indicators Movement" was born.

This paper takes inspiration from this last conclusion, in particular for what it concerns the digitalized-computerized evolution of our contemporary world. According to IBM, in 2012, 2.5 billion gigabytes of data were generated every day and roughly 90% of the data in the world today has been created in the last two years alone. About the 75% of these data is unstructured, coming from sources such as text, voice and video. This unstructured mass data is often called "Big Data" in order to denote complex sets of data. It is generally accepted that Big Data can be explained according to three versus (Laney 2001): (high) velocity (speed of data generation and use), (high) variety (range of data types and sources) and (high) volume (amount of data). Although in social sciences Big Data can be substantially considered to be process-produced data gathered tracking peoples' activities in different real or virtual environments, they can be distinguished from "large dataset" by the fact that the former have a high level of complexity and multidimensionality whereas the latter are merely datasets with many records. Complexity within Big Data came under the focus of academics and practitioners because many of them saw that it contains complex and possibly latent correlations among variables that may be difficult to observe using traditional surveys. So, it is not surprising that in the last years "Big Data" became a trending topic in international literature as an innovative source of information for complex phenomena. One of the most famous articles on the use of Big Data to model social phenomena was published by Ginsberg et al. (2009) where influenza epidemics were detected with high accuracy using the massive number of Google search queries gathered and analysed to reveal if there was the presence of flu-like illness in a population. Although other articles (e.g. Lazer et al. 2014) demonstrated that the model implemented in Ginsberg et al. (2009) could be much less accurate than what was initially stated, the widespread of the paper and the new opportunities that it opened for the analysis of complex phenomena gave rise to a wide debate (e.g.: Boyd and Crawford 2012) on the real usefulness of Big Data analysis because it became evident that it could be possible to use Big Data, available in real time, to study phenomena without direct observation or statistical surveys.

Big Data have a number of relevant pros that make them very interesting also for the definition of new social indicators or to integrate/substitute existing ones. First of all, being process-produced data, their cost is theoretically zero, being gathered not for statistical but for other purposes. Secondly, being generally electronic data, they are "hot data" that may be used in real-time to monitor phenomena of interest without the lag between observation and analysis that is typical problem in traditional statistical surveys. A good example to better understand this potential is given by "nowcasting" (e.g. Giannone et al. 2008), a set of techniques that combine data from multiple sources to provide relevant (mainly economic and monetary) statistics using all the possible information given at a certain time and updating these estimates as long as new information becomes available. The main goal of nowcasting is to provide estimates of these statistics much earlier than the final values are officially released, so that decision making can be made "real time" and not *ex-post*.

Anyhow, despite the intensively growing literature on Big Data, only a small share of academic articles directly deals with the complexity of social phenomena and in particular with the measurement of the quality of life and its complexity (*qolexity*). Although any effort to provide a complete review of the academic production of articles relating Big Data and Social Indicators would result to be vain because of the quickly growing number of articles, in this paper we try to size the dimension of the academic production of social indicators which made use of Big Data pointing out which are the contexts in which this production is more abundant. In particular, we made use of the Scopus database of peer-reviewed literature by Elsevier to find actual trends in academic literature.

## 2 Analysis of Metadata on Big Data Articles

The increasing interest of the scientific community in Big Data analysis is easily perceivable; however, Big Data applications are not evenly distributed across different research areas, with engineering and computer science being the subjects more involved by the "Big Data" revolution while social sciences applications still represent a limited share of the "Big Data" literature (Taylor et al. 2014; González-Bailón 2013). Taking a picture of the researches based on "Big Data" in general and for analysing social phenomena is a particularly hard task due to the great dynamism of the research on the topic which risks to seriously compromise the possibility of providing an up to date literature review on the topic. As a consequence, our approach to characterise Big Data literature makes use of the Scopus database combing different selected key-words useful to detect the trends and peculiarities of the Big Data research.

On the 7th September 2016 the number of articles reporting in the "Title, Abstract and Keywords" the term "Big Data" account to 21,691 (Table 1); 13,802 of these articles (63%) were published in the last two years (2015 and 2016). Among these articles, the ones classified in the field of "Social Sciences" represent just the 10% (2192 articles in total; Table 2).

A deeper analysis on the most popular keywords used in "Big Data" articles highlights the relative small relevance absorbed by researches in social sciences. Among the 21,691 selected articles, the 10 most common keywords turn out to be mostly connected to engineering and computer science. Obviously, the most common keyword is "Big Data" (72% of the articles); other common terms include "Data Mining" (11%), "Data Handling" (9%), "Algorithms" (9%), "Cloud Computing" (8%), "Digital Storage" (8%), "Information Management" (7%), "Artificial Intelligence" (7%), "Internet" (6%),

**Table 1** Number of articles published in Scopus database containing the term "Big Data" in the Title, Abstract or Keywords in the period 1957–2017 and selection of articles belonging to "Social Sciences" subject area

| Search fields | Year | | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| | Before 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | |
| "Big Data" | 122 | 30 | 88 | 648 | 2381 | 4618 | 8103 | 5699 | 2 | 21,691 |
| "Big Data" within Social Sciences subject area journals | 2 | 5 | 7 | 43 | 169 | 378 | 559 | 1028 | 1 | 2192 |

**Table 2** Number of articles published in Scopus database containing the term "Big Data" in the Title, Abstract or Keywords split by the main subject areas of interest

| Scopus subject area | Number of articles | % Total |
|---|---|---|
| Computer Science | 15,687 | 72 |
| Engineering | 4924 | 23 |
| Mathematics | 3452 | 16 |
| Social Sciences | 2192 | 10 |
| Medicine | 1848 | 9 |
| Decision Sciences | 1392 | 6 |
| Business, Management and Accounting | 1126 | 5 |
| Biochemistry, Genetics and Molecular Biology | 572 | 3 |
| Materials Science | 550 | 3 |
| Physics and Astronomy | 514 | 2 |
| Earth and Planetary Sciences | 399 | 2 |
| Energy | 332 | 2 |

The "% Total" values don't add up to 100% because articles may be classified in more than one subject area

"Distributed Computer Systems" (6%). Keywords more pertinent with social sciences involve only a minority of this vast literature: "Human" or "Humans" (1513 articles, 7% of the selected articles) and "Education" (373, 2%). "Commerce", "Healthcare", "Costs", "Energy efficiency", "Energy Utilization", "Transportation", "Health" and "Economics" involve each roughly the 1% of the selected articles.

Another search in the Scopus database for what it concerns the data types and sources of Big Data points out that several articles use data derived from social media (Table 3). Social media chatter may indeed become precious sources of public opinions and social cohesion trends; Twitter seems to be a particularly relevant social network in this context. Other common data sources are sensors, online searches, data derived from mobile phones and from satellites.

**Table 3** Number of articles using different data sources in the Scopus database

| Data source | Articles | Social sciences articles | Scopus query in the "Title, Abstract and Keywords" field |
|---|---|---|---|
| Satellite Images | 195 | 28 | "satellite" OR "satelliteimage" AND "bigdata" |
| Mobile phone | 197 | 33 | "mobilephone" OR "cellphone" AND "bigdata" |
| Online search | 65 | 3 | "websearch" OR "onlinesearch" AND "bigdata" |
| Social Media | 2037 | 301 | "socialmedia" OR "social network" OR "twitter" OR "facebook" OR "instagram" AND "bigdata" |
| Twitter | 512 | 87 | "Twitter" AND "bigdata" |
| Facebook | 251 | 46 | "Facebook" AND "bigdata" |
| Online search extended | 373 | 57 | "onlinesearch" OR "websearch" OR "blogs" OR "news" AND "bigdata" |
| Sensors | 1445 | 88 | "sensors" AND "bigdata" |
| Text Mining | 222 | 22 | "textmining" AND "bigdata" |

## 2.1 Big Data and SDGs

These preliminary considerations paved the way to a deeper analysis of the most common Big Data use in social sciences, in particular relating the sustainable development goals (SDGs). The increasing attention towards these two apparently distinct hot topics (i.e. "Big Data" and "SDGs") may indeed become an opportunity for exploiting the potentiality of "Big Data" analysis for supporting the progress in SDGs (United Nations 2015). We proceeded in two main steps:

1. we identified a certain number of keywords able to capture the main issues connected to each SDG[1] (Table 4);
2. using the Scopus database, we selected, for each SDG, those articles satisfying two conditions:

   a. containing the term "Big Data" in "Title, Keywords or Abstract";
   b. containing at least one of the selected keywords identified for each SDG in "Title, Keywords or Abstract".

Of course, this procedure has some limitations as it does not allow us to detect those articles not mentioning the term "Big Data" but it can anyway provide a good approximation of the relevance of different topics in Big Data literature.

As showed by Table 4, some SDGs seem to be more involved than others in Big Data applications. Industry, Innovation and Infrastructure (SDG9), Health and Well-being (SDG3), topics connected to Sustainable and Smart Cities (SDG11) and Quality Education (SDG4) are the most relevant topics in Big Data literature. They also represent booming fields of research, as more than the 60% of the existing articles have been published in the last two years. "Quality Education" (SDG 4), "Peace, Justice and Strong Institutions"

---

[1] We ignore the last SDG (i.e. SDG 17 = "Partnership for the goals") which is mainly focused on the cooperation in achieving the others 16 SDGs.

**Table 4** Number of articles published in the Scopus database containing the term "Big Data" in the "Title, Abstract or Keywords" by SDGs on the basis on the additional keywords

| | SDG | Number of articles | % Articles in social sciences | % Articles 2015–2017 | Scopus query in the "Title, Abstract and Keywords" additional to "Big Data" |
|---|---|---|---|---|---|
| 1 | No Poverty | 14 | 14 | 57 | "Poverty" OR "socio-economic level" |
| 2 | Zero Hunger | 132 | 11 | 67 | "Food security" OR "hunger" OR "malnutrition" OR "agriculture" |
| 3 | Good Health and Well-Being | 1799 | 8 | 64 | "Health"OR"influenza"OR"epidemics"OR"vaccine"OR"well-being"OR"wellbeing" |
| 4 | Quality Education | 839 | 26 | 73 | "Education" OR "literacy" |
| 5 | Gender Equality | 110 | 11 | 64 | "Gender gap" OR "gender disparities" OR "gender" OR "women" |
| 6 | Clean Water and Sanitation | 29 | 21 | 62 | "Clean water" OR "water quality" OR "water pollution" OR "drinking water" |
| 7 | Affordable and clean energy | 346 | 7 | 73 | "energy consumption" OR "renewable" OR "clean energy" OR "sustainable energy" |
| 8 | Decent work and economic growth | 183 | 19 | 71 | "Working conditions" OR "unemployment" OR "GDP" OR "economic growth" OR "employment" OR "tourism" |
| 9 | Industry, Innovation and Infrastructure | 3308 | 10 | 59 | "Innovation" OR "industry" OR "infrastructure" |
| 10 | Reduced inequalities | 254 | 15 | 64 | "inequality" OR "migration" OR "economic inequality" |
| 11 | Sustainable cities and communities | 1428 | 14 | 61 | "Sustainable cities" OR "smart cities" OR "communities" |
| 12 | Responsible Consumption and production | 4 | 50 | 75 | "Responsible Consumption" OR "Responsible production" OR "social cost" OR "green production" |
| 13 | Climate Action | 169 | 12 | 72 | "Climate Action" OR "emissions" OR "climate change" |
| 14 | Life below water | 332 | 13 | 69 | "Waters" OR "sea" OR "ocean" |
| 15 | Life on land | 137 | 20 | 72 | "Land" OR "terrestrial ecosystem" |
| 16 | Peace, Justice and strong institution | 423 | 26 | 61 | "Peace" OR "justice" OR "institutions" |

(SDG 16), "Clean Water and Sanitation" (SDG 6), "Life on land" (SDG 15) "Decent work and economic growth" (SDG8) represent those research areas in which articles classified as "Social Sciences" represent a relevant percentage on the total (from 26% to 19%). Generally, Big Data tend to be used to monitor those phenomena that are correlated to Big Data production of some kind or to provide alternative measures of sectors for which other traditional statistics are already available (United Nations 2015). A deeper analysis of "Big Data" based articles focused on SDGs and social world put well in evidence the great variety of applications present in the current literature which tends to be different in the source of data used and in the geographic area of application.

It is worth noting that Big Data applications have generally different aims in developed rather than developing countries. On one side, in developed countries Big Data seem indeed to be a good information base to create reliable proxies of social indicators and to complete official statistics analysis. Several researches in developed country are based on large volumes of social media data used to check the presence of correlations with many socio-economic characteristics and to predict a vast range of social trends. Among the others, Yazdani and Manovich (2015) used Tweeted images during one year in 20 USA cities to predict a number of socio-economic characteristics finding out correlations with self-reported social well-being, median housing prices, income, and education levels. The use of social media data has also been found useful in predicting emerging political trends in Germany through tweets (Rill et al. 2014) and detecting tourism preferences (Chang and Chu 2013). Also GPS data are catching increasing attention. Marchetti et al. (2015) used the GPS data from private vehicles finding out a good correlation between mobility and poverty level in an Italian region while Cord et al. (2015) used GPS data to explore the preferences of people in terms of recreational places to visit. Several applications have also been proposed in the energy sectors (Zhou and Yang 2016).

On the other side, in developing countries the use of Big Data may be a viable alternative to traditional surveys and to get timely and localized information (for instance using satellite images or mobile phone records) and to obtain estimations of local impacts of shocks (e.g. food crises, natural disasters) not immediately trackable(e.g.: United Nations Global Pulse 2014). The literature using Big Data in this context is focused on poverty and socio-economic status (Mao et al. 2015; Blumenstock et al. 2015; Smith-Clarke et al. 2014), crops production, drought and its impact on food security (Dutta et al. 2014). Mobile phone records seem the most applied sources of information across the developing world, as they are used daily to transfer money, buy and sell goods and communicate (United Nations Global Pulse 2013). Among the others, Mao et al. (2015) and Blumenstock et al. (2015) used mobile call log data to predict the income level of different regions respectively of Côte d'Ivoire and Rwanda showing their potentiality in monitoring remote and inaccessible regions. Elvidge et al. (2009) used instead night-time satellite images of a global map to build up a poverty index computed as the ratio between the population count and the brightness of satellite observed lighting (DMSP night-time lights). In addition, cell phone records represent useful means to monitor and predict the evolution of epidemics and flu in developing countries. Wesolowski et al. (2012) used mobile phone call data records to build up a malaria risk map, estimating how the malaria parasite moves across Kenya. In addition, the relevant diffusion of internet and social media also in developing countries is leading to an increase in the real-time information streams in social media data usable for social researches. Among the others, Yuan et al. (2013) used online search query data from Baidu to model and monitor influenza spread in China. Online searches (measured through Google trends indexes) have also been used to predict social unrest in the South of America (Manrique et al. 2013).

# 3 Discussion and Conclusions

A straight analysis of the metadata of the Scopus database of academic articles on Big Data outlines that most of the existing literature is focused on software and computational issues whilst articles that are specifically focused on statistical issues and on the procedures to build social indicators from Big Data are a much smaller share of this vast production. Nevertheless, the works that focus on these topics show promising results because in developed countries Big Data seem to be a good information base to create reliable proxies of social indicators, whereas in developing countries their use may be a viable alternative to traditional surveys. However, traditional statistical surveys and Big Data have relevant fundamental differences that it is worth recalling.

During the years, National Statistical Institutes defined a number of standards for data quality (e.g.: professional independence, impartiality and objectivity, accuracy and reliability, etc.) to provide the stake-holders with independent high quality information on the economy and society. The European Statistics Code of Practice (Eurostat 2012) is a relevant outcome of this process. It is based on *15 Principles covering the institutional environment, the statistical production processes and the output of statistics. Statistical authorities, comprising the Commission (Eurostat), National Statistical Institutes and other national authorities responsible for the development, production and dissemination of European Statistics, together with governments, ministries and the European Council, committed themselves to adhere to the Code.* These high standards of the whole statistical process are not guaranteed by most of the institutions that are actually collecting and using Big Data and the statistical quality of Big Data itself is still a mostly unexplored topic in international literature.

It is also worth noting that one of the key elements that let the "Social Indicators Movement" reach many results of its original agenda is that it worked hardly on constructs according to which surveys could be designed and social indicators defined. Understanding the cause-effects mechanisms in social frameworks (or, for instance, elasticities in economic modelling) is fundamental to let decision makers take the right choices having all the available information. A number of works show that Big Data could approximate very well some social indicators but, as it was already stated ahead, Big Data are process-produced data, and an overlap of the process that produces these information and the construct that theoretically funds the indicator that Big Data well approximate seems to be the exception rather than the rule. Therefore, there is the possibility that using Big Data we have cheaper and faster estimates of a relevant social indicator (e.g. unemployment) but we may not have all the complementary information that may help decision makers to address occupational issues. At the same time, Big Data are available in a sort of continuum and they may provide timely information. Proxying for unobservable variables by using, for instance, internet searches (e.g. using Google Trends[2]) may provide estimates of synthetic indicators in advance compared to official statistics (in the logic of nowcasting) but the availability of a measure of coverage of the indicator construct by Big Data covariates would add much to actual discussion to interpret the quality of the information provided by Big Data itself and to extend the indicator construct itself.

All these topics are rarely discussed in articles dealing with Big Data. The most common criticism is that being process-produced data, Big Data can't be considered to be representative of a whole population but a strategy to address such undercover issues is

---

[2] https://www.google.com/trends/.

missing and should be designed. However, we think that there are even more relevant issues that should be further discussed. First of all, Big Data are private process-produced data whose access by national statistical offices is rarely possible, while the intrinsic value of the information contained in them has a social importance that should be shared with the whole community. Secondly, although Big Data could introduce real-time indicators (a new idea of timely statistics), they totally lack a sort of quality standards. Moreover, although Big Data have really complex structures, they still represent partial observations of the social framework and so they may provide only partial information on a specific phenomenon of interest. The cause-effect mechanism cited above may be more clearly measured by a smaller (but more complete) traditional survey rather than a very large, but incomplete Big Data set. If some Big Data are supposed to measure a social phenomenon, the overlap between them and the phenomenon itself should be measured and monitored overtime. Finally, the Big Data revolution raises a set of ethical issues related to privacy, identity, confidentiality and transparency that are still open (e.g. Richards and Kings 2014; King 2011) and new rules on the flows of personal information in our society are needed: on one hand we have that most of Big Data are collected and analysed without a true knowledge or consent by the people on which the information is collected and, on the other, official statistics office have to deal with strict procedures and norms on data anonymization (particularly for health data) that affect data use and analysis. For instance, the effects on Big Data collection and analysis of the recently approved EU General Data Protection Regulation (Regulation (EU) 2016/679) on "the protection of natural persons with regard to the processing of personal data and on the free movement of such data" are yet to be discussed and evaluated.

Concluding, we think that the effects of the Big Data revolution on social indicators is still far from being visible because statistically rigorous works that really discuss on how Big Data may be used to improve statistical analysis of social phenomena are still few (e.g.: Marchetti et al. 2015), whereas the largest share of the available works are just something more than toy examples that explore the capability of Big Data to approximate social indicators, not telling anything more (if not even less) than the indicators they are trying to emulate. The minimal result that we may expect from using Big Data to build social indicators is a reduction of the frequency of official statistics surveys with Big Data based estimates for the periods in which these estimates are not carried out. But in order to have reliable Big Data social indicators, which add something new to the theories of Social Sciences, the international statistical community should start discussing these three preliminary key points: i) how it is possible to have a wider diffusion of Big Data (in particular for official statistics offices); ii) how Big Data quality measures can be defined; iii) how the normative on privacy and transparency should change in the Big Data era.

# References

Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science, 350*(6264), 1073–1076.

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

Chang C.C. & Chu K.H. (2013). A recommender system combining social networks for tourist attractions. In *5th International IEEE Conference on Computational Intelligence, Communication Systems and Networks (CICSyN)* (pp. 42–47).

Cord, A. F., Roeßiger, F., & Schwarz, N. (2015). Geocaching data as an indicator for recreational ecosystem services in urban areas: Exploring spatial gradients, preferences and motivations. *Landscape and Urban Planning, 144*, 151–162.

Dutta, R., Morshed, A., Aryal, J., D'este, C., & Das, A. (2014). Development of an intelligent environmental knowledge system for sustainable agricultural decision support. *Environmental Modelling and Software, 52*, 264–272.

Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B., et al. (2009). A global poverty map derived from satellite data. *Computers & Geosciences, 35*(8), 1652–1660.

Eurostat. (2012). *European statistics code of practice—Revised edition 2011*. Product Code: KS-32-11-955, release date: Jan 01, 2012. doi:10.2785/18474.

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics, 55*(4), 665–676.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature, 457*(7232), 1012–1014.

González-Bailón, S. (2013). Social science in the era of big data. *Policy and Internet, 5*(2), 147–160.

King, G. (2011). Ensuring the data-rich future of the social sciences. *Science, 331*(6018), 719–721.

Land K.C. & Michalos A.C. (2015). Fifty years after the social indicators movement: has the promise been fulfilled? An Assessment and an Agenda for the Future, forthcoming in *Social Indicators Research*. Pre-print available at: http://www.miqols.org/howb/wp-content/uploads/2016/06/LandAndMichalos-50YearsPaper-Draft7.pdf.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note, 6*, 70.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of google flu: Traps in big data analysis. *Science, 343*(6176), 1203–1205.

Manrique P., Qi H., Morgenstern A., Velasquez N., Lu T. C. & Johnson N. (2013). Context matters: improving the uses of big data for forecasting civil unrest: emerging phenomena and big data. In *IEEE International Conference on Intelligence and Security Informatics (ISI)* (pp. 169–172).

Mao, H., Shuai, X., Ahn, Y. Y., & Bollen, J. (2015). Quantifying socio-economic indicators in developing countries from mobile phone communication data: applications to Côte d'Ivoire. *EPJ Data Science, 4*(1), 1–16.

Marchetti, S., Giusti, C., Pratesi, M., Salvati, N., Giannotti, F., Pedreschi, D., et al. (2015). Small area model-based estimators using big data sources. *Journal of Official Statistics, 31*(2), 263–281.

Richards, N. M., & King, J. H. (2014). Big data ethics. *Wake Forest Law Review, 49*, 393–432.

Rill, S., Reinel, D., Scheidt, J., & Zicari, R. V. (2014). Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis. *Knowledge-Based Systems, 69*, 24–33.

Smith-Clarke C., Mashhadi A. & Capra L. (2014,). Poverty on the cheap: Estimating poverty maps using aggregated mobile communication networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 511–520).

Taylor, L., Schroeder, R., & Meyer, E. (2014). Emerging practices and perspectives on Big Data analysis in economics: Bigger and better or more of the same? *Big Data and Society, 1*(2), 1–10.

United Nations. (2015). *Global sustainable development report*. Available at: https://sustainabledevelopment.un.org/globalsdreport. Accessed September 01, 2016.

United Nations Global Pulse. (2013). *Big Data for Development: A primer*. http://www.unglobalpulse.org/sites/default/files/Primer%202013_FINAL%20FOR%20PRINT.pdf. Accessed September 01, 2016.

United Nations Global Pulse. (2014). *Mining indonesian tweets to understand food price crises*. Jakarta. http://www.unglobalpulse.org/projects/social-media-social-protection-indonesia. Accessed September 01, 2016.

Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W., et al. (2012). Quantifying the impact of human mobility on malaria. *Science, 338*(6104), 267–270.

Yazdani M. & Manovich L. (2015). Predicting social trends from non-photographic images on Twitter. In *IEEE International Conference on Big Data* (pp.1653–1660).

Yuan, Q., Nsoesie, E. O., Lv, B., Peng, G., Chunara, R., & Brownstein, J. S. (2013). Monitoring influenza epidemics in China with search query from Baidu. *PLoS ONE, 8*(5), e64323.

Zhou, K., & Yang, S. (2016). Understanding household energy consumption behavior: The contribution of energy big data analytics. *Renewable and Sustainable Energy Reviews, 56*, 810–819.