Big data: Issues for an international political sociology of the datafication of worlds[1]

Madsen, A.K, Flyverbom, M., Hilbert, M. and Ruppert, E.

In *International Political Sociology* (forthcoming 2016)

Abstract

The claim that big data can revolutionize strategy and governance in the context of international relations is increasingly hard to ignore. Scholars of international political sociology have mainly discussed this development through the themes of security and surveillance. The aim of this paper is to outline a research agenda that can be used to raise a broader set of sociological and practice-oriented questions about the increasing datafication of international relations and politics. First, it proposes a way of conceptualizing big data that is broad enough to open fruitful investigations into the emerging use of big data in these contexts. This conceptualization includes the identification of three moments contained in any big data practice. Secondly, it suggests a research agenda built around a set of sub-themes that each deserve dedicated scrutiny when studying the interplay between big data and international relations along these moments. Through a combination of these moments and sub-themes, the paper suggests a roadmap for an international political sociology of the datafication of worlds.

INTRODUCTION

The world's technological capacity to store information grew from 2.6 (optimally compressed) exabytes in 1986, over 55 in 2000, to 300 in 2007 (Hilbert and Lopez, 2011). Extrapolating these tendencies with complementary sources (e.g. Gantz and Reinsel, 2012; Turner et al., 2014), we estimate them to have reached 5,000 exabytes by 2014 (or 5 optimally compressed zettabytes). This is equivalent to some 4,500 piles of double printed books of 125 pages from planet Earth to the sun. The world's technological capacity to compute information has grown even 2-3 times faster than our capacity to store information (Hilbert, 2014a). Our computers are so powerful that if all 2.5 million students of the United Kingdom were to calculate from the big bang until now (without break), they could merely execute half as many basic calculations as computers can execute in only one second (Hilbert, 2014b). This situation provides both massive input (storage of events) and powerful computational tools (analysis of events) to derive intelligence for informed decision making, ranging from the everyday conduct of ordinary citizens to transnational governance.

In our daily lives, most of us are already experiencing how digital traces and algorithms shape how we think and act. Results from Google searches decide what we see and do not see (Madsen, 2016), Facebook aims to control our mood by tweaking the content that shows up in our daily feed (Booth, 2014), and the prices we are offered in online shops depend on our location, device and most recent online activities (Hannak et al.

[1] Accepted Manuscript version March 2016.

2013; Hannak et al. 2014). Looking beyond such everyday situations, it is evident that data and algorithmic operations are increasingly important in relation to political practices such as urban governance (Kitchin, 2014), anti-corruption (Hansen & Flyverbom, 2014), international mobility (Mocanu et al. 2013), and international development (Hilbert, 2013).

In short, big data is increasingly used to monitor, know, and govern populations. For this purpose analysts have typically utilized one of three sources of data. First, *cell phone logs* have been used to track real-time movements of humans, such as patterns of migration in Rwanda (Blumenstock 2012). Secondly, *internet searches* have been used to predict peoples wellbeing. For instance, Google searches have been used to predict dengue infections in Singapore and Thailand (Althouse and Ng, 2011). Thirdly, traces on various types of *social media platforms* have been used to track humans and predict their behavior. Geo-tagged pictures of Flickr have, for instance, been used to monitor human movements in situations of disease outbreaks (De Choudhury et al., 2010) and the sentiment of tweets have been used to predict economic crises (Global Pulse, 2011).

What to make of such big data practices? How to understand sociologically and politically their significance? An initial look at how the field of international political sociology (IPS) has treated this topic suggests that the interesting questions concern privacy, surveillance and security (Stevens 2015; Bauman et al. 2014 and Amoore 2014). More particularly, it seems that the Snowden leaks and the PRISM programme have been the most important important points of reference when theorizing big data. The literature has to a large extent focused on how intelligence agencies and the state are beginning to reassert sovereign power over new types of data and how they use this power to profile people as well as trace potential terrorists. This frames data-driven practices of governance as being of theoretical interest because they raise questions about, for instance, the power-relation between IT and telecompanies and the state.

While acknowledging the importance of this line of work, we want to argue that big data raises questions for IPS that go far beyond this focus on surveillance and security. We want to introduce a set of questions and focal points that make it possible for IPS to engage with big data as a wider political and social phenomenon. As a supplement to the issues already discussed in the literature, we suggest a need to focus on issues such as the formation of political subjectivities, the valuation of data and the epistemological and organizational turmoils accompanying new data practices.

The paper will take two steps in supporting a research agenda along these lines. First, Part I will propose a way of defining big data that unpacks it as a heterogenous phenomenon. For that purpose, it introduces 'three moments' of big data practices. The first concerns the datafication of daily life, the second concerns the production of patterns and predictions and the third concerns the making of new modes of governance. Thinking along these moments lays a foundation for asking a broad set of questions about data-driven, algorithmic forms of governance.

Secondly, Part II will use the suggested definition as a stepping-stone for deeper discussions of four sub-themes that we believe to be especially worthy of scholarly investigation. In sub-theme 1, Mikkel Flyverbom ponders the changed modes of conduct in an emerging datafication-governance nexus. In sub-theme 2, Evelyn Ruppert argues for the relevance of inquiring deeper into the political subjectivities of big data practices. In sub-theme 3, Anders Koed Madsen focuses on the dynamics of data

valuation involved in big data practices and Martin Hilbert ends the paper with reflections on the limits of big data and the importance of theory in knowledge production in sub-theme 4.

When combined, we believe that these moments and sub-themes can lay the foundation for a roadmap for future research on big data within the field of international political sociology. We have summaried some of the most important elements in this road map in the table below. It suggests relevant questions to ask in relation to each moment and it lists references to the most central works in the theoretical lineage that have motivated these questions. Finally, it lists cases in the context of big data and governance where this specific combination of research questions and theory has proved to be especially relevant. The rest of this paper will provide the background of the suggested roadmap.

	Sub-questions broken into moments (w/ central concepts highlighted)			**Theoretical lineage**	**Illustrative cases**
	Moment 1	**Moment 2**	**Moment 3**		
Sub-theme 1: **How is conduct shaped in the datafication-governance nexus?**	Through which **material infrastructures** are data sources produced and shaped?	Which **patterns of inclusion** and exclusion are at play when data is made 'algorithm-ready'?	To what extent does new data come with **calculative rationalities** that unsettles established ways of doing governance? What is the relation between **algorithmic transparency** and control in datafied governance?	Governmentality literature: Foucault (1983) Miller & Rose (1990) Dean (1999) Scott (1998)	Profiling of terrorists Google Flu Trends
Sub-theme 2: **Who are the political subjects of Big Data?**	How do platforms configure **everyday conduct** but at the same time create possibilities for digital citizens to act?	How do algorithms ignore the myriad acts of subjects who do not simply obey or submit but subvert their workings?	How do subjects make **rights claims** and become **digital citizens** in and by not only what they do through the Internet but what they say?	Power and subjectivity: Foucault (1983) Balibar (1991) Austin(1962) Franklin (2013)	Acts of Snowden and Manning Daily acts of blocking and filtering, encrypting communications, creating multiple and anonymised and shared identities,

	If what subjects are saying and doing through the Internet is changing **political subjectivity** then what does this mean for the data that is generated?				deploying bots, gaming trending algorithms, etc.
Sub-theme 3: **How are dynamics of valuation influencing Big Data projects?**	How is the **exchange value** of data established when organizations and companies form partnerships?		How is the **epistemic value** of new data settled in organizations that are in the process of changing practices of decision-making?	Valuation Theory: Muniesa et al. (2007) Espeland & Stevens (1998) Stark (2011)	Copenhagen City Data Market UN Global Pulse
Sub-theme 4: **How is theory and history imposing limits on the power of Big Data?**	Is it useful and possible to create big data beyond the empirical digital footprint of big data? Which kind?	Which types of big data projects assume stationarity and therefore are deceptive or limit political flexibility for future change?	How should traditional (empiricial) big data analysis be combined with theoretical envisioning of changing histories to be politically relevant for new modes of governance?	Critiques of econometrics: Lucas (1976) Goodhart (1976) Campbell (1976)	UN Environment programme City of Portland simulation

Table 1: A roadmap for future research on big data in international political sociology

Part I: Big data, a heterogeneous phenomenon.

Big data is a phenomenon that has been defined and conceptualized in various ways during the last five years (De Mauro et al. 2014). There is no consensus about what demarcates big data from other types of data, but a widely used definition has been Douglas Laney's '3 v's':

> Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization (Laney, 2012)

According to Laney's definition, big data practices are characterized by the technical attributes of the data they involve. Data-points are numerous and they come from

various sources in a pace that is sometimes referred to as real-time. This definition also entails that the problems raised by big data practices have technical solutions. For Laney, the pressing question is to find the right processing techniques to handle data with these new characteristics.

Writers such as Victor Meyer-Schönberger and Kenneth Cukier have challenged these largely technical definitions. Instead of focusing on the nature of data, they define big data with a focus on the new dynamics of knowledge production and valuation that are changed with the introduction of new data practices:

> Big Data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments and more (Mayer-Schönberger & Cukier 2013)

In line with our introductory comments above, we agree that the role of big data in transnational governance is not a purely technical phenomenon. It must be approached as a heterogeneous phenomenon that can be studied by attending to the interplay between social and technical relations in the way suggested by Meyer-Schönberger and Cukier. However, we want to add that this interplay can be fruitfully studied by attending to the following three moments of big data practices, 1) The datafication of daily life, 2) the production of patterns and predictions and 3) the making of new modes of governance.

1st moment: The datafication of daily life

A condition for any data practice is the existence of a dataset to be analysed. Data need to be produced in one way or another. The production of big data varies from more traditional data practices in the sense that it is tightly connected to the way people act. Surveys, focus groups and other traditional data-collection methods deliberately construct a specific situation in which relevant data is produced. This is not the case with big data where the data sources are telephone logs, web searches and social media feeds.

It is essentially a digital footprint almost inevitabley left behind by every digital step taken. The same as with footprints, these datapoints are produced as byproducts and rapidly increase with the increased datafication of people's daily lives. This raises questions about the way interplays between social and technical relations at the moment of production come to set the direction for big data practices.

Table 1 suggests a list of questions that we believe to be important for IPS in this regard. These questions illustrate the need to look at the interrelation between platforms, practices and data when conducting inquiries into the first moment of big data practices. More specifically, they indicate that a central analytical challenge is to formulate an alternative to the popular conceptualization of data as a raw product with an intrinsic value. Data is always produced through the infrastructures and acts that shape it, and its further analytical use is often conditioned by the value ascribed to it by its owners.

2nd moment: The algorithmic production of patterns and predictions

The second moment concerns this further analytic use in the sense that it involves the production of patterns and predictions on the basis of available data. In relation to transnational governance this could for instance be the work of showing that specific

Google searches correlate with flu outbreaks and translating this pattern into a predictive model for flu detection (Ginsberg et al 2009). The basic idea always is that some pattern in the digital footprint relates to something interesting of the real world, which therefore can be understood, traced and predicted through the digital pattern.

This strategy of producing patterns and predictions through correlations is central to the second moment of big data practices. We want to argue that it raises important questions about the role of big data algorithms in knowledge-production. Big data algorithms are here undertood as generalized procedures for turning dis-organized data-inputs into manageable outputs through series of logical rules (Flyverbom and Madsen, 2015). The strategy of quickly finding correlations in big data sets entails that such procedures become increasingly central to developments in governance and politics. They begin to shape what we aspire to and seek solutions to in societies.

Some applaud this and focus on the potentials for more timely, precise and rational forms of governance (O'Reilly, 2013), while others fear that trusting blindly in data and technological fixes will mean the 'death of politics' (Morozov, 2013). No matter what position one takes on this normative issue, we argue that big data algorithms, as defined here, must be a central concern when theorizing the second moment of big data practices. Table 1 suggests that relevant questions to pose in relation to this second moment could focus on the connection between predictions and the empirically detected past or the relation between algorithms and (the reconfiguration of) politics.

3rd moment: The making of new modes of governance

Studying big data with a focus on algorithmic operations may be a useful starting point, but it is also limited in its grasp and conceptualization of the many social and technical relations that constitute it. It is insufficient to focus on the technical capacities and character of the sources involved. The reason being that big data practices are not autonomous, technological processes with predictable effects, or the "inevitable consequence of a technological juggernaut with a life of its own entirely outside the social" (Zuboff, 2015: 75).

We need to leave deterministic frameworks and reflect on the larger transformation that big data is part and parcel of. This includes understanding how the abovementioned patterns and predictions begin to influence new modes of governance. This is what we refer to as the third moment of big data practices and we suggest that this form of influence is conditioned upon a certain combination of technological features, practical uses and meanings ascribed to data in specific situations (Madsen, 2013). Any successful attempt to conceptualize the role of big data in transnational governance must ask questions about the complex, entangled nature of technological developments.

In relation to this third moment, Table 1 suggests a range of relevant questions to ask. How is conduct is shaped in the datafication-governance nexus? Does digital data blur traditional demarcation about what is foreign and what is domestic? Does the call for working with varied data sources in big data practices blur the boundaries between what is a reserach design choice and what is a commercial design choice? Is there a danger that statistical analyses of empirical data, can lead to path-dependency and polarization?

Part II: Four sub-themes for a future research agenda on big data in transnational politics

Part I sketched the contours of a research agenda that defines big data as a heterogeneous phenomenon that can be studied by attending to the interplay between social and technical relations. Furthermore, it suggested three moments of big data practices where this interplay can be fuitfully explored, and Table 1 indicated central questions to ask for each of these moments.

Our argument is that an exploration of big data along these lines will help move from treating this phenomenon as a technical question of method into a political sociology that interrogates what big data means as a practice and what is at stake for international relations in 'the rise of big data'. The second part of this paper specifies the direction of such a research agenda by inquiring into four theoretical sub-themes that we believe could be central to a broader political sociological agenda for looking at big data and governance. Each of these sub-themes are reflected in the questions posed in table 1.

Sub-theme I: Big Data, datafication and (post-)politics[2]

With the growing reliance on digital traces, we need a critical focus on how datafication conditions particular forms of knowledge, politics and governance. A focus that takes us beyond both the laudatory accounts of big data as 'truths' or the 'end of theory' and the more gloomy concerns about the automation of governance and 'algorithmic regulation' (Morozov, 2014). Just like earlier technological developments such as electrification and digitalization did not have clear-cut or unitary effects, datafication will have multiple, conflicting and surprising consequeces for social and political formations. To grasp these, we need relatively open and curious approaches to governance and datafication.

The governmentality literature (Miller and Rose, 1990; Dean, 1999) foregrounds a specific set of issues that help to understand the workings of big data in the context of governance. Drawing on the work of Foucault (1983), such accounts understand governance as a complex set of activities, calculations and reasonings that shape conduct. Thus, governance is not necessarily tied to the state or formal power positions, but constituted through multiple practices, 'governmental techniques' and 'political rationalities' (Miller and Rose, 1990). By engaging insights from sociological accounts of data and algorithms, governmentality studies and related conceptions of governance, we can approach the nexus of datafication and politics by focusing on the resources, knowledge production techniques, temporal orientations, and rationalities at work.

Governance is knowledge-intensive work and big data provides new foundations and forms of efforts to steer the conduct of others. Whereas existing forms of knowledge production underpinning governance rely on data sources and calculative operations that can be defined, sorted and managed quite easily by established procedures, big data analyses involve more messy and unstructured types of materials and rather different operations. This requires new types of skills, unfamiliar understandings of validity and knowledge and different types of relations to data producers. The intersection of big data and governance deserves scrutiny along the lines of at least three dimensions: sources of knowledge, calculative rationalities, and temporal

[2] Sub-theme 1 was written by Mikkel Flyverbom

orientations, as each of these features of big data have implications for our understandings and the practical workings of governance.

Sources and knowledge practices

It is a key feature of big data that the sources it relies on have multiple origins and take very different shapes. When inquiring into the first moment of digital big data practices it, for instance, becomes evident that digital traces are produced by a wide range of objects, activities and platforms. This means that they come with very different information attached – there are big differences between GPS signals, tweets and social media scrapings.

Such questions about data variety are discussed in existing accounts, but the primary focus is often on the nature of data sources, not the contexts and material infrastructures that they are produced or shaped by. Platforms such as Twitter produce very particular kinds of data and meta-data, such as length of messages (maximum 140 characters), the location of the author and the language used. But accessing such data is only possible through the so-called Application Program Interfaces (APIs) set up by data holders, and it is rarely possible to know what part of the data these actually give access to. Such factors complicate both the idea that big data analyses tap into the 'total information' available about a given topic and the promise that they can 'speak for themselves' and give us the full picture.

Furthermore, we tend to forget that before we start to analyze big data, algorithms and databases need to be connected, and that there are important patterns of inclusion and exclusion involved. As Gillespie (2014) points out, not everything is 'algorithm-ready', and failing to account for the resulting omissions would be "akin to studying what was said at a public protest, while failing to notice that some speakers had been stopped at the park gates". These complications add to the more well-known, but equally important point that digital divides still exist. Big data is only as big as what digital infrastructures, telecom prices and local capacities make possible, and these conditions need to be kept in mind (Hilbert, 2016).

The realization that data does not speak for itself makes it pertinent to investigate the second moment of big data practices, which brings up the issue of how analyses are crafted. Big data often considered a break with traditional forms of knowledge production such as random sampling and deductive hypothesis-testing on limited samples. Instead, big data promises analyses of all available data (Hilbert, 2016) – both in terms of inclusion and granularity – and more inductive forms of pattern recognition. These intersections and differences between big data and more established ways of producing knowledge are central (Mayer-Schönberger & Cukier, 2013; Hansen & Flyverbom, 2014) and remind us that one of the foundations of governance and power is knowledge production (Scott, 1998).

All kinds of knowledge production involve choices about sources, methods and goals. This is obvious when we look at more traditional forms of knowledge used for political purposes, such as narratives or numbers, and their shortcomings are very well studied and understood. Along similar lines, we need more fine-grained investigations of the practices, worldviews and material objects at work in big data analyses – what Amoore & Piotukh (2015) term the 'little analytics of big data'. Also, it is important for scholars interested in big data to pay attention to the intricate relations between what we can think of as seeing, knowing and governing (Flyverbom, Leonardi, Stohl & Stohl, 2016).

This implies in particular that we remember that all types of knowledge production and visualization have implications for what we consider important and possible to govern.

Calculative rationalities

Big data unsettles established ways of doing governance by affording new (dis)connections between activities on 'the ground' and favoring particular political rationalities. When inquiring into the third moment of big data practices it is important to understand such unsettlements. For instance, by introducing algorithmic operations and calculations into governance efforts, we move closer to future scenarios such as intelligent refrigerators that will order new milk before we run out, cities that will arrange busses exactly when and where they are needed, and drones that will take out terrorists before they commit attacks.

Such forms of 'algorithmic regulation' (O'Reilly, 2013) are considered a valuable source of innovation in governance and politics by some, in particular because they offer more data-driven, 'rational' and less idiosyncratic forms of policy-making. As a policy director at Google put it, datafication allows for 'policy by numbers' rather than 'policy by emotions' (interviews, Google, 2011). Similarly, O'Reilly (2013) suggests that big data may "reduce the amount of regulation while actually increasing the amount of oversight and production of desirable outcomes".

But others warn us strongly against trusting technology and data as an engine of policy-making and governance. The main arguments are that such algorithmic forms of governance will lead to the 'death of politics' where technocratic and measurement-oriented 'Silicon Valley logics' replace politics, history and experience with a naïve belief in data and algorithms (Morozov, 2014). As big data are not able to predict a changing world (see sub-theme 4), such developments, it is feared, will undermine democracy, the welfare state and the long-term focus on enabling human well-being. Similarly, Zuboff (2015: 81) warns us against this "ubiquitous networked institutional regime that records, modifies, and commodifies everyday experience from toasters to bodies, communication to thought, all with a view to establishing new pathways to monetization and profit. Big Other is the sovereign power of a near future that annihilates the freedom achieved by the rule of law".

Even if we have less gloomy expectations about the effects of big data, these arguments accentuate the need explore the datafication of governance from more agnostic starting points than the forms of 'technophobia' and 'dataphilia' that currently shape public discourse about the topic. Datafication naturally pushes the frontier between political and technocratic decision making. By delivering convincing evidence, big data approaches promise to transfer several questions that were formally in the domain of political deliberation into the domain of technocratic decision making. Obviously, this can go too far and dig data and algorithmic approaches can re-articulate heated political controversies as administrative or technological matters, i.e. contribute to 'post-political' forms of governance (Garsten & Jacobsson, 2013).

Think for instance of the difference between treating terrorism as a historical, socio-economic issue (which would look at its foundations and causes) and approaching it as an informational problem (focusing on how to compile enough signals to predict when a terrorist is about to strike). Future research should pay attention to the calculative rationalities and conceptions of the relationship between knowledge and politics that underpin different interpretations of the effects of datafication for governance (Hansen and Flyverbom, 2014). This is particularly important because datafication reframes 'key

questions about the constitution of knowledge, the processes of research, how we should engage with information, and the nature and categorization of reality' (Boyd and Crawford, 2012: 665).

Another relevant component of the calculative rationalities shaping datafied governance concerns the lack of transparency when it comes to the operations and worldviews involved. Big data analyses are crafted at a distance from human experience, and the mechanisms and operations involved are often completely illegible and "beyond immediate social inspection and control" (Kallinkos and Constantiou, 2015: 72). One suggestion is that we need more 'algorithmic transparency' (Pasquale, 2015) and similar ways of assessing the worldviews and rationalities driving algorithmic forms of governance. Taken together, these discussions highlight the epistemological and ontological foundations of the datafication of governance and deserve further examination if we want to understand the social and political implications of these developments.

Temporal orientations

A third issue, which is also related to the third moment of big data practices, concerns the way datafication reconfigures the orientation and significance of time in governance. Although velocity is often mentioned as a defining feature of big data, such questions point to a more foundational matter. The temporal orientations afforded by big data differ markedly from more traditional forms of knowledge production involved in governance. This takes various forms. For instance, moving from household surveys and similar sources of knowledge to big data aggregations may reduce the time lag between the start of a trend and the response by governments and other authorities. Efforts such as Google Flu Trends that relies on search queries as a way to predict flu outbreaks, and the work of UN Global Pulse to detect food crises based on tweets and other digital traces point to these possibilities when it comes to speeding up analytical operations to real-time (Global Pulse, 2011).

But big data also alters our conception of time in more fundamental ways. Whereas governance usually relies on standardized principles, agreements and laws that are developed and institutionalized over time, and adapts rather slowly and in largely reactive ways, big data offers more proactive and anticipatory approaches. Assessing what might happen next and preparing for future events is a key component of governance, and big data promises novel ways of producing foresight that can be used to take timely action. Anticipation, and temporal orientations more generally, are a central, but under-researched issue in governance and politics, and datafication processes invite us to give more attention to emergent ways of making sense of and relating to future developments. The point is not that big data can be used to predict the future per se, but rather that this phenomenon challenges and expands our temporal orientation in governance efforts.

Along these lines, datafied governance can also be seen as a form of continuous testing, where new signals and emergent correlations can be used to optimize and test the potential value of an effort or focus area. This dimension of the datafication-governance nexus reminds us that questions of big data and anticipation deserve more attention. The increasing reliance on digital traces in risk calculations, predictive policing, global foresight and similar attempts to govern 'imagined futures' (Beckert, 2016) makes it pertinent to explore how datafication facilitates new forms of anticipatory governance. We will return to the foundation of such anticipations in sub-theme 4.

Sub-theme II: Who are the political subjects of big data?[3]

That so much of social and political life is being conducted through the Internet calls for critical framings of how power relations are ever more entwined with digital life and data. In relation to this we would like to argue that the multiple ways that subjects act through the Internet ought to be our starting point for analysis. It is an argument that recognises that subjects cannot act in isolation but only in relation to the mediations, regulations and monitoring of the platforms, devices, and algorithms or more generally the conventions that format, organize and order what they do, how they relate, act, interact, and transact through the Internet.

Understanding these relations is especially called for as governments, international organisations and corporations increasingly appropriate big data to monitor, know, and govern populations. How are the subjects of big data conceived? Principally, they are often assumed as data subjects because they (should) receive data protection (see eg. Franklin 2013; Crowe 2013 and Glennie 2013). This is well illustrated in the focus of government programmes on questions of ethics and privacy where solutions involve ensuring anonymisation and the non-disclosure of identities. Political concerns are then defined in closely related terms such as the potential sensitivity of topics, and perceptions or risks of surveillance.

Such legal framings often implicitly or explicitly regard people as passive subjects in need of protection. Of course, this is important and people are rarely aware or knowledgeable about the kinds of data collected by the tracking of their activities. Yet, it is also important to remember that people are not simply data subjects but conduct themselves as performative rights-claiming subjects. That is, making rights claims involves not only legal and imaginary but also performative forces. This is a conception of political subjectivity advanced in critical citizenship studies, which conceive of the citizen beyond its modern configuration as simply a member of the nation-state. Instead, citizenship is seen as a site of contestation or social struggle rather than made up of bundles of given rights and duties.

Here we take up this conception in relation to how subjects become digital citizens when they make rights claims in and by what they do or say through the Internet (Isin and Ruppert, 2015).[4] In other words, who they become as political subjects—or subjects of any kind, for that matter—is neither given or determined but enacted by what subjects do in relation to others and things. This then has consequences for the very data that governing authorities seek to collect and analyse and act upon. For this reason, inquiring about who are political subjects of big data is relevant in relation to all three moments of big data practices.

---

[3] Sub-theme 2 was written by Evelyn Ruppert. Funding from a European Research Council (ERC) Consolidator Grant (615588) supported the research and writing of this section.

[4] The argument in this section draws from Isin and Ruppert's book, which contends that studies of the Internet and empirical analyses of specific digital platforms are proliferating, yet we lack concepts for framing and interpreting what these mean for political subjectivities and being digital citizens. The objective of the book then is to focus on theorizing what is referred to as digital acts and digital citizens with the understanding that such theorizing is necessary to clear the ground for more detailed empirical investigations.

It is instructive to first consider this conception in relation to the imaginaries of the Internet that have for over twenty years shaped how we think of the subject. One celebrates the libertarian possibilities while another laments the obedient qualities of acting through the Internet. While the latter have been a good correction to former, they have replaced sovereign subjects with obedient ones. In this way, they reflect a reversal of the understanding of power advanced in modern political theory, which posits a divide between modernity and tradition where a subject to power (tradition) was replaced by a subject of power (modernity). Modern political theory upheld the arrival of the latter as the displacement of the former. In contrast, critical political theory questions both the divide and displacement and asserts that a subject is a composite of multiple forces, identifications, affiliations, and associations. The subject itself is a site of myriad forms of power (sovereign, disciplinary, control) that embodies composite dispositions (obedience, submission, subversion).

This is an understanding of power advanced by Etienne Balibar (1991) through his reading of Michel Foucault. Balibar conceived of the citizen as not merely a subject to power or subject of power but one who embodies both. Balibar argued that being a subject to power involves domination by and obedience to a sovereign whereas being a subject of power involves being an agent of power even if this requires participating in one's own submission. However, it is this participation that opens up possibilities of subversion and this is what distinguishes the citizen from the subject: she is a composite subject where all three forms of power are always-present dynamic potentialities.

Such a conception moves us away from how we are being 'liberated' or 'controlled' to the complexities of 'acting' through the Internet. This understanding of subjectivation stands against that of interpellation, which assumes that subjects are always and already formed and inhabited by external forces. Of course, one cannot act in isolation but only in relation to the mediations, regulations and monitoring of the platforms, devices, and algorithms or more generally the conventions that format, organize and order what subjects do, how they relate, act, interact, and transact through the Internet. And as Bigo argues, if the conduct of data subjects conflicts with behaviour targeted by algorithms, then their digital rights will come into question (Bigo, 2013). However, such participation involves the play of obedience, submission and also subversion and the possibility of new forms of subjectivity that are often reflexive of the consequences of making rights claims by acting through the Internet.[5] It is this form of participation that distinguishes the subject from the digital citizen.

How then do subjects make rights claims and become digital citizens? Words are of course one way that they make claims to rights such as speech, access, and privacy. As Austin (1962) famously argued, language is a means of social action: people *do* things with words. In this regard he defined five classes of speech acts that can have performative force: judgments, decisions, commitments, acknowledgements, and clarifications. However, these do not account for the claims of subjects who articulate 'I,

[5] Bauman et al (2014), in relation to the Snowden revelations, ask how surveillance strategies might not only be resisted but how subjects might adjust their conduct and develop 'new forms of subjectivity that is more reflexive about the consequences of their own actions' (124). They also note that subjects can play a variety of games and resignify the meanings of sovereignty and citizenship, security, and liberty.

we, they have a right to'. Claims are thus a sixth speech act and key to becoming a citizen.

To be sure citizens make rights claims through what they say as many individual and collective declarations attest. Chelsea Manning (2015), for example, says 'We're citizens, not subjects. We have the right to criticize government without fear'. Many citizens also call upon authorities for the inscription of digital rights through regulations and legislation to give them legal force. But citizens not only make digital rights claims in or by what they say about those rights in declarations and manifestos (imaginary) and when they call upon authorities for the inscription of those rights (legality) but also by acting through the Internet (performativity). However, what subjects do through the Internet involves not only doing things with words but also the reverse of Austin's principle: they also say words with things. Subjects make claims through their deeds, by doing words with the things that make up the Internet.

As Franklin argues, declarations, case law, technical standards, and international treaties, while having best intentions, do not attend to how 'ordinary people use the internet in their everyday life' (Franklin, 2013). From downloading, uploading, forwarding, and blocking to encrypting and cloaking their actions, digital citizens make claims to rights such as to access, share or make private what they do through the Internet. While much attention is reserved for whistleblowers and hactivists as the vanguards of Internet rights, there are many more anonymous political subjects of the Internet who not only make rights claims by saying things but also by doing things through the Internet. This affords an opportunity to understand how the everyday social life of communicating, interacting and networking are part of struggles and contestations over the emergence of a new political subjectivity. So when we study conventions such as microblogging we can ask: how do such platforms both configure everyday conduct and at the same time create possibilities for digital citizens to act? What are the possibilities of subjects thinking, speaking, and acting differently, of challenging and resignifying conventions of the Internet and thereby enacting digital rights through what they do and not only say?

If what subjects are saying and doing through the Internet is changing political subjectivity and how they understand themselves as political subjects then what does this mean for the data that is generated in the first moment of big data practices? Bringing the political subject to the centre of concern offers a way to investigate how their acts matter in the making of data. This challenges interpretations of big data as measures of 'actual' behaviour, which reduce all conduct to clicks and movements. Rather the data generated through the Internet is not apart from the multiple ways of acting that can be strategic, subversive, and inventive; how to attend to these and their consequences for the making and interpretation of big data is one of many theoretical and empirical challenges.

We woud like to argue that the multiple ways that subjects act through the Internet ought to be our starting point for analysis. With ever more tracing and tracking and selling and trading of data, how subjects are making rights claims by blocking and filtering, encrypting communications, creating multiple and anonymised and shared identities, deploying bots, gaming trending algorithms, and so on can be the question we start with. Even for those subjects who obey and adhere to conventions, how have their acts and the digital data they generate been configured by variations in what they do through the Internet? How do they act in multiple and not only pre-formed ways? To

pose these qustions is not to ignore dominant forms of subjectivation but even these call for closer analysis in relation to how subjects are formed and act. To reduce all forms of acting to one unexamined mode is to also to overlook the potentialities and possibilities of acting otherwise and that acting through the Internet is not homogenous and universal but fragmented, multiple, and agonistic.

These questions are especially critical given the tendency to treat digital data as raw. We now have myriad studies that challenge this by accounting for the relations between and among people and technologies that come to make it up (Ruppert et al, 2013). But we also need a framing to understand how these involve power relations and embodied subjects and citizens who act through the Internet and whose acts are a part of the relations that make data.

Another challenge in answering the question of who is the political subject of big data is that digital life flows across national regulatory jurisdictions, that the corporations who collect and own this data are transnational, and that the rights claims of citizens increasingly traverse multiple legal and usually national orders. This raises interesting issues in relation to the third moment of big data practices. In relation to security, Bauman et al. argue that big data blurs what is 'domestic' and what is 'foreign,' and reconfigures the boundaries of the sovereign state and turns it into a site of political struggles, resistance and dissent (Bauman et al, 2014). Because the acts of subjects traverse so many borders and involve a multiplicity of legal orders, identifying who is this political subject has become a fundamental challenge.

So far, describing this transversal political subject as a global citizen or cosmopolitan citizen has proved challenging if not contentious (see e.g. Archibugi 2008 and Schattle, 2008). Indeed, sovereign regional or national legal orders and their understanding of rights are not capable of addressing this political subjectivity. However, securing digital rights requires not only their legal inscription but also recognition of how their daily enactment is also constitutive of rights claims and the formation of this new political subjectivity. Such an understanding moves beyond legally securing the rights of data subjects to recognising the claims and struggles of citizens for rights over the data generated by their acts through the Internet.

Sub-theme III: Dymanics if valuation in big data governance[6]

We have already argued that one of the central characteristics of big data is that it is varied. In fact, the possibility to cross-aggregate data from various digitized sources is one of the characteristics that makes big data projects distinct from other types of analysis. This means that their success is dependent upon a successful construction of a digitized space where a variation of data points can cross-fertilize each other.

The creation of such a space naturally poses several technical challenges. In order to conduct analyses across data from distinct sources there is a need to ensure that data is stored in compatible formats, that meta-data structures are consistent, that there is enough computational power to run complex analyses and so forth. Any big data project needs to find ways of handling such technical challenges.

However, technical challenges are only a fraction of the challenges that arise from the ambition to cross-aggregate data from varied sources. Most importantly, big data

[6] Sub-theme 3 was written by Anders Koed Madsen

projects often involve an ambition to cross-fertilize data points from sources owned by different organizations and companies. For instance, when using data to respond to disasters such as a flood, one needs data from weather stations, sewage companies, public administrations and so forth, In other words, governance oriented big data projects most often require the establishment of novel data-partnerships between the involved parties.

Studies of recent big data projects indicate that challenges stemming from such partnerships are harder to solve than clearly defined technical problems (Madsen, 2013). More specifically, it seems that emerging partnerships raise complex questions about the valuation of data. How can different data points be compared and valued? How can incentives be created for different partners to share data? How is data to be payed for by potential users? What is the epistemic value of different data sources when making decisons? What is the relation between epistemic value, knowledge and action?

A research agenda for big data in the context of international political sociology must include ways of inquiring into such questions. They are crucial for understanding how big data creates transversal connections, i.e. connections across instituted agencies and (data) boundaries, and for understanding how value judgements are enacted in the transformation of 'data' into 'big data'. We want to argue that a fruitful startingpoint from which to iquire into such questions is to take inspiration from a recent substring of economic sociology – valuation studies.

Valuation studies have challenged the idea that assets have intrinsic value and emphasized the fact that valuation is something that is obtained through practice (Muniesa et al., 2007; Espeland & Stevens, 1998; Stark, 2011). However, studies in this tradition have mainly focused on the function of market devices. That is, technologies that enable things to be comparatively priced. However, the practice of valuation can also be more broadly conceived as an act of ordering that establishes a foundation for the prioritization of various kinds of scarce resources (Madsen, 2015).

A research agenda with theoretical roots in this line of work can therefore underpin inquiries into the way valuation occurs in relation to less tangible objects such as data. Most importantly, it can prompt researchers to problematize the idea of data as having intrinsic value, and instead turn to the social, organizational and political construction and production of data-points as valuable objects. Such a focus must also include sensitivity towards possible frictions between different ways of sorting, valuating, and attaching legitimacy to data in a specific developmental context.

The importance of this line of research is best illustrated through a brief look at recent projects that use big data as a tool of governance. We want to provide two such examples. The first example emphasizes the need to establish an *exchange value* of data in a Copenhagen-based Smart City project that required previously unrelated organizations and companies to share data. The second example illustrates the problem of determining the *epistemic value* of tweets as a new data source to inform decision-making in the UN. Accordingly, these two examples draw our attention to relevant dynamics of valuation at the first and third moment of a big data practice.

Establishing exchange value in a City Data Market in Copenhagen

The municipality of Copenhagen has recently launched a strategy that includes an ambition of being at the forefront of the Smart City movement. That means using big

data analysis to improve the response of the public administration when it is faced with problems such as traffic jams and floodings. Therefore, in 2014, the city established a specific governing body for smart city projects across all sectors in the city. This body was named Copenhagen Solutions Lab (CSL) and it was instructed to have focus on creating triple helix partnerships. In other words, the implementation of innovation and smart city development was to be done in close collaboration with knowledge institutions and companies as well as citizens.

CSL quickly came to the conclusion that in order to make any useful governance tools there was a need to establish a database with an interface that allows distinct actors to collaborate and share data. For instance, in order to visualize the flow of water after a heavy rainfall the municipality needs sewage data form a private contrator, weather data from a ministerially owed metrological institute and behavioral data from citizens. In order to guide parents to the nearest playground they need data for the various private contractors running the playgrounds. In order to intervene in traffic congestion they need data from the semi-private Danish Railways, the bus company MOVIA and so forth. In short, no governance related initiative could be successfully conducted within the existing municipal database.

The chosen solution was to build a data marketplace that would enable private and public data to be stored in a common database on top of which hackers and app-developers could build useful applications. In other words, the big data ambitions of Copenhagen sparked a process that can only be understood as a form of market creation and valuation. Furthermore, the most complicated questions emerging from this process were not the ones addressing how to build the technical infrastructure. They concern the complex practices of data-valuation.

For instance, in order to function the marketplace must be able to work with several layers of data openness. There will inevitably be projects where private data owners agree to share data with each other and the municipality but not with a broader public. Making such layers is not a technical problem but rather a problem in establishing a trust among the partners that the interface and the partners in the market respect the *strategic value* of their data.

A related matter concerns the need to establish an *exchange value* of data. For instance, what is real-time updated sewage data worth compared to a data from a coffeeshop showing ques in front of their store? Is the temporality of data (how often is it updated?) a factor to include in such a valuation or should other criteria be applied? When such questions are answered the next one emerges: How are users to pay for data? Can you make a subscription to a data feed? Can subscription prices vary with the kind of data requested? Can you only pay with money or can you pay with data – or perhaps analytical skills and hardware power?

The City Data Market has not yet found the right solutions to such questions. It takes time to build a market. But the character of the problems is illustrative of the argument that a research agenda for inquiring into big data in International relations must focus on dynamics related to data valuation in the first moment of big data practices. The way such challenges are handled influence the data available for governance before it even enters the second or third moment. The Copenhagen example shows the importance of attending to the *strategic value* as weel as the *exchange value* when organizing data

partnerships. The next case will supplement with the need to understand the *epistemic valuations* of data in the third moment of a big data practice.

Re-thinking epistemic values in the UN's crisis response

In 2009 The United Nations established Global Pulse – a flagship initiative under the Secretary General with the aim of exploring the potentials of using big data as a possibility to accelerate discovery, development and innovation for sustainable development and humanitarian action. The idea was that big data – such as online traces – could be used as a real-time indicator of public well-being and act as a feedback mechanism on the workings of policy interventions (Madsen, 2013).

The initiative is organized around public-private partnerships. Unsurprisingly, it has also faced problems about the *strategic value and exchange value* of data. For instance, it has promoted the concept of 'data philanthropy' to indicate sharing data as an act of social responsibility that can ultimately have an effect on brand value (Kirkpatrick, 2011). However, since we have already touched upon these forms of valuation above, we will focus on how the Global Pulse have also raised questions about the *epistemic value* in the third moments of a big data practice.

Such valuation dynamics stems from the fact that much of the data that the initiative works with data that has quite different characteristics from the kind of data that is usually judged as having high epistemic value in the organization. For instance, one project of Global Pulse has been to use sentiment in tweets to predict prices of rice in Indonesia. This is a very different mode of knowledge-production than a technology like the household survey that has traditionally been used to monitor crisis-prone populations. It is quite simply a different kind of foundation from which to make decisions about crisis management (Madsen, 2015).

The fact that tweets have different affordances than survey responses can be illustrated by three examples. First, when Twitter data comes within the reach of the UN it has already been formatted in specific ways. It comes in chunks of 140 characters and it is ordered though specific user-generated meta-data such as hashtags (#) and replies (@). Such choices about data formatting are beyond the control of the intended users in the UN. Second, it is produced on an interface that is not designed with the intention to produce reliable crisis-signals. To the contrary it is designed to make people communicate in specific event-driven ways that is of value to Twitter, Inc. (Rogers 2013). Third, the flow of data obtained from the API is so large and fast paced, that parts of the data analysis need to be automated. A human expert cannot routinely oversee the data behind the visualization because such a check would require an excessive amount of human brainpower.

The result of these conditions for data production is that their *epistemic value* is not given – Twitter is a source of data that need to be valuated anew by decision-makers in the organization. As the director of Global Pulse puts it: "It is not just about getting the data; it is also […] about the organizational capacity to facture a snapshot of these types of information in the context of their on-going policy development planning" (Madsen 2012). This capacity is dependent on whether or not the organization recognizes the *epistemic value* of the data. That is, whether is is deemed legitimate enough to base decisions upon. Accordingly, the case of Global Pulse illstrates that the relevent valuations processes to pay attention to in the context of big data is not just about *exchange value* and *strategic value* of data. A reserach agenda must also inquire into the

way the *epistemic value* of data is established and think about the way such valuation influces the existing modes of governance.

Sub-theme IV: Big data creating futures: the importance of theory driven data production[7]

As powerful as empirical data analysis might be, there is a fundamental limit to the power of big data. The reason is not technological or methodological, and cannot be easily solved through any kind of technological or scientific progress. Instead, it leads us deep down the rabbit hole of the philosophy of science. All data (big or not) are from the past — or, at best, the 'real-time past' (as soon as something is recorded and becomes data, the recorded activity has past). So any data analysis can only tell us about what has already happened. It cannot tell us about things that have never happened. This is an important point to remember when inquiring into the second moment of big data practices. Patterns and predictions are naturally path-dependent when they are based on empirical data, which is always from the past.

In cases where the past, present and future follow the same logic, this is extremely useful. For example, in physics, when the sun has been rising for the past millennia, and there is not a significant change, past data allows us to predict that the sun will be rising tomorrow again. The technical term for this omnipresent assumption of using data for prediction is 'stationarity' and it is always in the fine-print of every predictive statistical analysis of this type. However, if significant changes occur in the dynamic of the system, empirical statistics from the past are, at best, limited, if not deceiving.

We can think about it this way. It is true that Facebook, Google and Amazon can predict your future behavior better than any psychologist —if your future behavior follows the same logic as your past behavior. If you fall in love or get divorced, if you change your job or the country where you live, predictions from past data will at best be limited, if not deceiving. For example, data from the past alone cannot tell us what it would be like to live in a world without pollution, without hunger, without wars. How could it? There is no empirical evidence to feed data analysis.

What makes matters worse is that most of what we do is to work hard to create futures that are different from the past. Most public and private sector activities aim at creating futures distinct from the past in the best case, we want to create 'a better world'. We do so by designing or implementing public policies, private business strategies or any kind of intervention into the current trajectory. Our very actions usually destroy the stationarity of the time series of history. This argument is known as the 'Lucas critique' in economics (Lucas, 1976), as 'Goodhart's law' in finance (Goodhart, 1976) and as 'Campbell's law' in education (Campbell, 1976), all dating back to 1976.

Nobel Prize-winning economist Robert Lucas criticized colleagues who used sophisticated statistics to make economic predictions ('econometrics') in order to inform policy making. He argued that no useful information can emerge from such analysis because "any change in policy will systematically alter the structure of econometric models" (Lucas, 1976). In other words, using the conclusions from our data analysis to intervene in the system's dynamic changes the system and therefore leads to a different system. Data from the past cannot tell us how this new system will behave.

[7] Sub-theme 4 was written by Martin Hilbert

Given the complexity and uniqueness of each context, the result of such interventions is almost always novel and unique for each case.

One way to go about it is to take similar cases, from somewhere else or from a different scale, and to extrapolate to other cases. For example, we do not know what an industrialized country (A) would be like that is currently non-industrialized, but we could assume that we can statistically extrapolate it from a version of the past development of another country (B) that went through industrialization. This will work if we fulfill the assumption 'ceteris paribus' (i.e. "all other things being equal"), another omnipresent assumption of the fine-print of statistical analysis. Now, being reasonable, it is likely that between country A and country B there are 'other things not being equal' besides their degree of industrialization. Therefore, the explanatory power of this kind of statistical extrapolation might be very limited, if not deceiving as well. History is full of examples. For example, one of the main arguments in favor of the transitions from horse carriages to cars in cities was that it will be better for the health of people, as horse remainings represented a severe health hazard to modernizing cities. There was no empirical data available to predict that over a century later the atmosphere would be saturated with car exhaust, and even big data analysis would have had severe problems in predicting this. In order to make predictions in a changing world, theory is necessary.

The good news is that the digital revolution does not only change empirical data science, but also theory-driven modelling that allows us to explore scenarios that never existed, for example through simulation (Hilbert, 2015a). Engineers have used computers for a long time to simulate buildings and bridges that never existed. This also provides data, but data that is invented by a model that simulates futures that have never existed in empirical reality. It can create lots of data, in this sense, it can also be big data, but the idea is very different of what usually is understood to be the digital big data footprint (Hilbert, 2015b). What produces data is not empirical reality, but theoretical models. These models create realities that only exist "in theory". They can be better or worse, or simply different from what has ever existed. They are extremely creative and require a different set of skills than data analysis.

In a computational science framework, building models with digital means is very similar to designing a video game. They can be as visual as prominent video games like SimCity (Hilbert, 2014b). Some branches of science, industry and governments are working on creating computer simulation models at high speed. For example, the UN Environment Programme has teamed up with Microsoft Research for the past three years to create a computer model that simulates all ecological life on Earth (Purves et al., 2013). The city of Portland has simulated the daily behavior of its 1.6 million residents over 180,000 locations in a one-to-one fashion, in order to optimize the roll-out of a new light-rail infrastructure and to simulate epidemics (Barrett et al., 2005). Such simulations allow to explore hypothetical 'what-if' questions. They can be used to adjust for local particularities and today's computational power allows for an impressive amount of detail and sophistication in such models (for an introduction see Gilbert and Troitzsch, 2005; or Wilensky and Rand, 2015).

This leads us to another point that relates to the second moment of big data practices - the claim that big data leads to "the end of theory" (Anderson, 2003) have to be qualified. It is true that machine learning and deep learning approaches are incredibly effective when working with stationary sources, such as physical systems or natural language translation. For example, machine learning has solved the long-standing

challenge of natural language translation better than any theory driven approach could have (Halevy, Norvig and Pereira, 2009). However, the ambition of humanity goes beyond repeating past patterns.

The importance to realize that big data can only work with past patterns is also a concern with much urgency. Decision based on data from the past can only replicate past patterns into the future and therefore has the tendency to lock us into old patterns from the past. It often even accentuates it, creating self-fulfilling path dependencies. For example, a largescale study by Facebook showed that Facebook's big data based content selection algorithm suppresses the exposure to diverse content by 8% for self-identified liberals and by 5% for self-identified conservatives (Bakshy et al., 2015). Our innate drive to expose ourselves to like-minded content out of comfortability and familiarity only reduces our exposure to less diverse content by 6% (for liberals) (Bakshy et al., 2015). In other words, the big data Facebook algorithm doubles our path dependency, locking us into our past patterns with twice the force that we choose to. As this reinforces the past, it can lead to polarization and even extremism, which is very relevant from a perspective of international political sociology.

In order to create change, in order to create better people and better societies, we need to break free from old patterns. This requires more than data from the past. It requires something qualitatively different. It requires theory. Theories that model futures that have never been. Now these theories are rooted in the past, but project a future that is significantly different from it. Serious explorations of what could and should happen "in theory" are extremely creative acts. Of course, we have to remember that "all models are wrong, but some are useful" (Box and Draper, 1987). This applies to thoereticlly derived models, or simply some form of intuitive 'Weltanschauung'. So there will always be many models, of complementary aspects of reality, complementary visions, aimed at creating different aspects of better futures. So since no model will ever be perfect, there is a lot of work to. This is good news for the sciences. In contrary to a future vision where the scientific process has been automated by machine- and deep learning algorithms, it shows that there will be a great demand for new generations of visionaries and visionary scientists in the digital age, constantly figuring out what futures we would like to live in, how would this be achievable, and how would be get there, "in theory".

If big data is to be useful for political sociology, it need to to consider that human history is not automatically stationary, but changing. This requires to go beyond the traditional data paradigm of pure data analysis, and requires the creative combination of both theories that enable to forecast and project unprecedented scenarios, and (big) data to calibrate the models and to root them into empirical reality.

Conclusion

So, what are the most pressing theoretical issues for an international political sociology addressing the current dynamics of datafication of governance? How can we productively discuss the potentials and challenges for transnational politics in a time where its knowledge-foundations seem to be shifting?

One answer suggested by this paper is that we have to be careful to resist the temptation to reduce discussions of big data to a technical issue to be handled by computer scientists. Also, we need to refrain from treating it as a homogenous phenomenon with a set of identifiable and stable features that will produce predictable

outcomes. The overall theoretical aim of this paper has been to set the contours of a research agenda that can stimulate different types of inquiry into the relation between big data and transnational governance.

Our discussions of the intersection of datafication, governance and politics have highlighted the need to revisit widespread assumptions about big data, knowledge and governance. Rather than assume that data 'speaks for itself' or is the same thing as knowledge, we must investigate the sorting processes and analytical operations involved in the production of knowledge (Flyverbom & Madsen, 2015). Also, datafication rests on and facilitates particular forms of knowledge and rationalities, and these in turn have significant consequences for what we see and act on because they guide our attention in subtle ways. Finally, big data developments have important but somewhat under-theorized ramifications for temporalities, anticipation, and path-dependency that ought to be ecamined by sociologist and political scientists with an interest in this emergent phenomenon.

More specifically, we have argued that these themes can productively be studied by looking at socio-technical relations in three decisive moments of big data practices: 1st, the datafication of daily life, 2nd, the the production of patterns and predictions and 3rd, the making of new modes of governance. We take such relations to be situated and we, therefore, treat big data as a complex phenomenon that obtains different shapes in different contexts.

We have also identified four sub-themes that researchers could pay attention to when studying this 'taking shape' in specific contexts of their interest. More specifically, we argue that future research on big data and international relations could productively focus on a) the datafication-governance nexus, b) the enactment of political subjects, c) valuation-dynamics and data-partnerships, d) the role of theory and history in setting limits on the power of big data

Each of these sub-themes has been extensively discussed above and table 1 in the beginning of the paper has already summarized some of the main points of these discussions. It is our belive that this table suggests relavant questions and theoretical lineages to explore when studying big data in the context of international political sociology.

References

Amoore, L. (2014) Security and the Claim to Privacy. *International Political Sociology*, 8: 108–112.

Amoore, L., & Piotukh, V. (2015). Life beyond big data–governing with little analytics, *Economy and Society*, 44(3): 341-366

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine*, Retrieved from: http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

Althouse, B. M., Ng, Y. Y., & Cummings, D. A. (2011). Prediction of dengue incidence using search query surveillance. *PLoS Negl Trop Dis*, *5*(8): e1258.

Archibugi, D. (2008). *The Global Commonwealth of Citizens: Toward Cosmopolitan Democracy*, Princeton University Press

Austin, J L. (1962) *How to Do Things with Words*, Oxford University Press.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239): 1130–1132.

Balibar, E. (1991) Citizen Subject, in *Who Comes after the Subject?*, ed. Eduardo Cadava, Peter Connor and Jean Luc Nancy, Routledge: 33-57.

Barrett, C. L., Eubank, S. G., & Smith, J. P. (2005). If Smallpox Strikes Portland, *Scientific American*, 292(3): 54–61.

Bauman, Z., D. Bigo, P. Esteves, E. Guild, V. Jabri, D. Lyon, and R. B. J. Walker. (2014). After Snowden: Rethinking the Impact of Surveillance, *International Political Sociology,* 8(2): 121-144.

Beckert, J. (2016). Imagined Futures: Fictional Expectations and Capitalist Dynamics, Cambridge MA, Harvard University Press.

Bigo, D. (2013). The Transnational Field of Computerised Exchange of Information in Police Matters and Its European Guilds, in *Transnational Power Elites: The Social and Global Structuration of the EU*, ed. Niilo Kauppi and Mikael Rask Madsen, Routledge: 155-182.

Blumenstock, J. E. (2012). Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda. *Information Technology for Development*, *18*(2): 107-125.

Booth, R. (2014). Facebook reveals news feed experiment to control emotions, *The Guardian*, Retrieved from: http://www.theguardian.com/technology/2014/jun/29/facebook-users-emotions-news-feeds

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. Wiley Publishing.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. *Occasional Paper Series, #8*. Retrieved from http://eric.ed.gov/?id=ED303512.

Crowe, A. (2013). The Promise, and Problems, of Mobile Phones in the Developing World. *OpenDemocracy*, 1. Retrieved from: http://bit.ly/1WsiINZ.

Dean, M. (1999). *Govemmentality: Power and rule in modem society*, SAGE.

De Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R., & Yu, C. (2010). Automatic construction of travel itineraries using social breadcrumbs. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, ACM: 35-44

De Mauro, A., Greco, M., & Grimaldi, M. (2014). What is Big Data? A Consensual Definition and a Review of Key Research Topics. Presented at the *4th International Conference on Integrated Information*, Madrid. http://doi.org/10.13140/2.1.2341.5048

Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a social process. *Annual review of sociology* (24): 313-343.

Flyverbom, M. & Madsen, A. K. (2015) Sorting Data Out: Unpacking Big Data Value Chains And Algoritmic Knowlegde Production in *Die Gesellschaft der Daten: Über die digitale Transformation der sozialen Ordnung*, ed. ed. Süssenguth, F., Transcript Verlag: 123-144.

Flyverbom, M.; Leonardi, P.M.; Stohl, C. & Stohl, M. (2016). The Management of Visibilities in the Digital Age, introduction to special issue, *International Journal of Communication.*

Foucault, M. (1983). The Subject and Power. In *Michel Foucault: Beyond Structuralism and Hermeneutics*, ed. Dreyfus, H.L and Rabinow, P., University of Chicago Press: 208-226.

Franklin, M.I. (2013). *Digital Dilemmas: Power, Resistance, and the Internet*, Oxford University Press.

Gilbert, N., & Troitzsch, K. (2005). *Simulation for the Social Scientist* (2 edition), Open University Press.

Gillespie, T. (2014). The Relevance of Algorithms. In *Media technologies: Essays on communication, materiality, and society*, ed. Gillespie, T., Boczkowski, P.J. and Foot, K.A, MIT Press: 167-195

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

Glennie, J. (2013). A Development Data Revolution Needs to Go Beyond the Geeks and Bean-Counters. *The Guardian Poverty Matters blog*, Retrieved 4 May 2015 from: http://bit.ly/1LdARvF.

Global Pulse. (2011). *Twitter and perceptions of crisis related stress.* Retrieved January 11 2012), from: http://www.unglobalpulse.org/projects/twitter-and-perceptions-crisis-related-stress

Goodhart, C. (1976*). Problems of Monetary Management: The UK Experience*, The Public Affairs Center, Dartmouth College.

Halevy, A., Norvig, P., & Pereira, F. (2009). The Unreasonable Effectiveness of Data. *IEEE Intelligent Systems*, 24(2): 8–12.

Hannak, A., Sapiezynski, P., Molavi Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., & Wilson, C. (2013). Measuring Personalization of Web Search. In *Proceedings of the 22Nd International Conference on World Wide Web,* Republic and Canton of Geneva, Switzerland: 527–538

Hannak, A., Soeller, G., Lazer, D., Mislove, A., & Wilson, C. (2014). Measuring Price Discrimination and Steering on E-commerce Web Sites. In *Proceedings of the 14th ACM/USENIX Internet Measurement Conference (IMC'14)*. Vancouver, Canada.

Hansen, H.K and Flyverbom, M. (2014). The politics of transparency and the calibration of knowledge in the digital age, *Organization.*

Hilbert, M. (2013). Big Data for development: From information to knowledge societies. Retrieved June 15 2013 from: http://ssrn.com/abstract=2205145.

Hilbert, M. (2014a). How much of the global information and communication explosion is driven by more, and how much by better technology? *Journal of the Association for Information Science and Technology*, 65(4): 856–861.

Hilbert, M. (2014b). Big Data requires Big Visions for Big Change. London: TEDxUCL, x=independently organized TED talks. Retrieved from: www.youtube.com/watch?v=UXef6yfJZAI.

Hilbert, M. (2015a). DT&SC 7-1: Introduction to e-Science (Vol. Digital Technology & Social Change). University of California, Davis. Retrieved from: https://youtu.be/9x3d75ZMuYU.

Hilbert, M. (2015b). e-Science for Digital Development: ICT4ICT4D. Centre for Development Informatics, SEED, University of Manchester. Retrieved from http://www.seed.manchester.ac.uk/medialibrary/IDPM/working_papers/di/di-wp60.pdf.

Hilbert, M. (2016). Big Data for Development: A Review of Promises and Challenges. *Development Policy Review*. 34(1): 135-174.

Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, 332(6025): 60–65.

Hutchby, I. (2001) Conversation and Technology: From the Telephone to the Internet, Polity.

Isin, E. and Ruppert, E. (2015) *Being Digital Citizens*, Rowman and Littlefield International.

Kallinikos, J., & Constantiou, I. D. (2015). Big data revisited: a rejoinder. *Journal of Information Technology*, *30*(1): 70-74.

Kitchin, R. (2014) The real-time city? Big data and smart urbanism, *GeoJournal*, 79(1): 1-14.

Kirkpatrick, R. (2011). Data Philanthropy: Public & private sector data sharing for global resilience, Available at http://www.unglobalpulse.org/blog/data-philanthropy-public-private-sector-data-sharing-global-resilience

Laney, D. (2012), The importance of big data: A definition, Gartner.

Lucas Jr, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1: 19–46.

Madsen, A.K (2016), Beyond the bubble, *Mediekultur.*

Madsen, A. K. 2015. Tracing data –paying attention, In *Making Things Valuable,* (Eds. Kornberger et al.), Oxford University Press.

Madsen, A. K. (2013). *Web-visions: Repurposing digital traces to organize social attention*. PhD series 26.2013, Copenhagen Business School.

Madsen, A. K. (2012). *Interview with Robert Kirckpatrick.* Unpublished manuscript.

Manning, Chelsea E. (2015). We're Citizens, Not Subjects. We Have the Right to Criticize Government without Fear. The Guardian, Retrieved May 6 2015 from: http://bit.ly/1GORENO.

Mayer-Schönberger, V. & Cukier, K. (2013) Big data: A revolution that will transform how we work, live and think, Mariner Books.

Miller, P., & Rose, N. (1990). Governing economic life. *Economy and society*,*19*(1): 1-31.

Mocanu, D., Baronchelli, A., Perra, N., Gonçalves, B., Zhang, Q., & Vespignani, A. (2013). The Twitter of Babel: Mapping World Languages through Microblogging Platforms. *PLoS ONE*, *8*(4): e61981.

Morozov, E. (2014). *The rise of data and the death of politics*, The Guardian, Retrieved from: http://www.theguardian.com/technology/2014/jul/20/rise-of-data-death-ofpolitics-evgeny-morozov-algorithmic-regulation.

Muniesa, F., Millo, Y., & Callon, M. (2007). An introduction to market devices. *The Sociological Review, 55*(s2): 1 12.

O'Reilly, T. (2013). Open data and algorithmic regulation, in *Beyond Transparency: Open Data and the Future of Civic Innovation,* ed. by Goldstein, B and Dyson, L., Code for America.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press.

Purves, D., Scharlemann, J. P. W., Harfoot, M., Newbold, T., Tittensor, D. P., Hutton, J., & Emmott, S. (2013). Ecosystems: Time to model all life on Earth. *Nature*, 493(7432): 295–297.

Ruppert, E., Law, J., and Savage, M. (2013). Reassembling Social Science Methods: The Challenge of Digital Devices, *Theory, Culture & Society*, Special Issue on The Social Life of Methods 30(4): 22-46.

Rogers, R. (2013). Debanalizing twitter: The transformation of an object of study. *Proceedings of the 5th Annual ACM Web Science Conference:*356-365.

Schattle, H. (2008). *The Practices of Global Citizenship*, London: Rowman & Littlefield

Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*, Yale University Press.

Stark, D. (2011). What's valuable? In ed. Aspers, P. and Beckert, J., *The worth of goods: Valuation and pricing in the economy*, Oxford University Press: 319-339

Stevens, Tim. (2015) Security and Surveillance in Virtual Worlds: Who Is Watching the Warlocks and Why? *International Political Sociology*, doi: 10.1111/ips.12094

Wilensky, U., & Rand, W. (2015). *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*, The MIT Press.

Zuboff, S. (2015) Big other: Surveillance capitalism and the prospects of an information civilization, *Journal of Information Technology* 30: 75-89.