# The Role of Big Data in Addressing Societal Challenges: A Systematic Mapping Study

Farzana Quayyum<sup>1</sup>, Ilias O. Pappas<sup>1,2</sup> [0000-0001-7528-3488] and Letizia Jaccheri<sup>1</sup> [0000-0002-5547-2270]

<sup>1</sup> Norwegian University of Science and Technology (NTNU), Trondheim, Norway <sup>2</sup> University of Agder, Kristiansand, Norway farzanaq@stud.ntnu.no; ilias.pappas@uia.no; letizia.jaccheri@ntnu.no

**Abstract.** Big data has recently become the focus of academic and corporate investigation due to its high potential in generating business and social value. We have done a systematic mapping of the literature related to big data and its applications leading to social change through the lens of social innovation. The search strategy initially resulted in 593 papers, and after applying inclusion exclusion criteria a total of 156 papers were mapped; 59% of which were identified as empirical studies. This mapping investigated the publication frequency of the studies, research approach and contributions, research areas and article distribution per journal. We also address some challenges found from the mapping associated with the research topic. This mapping study will offer the basis for a reflection process among the researchers in this field and will allow us to develop a research agenda and roadmap of big data and its applications leading to social change.

**Keywords:** Big Data, Data Analytics, Social Innovation, Social Good, Social Change, Societal Transformation, Systematic Mapping Study.

### 1 Introduction

The evolution of Information and Communication Technology drives the digitalization process in many aspects of peoples' daily lives, which generates huge amount of data every moment from a growing number of sources. In the last decade, the use of big data and their analytics has earned a lot of attention. In various fields of science, technology, and business, the merit of big data is undeniable. But from the social perspective, the potential use of big data is yet to be figured out by social sector organizations [1]. Several definitions of big data exist, and they typically refer to the 'three Vs'' that characterize big data: volume, velocity, and variety, which have been extended including more characteristics of big data analytics exist [2], and they roughly refer to the combination of the data itself, the analytics applied to the data, and the presentation of results to generate value [3]. Thus, here we are interested in big data and their analytics.

The potential of big data and analytics to generate social value seems clear [1], however the main focus of big data analytics research has been on business value [4]. Combining big data analytics with social innovation can be the solution to address this gap [5,6]. Social innovation is defined as a novel solution to a social problem that is more effective, efficient, sustainable, or just than existing solutions and for which the value created accrues primarily to society as a whole rather than private individuals [7]. Social innovation can generate social good and lead to social change. Social good is typically defined as an action that provides some sort of benefit to the people of the society. The concept of social change refers to addressing the root causes of societal problems and changing them.

The terms social innovation, social good, social change, and societal transformation are related to each other. During this study our focus was on the applications of big data that have social impact and address social problems or challenges; so, to keep a broad and wide scope in this mapping review study we use all these terms.

Systematic literature reviews on big data applications have been conducted and investigate, among other things, big data dynamic capabilities [2], the operation and strategic value of big data for business [8], the impact of big data on business growth [9], the social economic value of big data [10]. Furthermore, there is an increasing number of studies that address both the business and social impact of big data as well as ways on how big data analytics can solve societal challenges, with evident examples the following recent special issues [1,11]. However, to best of our knowledge there is no systematic mapping or literature review that focuses solely on how big data and their analytics can lead to societal transformation and social good.

A systematic mapping can help us to understand what conditions can enable successful solutions, combined with strategies, tactics, and theories of change that lead to lasting impact [5,12,13]. Furthermore, this mapping will allow capturing the needed capabilities, resources, and conditions that the big data actors need to develop or acquire in order to manage big data applications, increase social value and solve societal challenges and create a sustainable society. To contribute to the creation of sustainable societies, we have done this systematic mapping of the literature related to big data and their applications leading to social innovation and thus societal transformation.

The objective of this study is to offer a map of the research that has being done, thus offering the basis to develop a research agenda and roadmap of big data and analytics and their applications leading to societal transformation and change. We have followed the standardized process for systematic mapping studies [14]. Based on the primary search with search strings, a total of 593 unduplicated papers was retrieved. After applying some exclusion criteria, the number was reduced to 165 (based on titles), then 153 (based on abstracts) and finally 146 were selected from the search and later 10 more papers were added manually from Google scholar.

The relative newness and growing interest in the research field, argues the need for a mapping study to identify the focus and quality of research in using big data analytics for social challenges. To provide an up to date overview of the research results within the field, we came up with the following research questions: RQ1: How the research about 'big data and social innovation' has changed over time (in the last decade)?

RQ2: How much of the research is done based on empirical studies and what type of empirical studies?

RQ3: What are the challenges or barriers for successful implementation of big data for societal challenges?

The paper proceeds as follows: Section 2 introduces the background of this study, Then, section 3 explains the detailed procedure of the research method, section 4 presents the results and findings of the mapping study, section 5 discusses the findings in relation to the research questions, section 6 concludes the paper presenting the implications of this study.

### 2 Background

### 2.1 Big Data

The digital and connected nature of modern-day life has resulted in vast amounts of data being generated by people and organizations alike. This phenomenon of an unprecedented growth of information and our ability to collect, process, protect, and exploit it has been described with the catchall term of Big Data [15]. Literature identifies 'big data' as the 'next big thing in innovation' [16], the next frontier for innovation, competition, and productivity" [17]. The rationale behind such statements is that the 'big data' is capable of changing competition by "transforming processes, altering corporate ecosystems, and facilitating innovation" [18]. It can be acknowledged as a key source of value creation. Beyond improving data-driven decision making, it also is crucial to identify the social value of big data [4], and what are the role of big data and potential impact of it in the society.

### 2.2 Social Innovation

The term social innovation has largely emerged in the last few years and there is much discussion about it now. The field of social innovation has grown up primarily as a field of practice, made up of people doing things and then, sometimes, reflecting on what they do [19]. The term social innovation has not any fixed boundaries, as it cuts across many different sectors like public sector, the social benefit sector, technology sector, and many others. The social innovation process has been described by scholars in multiple contexts as it needs to be multidisciplinary and cross social boundaries, for its impact to reach more people [20,21,22]. Social innovations are ideas that address various social challenges and needs.

#### 2.3 Big Data and Social Innovation

Big data contains a wealth of societal information and can thus be viewed as a network mapped to society; analyzing big data and further summarizing and finding clues and laws it implicitly contains can help us better perceive the present [23]. Data are an important element of social innovation. To initiate any innovative steps or to address any social challenge, data are needed. A deliberate and systematic approach towards social innovation through big data is needed as it will offer social value [5]. Since more data become available at a smaller cost, big data can be used as actionable information to identify needs and offer services for the benefit of the society and ensure aid to the individuals and society that generate them [13].

Following the importance of big data and social innovation, further work is needed to better define and understand how well society can benefit from big data to increase social value and lead to social good [4,6]. From this mapping, we can contribute to the field of big data research from a social perspective. While presenting an overview of the present research status, we also want to identify if there are any obstacles and challenges in using big data analytics that the stakeholder might face in their way to employ big data for their social innovative solutions. Big data can empower policymakers and entrepreneurs to provide solutions for social problems [6]. Identifying possible challenges and having a clear picture of the big data research in the social sector can also help stakeholders to prepare beforehand, to take advantage of the big data that are available, filter them and proceed to decisions that will help them innovate for social good and change.

### 3 Research Methodology

A systematic mapping study was undertaken to provide an overview of the research available in the field of big data analytics and social innovation leading to societal transformation, following the standardized process for systematic mapping studies [14] as illustrated in Fig. 1; along with guidelines from [24].

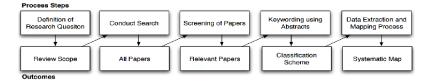


Fig. 1. The Systematic Mapping Study Process [14]

### 3.1 Data Sources and Search Strategy

In our primary search, we collected papers from all kind of sources including journals, conference papers, books, reports etc. This review was conducted in August 2018 and publications were searched from 2008 and onwards. We selected this timeframe as it is the time when these terms like big data and analytics, social innovation got the

momentum. The systematic search strategy consisted of searches in seven online bibliographic databases which were selected based on their relevance with our search topic and these databases are also well known for good quality literature resources in the field. To obtain high-quality data, we searched in the following databases – Scopus, ISI Web of Science, ACM Library, IEEE Xplore, SAGE, Emerald and Taylor & Francis. Then initial searches in the databases were conducted based on identified keywords (Table 1) related to this topic. The used search strings were:

Table 1. The keyword combination for initial search

"Big Data" OR "Data Analytics"	AND	"Social Innovation"
		"Societal Transformation"
		"Social Good"
		"Social Change"

#### 3.2 Study Selection

The study selection process is illustrated in Fig. 2, along with the number of papers at each stage. Searching the databases using the search string returned 593 papers, resulting in 465 unduplicated papers. These were imported into EndNote X8. Due to the importance of the selection phase in determining the overall validity of the literature review, a number of inclusion and exclusion criteria were applied. Studies were eligible for inclusion if they were focused on the topic of big data and data analytics, and their applications to foster social innovation, and lead to social impact, change and transformation. We used "big data" and "data analytics" separate to broader our search as several studies employ big data analytics techniques but do not use the term big data.

The mapping included research papers published in journals, conference proceedings, reports targeted at business executives and a broader audience, and scientific magazines. In progress research and dissertations were excluded from this mapping, as well as studies that were not written in English. Given that our focus was on the social innovation and societal transformation that big data entails, we included quantitative, qualitative, and case studies. Since the topic of interest is of an interdisciplinary nature, a diversity of epistemological approaches was opted for.

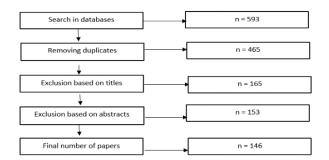


Fig. 2. The study selection process

#### 3.3 Manual Search

Following the systematic search, a manual search was also conducted. Google Scholar was used to searching for papers manually. At this stage total 10 papers from Google scholar was added to our EndNote library and the final number of papers became 156.

### 3.4 Data Extraction

After the mapping, we finally ended up with 156 papers. We performed a systematic analysis and extracted data from the abstracts of the papers that we need to answer our research questions. We extracted data regarding the - publication frequency, publication source, research area, research type, empirical evidence and contribution type.

# 4 Results and Findings

*RQ1:* How the research about 'big data and social innovation' has changed over time (in the last decade)?

**Publication Frequency.** The analysis shows that relevant papers are published from 2012 or later, with their frequency increasing yearly. The study was conducted in August 2018, so the year 2018 is not complete. The findings (Fig. 3) verify that the momentum or applications of big data are becoming increasingly popular.

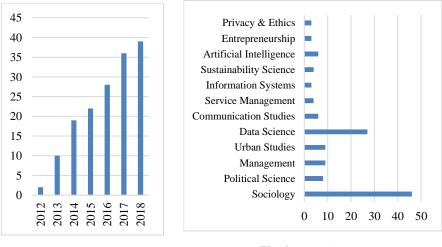


Fig. 3. Publication frequency

Fig. 4. Research areas

**Research Areas.** Next, we examined the research sectors of the published articles, to give an overview of the general categories. The findings are shown in Fig. 4.

**Publication Sources.** As mentioned in section 3, our mapping includes research papers published in academic outlets; but we have considered reports also (e.g., Hitachi reviews) because a lot of evidence is published by companies and a lot of work on social innovation and big data is done by companies as well.

We have tried to figure out how many of the relevant scientific papers are published in journals, how many as conference papers and from other sources. The statistic is given in Fig. 5.

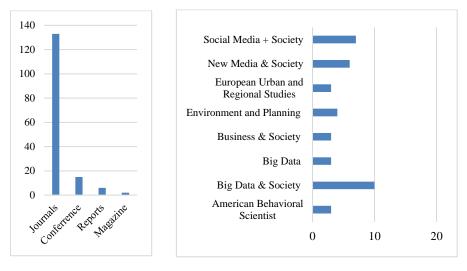


Fig. 5. Sources of publication

Fig. 6. Journals with a higher number of relevant publications

Along with the sources of the relevant papers, we have also searched for the journals who published maximum number of papers about our research topic i.e. big data and social innovation. Here in Fig. 6, we mention a few journals with maximum number of published papers from our review.

# *RQ2:* How much of the research is done based on empirical studies and what type of empirical studies?

**Empirical Evidence.** We primarily classified our reviewed papers as empirical and non-empirical papers. Non-empirical papers are conceptual papers. From the study, we see that majority (59%) of the paper s are based on empirical evidence. With this finding (Fig. 7), we also get the answer to our second research question.

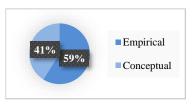


Fig. 7. Empirical and non-empirical ratio

We then classified the empirical papers based on the type of study. The research types that have been assessed followed the guidelines from [25] include: (1) survey, (2) design and creation, (3) experiment, (4) case study, (5) action research, and (6) ethnography. We have also included 'Discussion' as a research type, inspired by [26]. We have added this last method as we felt that some papers are more suitable to categorize as a discussion paper. Discussion papers are also known as 'Expert opinion'.

After deciding about the research types, we counted the numbers for each type. The following figure shows which research types of the studies found from our mapping. Only the papers providing empirical evidence (92 papers) were included in Fig. 8, covering a total of 7 research methods.

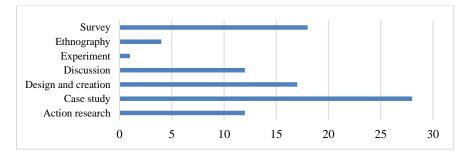


Fig. 8. Empirical evidence (research type)

**Contribution Type.** Every research paper has some contribution to the advancement of research in the relevant field by providing something new. To illustrate which types of contributions that have been made within the research area between, Fig. 9 was made. The figure shows the contribution type of papers. All 156 primary papers selected finally, in our mapping study are considered in this figure.

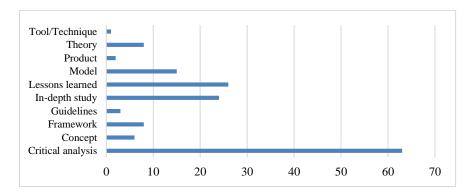


Fig. 9. Contribution types

We differ between 10 contribution types. Based on [25], we define six different knowledge outcomes including (1) product, (2) theory, (3) tool/technique, (4) model, (5) in-depth study and (6) critical analysis. We also adapt some more knowledge outcomes or contribution types since some contribution types from [27] can describe the contribution of some papers more precisely; including (1) framework, (2) lessons learned, (3) tool/guidelines and (4) concept.

# *RQ3:* What are the challenges or barriers to the successful implementation of big data for societal challenges?

Studying the title and abstract of all 156 papers, it has been found that only 3 papers mentioned challenges regarding employing big data in their studies. The challenges we find from this study are mentioned below:

- Open data and privacy concern [28]
- Challenge around obtaining data [29]
- The prominence of marketing-driven software [30]
- The interpretation of unpredictability [30]

So little evidence is not enough to generalize a fact for all and answer a research question like what the challenges or barriers for the successful implementation of big data for societal challenges are. So, we believe there is further research scope on this issue.

# 5 Discussion

# *RQ1:* How the research about 'big data and social innovation' has changed over time (in the last decade)?

From this mapping we have presented an overview of the research status on big data and social innovation that has been done in the last decade. As we could not find any prior systematic study on this topic, we cannot compare the results. But this mapping will now help other researchers to understand to social potential of big data research and applications. Our study proves that terms like big data and social innovation gained the attention of academic and business communities later than in 2010. It can be also seen that the number of researches and publications are increasing every year since then, which proves the importance and increasing attention big data and social innovation is getting day by day. Another study on big data [8] also stated that, "With regard to the literature review, 'big data' relevant journal articles have started appearing frequently in 2011. Prior to these years, the number of publications on the topic was very low. Publications on 'big data' related topics started only in 2008 (with 1 article) and then a steady increase in the number of publications in the following years".

From this mapping, we can see that many fields including social science, political science, information systems, urban management, communication, healthcare sector adapted big data for their applications. In the results section, we have presented the fields with major number of research studies, but there are also research fields we have

found form the mapping where big data is being used; like- education, journalism, tourism, etc. Here notable that all these papers with applications of big data in different fields are directly or indirectly related to various social issues; which proves that big data applications have a big potential to be used for the good of the society not only for business or technology.

# *RQ2:* How much of the research is done based on empirical studies and what type of empirical studies?

In our systematic mapping, more than half of the papers (59%) provide empirical evidence. As there was no previous mapping on this topic, we cannot say how much empirical work was done before. But when 59% of the studies are empirical it proves that the researchers of this field are contributing much. With their contributions, the quality of research is also improving. The major contribution of the research papers from our mapping was a critical analysis, both empirical and non-empirical. When analyzing different topics, the authors also presented their insights, research agenda, guidelines for future research, what lessons they learned and their opinions. The empirical studies also presented models, frameworks and tools that can be used in future research.

# *RQ3:* What are the challenges or barriers to the successful implementation of big data for societal challenges?

In article [28], the authors reflected on various case related to big data challenges, including the challenge of maintaining data privacy and ethics when using all forms of big data for positive social change. The authors recommended exploring new formats for educating people about privacy/data protection risks to overcome data privacy challenges and to use templates to evaluate open data sources. In [29] authors investigate how the challenges around obtaining data to enforce new regulations are addressed by local councils to balance corporate interests with the public good. The authors stated that triangulating different sources of information is not always straightforward as the publicly available data might be partially obscured. In their case study, the authors recommend about platform economy to overcome the challenges regarding data collection. In [30], the authors examine the dominance of marketing-driven commercial tools for predictive analytics of data and their effectiveness to analyze data for completely different purposes such as law enforcement. Another challenge that [30] mentions is, the notions of predictability and probability remain contentious in the use of social media big data. The authors reflected upon the challenges and points to a crucial research agenda in an increasingly datafied environment.

### 5.1 Use of Keywords

We found some research papers relevant to our study, but they have not been included in the mapping as they do not use the keywords we searched with. For example, [31] use mobile call data to predict the geographic spread and timing of epidemics, and indeed they address a social challenge and has a significant societal impact. However, they do not use keywords regarding data analytics and societal impact, maybe because their focus is mainly on modeling and technical aspects. Instead their keywords include human mobility, mobile phones, epidemiology, dengue etc. Considering the importance of social implications of big data research as well as the interest of publication venues in contributing to societies [1], we suggest that future papers should take into account and report such implications in their abstract and keywords. We should note that indeed many papers discuss social implications, however they do not mention them in their abstracts, raising the need for a systematic literature review in the area. Thus, a more detailed analysis of the research articles can lead, among other things, to new combinations of keywords that will be able to better capture the current status regarding the impact of big data analytics on societal challenges.

#### 5.2 Limitation of the Study

Systematic mapping study approach is not without limitations [32]. For the validity of this review, threats of retrieval of papers need to be considered. Even though a systematic approach was used during this mapping, the selection of papers dealing with "big data" that we have included was based on our subjective judgment. Another limitation is, we have used only titles and abstracts to extract data. So, the categorizing and data extraction process depends on the quality of the abstracts. ICT-related research publications often do not use structured abstract [33] which results in poor accuracy when classifying papers based solely on abstracts. Also, following the standard procedure of systematic mapping, we could not include research papers in our study which do not have the keywords we searched with; even though some papers might be relevant to our topic.

### 6 Implication for Research and Practice

This systematic mapping study extends the big data research in several ways. Our work contributes to the social perspective that emphasizes the importance of adoption and applications of big data. This study can guide other researchers of this field to develop their agenda and roadmap for future research. The findings of this research show the type of contributions big data research is making in the industry; based on that future researchers can think what type of contributions we are lacking and make their research agenda on that. In this research, we have also identified the challenges of big data adoption in the social sector. Future researchers can explore more about these challenges and can investigate if there are other challenges. There is possible future research potential to address and propose solutions for these challenges; so that employing big data can be easier and more efficient for the stakeholders.

### 7 Conclusion

This paper presents findings of a systematic mapping study that researchers, social innovators, social entrepreneurs and all other stakeholders can use to unlock the power of big data for the benefit of the society. We have presented the current status that shows how research into big data and social innovation has increased over the last decade, attracting significant attention from across a wide array of disciplines. We have identified the major research areas where big data is getting significant attention; so future researchers can explore more about the impact of big data in those areas. This study also proves that the empirical ground of research in this field is strong; research is not only limited to case studies, but also other forms of research is being done like action research, critical analysis, designing and creating new products, etc. The key contribution this paper has made is offering the basis for a reflection process among the researchers in this field.

### References

- Pappas I.O., Mikalef P., Giannakos M.N., and Krogstie J.: Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies. Inf Syst E-Bus Manage 16: 479, (2018).
- Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and e-Business Management*, 16(3), 547-578.
- Gantz J, Reinsel D (2012) The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east. IDC iView IDC Anal Future 2012:1–16.
- Agarwal, R. & Dhar, V.: Editorial—Big data, data science, and analytics: The opportunity and challenge for IS research. Information Systems Research, 25, 443-448 (2014).
- Pappas I. O., Jaccheri L., Mikalef P. and Giannakos M. N.: Social Innovation and Social Entrepreneurship through Big Data: Developing a Research Agenda. 11th Mediterranean Conference on Information Systems (MCIS) (2017).
- 6. Desouza, K. C., and Smith, K. L.: Big data for social innovation. Stanford Social Innovation Review, Summer (2014).
- Phills J. A., Deiglmeier K. and Miller D. T.: Rediscovering social Innovation, Stanford Social Innovation Review 6(4), 34-43, (2008).
- Wamba S. F., Akter S., Edwards A., Chopin G. and Gnanzou D.: How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study, International Journal of Production Economics, Volume 165, Pages 234-246, July (2015).
- de Camaro Fiorini, P., B.M. Roman Pais Seles, C.J. Chiappetta Jabbour, E. Barberio Mariano and AB. L. de Sousa Jabbour: Management theory and big data literature: From a review to a research agenda. International Journal of Information Management 43: 112-129 (2018).
- Gunther, W. A., M. H. Rezazade Mehrizi, M. Huysman and F. Feldberg: Debating big data: A literature review on realizing value from big data. The Journal of Strategic Information Systems 26(3): 191-209 (2017).
- 11. Tseng F. M. and Harmon R. R.: The impact of big data analytics on the dynamics of social change, Technological Forecasting and Social Change, Vol 130, (2018).

- Cajaiba-Santana, G.: Social innovation: Moving the field forward. A conceptual framework. Technological Forecasting and Social Change, 82, 42-51, (2014).
- Zicari R. V.: Big data: Challenges and opportunities. AKERKAR, R. (ed.) Big data computing. Boca Raton, FL: Chapman and Hall/CRC, (2014).
- Petersen K., Feldt R., Mujtaba S., and Mattsson M.: Systematic mapping studies in software engineering. In EASE, volume 8, pages 68–77, (2008).
- Cuquet M., Vega-Gorgojo G., Lammerant H, Finn R., Umair ul Hassan: Societal impacts of big data: challenges and opportunities in Europe, Cornell University Library, (2017).
- Gobble M. M.: Big data: the next big thing in innovation, Res. Technol. Manag., 56 (1), pp. 64-66, (2013).
- Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., and Byers A. H.: Big Data: The Next Frontier for Innovation, Competition and Productivity, McKinsey Global Institute (2011).
- 18. Brown B., Chul M., Manyika J.: Are you ready for the era of 'big data'?, McKinsey Q., 4 (2011), pp. 24-27+30-35, (2011).
- Mulgan G.: The theoretical foundation of social innovation, Social Innovation, Springer, (2012).
- 20. Mulgan G.: The process of social innovation. Innovations, 1(2), 145-162, (2006).
- 21. Herrera M. E. B.: Creating competitive advantage by institutionalizing corporate social innovation. Journal of Business Research, 68(7), 1468-1474, (2015).
- Murray R., Caulier-Grice, J., and Mulgan, G.: The open book of social innovation. London, UK. National endowment for science, technology and the art (NESTA), (2010).
- Jin X., Wah B. W., Cheng X., Wang Y., Significance and challenges of Big Data Research, Big Data Research, Volume 2, Pg. 59-64 (2015).
- Kitchenham B.: Procedures for performing systematic reviews. Keele, UK, Keele University, 33:1–26 (2004).
- Oates B. J.: Researching Information Systems and Computing: Sage Publications Ltd., (2006).
- Zannier C., Melnik G., and Maurer F.: On the success of empirical studies in the international conference on software engineering, ICSE '06 Proceedings of the 28th international conference on Software engineering, Pages 341-350, (2006).
- Mathiassen L., Chiasson M., and Germonprez M.: Style Composition in Action Research Publication. MIS Quarterly, 32(2), pp. 347-363, (2012).
- Taylor L., Cowls J., Schroeder R., Meyer E. T.: Big data and positive change in the developing world. Policy and Internet 6, 418-444, (2014).
- 29. Ferreri, M. and Sanyal, R.: Platform economies and urban planning: Airbnb and regulated deregulation in London, Urban Studies, Volume: 55 issue: 15 (2018).
- Dencik L., Hintz A., Carey Z.: Prediction, pre-emption, and limits to dissent: Social media and big data use for policing protests in the United Kingdom. New Media & Society 20, (2017).
- Wesolowski A., Qureshi T., Boni M. F., Sundsoy P. R., Johansson M. A., Rashed S. B., Engo-Monsen K. and Buckee C.O.: Impact of human mobility on the emergence of dengue epidemics in Pakistan, PNAS September 22, (2015).
- 32. Kitchenham BA, Budgen D, Pearl Brereton O (2011) Using mapping studies as the basis for further research—a participant observer case study. Inf Softw Technol 53:638–651.
- Kitchenham BA, Pearl Brereton O, and Owen S.: Length and readability of structured software engineering abstracts. IET Softw 2:37 (2008).