

# Big Data in Government: Discursive proto-institutionalization in European Metropolitan Governments

Research Proposal

14 August 2015

Basanta E.P. Thapa

PhD Fellow at the DFG Research Training Group “Wicked Problems, Contested Administrations”

University of Potsdam

*thapa@uni-potsdam.de*

## **Abstract**

Big Data Analytics is considered a major technological innovation for government and public administrations, but there is substantial disagreement about the scope of its impact. In the theoretical perspective of new institutionalism, Big Data in government is a case of proto-institutionalization whose specific social construction or ‘theorization’ is subject to discursive struggles among actor constellations and processes of translation and editing that render it compatible to institutional legacies. Comparing the discursive dynamics of the theorizations of Big Data Analytics in three European metropolitan government (Vienna, Amsterdam, and London) with qualitative data analysis and co-occurrence analysis, the patterns of theorization and the influence of actor constellations, contextual factors, and institutional legacies on the local understanding of Big Data in government are explored. The findings also shed light on the suggested emergence of a new digital public sector reform paradigm.

Contents

- Introduction..... 3
  - Big Data Analytics as a socio-technical assemblage..... 3
  - Big Data Analytics in government ..... 4
  - Big Data as the herald of a digital public sector reform paradigm?..... 5
- Puzzle & Research Objective ..... 6
- Theoretical framework..... 7
  - Institutional work and theorization..... 8
  - Translation and editing..... 9
  - Research gaps..... 10
- Research Questions..... 10
- Research Design ..... 11
  - Case selection..... 11
  - Data collection..... 13
  - Data analysis..... 14
- Work plan ..... 15
  - Slicing of cumulative dissertation..... 15
  - Timetable (tentative)..... 15
- Summary ..... 15
- References..... 16

## Introduction

The new technology of Big Data Analytics – the analysis of data in unprecedented volume, velocity and variety – is said to usher in “A Revolution That Will Transform How We Live, Work and Think” (Mayer-Schönberger and Cukier 2013). Enabled by the “datafication of the world” (Mayer-Schönberger and Cukier 2013), e.g. due to the proliferation of smartphones, social media and the Internet of Things, as well as advances in data storage and processing capacity, Big Data stands to vastly increase our ability to observe social and natural phenomena and spot (and predict) patterns in unprecedented granularity and real-time.

### Big Data Analytics as a socio-technical assemblage

Originally, Big Data is a technical term that specifies a certain kind of digital data storage and processing. However, as the hype around Big Data Analytics has moved beyond computer science, the term has increasingly become separate from technical specificities and is now used as a more general label for data-driven and data-intensive ways of decision-making and management.

The technical definition of Big Data is highly contested (Ward and Barker 2013). Typically, Big Data is characterized by the ‘3V’: *volume*, *velocity* and *variety*. *Volume* not only refers to the unprecedented scale of Big Data, but also to the idea of working with full datasets rather than partial samples, upending accepted practices of inferential statistics (Cukier and Mayer-Schönberger 2013). *Velocity* highlights that Big Data often means (near) real-time monitoring, e.g. with wirelessly connected sensors, smartphones or social media. *Variety* points towards the varied sources of Big Data, which can be both structured and unstructured data, such as traffic surveillance data and newspaper articles. In this regard, Big Data is made possible by the growing “datafication” (Mayer-Schönberger and Cukier 2013) of the world, i.e. the transformation of phenomena into digitally processable data.

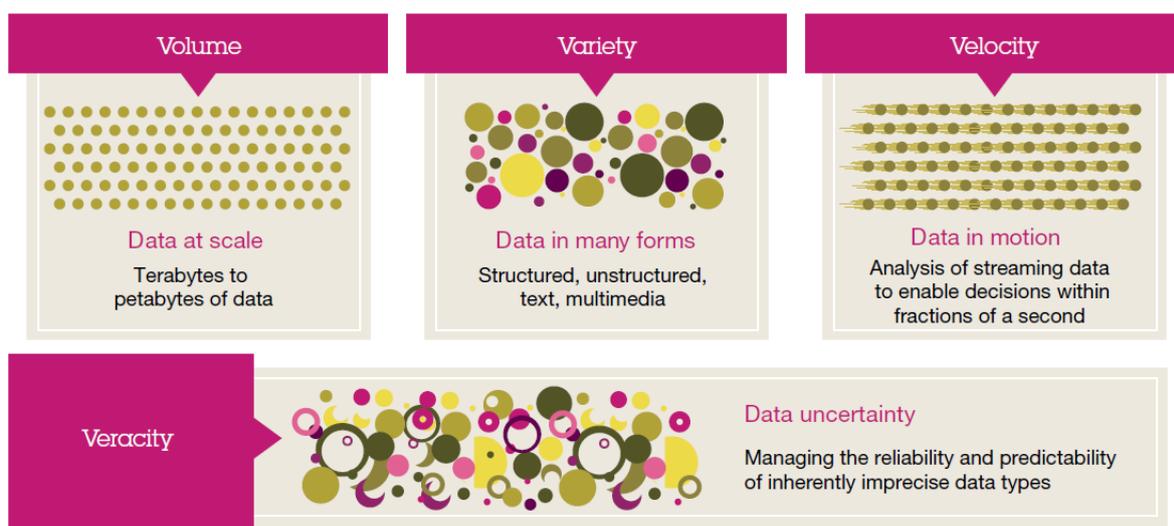


Figure 1: Four dimensions of Big Data (IBM Institute for Business Value 2012, p. 4)

However, Big Data's claim as a 'revolution' goes beyond data volume, "because more isn't just more – more is different" (Anderson 2008b). The explosion of sample sizes increases statistical reliability, which leads some authors to predict an "end of theory" (Anderson 2008a) where correlation trumps causation. Thus, the promise of data-driven 'predictive analytics' is to predict behavior based on observed correlations, even in absence of any causal model (Anderson 2008b). While this provocative view has been heavily criticized (Bollier 2010), it illuminates the data-driven approach to analysis and decision-making that has been inspired by the technical possibilities of Big Data Analytics.

As Big Data has become more widely discussed, it has partially detached from its technological foundations and come to generally label any kind of decision-making that is supported by digital data processing. Ruppert et al. (2015) therefore define Big Data as a "socio-technical assemblage", a highly contextual combination of technology and social practice. This openness to different socially determined uses is typical for general-purpose technologies like information technology, where the material technology hardly prescribes purposes and operational routines (Dolata and Werle 2007). Underlining this, Wang (2010) finds that socio-technical assemblages are typical among "information technology fashions", which therefore rather resemble management fashions (Abrahamson 1996) than technology diffusion (Rogers 2010). Thus, Big Data Analytics has to be considered more than just a specific statistical software and can be treated as a technologically-enabled social practice.

### Big Data Analytics in government

Governments and public administrations have taken longer to react to the Big Data hype, but are unequivocally listed among the sectors that are expected to be impacted the most (McKinsey Global Institute 2011). After all, government ranks among the most knowledge-intensive businesses (Lenk and Wengelowski 2004). Hence, governments, international organizations, consultancies, and scholars around the globe are abuzz with the potential uses of Big Data in government (Poel et al. 2015). According to Pollitt (2014), Big Data Analytics is among the most important technologies for the future of public administration, about whose effects "[o]pinions range from extreme optimism to dire pessimism" (Pollitt 2014, p. 19).

The interest in Big Data in government is understandable as politicians and public administrations face an intertwined challenge: In increasingly complex societies, they have to cope with more and more "wicked" problems (Rittel and Webber 1973) while at the same time experiencing an erosion of legitimacy (Kaase 1979). "Decisionmakers have to realize rationality despite complexity" (Schimank 2006). Tackling seemingly wicked problems with Big Data Analytics is an attractive option, as the technology promises to find solutions where humans fail (Goodspeed 2015). Further, 'data-driven' policies may be more easily communicated to the public, as Rüb & Straßheim (2012) point out the

rise of seemingly objective empirical “evidence” – particularly statistics and computer simulations – as a means of legitimizing policy. Given the uncertainty, ambiguity, and complexity of wicked problems, using seemingly objective decision-support systems like Big Data Analytics to legitimize decisions towards the public and towards the decisionmakers themselves is highly attractive (Rüb and Straßheim 2012).

However, the impact of Big Data on government can be assumed to go beyond its instrumental use for decision-support. Technological development have tremendous effects on the workings of public administration and the modes of governance (Pollitt 2011). This holds especially for information technologies: Scholars such as Porter (1996) and Scott (1998) have shown the historically close interrelation between the state’s ability to observe its society, to analyze these observations in a structured way, and the possible modes of governance and administration. For instance, Scott (1998) emphasizes that rendering idiosyncratic local contexts “legible” has been the precondition for their rule from a spatially and culturally distant central government, thus enabling the shift from feudal fiefdoms to nation states. In a similar vein, Weber (1922) presents “Herrschaft kraft Wissen” (“authority through knowledge”) and the necessary information collection and management as a central pillar of rational bureaucracy (Wolf 1988).

### Big Data as the herald of a digital public sector reform paradigm?

Big Data Analytics is hardly the first information technology to be introduced in public administration. However, Big Data in government is the first prominent manifestation of a new wave of IT- and data-driven innovations in policymaking, such as “data-driven policymaking” (Esty and Rushing 2007) “policy informatics” (Dawes and Janssen 2013) and “agile policymaking” (Parcell and Holden 2013). This latest push in digital government innovations feeds a long-standing expectation for information technology to fundamentally change government and public administration. For more than a decade, scholars of information technology in government such as Dunleavy et al. (2006) have heralded „digital-era governance“ as the successor of New Public Management as the dominant public sector reform paradigm, boasting: “New Public Management is dead – Long live digital-era governance” (Dunleavy et al. 2005). While traditional e-government has merely digitized existing administrative processes, concepts such as “infocracy” (Zuurmond 1998), “iGovernment” (Prins et al. 2011) suggest that the informatization and especially the systematic linking and analysis of governmental data sources gives rise to new ways of thinking and decision-making in government. This highlights the more general significance of the case of Big Data Analytics in government beyond its direct practical uses.

## Puzzle & Research Objective

As Pollitt (2011) states, the fundamental question studying technological change in public management always reads: What is the effect of technological change on administrative change?

Summarizing the dominant perspectives on the effects of information technology in government, Hood (2008) identifies two distinct motives in the extant answers to Pollitt's question: First, "transformative visions" that predict fundamental changes in the structures and processes of government, e.g. the "virtualization" of the state (Fountain 2001). These populate a continuum ranging between enthusiastic "techno-utopianism" (Wiig 2015) and "surveillance society" dystopias. Second, the motive of "dynamic conservatism" which predicts "technical modernization" (Capano 2003) without significant impact on the underlying social order.

These motives are also visible in in the nascent literature on Big Data in government:

- 1) "Transformative visions": Big Data in government as a paradigm-changing innovation that ushers in the age of "digital-era governance" (Dunleavy et al. 2005), enabling fundamentally new ways of decisionmaking and public management such as "agile policymaking" (Parcell and Holden 2013) and "policy informatics" (Dawes and Janssen 2013).

Here, we find critical perspectives warning against a surveillance state (Bannister 2005; Boellstorff 2014; Sciencewise 2014) as well as highly enthusiastic assessments. The latter is especially prominent in the consultancy literature on Big Data in the public sector, as Wiig (2015) works out with the example of the IBM Smart City concept.

- 2) "Dynamic conservatism": Big Data in government as a tool that does not affect the overall setup of government. Here, following the traditional e-government trinity of "faster, better, cheaper" (Sharma 2004), Big Data Analytics in government is exclusively seen as a way to increase efficiency, improve monitoring, or act as an early-warning system (Australian Government Information Management Office 2013; Deloitte 2011; IBM Institute for Business Value 2012; McKinsey Global Institute 2011; Yiu 2012).

Given the "embryonic state" (European Commission 2015) of Big Data in government, it is too early to examine its practice and impact. However, as governments around the world are actively evaluating Big Data, they establish the way they understand Big Data in government, shaping the sociotechnical assemblage and setting the tracks for their eventual practice of Big Data in government.

*Therefore, the overall objective is to understand how public administrations make sense of or construct Big Data Analytics in government.*

Further, the notion of an emerging digital public sector reform paradigm as well as different scholars (Dunleavy et al. 2006; Pollitt 2014) have argued for a global convergence of IT- and data-driven government. *Therefore, a secondary objective is to examine if different public administrations construct Big Data in government similarly or whether different contexts react differently.*

## Theoretical framework

Neo-institutionalist theory provides a well-developed theoretical framework to examine the emergence of Big Data as a new organizational practice. In contrast to the mostly rationalist theories in e-government research, new institutionalism assumes that organizational behavior is based on a 'logic of appropriateness' focused on achieving or maintaining legitimacy within an organizational field rather than maximizing their efficiency. In the case of Big Data Analytics, this allows a nuanced assessment as Big Data probably increases efficiency, but may, due to the heavy criticism directed at it, be considered inappropriate in many situations.

While e-government research has developed own theories inspired by institutionalist perspectives, particularly the Technology Enactment Framework (Cordella and Iannacci 2010; Fountain 2001), drawing on new institutionalism in organizational research provides a more mature theoretical framework. Orlikowski and Barley (2001) explicitly call for a stronger integration of information systems and institutional theory, highlighting the merits of institutionalist perspectives in addition to the mostly rationalist and technologically deterministic theories of information technology. Hood (2008) stresses that the "politics-of-instruments" perspective in policy analysis, which examines "what political, ideological, or cognitive processes lead to the choice of one policy instrument rather than another", is among the most important and most neglected in the study of information technology in government. In the same vein, Pollitt (2012) urges to study e-government issues with mainstream theories of public administration research, such as neoinstitutionalism, to end the "ghettoization" of e-government research. Applying neoinstitutionalist theory to the study of information technology, Nielsen, Matthiassen, and Newell (2008) have particularly shown the relevance of the concepts of "theorization" and "translation" for "IT institutionalization".

In institutional theory, institutions represent established social orders (Berger and Luckmann 1966), including management practices that have become taken-for-granted (Meyer and Rowan 1977). Institutionalization, the process of achieving taken-for-grantedness, has been one of the foci of institutional theory (Cooper, Ezzamel, and Willmott 2008). Of course, Big Data in government is far from being an institution or even a social practice on the road to institutionalization. However, given the ongoing hype around Big Data in government, it can be considered a proto-institution, i.e. "new practices, rules, and technologies" which "may become new institutions if they diffuse sufficiently"

(Lawrence, Hardy, and Phillips 2002, 281). Thus, “proto-institutions are candidates for institutionalization” (Zietsma and McKnight 2009).

### Institutional work and theorization

The process of proto-institutionalization is outlined by the framework of “institutional work”, which “describes the practices of individual and collective actors aimed at creating, maintaining, and disrupting institutions” (Lawrence, Suddaby, and Leca 2011). Institutional work allows greater consideration of agency in institutional processes, including actors beyond the heroic “institutional entrepreneur” (DiMaggio 1988; Lounsbury and Crumley 2007). In a nutshell, actors within an organization act out different strategies to demolish, maintain or create institutions, with the potential of several such conflicting or mutually reinforcing institutional projects at a time. With regard to Big Data Analytics in government, the creation of institutions is relevant (see Table 1 for a selection of possible strategies).

*Table 1: Forms of institutional work for creating institutions (Lawrence and Suddaby 2006, p. 221)*

<b>Forms of institutional work</b>	<b>Definition</b>
Advocacy	The mobilization of political and regulatory support through direct and deliberate techniques of social suasion
Defining	The construction of rule systems that confer status or identity, define boundaries of membership or create status hierarchies within a field
Vesting	The creation of rule structures that confer property rights
Constructing identities	Defining the relationship between an actor and the field in which that actor operates
Changing normative associations	Re-making the connections between sets of practices and the moral and cultural foundations for those practices
Constructing normative networks	Constructing of interorganizational connections through which practices become normatively sanctioned and which form the relevant peer group with respect to compliance, monitoring and evaluation
Mimicry	Associating new practices with existing sets of taken-for-granted practices, technologies and rules in order to ease adoption
Theorizing	The development and specification of abstract categories and the elaboration of chains of cause and effect
Educating	The educating of actors in skills and knowledge necessary to support the new institution

As Big Data in government is in an early stage of proto-institutionalization, its “theorization” (Strang and Meyer 1993) is of specific importance. In the context of institutional work, theorization refers to the “specification and explanation of characteristics and qualities of a practice as well as the definition of the ‘problems’ for which it is a suitable solution” (Walgenbach and Meyer 2008). Munir (2005) argues that technological innovations often act as exogenous ‘jolts’ to jumpstart institutional change, but may pass unnoticed if they are not rendered relevant to the organization through theorization. Given the nature of Big Data as a sociotechnical assemblage, the leeway for theorization – or social construction – is larger than for more material technologies.

Since actors within the framework institutional work act according to their interests and extant systems of meaning (or “cognitive-cultural institutions” (W. R. Scott 2000)), “several [actors] may engage in parallel institutional work, and find they are competing against, and impacted by, other actors sponsoring different arrangements” (Zietsma and McKnight 2009). Therefore, local mobilization structures and ‘domestic’ power constellations strongly influence which theorizations arise and eventually become dominant within an organization (Walgenbach and Meyer 2008).

Thus, an analysis of the theorization of Big Data Analytics in government has to take into account actor constellations and their respective interests and preferred systems of meaning.

### Translation and editing

However, actor involved in institutional work do not follow their interests unrestrictedly in the sense of rational choice theory. One of the fundamental insights of empirical research into institutionalization is that “novel organizational forms are most likely to become legitimated when they fit into the preexisting cultural beliefs, meanings, and typifications of an organizational community” (Ruef 2000), or if they are at least invoke “grand legitimizing myths” (Walgenbach and Meyer 2008) of our times, e.g. modernity, rationality, or progress.

Thus, when (proto-)institutions are adopted in a new organization, they are translated and edited to fit the local institutional setting: “Translation refers to the notion that ideas change when they travel from one context to another” (Boxenbaum and Strandgaard Pedersen 2009). According to Sahlin and Wedlin (2008), editing processes in institutional translation follow three rules: As a practice is dislocated from a setting, it is stripped of time- and space-bound features. Its logic of initiative and effect is reconstructed. And it acquires specific labels and is dramatized in a certain language. In other words: It is re-theorized.

Therefore, it can be assumed that the theorization of Big Data Analytics differs greatly among contexts. For instance, Boswell (2009) has found that the theorization and use of expert knowledge – not entirely unrelated to Big Data Analytics – differs significantly among policy fields, tasks of public

sector organizations and administrative traditions. However, “what initially appears as significant variation across organizations may, upon scrutiny, turn out to be only slightly different versions of a similar organizational form, and vice versa” (Boxenbaum and Strandgaard Pedersen 2009).

Nonetheless, institutional legacies have to be taken into account when analyzing convergence or divergence in the theorization of Big Data in government.

### Research gaps

Unsurprisingly, the emergence of Big Data Analytics in government has not been the subject of institutional analysis yet. However, Puschmann and Burgess (2014) more generally examine metaphors used in talking about Big Data and identify two distinct framings: Big Data “as a natural force to be controlled and as a resource to be consumed”. Ruppert et al.'s (2015) exploration of three use cases of Big Data in the United Kingdom leads them to abandon the purely technical definitions of Big Data and instead describe it as a “socio-technical assemblage”, which is significantly shaped by the meanings and practices attached to it by its users. Both studies highlight that institutional analysis can yield crucial insights to understand the practice and future of Big Data in government.

With regard to institutional theory, the proposed theoretical framework holds merit on several accounts: By looking inside organizations rather than an entire organizational field, it answers the lament that “the examination of internal organizational processes has been a neglected aspect of institutional analysis” (Lounsbury and Crumley 2007). Examining institutional work in proto-institutionalization aims to redeem that “we know little about the processes by which new institutional innovations emerge, compete, and resolve into shared logics and practices over time” (Zietsma and McKnight 2009) and that “we lack accounts of the process by which the assumptions that define institutional logics are contested and changed” (Suddaby and Greenwood 2005).

### Research Questions

Informing the research objectives with the proposed theoretical framework, three research questions are stated:

1. *How is Big Data Analytics theorized in public administration?*  
What clusters of theorization exist? Who are the key actors?
2. *How is the proto-institution of Big Data Analytics in government edited and translated?*  
Which existing local and global institutions and actors' interests do the theorizations link to?

3. *What are the patterns of theorization?*

What is the influence of institutional legacies and contextual factors? Is there institutional divergence or convergence?

## Research Design

The overall research design is an exploratory comparative multiple case study employing qualitative content analysis to analyze the discursive dynamics in the proto-institutionalization of Big Data Analytics in three European metropolitan governments.

According to Yin (2009), case study research is appropriate when “a how or why question is being asked about a contemporary set of events over which the investigator has little or no control”. Only qualitative case studies are able to produce the “thick analysis” (Collier, Brady, and Seawright 2010) necessary to reconstruct complex social processes. “The qualitative analysis of a few cases naturally has the advantage that it is possible to give a far more penetrating and variegated picture of the situation under study. It is thus feasible to follow in detail the processes which contribute to (or prevent) change in the respects of interest. Furthermore, such studies normally need not be as firmly structured beforehand as those quantitatively oriented. Through this more flexible approach the researcher is in a better position to find new trails and clues along the way. The scope for ‘discoveries’ is therefore greater.” (Hadenius 1992)

As this dissertation is mostly concerned with the theorization of Big Data Analytics in government, it is an interpretive study of such complex social processes in the Weberian sense of “Verstehen” (Muno 2009). It aims to trace and understand the meanings and motivations of the involved actors rather than coming up with external causal explanations in the sense of “Erklären”. Therefore, the research design is not geared towards statistical generalizability (or extrapolation) but analytical generalizability or, more modestly, plausibility (Maxwell 2008). Although some scholars deny the usefulness of case comparison for interpretive inquiries (Muno 2009), Scharpf (2002) holds that interpretive small-n comparisons are conducive to the “reconstruction” of social phenomena through “isolation and generalization” of common elements and mechanisms. Thus, interpretive small-n comparisons can achieve “contingent generalizability” (Scharpf 2002) that goes beyond the analytical generalizability of single case studies.

## Case selection

Three European metropolitan governments serve as comparative cases for the dissertation.

Metropolitan governments are promising cases to study Big Data Analytics for several reasons. For one, urban governments have been pioneers of embedding information technologies into public

administration for several decades (Kitchin 2013). Cities are therefore an ideal site to study these pioneering efforts: Most public services are produced and delivered at the local level, which provides an array of possible applications for information technology. High population density in metropolitan areas not only enables economies of scale in the application of technological infrastructure like sensor systems, but as well increases the likelihood of ‘wicked problems’, which may be tackled with digital solutions.

As different as the governance of metropolitan cities is organized across Europe, they are all characterized by a strong proximity of political leadership, public service delivery and the ‘problems’ the political-administrative system is aiming at (Lefèvre 1998). This typically results in strong political legitimacy and autonomy as well as wide-ranging jurisdiction of metropolitan governments (Sharpe 1995). This enables metropolitan governments to experiment more encompassingly with new technologies and policy instruments than other government units.

As Sassen (1991, 2002) argues with the concept of “global cities”, metropolises are the nodes of the globalized world. They are therefore not only most susceptible to new societal and technological trends, but also in a privileged position to shape these trends. Therefore, the discourse on Big Data in government in Europe’s metropolises is worthwhile analyzing.

The tentatively selected three metropolitan governments to be examined in this research project are:

- 1) Vienna, Austria
- 2) Amsterdam, Netherlands
- 3) London, United Kingdom

Case selection follows a purposeful sampling rationale to identify “information-rich cases” (Patton 1990), which means that the cases are deliberately selected for the presence of the phenomenon of interest and a variation in contextual factors. “Purposeful sampling for demographic homogeneity and selected phenomenal variation is a way a researcher working alone with limited resources can reduce the minimum number of sampling units required within the confines of a single research project, but still produce credible and analytically [...] significant findings” (Sandelowski 1995).

Restricting the case selection to ‘Northern’ EU member states ensures general homogeneity and thus comparability in terms of macroeconomic performance, technological development, government capacity and democratic governance. All three cities are the capitals as well as the largest urban agglomerations of their respective countries, with 2.6 million inhabitants in Vienna, 2.4 million in Amsterdam, and 13 million in London. They are all internationally renowned for their “Smart City” concepts (Caragliu, Del Bo, and Nijkamp 2011; Gil-Garcia, Pardo, and Aldama-Nalda 2013), which

explicitly highlight the use of Big Data Analytics in government. Initial research has confirmed that each of the three cities is actively using or evaluating urban analytics for different purposes, ensuring that the phenomenon can indeed be studied there.

This case selection achieves “phenomenal variation” (Sandelowski 1995) by purposefully covering different administrative traditions, from a still largely Weberian ‘Rechtstaat’ tradition in Austria (Hammerschmid and Meyer 2003) to a ‘public interest’ tradition that is strongly characterized by New Public Management in the United Kingdom, with the Netherlands as an ‘intermediate’ type inbetween the two (Kuhlmann and Wollmann 2013; Pollitt and Bouckaert 2011). Additionally, the general stance towards information technology innovations, specifically Big Data Analytics, varies from a more cautious approach in Austria to a readiness to experiment in the Netherlands and a clear focus on the possible benefits in the United Kingdom. Thus, the three cases provide promising conditions to observe the editing and translation in the proto-institutionalization of Big Data Analytics in government.

### Data collection

The theorization of Big Data Analytics can be best studied by examining the “organizational discourse” (Lawrence and Suddaby 2006) within each city administration. The reconstruction of this discourse draws on three major sources:

- 1) All accessible documents produced by the city administration dealing with the topic of Big Data Analytics, e.g. meeting minutes, policy documents, evaluation reports, press statements
- 2) Interviews with politicians and public managers of the city administrations who are involved in the evaluation processes or pilot projects.
- 3) Publications about Big Data in government in specialized outlets, e.g. public sector magazines, government technology blogs, and consultancy reports, which are identified as relevant by the interviewees.

For each city, about 12-20 interviews will be conducted in a semi-structured manner. Interviewees are selected both based on snowball sampling as well as the accessible documents. While the guiding questions ensure that different aspects, criticisms and examples of Big Data in government are visited in all interviews, the interviews are generally conducted in an open fashion to elicit the meanings and narratives of the interviewees.

Overall, triangulation is a lesser concern for this kind of interpretive approach. However, by combining interviews and documents from different sources, statements and positions can be more reliably reconstructed and attributed to the right actors.

Access may become an issue as Big Data in government is a potentially touchy subject: “In the case of organizations engaged in surveillance, practitioners are often keenly aware that any negative representation of their activities could invite public backlash, legal action, or dissolution of their organizations” (Monahan and Fisher 2014). However, the research design is flexible enough to substitute cases if access attempts fails.

### Data analysis

The collected data is analyzed with conventional qualitative content analysis within a discourse analytic framework adapted from Hajer (2006) to identify narratives, metaphors, argumentative patterns, etc. and map these to discourse coalitions.

Lawrence and Suddaby (2006) recommend discourse analysis to study instances of institutional work since “many forms of institutional work [...] involve practices of speaking and writing that are aimed at affecting the institutional context within which those practices occur.” Similarly, Kitchener (2002) explicitly stresses that theorization is narrative.

Adapting Hajer's (2006) ten-step process to the case of proto-institutionalization of Big Data Analytics in government, the analysis follows these steps:

- 1) *Desk research*: Survey of all available documents and media reports about the phenomenon in question to achieve a first overview and possibly a preliminary chronology.
- 2) *Helicopter interviews*: Expert interviews with three or four key informants who have an overview of the phenomenon in question from different perspectives to improve orientation.
- 3) *Document analysis*: First approach to structure concepts, ideas and categorizations, storylines, and metaphors. Map sites, process and actors of the discourse.
- 4) *Interview with key players*: Reconstruct their perspective on the discursive process and elicit their positions and interpretations on the phenomenon in question.
- 5) *Interpretation*: Discourse analysis of the documents and interviews combined.
- 6) *Second visit to the key actors*: Do they recognize and agree with the structuring of the discourse?

Unlike many variants of discourse analysis, Hajer's approach links the discourse to identifiable actors (as part of “discourse coalitions” (Hajer 1993)), which fits well with the actor orientation of institutional work. Distinguishing which actors champion which theorizations is crucial to understand the different proto-institutionalization processes.

The interpretation of the discourse is conducted with software-aided (e.g. MaxQDA or Atlas.ti) conventional qualitative content analysis (Hsieh and Shannon 2005). This means an inductive process of open coding where categories emerge from the text itself. Thus, the reconstructed theorizations

of Big Data Analytics in government are based on the actors' meanings rather than pre-existing theoretical frames (Kohlbacher 2006). However, in a later step, these categories may be grouped into theoretically informed meta-categories, e.g. public sector reform paradigms like New Public Management and Public Administration (Hyndman et al. 2014), to render the analysis compatible with extant literature.

Additionally coding contextual factors like actors, administrative traditions, policy fields, types of task (e.g. technical vs political) allows to trace institutional work, translation, and editing of the proto-institutionalization of Big Data in government through co-occurrence analysis (Harwood and Garry 2003; Kelle 2004).

Overall, the data analysis identifies patterns in the theorizations of Big Data in government, allowing conclusions about which institutional conditions and actors' interests lead to which understandings of Big Data in government.

## Work plan

### Slicing of cumulative dissertation

The dissertation is to be structured as a cumulative dissertation of 3-4 articles. A straightforward slicing option is to write one article laying out the theoretical framework, one analyzing the global discourse on Big Data in government and one to two articles on the local translations of Big Data in the examined metropolitan governments.

### Timetable (tentative)

1. Establishing access to selected cases (August – October 2015)
2. Desk research & helicopter interviews (August – December 2015)
3. Document analysis & interviews with key players (November 2015 – August 2016)
4. Content Analysis (March – December 2016)
5. Co-occurrence analysis (January – March 2017)
6. Consolidation and presentation of findings (March – June 2017)

## Summary

Big Data Analytics is considered one of the most relevant technological innovations for government and public administration. As a means of 'evidence-based' and 'objective' decision-support for government, Big Data promises to solve wicked problems by cutting through complexity and reducing uncertainty. However, projections of the scope of its impact on government range from mere technical modernization to paradigmatic revolutions.

To shed light on this issue, this dissertation aims to understand how public administrations make sense of Big Data Analytics in government. Drawing on neo-institutional theory, Big Data Analytics in government is considered a case of proto-institutionalization, i.e. a candidate for institutionalization that is still subject to struggles over its definition. Highlighting the institutional processes of ‘theorization’ and ‘translation’, the influence of actor constellations and locally pre-existing institutional arrangement in the discursive construction of Big Data in government is stressed.

These processes are then examined in three European metropolitan governments: Vienna, Amsterdam, and London. Structuring the discursive theorization through qualitative content analysis of accessible government documents, interviews with key actors, and relevant third-party publications, clusters and patterns of theorization are worked out. Further, co-occurrence analysis allows the systematic mapping of specific theorizations to actors, contextual factors like policy fields and types of tasks, as well as institutional legacies such as administrative traditions.

Thus, in showing how and why Big Data Analytics in government are theorized in distinct manners, this dissertation is among the first systematic empirical inquiries to yield a deeper understanding of the practice of Big Data in government and to make qualified statements about the scope of the entailing administrative change.

On a conceptual level, this dissertation furthers the integration of e-government research into mainstream public administration research by applying neoinstitutionalism as a mainstream theory to an e-government phenomenon. Within institutional theory, it contributes to the study of proto-institutions and institutionalization processes within organizations.

## References

- Abrahamson, Eric. 1996. “Management Fashion.” *Academy of Management Review* 21(1): 254–85.
- Anderson, Chris. 2008a. “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.” *Wired Magazine*.
- . 2008b. “The Petabyte Age: Because More Isn’t Just More—more Is Different.” *Wired Magazine*: 106–20.
- Australian Government Information Management Office. 2013. *The Australian Public Service Big Data Strategy*.
- Bannister, Frank. 2005. “The Panoptic State: Privacy, Surveillance and the Balance of Risk.” *Information Polity: The International Journal of Government & Democracy in the Information Age* 10(1,2): 65–80.

- Berger, Peter L., and Thomas Luckmann. 1966. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*.
- Boellstorff, Tom. 2014. "Die Konstruktion von Big Data in Der Theorie." In *Big Data: Analysen Zum Digitalen Wandel von Wissen, Macht Und Ökonomie*, ed. Ramón Reichert. , 105–31.
- Bollier, David. 2010. *The Promise and Peril of Big Data*. Aspen Institute.
- Boswell, Christina. 2009. *The Political Uses of Expert Knowledge: Immigration Policy and Social Research*.
- Boxenbaum, Eva, and Jesper Strandgaard Pedersen. 2009. Institutional work: Actors and agency in institutional studies of organization *Scandinavian Institutionalism—a Case of Institutional Work*. eds. Thomas B. Lawrence, Roy Suddaby, and Bernard Leca. Cambridge University Press Cambridge,, UK.
- Capano, Gilberto. 2003. "Administrative Traditions and Policy Change: When Policy Paradigm Matters. The Case of Italian Administrative Reform During the 1990s." *Public Administration* 81(4): 781–810.
- Caragliu, Andrea, Chiara Del Bo, and Peter Nijkamp. 2011. "Smart Cities in Europe." *Journal of Urban Technology* 18(2): 65–82.  
<http://www.tandfonline.com/doi/abs/10.1080/10630732.2011.601117> (July 22, 2014).
- Collier, David, Henry E. Brady, and Jason Seawright. 2010. "Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, eds. David Collier and Henry E. Brady. , 161–200.
- Cooper, David J., Mahmoud Ezzamel, and Hugh Willmott. 2008. "Examining 'Institutionalization': A Critical Theoretic Perspective." In *The SAGE Handbook of Organizational Institutionalism*, , 573–96.
- Cordella, Antonio, and Federico Iannacci. 2010. "Information Systems in the Public Sector: The E-Government Enactment Framework." *Journal of Strategic Information Systems* 19(1): 52–66.
- Cukier, Kenneth, and Viktor Mayer-Schönberger. 2013. "The Rise of Big Data: How It's Changing the Way We Think about the World." *Foreign Affairs* 92.
- Dawes, SS, and Marijn Janssen. 2013. "Policy Informatics: Addressing Complex Problems with Rich Data, Computational Tools, and Stakeholder Engagement." *Proceedings of the 14th Annual International Conference on Digital Government Research*: 251–53.
- Deloitte. 2011. *Deloitte Analytics –Insight on Tap: Improving Public Services through Data Analytics*.
- DiMaggio, Paul. 1988. "Interest and Agency in Institutional Theory." In *Institutional Patterns and Organizations: Culture and Environment*, ed. L. Zucker. , 3–22.
- Dolata, Ulrich, and Raymund Werle. 2007. "' Bringing Technology Back In': Technik Als Einflussfaktor Sozioökonomischen Und Institutionellen Wandels." In *Gesellschaft Und Die Macht Der Technik*, eds. Ulrich Dolata and Raymund Werle. Campus Verlag, 15–43.

- Dunleavy, Patrick, Helen Margetts, S Bastow, and J Tinkler. 2005. "New Public Management Is Dead—long Live Digital-Era Governance." *Journal of Public Administration Research and Theory* 16: 467–94.
- Dunleavy, Patrick, Helen Margetts, Simon Bastow, and Jane Tinkler. 2006. *Digital Era Governance: IT Corporations, the State, and E-Government*. Oxford University Press.
- Esty, Daniel C., and Reece Rushing. 2007. *Governing by the Numbers. The Promise of Data-Driven Policymaking in the Information Age*. Center for American Progress.  
<http://www.ncbi.nlm.nih.gov/pubmed/22212954>.
- European Commission. 2015. *Future-Proofing eGovernment for a Digital Single Market Future-Proofing eGovernment for a Digital Single Market*.
- Fountain, JE. 2001. *Building the Virtual State: Information Technology and Institutional Change*.
- Gil-Garcia, J. Ramon, Theresa a. Pardo, and Armando Aldama-Nalda. 2013. "Smart Cities and Smart Governments." *Proceedings of the 14th Annual International Conference on Digital Government Research - dg.o '13*: 296. <http://dl.acm.org/citation.cfm?id=2479724.2479728>.
- Goodspeed, R. 2015. "Smart Cities: Moving beyond Urban Cybernetics to Tackle Wicked Problems." *Cambridge Journal of Regions, Economy and Society* 8(1): 79–92.
- Hadenius, Axel. 1992. *Democracy and Development*. Cambridge University Press.
- Hajer, Maarten. 1993. "Discourse Coalitions and the Institutionalization of Practice: The Case of Acid Rain in Great Britain." In *The Argumentative Turn in Policy Analysis and Planning*, eds. Frank Fischer and John Forester. Duke University Press Durham, NC, 43–76.
- Hajer, Maarten a. 2006. "Doing Discourse Analysis: Coalitions, Practices, Meaning." *Words matter in policy and planning. Discourse theory and method in the social sciences*: 65–74.
- Hammerschmid, Gerhard, and Renate Meyer. 2003. "Current Initiatives and Central Actors in the Austrian Administrative Reform Process." In *EGPA Annual Conference, Oeiras/Portugal, 3 to 6 September 2003*, , 1–22.
- Harwood, Tracy G., and Tony Garry. 2003. "An Overview of Content Analysis." *The Marketing Review* 3(4): 479–98.
- Hood, Christopher. 2008. "The Tools of Government in the Information Age." In *The Oxford Handbook of Public Policy*, eds. Michael Moran, Martin Rein, and Robert E. Goodin. , 469–81.
- Hsieh, Hsiu-Fang, and Sarah E Shannon. 2005. "Three Approaches to Qualitative Content Analysis." *Qualitative health research* 15(9): 1277–88.
- Hyndman, Noel, Mariannunziata Liguori, Renate E. Meyer, Tobias Polzer, Silvia Rota, and Johann Seiwald. 2014. "The Translation and Sedimentation of Accounting Reforms. A Comparison of the UK, Austrian and Italian Experiences." *Critical Perspectives on Accounting* 25(4-5): 388–408.
- IBM Institute for Business Value. 2012. *Analytics: The Real-World Use of Big Data*.

- Kaase, Max. 1979. "Legitimitätskrise in Westlichen Demokratischen Industriegesellschaften: Mythos Oder Realität." In *Wertwandel Und Gesellschaftlicher Wandel*, eds. Helmut Klages and Peter Kmiecik. , 328–50.
- Kelle, Udo. 2004. "Computer-Assisted Qualitative Data Analysis." In *Qualitative Research Practice*, Sage Publications London, 473–89.
- Kitchin, Rob. 2013. "The Real-Time City? Big Data and Smart Urbanism." *GeoJournal* 79(1): 1–14.
- Kohlbacher, Florian. 2006. "The Use of Qualitative Content Analysis in Case Study Research." *Forum: Qualitative Social Research* 7(1).
- Kuhlmann, Sabine, and Hellmut Wollmann. 2013. *Verwaltung Und Verwaltungsreformen in Europa*.
- Lawrence, Thomas B, C Hardy, and N. Phillips. 2002. "Institutional Effects of Interorganizational Collaborations: The Emergency of Proto-Institutions." *Academy of Management Journal* 45(1): 281–90.
- Lawrence, Thomas B, R. Suddaby, and B. Leca. 2011. "Institutional Work: Refocusing Institutional Studies of Organization." *Journal of Management Inquiry* 20(1): 52–58.
- Lawrence, Thomas B, and Roy Suddaby. 2006. "Institutions and Institutional Work." In *The SAGE Handbook of Organization Studies*, eds. Stewart R Clegg, Cynthia Hardy, Tom Lawrence, and Walter R Nord. , 215–54.
- Lefèvre, Christian. 1998. "Metropolitan Government and Governance in Western Countries : A Critical Review." *International Journal of Urban and Regional Research* Volume 22(Issue 1): pages 9–25.
- Lenk, Klaus, and Peter Wengelowski. 2004. "Wissensmanagement Für Das Verwaltungshandeln." In *Wissensmanagement in Politik Und Verwaltung*, eds. Werner Jann and Thomas Edeling. , 147–65.
- Lounsbury, M., and E. T. Crumley. 2007. "New Practice Creation: An Institutional Perspective on Innovation." *Organization Studies* 28(7): 993–1012.
- Maxwell, Joseph A. 2008. "Designing a Qualitative Study." In *Handbook of Applied Social Research Methods*, eds. Leonard Bickman and Debra Rog. , 214–53. [http://www.corwin.com/upm-data/23772\\_Ch7.pdf](http://www.corwin.com/upm-data/23772_Ch7.pdf) (October 5, 2012).
- Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Eamon Dolan/Houghton Mifflin Harcourt.
- McKinsey Global Institute. 2011. *Big Data: The next Frontier for Innovation, Competition, and Productivity*.
- Meyer, John W., and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83(2): 340.
- Monahan, Torin, and Jill a Fisher. 2014. "Strategies for Obtaining Access to Secretive or Guarded Research Sites." *Journal of Contemporary Ethnography* (Monahan): 1–28.

- Munir, Kamal A. 2005. "The Social Construction of Events: A Study of Institutional Change in the Photographic Field." *Organization Studies* 26(1): 93–112.
- Muno, Wolfgang. 2009. "Fallstudien Und Die Vergleichende Methode." In *Methoden Der Vergleichenden Politik- Und Sozialwissenschaft*, eds. Susanne Pickel, Gert Pickel, Hans-Joachim Lauth, and Detlef Jahn. , 113–31.
- Nielsen, Jeppe Agger, Lars Matthiassen, and Sue Newell. 2008. "Theorization and Translation in Information Technology Institutionalization: Evidence from Danish Home Care." *MIS Quarterly* 32(1): 67–96.
- Orlikowski, Wanda J., and Stephen R. Barley. 2001. "Technology and Institutions : What Can Technology Research on Information Technology and Research on Organizations Learn From Each Other?" *MIS Quarterly* 25(2): 145–65.
- Parcell, Jacob, and SH Holden. 2013. "Agile Policy Development for Digital Government: An Exploratory Case Study." *Proceedings of the 14th Annual International Conference on Digital Government Research*: 11–17.
- Patton, Michael Quinn. 1990. *Qualitative Evaluation and Research Methods*.
- Poel, Martijn, Ralph Schroeder, Jérôme Trepermann, Mor Rubinstein, Eric Meyer, Bea Mahieu, Chiel Scholten, and Marina Svetachova. 2015. *Data for Policy: A Study of Big Data and Other Innovative Data-Driven Approaches for Evidence-Informed Policymaking (Draft)*.
- Pollitt, Christopher. 2011. "Mainstreaming Technological Change in the Study of Public Management." *Public Policy and Administration* 26(4): 377–97.
- . 2012. *New Perspectives on Public Services: Place and Technology*. Oxford University Press.
- . 2014. *Future Trends in European Public Administration and Management: An Outside-in Perspective*.
- Pollitt, Christopher, and Geert Bouckaert. 2011. *Public Management Reform: A Comparative Analysis—New Public Management, Governance, and the Neo-Weberian State*. 3rd ed.
- Porter, Theodore M. 1996. *Trust in Numbers : The Pursuit of Objectivity in Science and Public Life*. Princeton, N.J.: Princeton Univ.Press.
- Prins, Corien, D. Broeders, H. Griffioen, A.G. Keijzer, and E. Keymolen. 2011. *iGovernment*. Amsterdam University Press.
- Puschmann, Cornelius, and J Burgess. 2014. "Metaphors of Big Data." *International Journal of Communication* 8: 1690–1709.
- Rittel, HWJ, and MM Webber. 1973. "Dilemmas in a General Theory of Planning." *Policy sciences* 4(2): 155–69.
- Rogers, E. M. 2010. *Diffusion of Innovations*.

- Rüb, Friedbert W., and Holger Straßheim. 2012. "Politische Evidenz – Rechtfertigung Durch Verobjektivierung?" In *Der Aufstieg Der Legitimitätspolitik. Rechtfertigung Und Kritik Politisch-Ökonomischer Ordnungen*, , 377–98.
- Ruef, Martin. 2000. "The Emergence of Organizational Forms: A Community Ecology Approach." *American Journal of Sociology* 106(3): 658–714.
- Ruppert, Evelyn, Penny Harvey, Celia Lury, Adrian Mackenzie, Ruth McNally, Alice Stephanie Baker, Yannis Kallianos, and Camilla Lewis. 2015. *Socialising Big Data: From Concept to Practice*.
- Sahlin, Kerstin, and Linda Wedlin. 2008. "Circulating Ideas: Imitation, Translation and Editing." In *The SAGE Handbook of Organizational Institutionalism*, , 218–42.
- Sandelowski, Margarete. 1995. "Focus on Qualitative Methods: Sample Size in Qualitative Research." *Research in Nursing & Health* 18: 179–83.
- Sassen, Saskia. 1991. *The Global City*. <https://op.kulib.kyoto-u.ac.jp/webopac/catdbi.do?pkey=BB02333006>.
- . 2002. *10 Secretary Global Networks, Linked Cities*. <http://www.amazon.com/dp/0415931630>.
- Scharpf, Fritz W. 2002. "Kontingente Generalisierung in Der Politikforschung." In *Akteure – Mechanismen – Modelle: Zur Theoriefähigkeit Makro-Sozialer Analysen*, ed. Renate Mayntz. , 213–35.
- Schimank, Uwe. 2006. "Rationalitätsfiktionen in Der Entscheidungsgesellschaft." In *Zur Kritik Der Wissensgesellschaft*, eds. Dirk Tänzler, Hubert Knoblauch, and Hans-Georg Soeffner. , 57–81.
- Sciencewise. 2014. *Big Data: Public Views on the Collection, Sharing and Use of Personal Data by Governmnet and Companies*.
- Scott, James C. 1998. *Seeing Like a State*. Yale University Press.
- Scott, William Richard. 2000. *Institutions and Organizations*.
- Sharma, Pankaj. 2004. *E-Governance*. APH Publishing.
- Sharpe, Laurence James. 1995. *The Government of World Cities: The Future of the Metro Model*. Wiley.
- Strang, David, and John W. Meyer. 1993. "Institutional Conditions for Diffusion." *Theory and Society* 22(4): 487–511.
- Suddaby, Roy, and Royston Greenwood. 2005. "Rhetorical Strategies of Legitimacy." *Administrative Science Quarterly* 50(1): 35–67.
- Walgenbach, and Meyer. 2008. *Neoinstitutionalistische Organisationstheorie*.
- Wang, Ping. 2010. "Chasing the Hottest IT: Effects of Information Technology Fashion on Organizations." *MIS Quarterly* 34(1): 63–85.

- Ward, Jonathan Stuart, and Adam Barker. 2013. "Undefined By Data: A Survey of Big Data Definitions." *cs.DB*. <http://arxiv.org/abs/1309.5821> (January 21, 2014).
- Weber, Max. 1922. *Grundriss der Sozialökonomik ; 3. Abt Wirtschaft Und Gesellschaft*.
- Wiig, Alan. 2015. "IBM's Smart City as Techno-Utopian Policy Mobility." *City* 19(2-3): 258–73. <http://www.tandfonline.com/doi/full/10.1080/13604813.2015.1016275>.
- Wolf, Rainer. 1988. "„Herrschaft Kraft Wissen" in Der Risikogesellschaft." *Soziale Welt* 39(2): 164–87.
- Yin, Robert K. 2009. *Case Study Research: Design and Methods*. Sage.
- Yiu, Chris. 2012. *The Big Data Opportunity: Making Government Faster, Smarter and More Personal*. Policy Exchange.
- Zietsma, Charlene, and Brent McKnight. 2009. "Building the Iron Cage: Institutional Creation Work in the Context of." In *Institutional Work: Actors and Agency in Institutional Studies of Organizations*, eds. Thomas B. Lawrence, Roy Suddaby, and Bernard Leca. , 143–77.
- Zuurmond, A. 1998. "From Bureaucracy to Infocracy; Are Democratic Institutions Lagging Behind?" In *Public Administration in an Information Age: A Handbook*, eds. I.Th.M. Snellen and W.B.H.J. van de Donk. , 259–71.