Big Data in International Relations: Big Data in Conflict Prediction and Prevention

Master thesis

Author: E. Toth-Szollos
Supervisor: Prof. dr. A.W.M. (André) Gerrits

24 June 2015

Master International Studies
Leiden University

Table of contents

Abstract

With the abundance of data and more sophisticated analytical capabilities, big data is receiving increasing attention in international relations research. This thesis intends to explore the relevance of real-time big data analysis in violent conflict prediction and prevention by examining the state of the art of big-data based social science research, assessing the value of open data and elaborating on the question of access and the interests of stakeholders in data-sharing across sectors. Finally, three data sharing models will be visualised and the privacy vs. security dichotomy will be addressed.

Introduction

As data-mining and algorithms on human behaviour become more developed and nuanced, vast investments are being made in the commercial sphere to exploit the power of big data. Increasingly, big data is also being used in natural sciences such as climate change research (Fagmous&Vipin, 2014), sustainable development research (Gijzen, 2013), and genomics and biology (Marx, 2013). In the field of economics, Piketty was pioneering with using large datasets in his groundbreaking book on inequality (see Piketty, 2014). In humanitarian relief efforts and international development, big data is used in the field of public health, food security, poverty alleviation, disaster relief, etc (Kirkpatrick, 2014, p. 5, Meier, 2015, pp. 19-22). However, big data-based research and predictive analytics in social sciences are still at an initial stage. This thesis contributes to knowledge by exploring a relatively less used application of big data in the field of social sciences, namely the potential of big data to predict violent mass conflicts and trigger preventive mechanisms through cooperation between private corporations, governments and non-governmental organisations (NGOs).

The importance of big data in today's economy is demonstrated by the dynamic growth of the field as well as investments made in it: big data technologies and big data related services are estimated to grow globally to 16.9 billion USD in 2015. This growth demonstrates a compound annual growth rate of 40%, which is seven times bigger than the growth of the global information and communications technology (ICT) market in its entirety (European Commission, 2014, p. 2). Already in 2011, the United States Department of Defense spent approximately 28 million USD on human dynamics and computational social science related projects (Weinberger, 2011, p. 566). In August 2014, the

United Nations General Secretary created an Independent Expert Advisory Group on the Data Revolution for Sustainable Development which illustrates the UN's shifting focus towards big data. Among others, the aim of this group is to explore opportunities arising with the surge of new public and private data providers, complement conventional statistical systems and ensure accountability on all levels (United Nations Press Release, 29 August 2014). In October 2014, the European Union announced a public-private partnership with the European data industry that is worth 2.5 billion euro. This project aims at joining forces to boost innovation and support research by harnessing the power of big data on areas such as energy efficiency, manufacturing and healthcare (European Commission Press Release, 13 October 2014). Notable investments in academia include the establishment of the Stanford Peace Lab and the Peace Informatics Lab at Leiden University.

As a result of the abundance of data, the stance towards data processing and data protection is increasingly contentious in society. Milan defines the term data activism as the awareness of new social practices rooted in technology and data, and distinguishes two kinds of data activism: re-active and pro-active data activism. Re-active data activism resists massive data collection by corporations and government entities and regards it as a threat to civil liberties. On the other hand, pro-active data activists assert that big data promises great opportunities for collective action and the creation of social good (Milan, 2015). This thesis takes a strong pro-active data activist approach and intends to create a model where big data can be exploited widely for the creation of social good, while limiting concerns about privacy to the minimum.

In order to narrow down the scope of research, this study will focus on the possible usage of data collection and analytics by large telecommunications companies and social media platforms in the area of conflict prevention and peace studies. Therefore, lone-wolf terrorism and the controversies of data collection on individuals is not a key focus of study, as this research focuses on real-time analysis of group dynamics and extremist tendencies with help of big data.

Research question

The main research question of this thesis is the following:

- What is the value and potential of the real-time analysis of big data for the purposes of violent conflict prediction and prevention?

The sub-questions pertaining to the research question are the following:

- What is the state of the art of big data based research in social sciences? What are the criteria of efficient conflict prediction and how can big data contribute to success?

- What is the real value of open source data in real-time analytics and research? How is data accessibility impeding the big data for peace project? How does cooperation or the lack of cooperation affect the potential of big data-based conflict prediction and prevention?

- How can data-access related mechanisms be coordinated with regard to conflict prediction and prevention?

By posing and attempting to answer these questions, this study aims at understanding the implications of an emerging trend in international relations. An overview of the first sub-question, the status quo will be given in the literature review. Subsequently, the accessibility issue will be discussed through the theoretical framework of complex interdependence. Finally, an analysis will be given on cooperation and a conclusion will be made. The key learning of this thesis shall be a better understanding of the relevance of big data in international relations and particularly peace studies research.

Literature review

Big data and prediction in social sciences

This section will locate big data technologies and predictive analytics in the realm of social sciences and examines their relevance in international relations. In addition, major questions about the feasible, efficient and ethically acceptable application of big data will be touched upon.

It is becoming increasingly accepted that in critical decision making, algorithms often do better than humans (Kahneman, 2012, p. 229). In this respect, big data has been a popular buzzword in recent years in both commercial and scientific circles. In the private sector, big data is primarily used for customer behaviour analysis with a purpose of consumer influence and persuasion, as well as optimisation of operations (Mayer-Schönberger & Cukier, 2013, pp. 57-59). Natural sciences increasingly operate with large datasets and their predictive value. Big data is increasingly being used in social science research as well, but is still barely used in policy design.

It has been generally acknowledged that big data has the potential to fundamentally change key methodologies of social sciences. Through the analysis of interactions of individuals on online social platforms, data analysts can explore existing trends and what is actually happening in real life

communication, as opposed to attempting to simulate reality with questionnaires. As such, instead of conducting a small-scale experiment using surveys or questionnaires to explore individual's opinion on a given issue, big data mining can lead to research findings that are closer to reality (Rudder, 2014, pp. 10-11). Patterns and trends can be found by analysing large amounts of data, with the advantage that participants are not influenced by phenomena such as response bias, defined as the deviation from the correct, honest accurate response by giving socially desirable answers or by faking good or bad answers and dissimulating the report (Furnham, 1986, p. 385). As opposed to the kind of "manual data collection" conducted via surveys and questionnaires, having an entire dataset which represents real-life communication offers access to non-biased sources, but at the same time poses new challenges, discussed later in this chapter.

Moreover, big data is not solely suitable to analyse past or current events and trends, it also provides scientists with substantial predictive capabilities. The predictive potential is already being exploited on a large scale in private companies (Mayer-Schönberger&Cukier, 2013, pp. 57-59). Mises alludes to the idea of prediction in social sciences when he juxtaposes natural sciences with sciences dealing with human action and society. He argues that there is a lack of regularity in human action, hence specific outcomes cannot be predicted: "It is not a shortcoming of the sciences of human action that they have not succeeded in discovering determinate stimulus-response patterns. What does not exist cannot be discovered" (Mises, 1957, p. 9). In his book he examined prediction in a socially complex environment and asserted that specific outcomes could not be predicted in social science with accuracy, therefore, he concluded that social science should aim for prediction of patterns. The approach to human action that Mises took predated rational choice theory, which primarily focuses on human nature governed by intention and purpose (Smith, 1998, p. 155). As demonstrated above, pattern conceptualisation and prediction based on sequences and structures have a long tradition in social sciences. However, an important difference in regular research and big data analytics is that by mining data, no causalities can be found. Instead, patterns in data can be identified through the use of algorithms (Mans, 2014, p. 7).

The excessive and real time use of large datasets opens up a novel way of doing predictive and scientific research in a socially complex environment. Prediction in social science has been accepted as a scientific activity, and scientific attempts such as future studies or futurology were recognised as valuable additions to the scientific community already in the 1980s (Wilcox, 1983, p. 202). Fu-

ture studies presently is categorised as design science (Niiniluoto, 2001), and it has an ontological typology that separates its predictive and prognostic disciplines, which are of explanatory nature, from non-scientific scientific branches, such as science fiction and utopias/dystopias (Bergman et al., 2010, p. 859). However, the emergence of big data is changing this trend and points to more elaborate predictions within the realms of social science and more significance for computational social science (Ch'Ng, 2014, p. 2). In essence, even though prediction and future-related research have been acknowledged in social science, big data is bringing a new edge to the assessment of prediction in social science. As a consequence, the emerging multidisciplinary field of 'human dynamics' and 'computational social science' bring together research fields such as social-network analysis, political forecasting and complexity science (Weinberger, 2011, p. 566).

Challenges in the application of big data for conflict prevention

In the early 2000s, Laney defined three major challenges in managing large datasets: volume, velocity and variety (Laney, 2001). In 2014, Khan et al. added three more variables and asserted that value creation from big data depended on six factors: volume, velocity, veracity, validity, variety and volatility (Khan et al., 2014, fig. 1), terms which are key to understand the applicability of big data in social science. *Volume* describes the challenge posed by the sheer amount of data, which can be demonstrated by the following example: as per 2013, Google processed approximately 24 petabytes of data on a daily basis. To translate this into more comprehensible terms, this quantity is thousands of times that of printed material located United States Library of Congress (Mayer-Schönberger & Cukier, 2013, p. 8). Closely connected to volume is *velocity*: with the staggering amount of data being produced in high speed, tailored methods are required to ensure quick data processing and filtering. *Veracity* denotes the truthfulness of data, whereas *validity* refers to the accuracy and correctness of data in the context of its intended usage. It is vital to distinguish veracity from validity in a specific context (Khan et al., 2014, p. 3). *Variety* describes the idea that data is present in different formats which are often inconsistent and incompatible and require alignment and structure (Laney, 2001, p. 2). Moreover, from a different perspective, variety may also stand for the notion that combining data from multiple sources is a pivotal prerequisite to successful big data-based research (Manyika et al, 2011, p. 118). Valuable conclusions can only be drawn if data comes from varied sources and results are in line with each other and can be shared across different parties. Finally, *volatility* is crucial due to the volume, variety and velocity of data: retention of data is cum-

bersome due to its size, diverse forms and the speed with which it is produced, which is why the dataset is liable to change rapidly (Khan et al., 2014, p. 3).

Moving forward, the three major questions about big data that are relevant to this study are access, ethics and regulation. These are the main factors that may impede the widespread application of big data for purposes related to international relations, and more specifically to the identification and eventual prevention of extremist violence in society. To start with, the major access issue is the fact that social media data, scripts of social exchange and interaction are predominantly owned by private companies. Moreover, messages sent via popular applications such as iMessage, WhatsApp or Facebook are encrypted, and the private corporations owning the data are currently not likely to share information or datasets for reasons varying from institutional interests, competitive risk, business risk, and the privacy risk of individuals. Corporate data is presently being shared with entities outside the corporations such as the United Nations, however, this is done on an ad hoc basis and is not scalable in its current form (Hoffmann, Big Data for Humanity, 2014, 33:33-36:30).

Secondly, ethical questions such as privacy are prominently present in the current discourse on big data. Profiling activities, information collection and information retention methods of law-enforcement and intelligence agencies have been highly contentious in recent years, which has given a negative connotation to big data in the public. The dichotomy between security and privacy is increasing, as separating information about law-abiding citizens with no suspicious activities ensures that their privacy is respected, but is also conducive to the large-scale inefficiency of investigations. As argued by Schmidt and Cohen, the big data challenge lies in the integration of digital information owned by different actors such as law enforcement bodies and private companies, without violating the privacy of citizens (Schmidt & Cohen, 2014, pp. 268-269).

Currently, there are no scientific sources on the potential use of big data in the framework of corporate social responsibility (CSR). Open data publications have been a widely accepted and used way for private companies and institutes to share data for scientific or informative purposes. The aim of the open data movement is to democratise the process of information and knowledge creation instead of letting those possessing the data make use of their monopoly. Within the open data project, data mainly produced by US public agencies has been released and made available for free retrieval (Kitchin, 2014, p. 48). This data is suitable for social science research, but the variety of data is not

sufficient without more data from private companies. Moreover, open source data is usually posted retrospectively, hence it is not suitable for real-time analysis.

Thirdly, there is a significant lack of regulation on big data. The above example of the open data project is a positive example of regulation on data sharing, however, significant data is not being shared across sectors due to missing mechanisms and a platform to perform such action. Neutral bodies, such as the United Nations are often contacted by private companies when a possibility for creation of public good from big data is discovered, but this is only done as a result of the good will of individual companies (Kirkpatrick, Big Data for Humanity, 2014, 14:41-17:45). In addition to the good will, there may be a significant commercial interest behind these data-sharing initiatives which aim at positive image building or the development of a corporate social responsibility strategy.

Bringing big data and peace research together

One of the most common commercial sources of big data is Google Analytics. Google uses a software to track mouse clicks and information requests made by website visitors. Google shares this data with web managers to improve traffic on their websites — in this scenario, managers of websites that use Google's services get access to big data by paying to the company for their services. This data is primarily useful to determine the popularity of webpages and their sections, the type of information Internet users are interested in, the path they take to navigate to and from a website, the time spent on a website, etc. In this process, no personal information is stored about unique users, the sole use of the metric data is the establishment of trends (Kent et al. 2011, pp. 536-537). Even though this source of big data may be relevant to observe traffic of websites and the popularity of a website, and clicks leading to certain websites may help to identify extremist tendencies in a region, this study will specifically focus on social media research and written expression of extremist views. However, the basic idea of Google Analytics that aims at establishing trends as opposed to identifying or targeting individuals will serve as guiding principle in the research.

Mining social media data can shed light on the following questions: relationships within a network, frequency of communication, symmetry of communication, network analysis, and content analysis of communication (Russell, 2011, p. xiv). The first four elements can be confined to individuals and

networks and answer questions such as who knows whom, who are the influential people in a network, what are the dynamics between people communicating with each other. For the purpose of this study, the last item, content analysis of social media communication is the most relevant, i.e. how expressions of extremist views can be identified, filtered, aggregated and analysed in order to identify extremist waves at an early stage and prevent an outburst of violence. When doing research on written expressions of Internet users' opinion, several questions need to be taken into account: the language of the text, the accessibility of the text, the variety of data and the available methods to infer valuable information from the datasets. Out of these fours notions, access is the most relevant issue from an international relations perspective, as this question has wider implications on cooperation in the international community.

After the tragic massacres in Rwanda and Srebrenica, the international community had to recognise that conflict prevention could not be limited to solely diplomatic or military initiatives: new ways of early intervention methods must be applied which make use of the power of technology. This expectation towards the international community grew with the rise of data on social dynamics and the fact that data from social media does not only shed light on facts but also on what people think and feel. This kind of data can be harnessed from social media communication, analysis of SMS and text messages, satellite information and GIS mapping and news-scraping applications (Himelfarb, 2014, p.1).

There is a growing number of peace research projects that aim at bringing large amounts of data online and accessible to the public in order to harness the power of big data in peace studies (Leiden Peace Informatics Lab[1], Integrated Network for Societal Conflict Research[2], Global Peace Index[3]). However, most of these research projects are reactionary, focus on events that had already happened and big data is analysed to understand root causes and give a deeper historical background. Moreover, open source data has been used by Internet users for humanitarian purposes: crisis mapping after the Haiti earthquake in 2010, the typhoons in the Philippines in 2013, the mapping of the Syrian crisis since 2011 (Meier, 2015, pp. 18-21). These projects were based on real time analysis of

[1] http://www.peaceinformaticslab.org

[2] http://www.systemicpeace.org/inscrdata.html

[3] www.visionofhumanity.org/#/page/indexes/global-peace-index

data and accelerated the reaction to humanitarian catastrophes and facilitated the access to help to those in need. Finally, examples to predictive research includes Google's efforts to identify and map the H1N1 virus in 2009 or the ebola epidemic in 2014 by real-time analysis of symptom-related search terms entered into its search engine (Mayer-Schönberger & Cukier, pp. 2-4, The Economist, 2014). However, there is little evidence on scientific endeavours or civil projects in the domain of peace studies where the three elements, namely big data-based research, real-time analysis and predictive features were used together. Accordingly, Mans identifies three key issue areas in big data for peace projects: first, data analytics is not yet applied widely in the field of peace studies; second, the availability of data is expected to increase in the coming years; third, the validity of the big data concept in peace research must be proved in order to move forward (Mans, 2014, p. 3).

Definitions and theoretical framework

Definitions

The most recurring field specific terms are explained in this section. These are big data technology, big data, metadata and open data. Big data technology can be defined as a technology that aims at extracting valuable information from vast volumes of diverse data through high-velocity capture, discovery and analysis (Vesset et al, 2012, p. 1). Another key feature of the big data technology is its potential: the increased capacity to search, aggregate and cross-reference data sets of great size (Boyd & Crawford 2012, p. 663). For the purpose of this study, big data denotes large and diverse datasets, and big data technology refers to the various methods of value creation from large datasets. The term metadata refers to "the data about the data" which includes names and descriptions of specific fields of the content that help data analysts understand the composition of datasets and facilitate their interoperability and discoverability (Kitchin, 2014, pp. 8-9). Finally, the term "open data" is used for data made available to the public free of charge, for both commercial and non-commercial purposes (European Commission, 2014, p. 5).

In addition, for social science research purposes it is vital to distinguish two kinds of data: the first kind provides information on *"what people say,"* which can be publicly available in form of tweets, blog posts or news sites and publicly restricted as in the case of private messages. The second kind of data reflects *"what people do,"* which includes information such as mobile phone traffic information that is gathered for business intelligence and sales optimisation purposes and can also be

useful for the tracking of individuals after a natural disaster (Kirkpatrick, 2014, p. 4). For the purpose of research on mass behaviour, the first kind of data will be more relevant.

Theoretical framework

Cooperation and the role of non-state actors have broad literature in international relations theory (see Herber, 1996; Keohane, 2005; Perlman & Cunningham, 2011), out of which the theory of complex interdependence and institutional liberalism will serve as the cornerstone for theoretical analysis. Keohane and Nye, when putting the information revolution in the context of their theory of complex interdependence, argue that the information revolution affects power measured in terms of resources rather than behaviour. Moreover, they add that the information revolution, by reducing the costs of disseminating information, has added to corporations' marketability and has increased the value of commercial information. In addition, they highlight that the ability to disseminate free information has increased the agency of those sharing the information, and has enhanced their power for persuasion in world politics (Keohane & Nye, 1998). In the present context, these assertions are valid; however, the fact that companies do not only disseminate a vast amount of information, but also collect it and keep much of it to themselves, is vital. Power has shifted from the ability to dissemination to the ability to collect, filter, and share — or not share — information. Moving forward, to assess the cooperation or the lack of it between the affected actors in the realm of big data based analytics in peace studies, three factors must be taken into account: state interests, the influence of private interests and the role of ideas (Ovodenko & Keohane, 2012, p. 538). The three factors will provide the foundation for the analysis of the power and influence of the relevant actors.

The analysis will examine in detail the question of accessing datasets from diverse sources and the motivation to share or not to share data across sectors. Currently, as companies that possess data of interest to scientists usually do not share client data with the academic community (Savage & Burrows, 2007, p. 887), scholars opting for conducting research based on big data are forced to use the massive amounts of publicly available data, i.e. open data. As the practical use of open data is relatively low in real-time analytics, more widespread mechanisms for sharing data across sectors would be desirable and beneficial for humanity. As a prerequisite to this, interests of different actors must be converged and data privacy issues must be regulated in a favourable manner. Prior to the analysis, the following section will present a case study that will give an illustration on how these actors can affect big data-based real-time analytics and help to maintain peace.

Case Study

Big data is receiving an increasing amount of attention from international organisations, NGOs and social scientists. However, there is little practical evidence for value creation from peace-related big data research that resulted in proven and significant public good. In addition, while international organisations such as the UN and the EU clearly show an increased interest in big data through their investments, there has been no sustainable cross-sector sharing mechanism established either by these entities, or by governments. Moreover, a key limitation to social media based social science or peace research is the fact that it mainly focused on the situation of developed countries with a high smartphone penetration and social media activity. Typically, in these countries there is a low number of occurrences where extremist views and hate speech have resulted in violent uprisings or massacres in recent years.

In addition, the growing trend must be underlined that since the 1950s, wars and violent conflicts occur less often between states and more often between groups or tribes within a state (Pearlman & Cunningham, 2011, p. 3). The growing tendency of intra-state and intra-group militarised disputes is a vital consideration in conflict prevention through big data: from the perspective of access, the advantage of the limitation to one country is that there is a limited number of mobile phone or telecommunication providers, however, further factors may complicate the feasibility of research, such as the use of diverse local languages and certain features in communication such as abbreviations, which make language as data more difficult to quantify and measure (Kirkpatrick, 2014, p. 6). In short, big data for peace projects have more relevance in developing countries, therefore the case study will focus on one of these countries, namely Kenya.

The case of the 2007 and 2013 Kenyan elections

Kenya is a valid choice to study as it has a high penetration of mobile phones: in 2011, 63% of population had a mobile phone (Verjee, 2013, p. 2). High mobile phone activity is vital in order to ensure that there is enough data to be researched. However, previously established early warning and prevention projects in Kenya had had several drawbacks: lack of transparency and accountability, government inefficiency and a gap between warning and prevention. For a considerable amount of

time, Kenyan early warning and prevention projects were run through CEWARN, the Conflict Early Warning and Response Mechanism, a collaborative effort of seven East African states, including Kenya. The mechanism aimed at ensuring timely transmission of warning information on increasing violence from local to state level, but the sluggish system prevented NGOs from accessing information on time and the project largely failed (Mancini & O'Reilly, 2013, pp. 2-3). During the Kenyan elections in 2007, tribal conflicts ended in violence and the death of more than 1000 people. In this instance, text messages were used to cause fear and mobilise people to perpetrate acts of violence (Verjee, 2013, p. 2). In contrast, the Kenyan election of 2013 is considered as one of the major successes of conflict prevention through social media monitoring and data analytics (Himelfarb, 2013, p. 2).

A key player in monitoring the 2013 elections, the NGO Sisi Ni Amani ("We are Peace" in Swahili) was founded in 2011 in response to the 2007 election tragedy with the aim to combat violence and prevent tragic happenings similar to the 2007 election riots. The NGO cooperated with Safaricom, East Africa's largest mobile communication provider, which experienced hate speech through its network during the election violence in 2007 and had the ability to analyse data through filtering to demographic data, gender and geography (Verjee, 2013, p.1). The collaborative effort coordinated by the Kenyan NGO was based on monitoring and streamlining media reports, tweets and blog posts to identify rising extremist views, tension, hate speech and calls to engage in violent activities. The collected information was fed into an election-monitoring map. Based on this information, the local peace-building organisation sent out 50 million text messages via Safaricom at the peak of violence in order to prevent violence and the spread of harmful rumours. In addition, security forces were sent on-site to control polling stations (Himelfarb, 2014, p. 3).

The reason why this case study is applicable to this research is the presence of actors from government, private sector and NGOs. Mancini and O'Reilly define three key prerequisites to successful early warning and prevention projects: cooperation with local bodies and the accessibility to local input, effective partnerships between NGOs, mobile phone/data providers, and horizontal sharing of information as opposed to top-down government coordination. A strong prerequisite to successful prediction of possible conflict through civil, governmental and regional early-warning efforts is a relatively stable setting where governments do not restrict access to information flows for purposes of conflict prevention (Mancini & O'Reilly, 2013, p. 3). These were largely met in this case study.

The strong NGO involvement sheds light on the importance of the inclusion of the civil society in conflict prevention, as well as international bodies such as the UN and the EU. Finally, the Kenyan case is one of the few relatively widely cited successful instances of conflict prevention through the application of big data technologies.

In contrast, this case can also point to key limitations and lead to the need of more extensive analysis of telecommunications data or social media access related questions. First, while there may be a correlation between the conflict prevention activities and the peaceful outcome of the elections, causality cannot be proven: one must not be trapped by the logical fallacy of *post hoc ergo propter hoc*. Moreover, it is questionable whether in other violent conflicts the transmission of messages of peace can result in meaningful impact — the replicability of the project is cumbersome. In addition, it is a major question to what extent this kind of monitoring and real-time analysis can bring new insights and results that cannot be yielded from research done with more traditional methods. The tool may be used as a source of verification to previous assumptions established via other research methods and as a source to identify trends and possible outbreaks of violence, but cannot be expected to give straightforward answers to complex questions (Weinberger, 2011, p. 568). Nevertheless, Kirkpatrick argues that there is a false dichotomy between quantitative use of data in decision-making and qualitative assessment used by policymakers. As such, big data analytics is and can be a useful addition to peace research and be part of the solution, not the solution itself (Kirkpatrick, 2014, p. 6).

To conclude, the case study in the Kenyan setting has several promising features, notwithstanding the fact that causality is difficult to prove. It can be concluded based on the literature review, the theoretical framework and the case study that the limitations and the feasibility of the application of big datasets come down to three major factors: first, the nature of the conflict being analysed; second, data privacy and Internet regulations in the setting of a developed country; third, the lack of mechanisms to share data for research related purposes that have a predictive and preventive edge. The subsequent chapter deals with issues pertaining to access of data and data sharing mechanisms.

Analysis

The main focus of this section is on the source of data and the interest of different actors in data sharing. First, it will be discussed what the difference between the value of open data and anonymised data is. Then, the interest of sharing or accessing data will be examined from the perspective of different actors and two data sharing models (see Model 1 and Model 2) will be established based on regulations in the United States. Finally a third model will be established (see Model 3) to suggest a more efficient cooperation between actors on predictive and preventive peace research.

The value of open source data

The matter of open source data vs. anonymised data comes into question in peace research, as it is mostly the latter kind of data that can be used for real-time analysis and short-term prediction. As defined by the open source project, *"Knowledge is open if anyone is free to access, use, modify, and share it"* (Open Definition, 2015). There are several additional factors to be considered in the case of open data: the specific terms and conditions under which data is released, and whether the openly accessible data can be reused, reworked, redistributed, resold. In many cases, organisations may keep the primary access and the data itself but release metadata and allow redistribution only if value was added to the data (Kitchin, 2014, p. 50). The US government[4], the US Census Bureau[5], the CIA[6], the EU[7] all have their open data portals. As for online big data-based platforms relevant to this research, Google has its own Public Data Explorer[8], in addition to Google Trends[9], Google Finance[10] and Google Books Ngrams[11]. Facebook Graph[12] shows general trends on Facebook users' profiles and activities. However, from the perspective of this research, these open data sources do

[4] http://www.data.gov

[5] http://www.census.gov/data.html

[6] https://www.cia.gov/library/publications/the-world-factbook/

[7] http://open-data.europa.eu/en/data/

[8] http://www.google.com/publicdata/directory

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

[10] https://www.google.com/finance

[11] https://books.google.com/ngrams

[12] https://developers.facebook.com/docs/graph-api

not contain data that could contribute to real-time analysis of the expression of extremist thoughts, hate speech or violence. Twitter, which can be considered an open mass blog with wide accessibility to short and well-researchable texts may seem to be a usable source with a great deal of open data and real-time information. However, even though Twitter shows trending expressions and basic information on real-time trends, the external or regular user access to Twitter does not provide an opportunity for data scientists to sort, classify and evaluate data (Burgess & Bruns, 2015, p. 7). Data scientists must be able to organise their mass data and run queries in them, hence big data technologies cannot be applied on publicly available tweets if there is no access to data on a large scale. Therefore, Twitter is largely useless for nuanced real-time conflict prediction and prevention in a scientific way without in-house access: the feasibility of Twitter research without internal access to information is impossible due to the inability to access, filter, query information efficiently. In essence, corporations' data monopoly places the scientific community in a difficult situation where valuable data cannot be accessed for the conflict prevention purposes that are the focus of this study (Halavais, p. 590).

This study argues that while the open source data projects of government entities and corporations are relevant additions to the scientific community and humanity, they cannot be used for real-time analytics and therefore are not relevant for conflict and violence prevention purposes. Thus, focus must be shifted to data inaccessible for the public or the scientific community. Telecommunications and social media platforms collect data for their ongoing operations and business purposes, which is facilitated by the terms & conditions agreements which users click at without having read them in the majority of cases. These corporations have a vast in-house research team that analyse how customers use the products, services and platforms offered by the companies. However, the research done by corporations is in most cases "research that fails to support the researched" (Halavais, p. 589). In other words, the in-house analytical capabilities of corporations are only used for commercial interests, even though they could also be harnessed for the creation of peace and social good in general. Moving on, the interests of different actors will be examined in the equation.

The interests of actors in data sharing

First, the role and interests of states is twofold in the big data question: on the one hand, strong data privacy laws are important for citizens and therefore they should also be important for the state. On

the other hand, the will to create a data-driven economy is gaining momentum: this dichotomy is well-demonstrated by the EU's data-related policies. Regulation on processing personal data was first stipulated in the EU's 1995 directive (95/46/EC), and a General Data Protection Regulation (GDPR) is expected to supersede this document in 2015. On the one hand, strict data privacy rules are of pivotal interest to citizens, and good governance should include a respect for this need. On the other hand, free flow of data and a stress on value creation is increasingly present: in the EU's most recent piece of communication on the data driven economy, the first priority of the EU is the availability of good quality, reliable and interoperable datasets and an enabling infrastructure to value creation from big data (European Commission, COM(2014)442, p. 5). In addition, it is highlighted that in the Digital Single Market no inappropriate restrictions should hinder data flow across sectors, borders and languages (ibid. p. 6). In practice, the EU aims at facilitating the unhindered flow of data by public-private cooperations between universities/public research institutes and private partners and by promoting cooperation on R&D through facilitated access and transfer of knowledge and technology (ibid, p. 6). As for the developing world, there is often a lack of transparency and concerted action concerning data privacy, as also demonstrated by the case study in the previous section.

Second, private interests describe the stakes of corporations that collect, possess and analyse a great deal of data that users produce while using their services. Multinational corporations leverage big data to optimise the allocation of their resources, reduce waste, improve transparency and accountability, and facilitate research, development and innovation (Manyika et al., 2011, p. 97). In addition, there are several examples where these corporations share with law enforcement bodies their data on "dangerous individuals" who engage in suspicious activities (Schmidt & Cohen, 2014, pp. 268-269). Google, Facebook and Twitter all share substantial data as open data that can be used for research, however, it remains a question how real-time data could be shared with the scientific community and relevant government entities to facilitate research not only on past events, but also predictions about the future with big data.

The interests of corporations in sharing or not sharing data are threefold: their main interest is commercial and aims at achieving maximum profit. The monetisation of data analytics is an emerging phenomenon which includes the creation of profit from scientific data analytics based on social media data. This is a vital commercial growth area for corporations and social media platforms,

which is why it is key for these entities to keep their customers satisfied — they rely on their data inputs as it has become the commodity of their business (Burgess & Bruns, 2015, p. 7). As such, it is expected from corporations to ensure that the rights and wishes of their customers are taken into account in every business decision. However, companies have a third interest which is related to corporate social responsibility (CSR) and the creation of a positive image: CSR activities are increasingly popular with telecommunication companies who therefore perform "data philanthropy," i.e. the sharing of data for purposes that contribute to the creation of social good (Kirkpatrick, 2014, p. 5). In short, the primary interest of corporations is commercial and their behaviour is influenced by privacy regulations, but CSR and data philanthropy are of growing importance in their behaviour.

Third, the role of ideas denotes the interest of humanity as well as the agency of international organisations and NGOs. There are two key ideas to consider here: the ideal of strong and effective privacy regulations and the quest for security. The privacy dilemma is mainly a concern for re-active data activists, whereas pro-active data activists aim for the creation of more flexible data privacy regulations that allow for the creation of social good and increase security in society. As asserted by Kirkpatrick, a "new *social contract*" must be established around big data: citizens must be assured that big data can be used to improve services, to hold governments accountable and to create social good and not to do harm, to create corporate tyranny or a surveillance state. Therefore, civil society has to contribute to the shaping of a future where there are human-centered regulatory frameworks, standards, guidelines, so that big data can be used as a tool to empower individuals and communities with maximum value and minimum risk (Kirkpatrick, Big Data for Humanity, 2014, 48:10-49:45). As there is a controversy on big data in civil society at the moment, innovation in technology and business, and a practice of good governance needs to be established (Hoffmann, Big Data for Humanity, 2014,46:10-48:00). However, a prerequisite to this is a civil society that supports pro-active data activism and is open to the ideas of data-driven development, so that both governments and corporations see the benefits of their eventual data sharing programmes for the creation of social good. As an example to this approach, the flagship big data-related project of the United Nations, the UN Global Pulse brings together corporations from the private sector that are willing to commit their resources to "data philanthropy," as well as industry leaders, universities, research institutes, and non-profit organisations. One of the key aims of the UN Global Pulse is to find ways of making use of big data for global development without analysing private or confiden-

tial information. Accordingly, the UN Global Pulse advocates strongly for strong privacy protection regulations, policies and techniques (Kirkpatrick, 2014, pp. 5-6). The question of the privacy vs. safety dichotomy becomes clear in the mission statement of the UN Global Pulse where it underlines value creation without the collision with data privacy. The dichotomy refers to the fact that from individuals' perspective, the dilemma between security and privacy has been increasing (O'Neill, 2001).

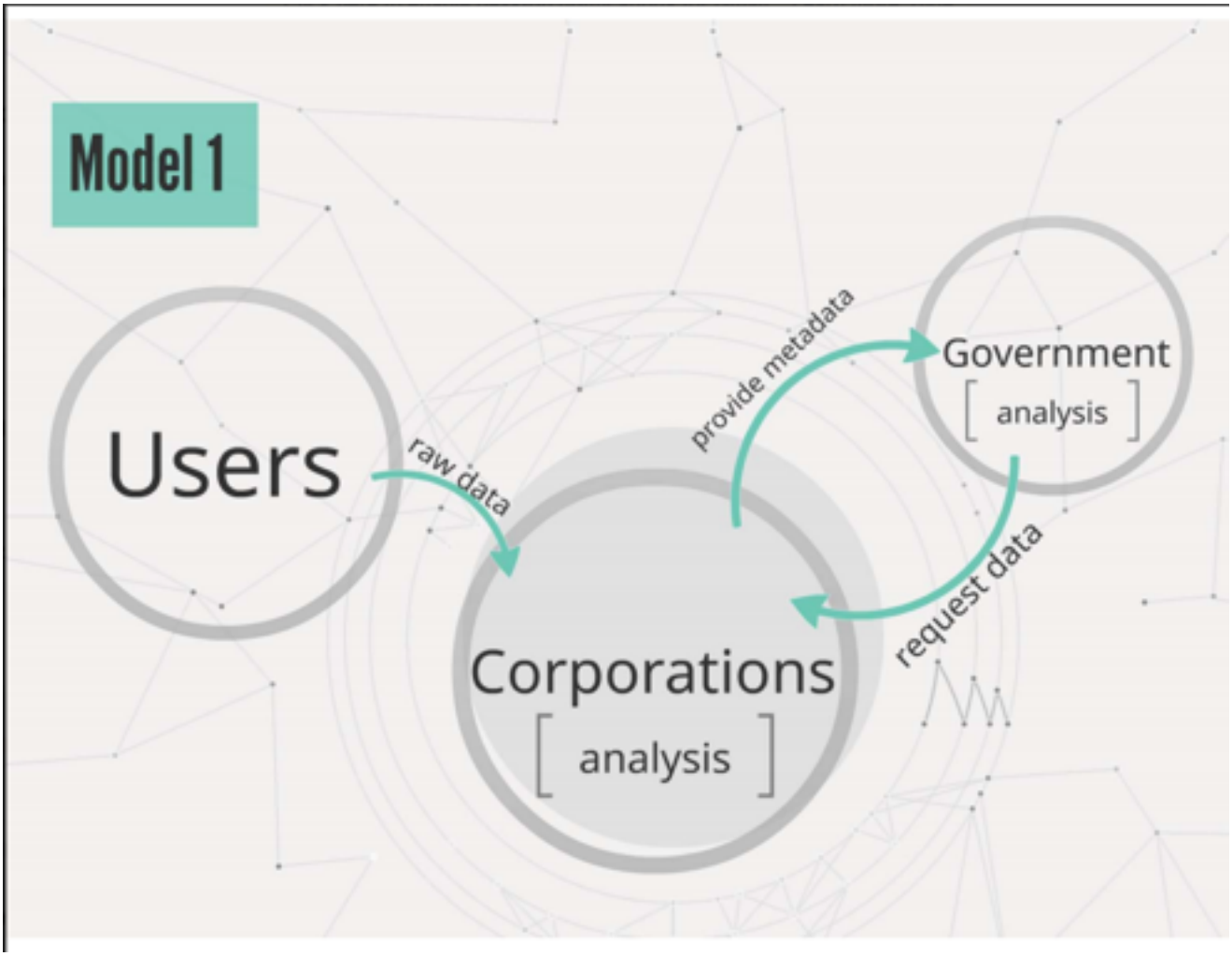
Data access-related mechanism

The above paragraphs have given a brief overview on the interests of the stakeholders in sharing or not sharing data. To understand issues pertaining to data access, the United States' legislation on data privacy and access regulations can serve as subject of analysis. Even though lone-wolf terrorism is out of the scope of this research, it is informative to give a quick overview on the two competing mechanisms of the United States government in using big data to identify possible terrorist threats in her territory. This approach will describe two existing mechanisms and establish a third one in order to clarify the privacy vs. security dichotomy. To visualise the questions, three models will be used: the first one is based on the USA Patriot Act, which can be considered as an initial reaction to the abundance of data; the second one is based on the USA Freedom Act, which can be interpreted as a data-reactionist approach to big data; the third model is the visualisation of a data-activist concept established here as a potential mechanism to use big data for conflict prediction and prevention related purposes by sharing real-time filtered data with additional actors. This hypothetical model includes an alert mechanism that is based on the assumptions that sharing real time filtered data does not collide with privacy regulations and can contribute to efficient conflict prevention.

The first model visualises how raw data produced by users of social media platforms, Internet service providers and telecommunications services was transmitted to corporations under the USA Patriot[13] Act passed in 2001. This established strengthened security controls which included enhanced surveillance procedures and provided the government with an authority to intercept wire, oral, and electronic communications to prevent criminal acts (Library of Congress, 2001, sec. 201-203). The

[13] short for "Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism"

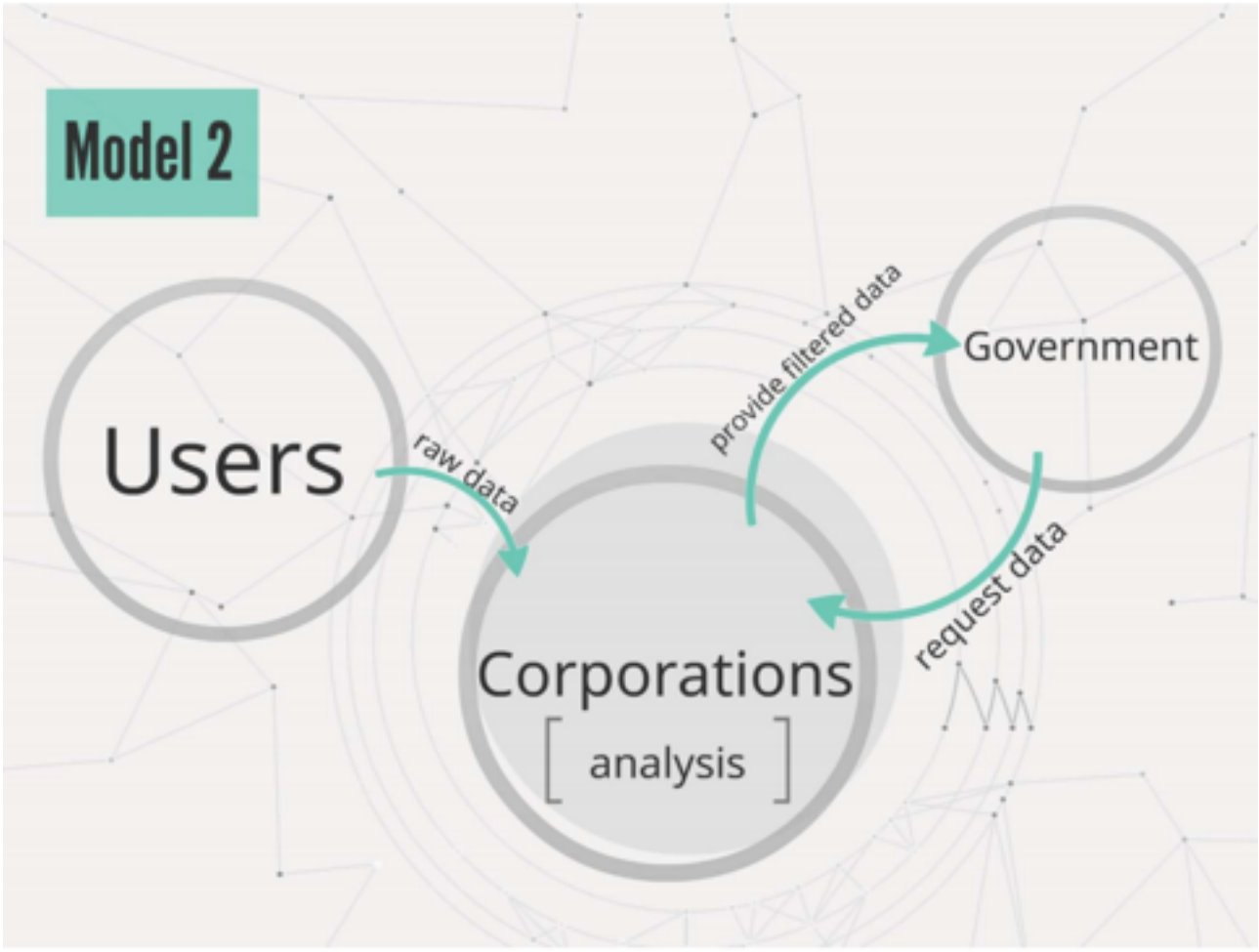
data was forwarded to government agencies upon their request in the format of metadata, which provided a broad overview of data for governments, as well as unfettered access to a great deal of data that did not pertain to individuals that have displayed suspicious behaviour. Even though companies had the analytical capabilities and may already analyse their data in-house, analysis was conducted by governments as well (see Model 1, based on US Congress, 2001).



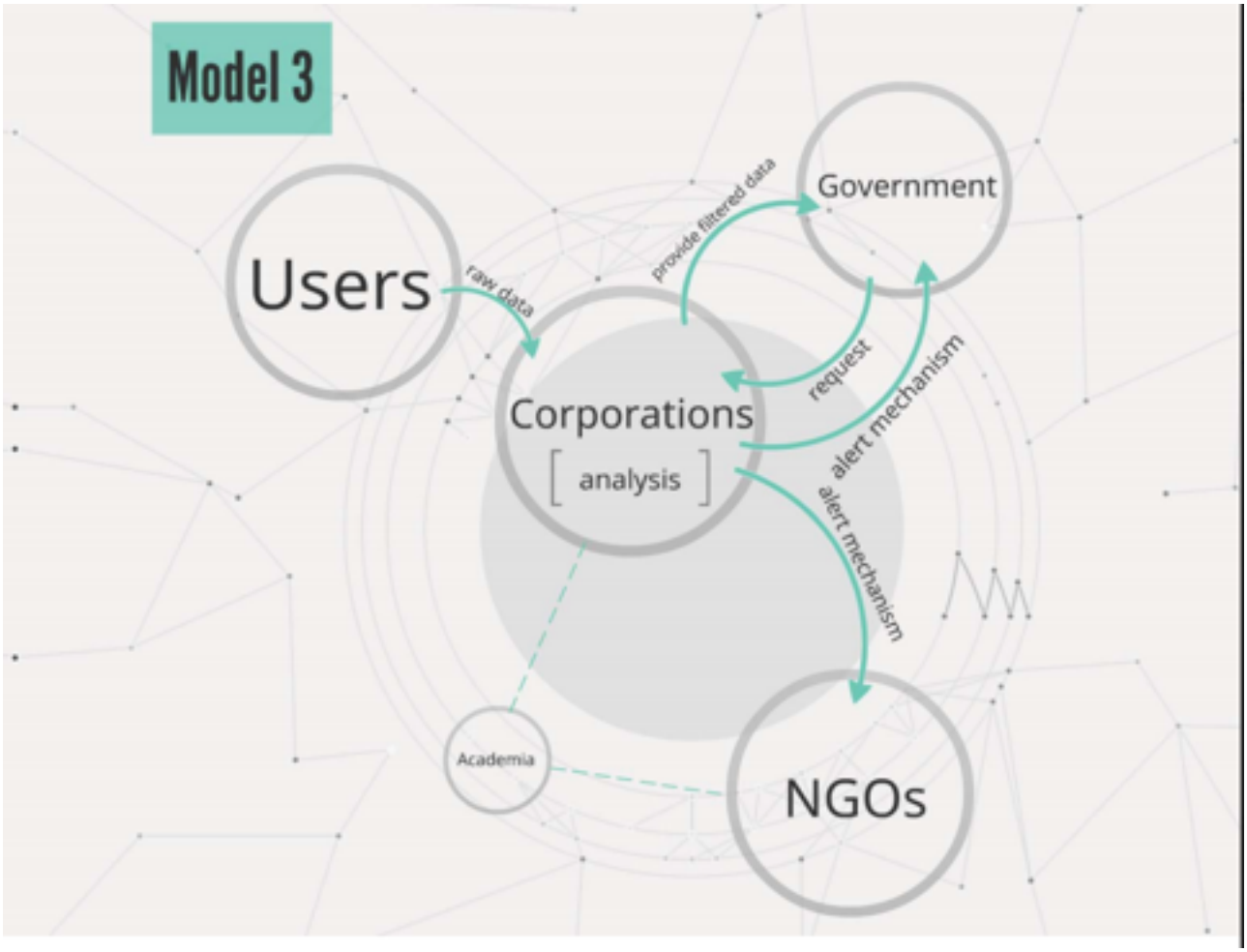
In contrast, the second model based on the USA Freedom Act[14] displays a more nuanced way of data regulation. Big data as a tool to monitor citizens' behaviour came under greater scrutiny after the revelations on the United States National Security Agency's excessive collection of metadata (see Pruner, 2014, McGowan, 2014). During the enforcement of the Patriot Act, the collected meta-

[14] short for Uniting and Strengthening America by Fulfilling Rights and Ending Eavesdropping, Dragnet-collection and Online Monitoring

data was used to monitor and analyse communication data of citizens for surveillance purposes. Government agencies were able to access and analyse communications of ordinary citizens, a practice the USA Freedom Act aims at changing: this piece of legislation gives the authority to corporations to analyse their own data instead of providing unfiltered data for government agencies to do so. As companies already possess data analysing capabilities and users have given their consent to their data storing and handling operations by agreeing to their Terms & Conditions, it makes more sense to restrict data analysing activities to companies only. Once raw data is analysed, companies may release the filtered data with governments as opposed to giving access to entire datasets in the form of metadata. This way data of law-abiding citizens is not shared with government authorities (see Model 2, based on US Congress, 2015).



The third model builds further on the existing data analytics capabilities of corporations and involves further parties that can contribute to conflict prevention. It is nearly identical to the Freedom Act with an additional alert mechanism that would allow corporations to take the initiative and inform relevant government agencies or NGOs to trigger action. As companies' data analytics capabilities are already available and in use due to regulations in the Freedom Act, it would be advised to consider using these capabilities for preventive purposes. Hence, if corporations identify expressions of violence in a significant magnitude, a real-time alert scheme between corporations and governments may resolve access related hindrances in the application of big data in conflict prevention. In addition, NGOs are included in this model as the cooperation between corporations, governments and NGOs could result in more efficient action. Finally, the inclusion of academia could deliver additional results — allowing data scientists and peace experts affiliated to entities such as the Stanford Peace Lab or Leiden University's Peace Informatics Lab to use real-time data could contribute to projects of higher impact (see Model 3).



Currently, the majority of national legislations restrict telecommunications companies from accessing or storing data on personal interactions — these laws are not futile, but it must be contemplated to what extent privacy and safety can be integrated and harmonised in future privacy regulations. In order to ensure that corporations can use data analytics to create social good and to prevent violent conflicts arising from extremist views, more powers would have to be given to telecommunications companies and social media platforms. However, this compromise may directly exacerbate the increasing fragmentation of the current international order and the weakening of the influence of the state. The feasibility of this "new social contract" and the allocation of more rights to corporations to store, access and analyse data would require a strong proactive data activist approach from citizens. It remains a question to what extent actors of global governance, the UN or the EU can influence this kind of evolution in the acceptance of big data based analytics. Even though projects such as the UN Global Pulse aim at painting a positive image on big data, both organisations are rather cautious about its more widespread use for research purposes.

Conclusion

To conclude, using big data for prediction and prevention of violence is a novel way of approaching early intervention and conflict. While there are already examples to big data based conflict prevention initiatives, such as that of the Kenyan elections in 2013, the lack of a sufficient number of samples does not allow for a clear assessment of the added value of these projects and the relevance of this method. In general, it appears to be difficult to detect causality between conflict prevention through the dissemination of messages of peace and the actual creation of peace. As for the potential of big data based conflict prediction and prevention, there is clearly a great deal of information to gain from large datasets that can provide useful insights on social dynamics. However, the potential can only be turned into reality if a sustainable and real-time data-sharing method is established between corporations, governments and NGOs. Once a proactive data-activist approach becomes more widespread in society, the potential of big data in preventive and predictive research may increase significantly. Consequently, a prerequisite to successful big data-based peace research and conflict prediction is a solution to the debate on privacy vs. security. Risks of cross-sector data sharing must be minimised in order to ensure real-time prediction capabilities for corporations as well as access for NGOs and governments to the relevant data.

Concerning the state of the art of big data based projects in social science, there are only a handful of examples that meet all criteria of the types of projects that are in the focus of this study (big data based, real time, involving cross-sector cooperation). As technical capabilities are already available from corporations' side, it is a sustainable mechanism, proper regulations and a more widespread acceptance for cross-sector data sharing that are missing for the proactive contribution of companies and more successful projects. Closely related to this, it is essential to underline that open source data is largely futile for real time analysis and conflict prediction, hence it is vital for relevant stakeholders to be able to access anonymised data sources. By visualising three data-sharing models, this study aimed at highlighting that sharing raw data or metadata with governments or NGOs is not a *sine qua non* of big data based peace projects: as shown by Model 3, using filtered data and establishing a proactive alert mechanism by companies may allow for value creation from data without risking non-compliance with privacy regulations. The balance between necessary regulation and productive use must be found by the international community and the civil society.

Bibliography

Bergman, Ann, Jan Ch Karlsson, and Jonas Axelsson. "Truth Claims and Explanatory Claims—An Ontological Typology of Futures Studies." *Futures* 42.8 (2010): 857-65. Web.

Boyd, Danah, and Kate Crawford. "Critical Questions For Big Data." *Information, Communication & Society* 15.5 (2012): 662-79. Web.

Burgess, Jean & Bruns, Axel. "Easy data, hard data : the politics and pragmatics of Twitter research after the computational turn." In *Langlois, Ganaele, Redden, Joanna, & Elmer, Greg (Eds.) Compromised Data : From Social Media to Big Data.* Bloomsbury Publishing, London, 2015. In Press, downloaded from Web <http://eprints.qut.edu.au/79308/> Accessed on 2 June 2015.

Ch'Ng, Eugene. "The Value of Using Big Data Technologies in Computational Social Science." (2014). *arxiv.org* Web.

Economist, The. *Big Data and Ebola: Call for Help.* 25 Oct. 2014. Web. 28 Oct. 2014.

European Commission. “COM(2014)442: Towards a Data-driven Economy.” Web. 2 July 2014. <http://ec.europa.eu/digital-agenda/en/news/communication-data-driven-economy>. Accessed on 11 May 2015.

European Commission. "European Commission and Data Industry Launch €2.5 Billion Partnership to Master Big Data." *EU Press Release Database*. Web. 13 Oct. 2014. <http://europa.eu/rapid/press-release_IP-14-1129_en.htm> Accessed on 11 May 2015.

European Parliament and Council of the European Union. "Directive 95/46/EC of the European Parliament and of the Council". *Official Journal L 281* (1995). Web. 1 Apr. 2015. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML>. Accessed on 11 May 2015.

Faghmous, James H., and Vipin Kumar. "A Big Data Guide to Understanding Climate Change: The Case for Theory-Guided Data Science." *Big Data* 2.3 (2014): 155-63. Web.

Fuchs, Christian, and Daniel Trottier. "Towards a Theoretical Model of Social Media Surveillance in Contemporary Society." *Communications* 40.1 (2015). Web.

Furnham, Adrian. "Response Bias, Social Desirability and Dissimulation." *Personality and Individual Differences* 7.3 (1986): 385-400. Web.

Gijzen, Hubert. "Development: Big Data for a Sustainable Future." *Nature* 502.7469 (2013): 38. Web.

Halavais, Alexander. "Bigger Sociological Imaginations: Framing Big Social Data Theory and Methods." *Information, Communication & Society* 18.5 (2015): 583-94. Web.

Herbert, Anne L. "Cooperation in International Relations: A Comparison of Keohane, Haas and Franck." *Berkeley Journal of International Law* 14.1 (1996): 222-38. Web.

“High-Level Panel Session: Big Data for Humanity” Robert Kirkpatrick, William Hoffmann, Jaap Van Der Henrik, Caroline Kroon, and Mark Nelson. *Leiden University Online Learning Lab*, 21

Aug. 2014. Web. 29 Oct. 2014. <https://www.youtube.com/watch?v=yYeuLHtQF6Q>. Accessed on 27 March 2015.

Himelfarb, Sheldon. "Can Big Data Stop Wars Before They Happen?" *Foreign Policy* (2014). Web. 25 Apr. 2014. <http://foreignpolicy.com/2014/04/25/can-big-data-stop-wars-before-they-happen/>. Accessed on 30 April 2015.

Himelfarb, Sheldon. "The Quiet Revolution." *Foreign Policy* (2013). Web. 3 Oct. 2013. <http://foreignpolicy.com/2013/10/03/the-quiet-revolution/>. Accessed on 30 April 2015.

Kahneman, Daniel. *Thinking, Fast and Slow.* London: Penguin, 2012. Print.

Kent, Michael L., Bryan J. Carr, Rebekah A. Husted, and Rebeca A. Pop. "Learning Web Analytics: A Tool for Strategic Communication." *Public Relations Review* 37.5 (2011): 536-43. Web.

Keohane, Robert O. *After Hegemony: Cooperation and Discord in the World Political Economy.* Princeton, NJ: Princeton UP, 2005. Print.

Keohane, R.O., and Joseph S. Nye. Power and Interdependence in the Information Age. *Foreign Affairs* 77.5 (1998). Web. 1 Apr 2015. <http://www.foreignaffairs.com/articles/54395/robert-o-keohane-and-joseph-s-nye-jr/power-and-interdependence-in-the-information-age>

Kirkpatrick, Robert. "A Conversation with Robert Kirkpatrick, Director of United Nations Global Pulse." *SAIS Review of International Affairs* 34.1 (2014): 3-8. Web. <https://muse-jhu-edu.ezproxy.leidenuniv.nl:2443/journals/sais_review/v034/34.1.kirkpatrick.pdf> Accessed on 2 May 2015.

Kitchin, Rob. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences.* Sage Publications, 2014. Web.

Khan, M. Ali Uddin, Muhammad Fahim Uddin, and Navarun Gupta. "Seven V's of Big Data - Understanding Big Data to Extract Value." *Conference of the American Society for Engineering Edu-*

cation (2014). Web. 1 Mar. 2015. <http://asee-ne.org/proceedings/2014/Professional%20Papers/113.pdf>. Accessed on 22 March 2015.

Laney, Doug. "3D Data Management: Controlling Data Volume, Velocity, and Variety." *Application Delivery Strategies* (2001): *MetaGroup*. Web. 1 Mar. 2015. <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>. Accessed on 22 March 2015.

The Library of Congress "USA FREEDOM Act." Web. 2 June 2014. <http://www.gpo.gov/fdsys/pkg/BILLS-113hr3361rfs/pdf/BILLS-113hr3361rfs.pdf>. Accessed on 28 May 2015.

The Library of Congress "USA PATRIOT Act." Web. 26 October 2001. <https://www.congress.gov/107/plaws/publ56/PLAW-107publ56.pdf>. Accessed on 28 May 2015.

Mancini, Francesco, and Mary O'Reilly. "New Technology and the Prevention of Violence and Conflict." *Stability: International Journal of Security & Development* 2.3 (2013): 1-9. Web. <http://dx.doi.org/10.5334/sta.cp>. Accessed on 2 May 2015.

Mans, Ulrich. "Peace Deals Revisited." *Leiden University Peace Informatics Lab* (2014). Web. <http://issuu.com/ulrichmans/docs/20140728_peaceinformatics__weebly_/0>. Accessed on 20 May 2015.

Manyika, James, Michael Chui, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. "Big Data: The next Frontier for Innovation, Competition, and Productivity." *McKinsey Global Institute* (2011): *mckinsey.com*. Web. <http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation>. Accessed on 29 October 2014.

Marx, Vivien. "Biology: The Big Challenges of Big Data." *Nature* 498.7453 (2013): 255-60. Web.

Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think.* Boston: Houghton Mifflin Harcourt, 2013. Print.

McGowan, Casey J. "The Relevance of Relevance: Section 125 of the USA Patriot Act and the NSA Metadata Collection Program." *Fordham Law Review* 82.5 (2014 April): 2399-2442. Web.

Mcguire, Steven. "Multinationals and NGOs amid a Changing Balance of Power." *International Affairs* 89.3 (2013): 695-710. Web.

Meier, Patrick. Digital Humanitarians: *How Big Data Is Changing the Face of Humanitarian Response.* Boca Raton, FL: Taylor & Francis Group, 2015. Print.

Milan, Stefania. "Data Activism: The Politics of Big Data." Web. 2015. <http://stefaniamilan.net/DATACTIVE>. Accessed on 31 May 2015.

Mises, Ludwig Von. *Theory and History; an Interpretation of Social and Economic Evolution.* New Haven: Yale UP, 1957. Print.

O'Neil, Michael. "Cyber Crime Dilemma: Is Possible to Guarantee Both Security and Privacy?" *The Brookings Review* 19.1 (2001): 28-32. Web.

Niiniluoto, Ilkka. "Futures Studies: Science or Art?" *Futures* 33.5 (2001): 371-377. Web.

Open Definition. Web. <http://opendefinition.org/od/> Accessed on 22 May 2015.

Ovodenko, Alexander, and Robert O. Keohane. "Institutional Diffusion in International Environmental Affairs." *International Affairs* 88.3 (2012): 523-41. Web.

Pearlman, W., and K. G. Cunningham. "Nonstate Actors, Fragmentation, and Conflict Processes." *Journal of Conflict Resolution* 56.1 (2011): 3-15. Web.

Piketty, T. 2014. *Capital in the Twenty-First Century.* London: Harvard University Press. Print.

Pruner, Clayton J. "No Service: An Insight into the Statutory Shortcomings of the NSA's Metadata Collection Program." *Journal of Legislation*, 40.2 (2014.Winter): 374-98. Web.

Rudder, Christian. *Dataclysm: Who We Are When We Think No One's Looking*. New York: Crown, 2014. Print.

Russell, Matthew A. *Mining the Social Web*. Sebastopol, CA: O'Reilly, 2011. Print.

Savage, Mike, and Roger Burrows. "The Coming Crisis of Empirical Sociology." *Sociology* 41.5 (2007): 885-99. Web.

Schmidt, Eric, and Jared Cohen. *The New Digital Age: Transforming Nations, Businesses, and Our Lives*. New York: Vintage, 2014. Print.

Smith, Mark J. *Social Science in Question: Towards a Postdisciplinary Framework.* London: Sage Publications, 1998. Print.

United Nations. "United Nations Secretary-General Appoints Independent Expert Advisory Group on the Data Revolution for Sustainable Development". 29 Aug. 2014. Web. <http://www.un.org/millenniumgoals/pdf/Press%20Release_Announcement_Secretary%20General's%20Independent%20Expert%20Advisory%20Group%20on%20Data%20Revolution.pdf> Accessed on 24 October 2014.

Verjee, Neelam. "Will Kenya's Elections Transform the Text Message from Deadly Weapon to Peace Offering?" *Quartz* (2013). Web. 2 Mar. 2013. <http://qz.com/58510/will-kenyas-elections-transform-the-text-message-from-deadly-weapon-to-peace-offering/>. Accessed on 2 May 2015.

Vesset, Dan, Henry D. Morris, and Gard Little. *Worldwide Big Data Technology and Services 2012-2015 Forecast.* Framingham, MA: IDC Report, 2012. Web. 1 Apr 2015. <http://ec.europa.eu/digital-agenda/en/news/worldwide-big-data-technology-and-services-2012-2015-forecast> Accessed on 24 October 2014.

Weinberger, Sharon. "Social Science: Web of War." *Nature* 471.7340 (2011): 566-68. Web. <http://www.nature.com.ezproxy.leidenuniv.nl:2048/news/2011/110330/full/471566a.html>. Accessed on 15 May 2015.

Wilcox, Larry D. "Futurology and the Social Sciences — Bloom and Boom or Gloom and Doom?" *International Social Science Review* 58.4 (1983): 202-10. Web.