

THE DARK SIDE OF DATA INTENSIVE PROJECTS: FUNCTION CREEP AND STAKEHOLDER CREEP

Research paper

Aaen, Jon, Aalborg University, Aalborg, DK, jonaaen@dps.aau.dk

Nielsen, Jeppe Agger, Aalborg University, Aalborg, DK, agger@dps.aau.dk

Abstract

In this article, we investigate the unintended consequences of initiating a large data analytic (DA) initiative in healthcare. DA is promising in terms of aiding decision making both for healthcare decisions and administrative decisions. However, while the (re)use of large amounts of data has lots of potential, it also has a dark side in terms of privacy, ethics, and control. Through the analysis of a longitudinal case study (2003-2015) of the emergence, expansion, and collapse of a large-scale DA project in the Danish healthcare sector, we demonstrate how high expectations for DA caused a surge of stakeholders and features that ultimately resulted in the project's demise. We call these surges stakeholder creep and function creep and theorize their emergence and consequences for DA in organizations.

Keywords: Data Analytics, Project Management, Function Creep, Stakeholder Creep.

1 Introduction

The idea of using large amounts of data and technologies that are able to uncover hidden information in these data (data analytics) is increasingly viewed as the engine behind an improved decision-making process (Brynjolfsson, Hitt, & Kim, 2011; Chen, Chiang, Storey, 2012, McAfee et al., 2012; Khan & Vorley, 2017). Data Analytics (DA), and its counterpart Business Intelligence, is promoted as today's "Digital Oil" (Yi, Liu, Liu, & Jin, 2014; Sivarajah, Kamal, Irani, & Weerakkody, 2017), and it is one of the most hyped technological innovations, which contains the seeds of new, valuable operational insights for private firms and public-sector organizations (Mayer-Schönberger & Cukier, 2013). In the healthcare sector, the increased use of health information technology (HIT) has produced massive amounts of patient data that come from medical records, such as electronic healthcare journals, as well as external data sources, such as social media. Consequently, from the dominant, technology optimist view, DA should be capable of processing an immense volume of data across a wide range of healthcare platforms to improve decision making and create value in healthcare organizations in terms of better diagnoses and improved courses of treatment (Srinivasan & Arunasalam, 2013; Raghupathi & Raghupathi, 2014; Margolis et al., 2014; Wang & Hajli, 2017). McKinsey estimated that better use of data could create new value of more than \$300 billion a year in the American healthcare system (Manyika et al., 2011).

Despite these optimistic and functionalistic perspectives on data analytics, this topic is not shy of critics. A less developed but equally important research stream focuses on "the dark side" of data use. Edward Snowden's exposure of Western intelligence agencies' systematic mass surveillance of citizens' digital activities served as a warning as to how the misuse of data can lead to comprehensive control and "big brother" societies (Greenwald, 2014). When Mayer-Schönberger & Cukier (2013) reflect on the dark side of data use, they refer to "dictatorship of data," in which the results of data analysis are accepted

uncritically without consideration of the validity of the underlying algorithms. Further, privacy has become much more difficult to protect given the numerous ways in which data may be reused, repurposed, and sold to other organizations. This often makes it impossible for users to give informed consent to “innovative secondary uses” that are not even imagined when the data is first collected (Mayer-Schönberger & Cukier, 2013). Paradoxically, the same qualities that make data so appreciated – their volume, their value for a plurality of stakeholders, and the capacity to be indefinitely reused – can also undermine privacy, obscure data-validity, have ethical implications for the citizens, and raise organizational and legal issues about control, responsibility, and transparency.

In contrast to this prevailing, enthusiastic view on big data and DA, this article contributes with new knowledge about the unintended consequences of DA in healthcare projects. We report the results of a longitudinal case study of the emergence, expansion, and collapse of a large-scale DA project in the Danish healthcare sector. DAMD (Dansk AlmenMedicinsk Database) was a Danish multi-million-dollar project aimed at collecting and analyzing healthcare data about all patients associated with general practitioners (which in Denmark is the entire population). Since DAMD seemed to be a panacea for a wide variety of goals and interests, it attracted more and more stakeholders, which had more and more requests for the DA system. The database was called a “golden egg” by health professionals (Fischer & Tynell, 2014), “indispensable” by the research community (Vestergaard et al. 2014), and a “game-changer” by the pharmaceutical industry (Lassen, 2013). Yet, the database was shut down in 2014 under considerable media and political attention when doubt was raised about whether the extensive data collection had the necessary legal basis. Hence, the reuse of health data for multiple purposes to serve different stakeholder interests ended up “muddying the waters” with less transparency, unclear organizational boundaries, and eventually illegal use of data. The fate of the DAMD project raises many questions: How can a seemingly successful data analytics project with so many potential benefits collapse? Was it a victim of its own success? In this paper, we investigate the following research question: *How can we understand the mechanisms that caused the failure of the DAMD initiative?*

In unraveling the mechanisms behind the DAMD collapse, our longitudinal case study reveals how divergent interests and high expectations for the DA initiative caused a surge of both stakeholders and requested features that ultimately resulted in demise of the data intensive project. We conceptualize the surge of requested system features as *function creep* and surges of stakeholders as *stakeholder creep* and discuss the existence of a sweet spot for governing DA projects with multiple stakeholders and continual request for new functionalities. We ground our theorizing of function creep and stakeholder creep in the DAMD case and connect to existing literature on IT and data governance to further develop and support our argument. Our study expands current research by adding to the scarce literature on the dark side of data intensive project.

2 Data Intensive Projects: Literature Background

Recent literature reviews show that research on DA and big data tend to focus either on the new opportunities and benefits that data provides organizations or the technical issues related to DA and big data, focusing on improving data models and algorithms (Ahangama & Poo, 2015; Saltz & Shamsurin, 2016; Günther et al., 2017). Against this backdrop, Günther et al., (2017) calls for scholarly papers that examine how different actors work with large amounts of data in practice and how to deal “with different stakeholder interests” (p. 10) to realize value from DA projects. Reaching a similar conclusion, Van den Broek & Van Veenstra (2015) suggest that “further research should look into how control over data can be organized in inter-organizational data collaborations in order to allow for data sharing in a responsible manner” (p. 11). Inspired by these scholarly calls, we zoom-in on a problematic implication of the (re)use of data in an inter-organizational setting with multiple stakeholder interests at play and data with high sensibility.

Günther et al. (2017) argues that two main characteristics of (big) data analytics are portability, i.e. the possibility to transfer and remotely access digitized data to new contexts, and interconnectivity, i.e. the possibility to synthesize data from various sources. As such, the management of data projects is often rooted in an increased awareness that data collected for one purpose can also be (re)used for many other

purposes (*ibid.*), which has resulted in DA approaches associated with “data lakes” or “data hubs” (Constantiou and Kallinikos, 2015; Miloslavskaya & Tolstoy, 2016). However, tapping into a data lake by reusing the data for multiple purposes to serve different stakeholder interests may not be unproblematic. In general, managing IT projects that involve multiple stakeholders with different functions and interests is considered notoriously difficult since “stakeholder groups are typically socialized differently and may have different perspectives” (Boonstra, Eseryel, & Offenbeck, 2017:1). Similar challenges of managing competing stakeholder interests are very likely to be relevant in data intensive projects in addition to ethical and legal discussions – especially in inter-organizational settings (Günther et al., 2017).

Typical projects are characterized as successful if they deliver exceedingly good results, finish before planned, and cost less than planned. Data intensive projects are very appropriate in terms of delivering more with a fixed quantity of resources, since data can be reused in multiple ways without deteriorating. Thus, for intensive data projects, it seems that the traditional parameters of specifications, cost, and time may behave in a nonlinear fashion. New specifications can emerge without resource input, the cost of adding one function is minimal, and new stakeholders can join without barriers. However, while involving more stakeholders and new functions in most projects would be considered positive characteristics, they bear the risk of slowly building up a dark side and ultimately backfiring in data projects. This downside to data projects will be demonstrated in this paper.

3 METHOD

3.1 Case and research setting

We ground our conceptualization of function creep and stakeholder creep in data intensive projects through the analysis of a longitudinal case study (2003 to 2015) of the use of huge amounts of health data in the Danish DA initiative DAMD. DAMD is interesting as an illustrative case study because it was characterized by high expectations from multiple stakeholders seeking to reuse data for numerous purposes. DAMD was a database that contained healthcare data on (almost) every patient in the primary healthcare sector in Denmark and was linked to the personal identification system (the centralized person register – CPR). The DAMD database was operated by a formal project organization called the Danish General Practice Quality Unit (DAK-E), which was governed by and financed with resources from the collective agreement between the Danish Regions (interest group for Danish hospitals) and the General Practitioners’ Organization (interest group for Danish GPs).

Before the database was shut down in 2014, it collected a very large amount of healthcare data about citizens’ visits to general practitioners (GPs) on a daily basis. The data were converted into quality reports that gave general practitioners an overview of their patients and was also included in many research projects (Thomsen, 2014). The database contained millions of personally identifiable pieces of information, but in the autumn 2014 the excessive data collection stopped when doubt was raised on the legality of the DAMD project. When the data controlling authority of the database (Southern Region of Denmark) in continuation of the shutdown of the database announced that the collected data would be erased, several authorities stepped in to preserve the database. As such, the Danish National Archives first determined that the database was worth preserving and wished to archive the database, but on May 12th 2015, the Danish Parliament decided that DAMD should be deleted.

3.2 Data collection

DAMD has been subject to substantial political attention and media publicity, which means that this case is well-documented in public reports and other documents. Hence, a large range of accessible documents about DAMD is available, written as well as audio files and videos. As listed in Table 1, we included multiple sources of documents collected from media outlets (newspaper and radio), debates, conferences, yearly reports, newsletters, governmental documents, press releases, meetings, open councils, statements, as well as evaluations and consultant reports. The collection of primary data occurred from 2014 to 2017 and was carried out by the first author using the ‘snowball method,’ which entailed

pursuing references in the different reviewed documents. This process took its starting point in a handful of central documents, such as the State's Serum Institute's reports and meeting logs and yearly reports from the formal project organization of DAMD (DAK-E). This improved our understanding of the research context and helped us map out key stakeholders, functionalities, data-flow, and the case chronology. In total, 278 minutes of video and 1843 pages of documents were included.

We supplemented the vast number of accessible documents with six stakeholder interviews, using a semi-structured interview guide. The interviews were conducted during 2017 and amounted to a total of seven hours. The interviewees included informants from the formal project organization of DAMD as well as other key stakeholders (GPs, researchers, and businesses). The questions we asked touched upon topics such as the chronology of the project, the stakeholders involved, and the functions embedded to the project.

Source	Quantity	Original data source
1. Plans and agreements	Collective agreements 3 (394 pages) Strategies 4 (122 pages) Application 1 (6 pages)	DAK-E, PLO, Danish Regions, Southern Region of Denmark, MedCom
2. Annual accounts and meeting logs	Annual accounts: 7(250 pages) Meeting logs: 20 (62 pages)	DAK-E
3. Information materials	Websites 5 (6 pages) Powerpoint presentations 3 (54 pages) Handouts 4 (16 pages) Newsletter 1 (2 pages) Instructional videos 5(18 mins)	DAK-E, PLO
4. Reports	Inquiries 3 (40 pages) Other reports 7 (516 pages)	SSI, National Archive, DAK-E, OECD, KORA, LeapCraft
5. Newspaper articles and media interviews	15 Articles (25 pages) 1 Radio interview (11 min.)	Newspapers
6. Conferences	Video broadcasts 10 (260 mins) Presentations 2 (44 pages)	PLO-conference on big data, The Danish Council on Ethics-conference on big data
7. Disclosed email correspondences (access to documents)	Akt 1 (157 pages) Akt2 (29 pages) Akt 3 (52 pages)	DAK-E, Danish Regions, Southern Region of Denmark
8. Political consultation	Speech 1(17 pages) Parliamentary hearing 3(51pages)	Minister of Health, Danish Patients
9. Interviews	6 (413 mins) * more interviews to be conducted	Informants from internal and external stakeholders
Sum	1843 pages 413 minutes interview recordings 278 mins (video)	

Table 1. Summary of data sources

3.3 Data analysis

To manage the complexity of the case and density of data, we started by reading through the collected empirical material to create a timeline of major events that took place between 2003 and 2015. Following

a temporal bracketing strategy (Langley, 1999), we traced the lifecycle of DAMD – from its emergence to its shutdown – and divided the DAMD project into four phases: emergence (2003-2005), formalization (2006–2009), expansion (2010-2013), and collapse (2014–2015). During the data coding, we used techniques for grounded approaches for analyzing qualitative data (Locke, 2001). When searching for mechanisms that caused the failure of the DAMD initiative, we started to see patterns of how the number of stakeholders involved and the functionalities ascribed to the project increased during the project’s lifecycle and how this affected the outcome of the DA initiative. For each of the four phases, we present an account of the added functionalities and involved stakeholders, and use displays (Miles & Huberman, 1994) to visualize how functions and stakeholders embedded in the project simultaneously increased during the DAMD lifecycle (see Figure 1, 2, 4 and 5). We use these findings to conceptualize function creep and stakeholder creep in large scale data-intensive projects.

4 THE DAMD DATA PROJECT

4.1 Phase I: Emergence (2003 – 2005)

The idea of DAMD originates from experimentations with electronically coding diagnoses in general practices in the 1990’s. Aligned with the vision of transforming the original paper-based patient journal system into a digital solution, these experimentations revealed three potential data analytics opportunities. First, the GPs were able to search and filter patient journals, which enabled an overview of each patient’s consultation history. Second, the diagnosis-coding was connected to a central knowledge hub that allowed doctors to access relevant and up-to-date information about a specific diagnosis or health issue. And third, the diagnosis-coding system provided data for quality development in general practices (Schroll, 2015). Based on these opportunities, the DAMD was initiated in 2003 as a pilot project.

4.1.1 First functionality: Patient overview and quality reports

The ambition was to develop a DA system that collected structured data from general practices to be processed into quality reports with patient overviews and relevant decision-making information. These reports were meant to be available to GPs. This required both datafication and digitalization processes. The first challenge was to determine the data input and it involved the development and implementation of a common classification system to ensure consistent coding among GPs based on a diagnosis-system (International Classification for Primary Care). The next challenge was to collect these data in an easy and time-saving manner. This involved the development of a “data capture module” that would be integrated with the many different IT systems used for patient administration among GPs. The development of the data capture module was initiated in 2004 with additional funding from the Ministry of Health and The Association of Danish Pharmacies. Pilot testing began in 2005 with 24 voluntary GPs.

The primary functionality of this data capture module was to automatically and continually capture data from the IT systems that are used locally in general practices and to integrate and structure the data in a central database (DAMD). The collected data consisted of unique patient identification numbers, dates, diagnoses, medicinal prescriptions, medical services, and laboratory codes. Most data were collected automatically, which limited any additional burden on the GPs, although supplementary information could be added through a pop-up mechanism. The data were then processed into quality reports (for patient overview and decision-making support) and delivered back to the local general practice. This dataflow defines the first and primary functionality of DAMD as visualized in Figure 1.

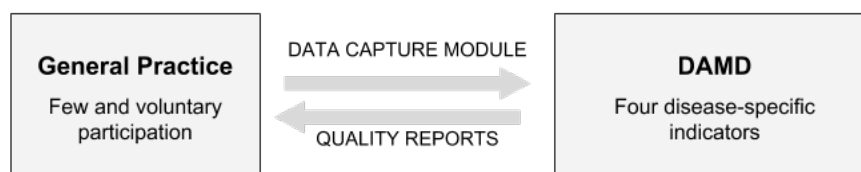


Figure 1. Dataflow related to function 1: local patient overview and quality reports

Overall, this phase was characterized by the bottom-up experimentation and development of a new system consisting of a data capture module based on four ICPC-diagnoses and an SQL-database to handle the patient information from general practices and improve the quality support for individual GP's. The participation was quite limited and voluntary among GPs.

4.2 Phase 2: Formalization (2006 – 2009)

In this next phase, the scope of the DA initiative increased. Inspired by the positive results from the pilot project, the collective agreement in 2006 between Danish Regions and General Practitioners' Organization (PLO) provided the foundation for transforming the pilot project into a formalized project organization to commence the large-scale implementation of the data capture module and the DAMD database. This resulted in the addition of new stakeholders and two new functionalities; the delivery of data to research projects and the export of data to the clinical quality database program as visualized in Figure 2.

4.2.1 Second functionality: Cross-sectorial quality monitoring

To be exempted from the legal requirement of patient consent for data collection, DAMD needed to be classified as a "Clinical Quality Database." This introduced the Danish Clinical Registries (RKKP) as a new stakeholder. RKKP constitutes the infrastructure of National Clinical Quality Databases, and is mandated by law, regulated by the government, and financed and owned by the Danish Regions. By exporting data to RKKP, DAMD created a valuable DA opportunity for supplementing existing national and regional quality databases with data from general practices to thereby enable cross-sectorial quality monitoring of the entire course of patient care. For the GPs, this arrangement further ensured an easier legal compliance as they (and other health professionals) were already required to deliver relevant data to the clinical quality databases – and now DAMD could carry out this obligation on their behalf in an efficient and time-saving manner.

4.2.2 Third functionality: Deliver data to research projects

The research communities had high expectations to the DA potential of DAMD because of the information rich data. This attracted interests from Danish universities and research units for general practices, who since the beginning of DAMD had sought to influence the operation of DAMD and the data availability. To avoid conflicts of interest, the Quality and Research Committee was established to handle applications from research projects seeking access to DAMD. The procedure involved several steps. First, researchers had to be authorized as potential *data processors* by the Danish Data Protection Agency. This meant that research results could only be published if the included data were aggregated or anonymized. The applications were then processed by The Quality and Research Committee (*data controller*) before DAK-E requisitioned the final permission from the Danish Data Protection Agency to transmit the data. This process includes the third functionality of DAMD, namely to deliver data to research projects.

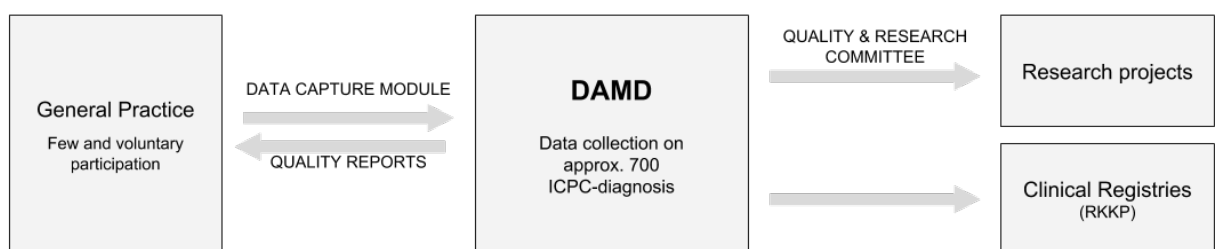


Figure 2. Dataflow with function 2 (cross-sectorial quality monitoring) and function 3 (deliver data to research projects) added. Applications for data use in research projects (*data processor*) were handled by the Quality and Research Committee (*data controller*).

In this phase, the scope of the data collection increased from the four specific diseases to include all approximately 700 ICPC-diagnoses. In return, the data was to be exported from DAMD to the clinical registries (RKKP) to enable cross-sectorial linkage of data as well as deliver data for research projects.

4.3 Phase 3: Expansion (2010– 2013)

Up until this phase, the GPs participated voluntarily, however, the new collective agreement between Danish Regions and General Practitioners' Organization in 2010 made it obligatory for GPs to install the data capture module for collecting and transmitting data to the DAMD database. This resulted in a rapid expansion of GPs connected to the DAMD database with more than 96% of Danish GPs registered as users by 2013 (Figure 3). This significantly increased the size and quality of the database, as it covered data from (almost) the entire Danish population of patients in the primary healthcare sector. This attracted further interests in the database seeking new potential usages of the huge amount of data.

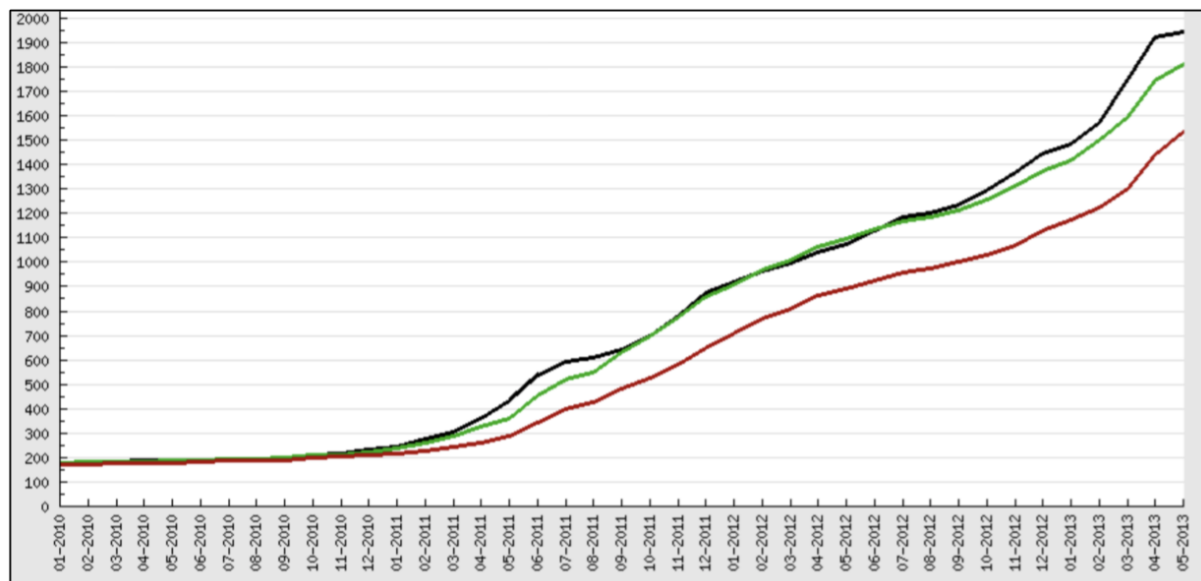


Figure 3. GP participation in DAMD from 2010–2013. Registered users (black), User who have the data-capture module installed (green), and users who have completed more than five pop-ups (red). Source: Paulsen & Thomsen (2013)

4.3.1 Fourth functionality: Deliver patient's own data

With data on almost every patient from Danish general practices being collected in the DAMD database, new data analytics potentials emerged in regards to empowering patients with better self-care opportunities through self-monitoring of their own health data. This functionality was sought to be realized with new plans and stakeholder agreements that ensured that DAMD data should be available to patients on their profile on the Danish national e-health platform (sundhed.dk). Thereby, patients were able to supervise their own medical history in terms of treatments, diagnoses, prescriptions, and so on. This new data processing routine of transferring data from DAMD to sundhed.dk was facilitated by a taskforce that consisted of the DAMD Steering Committee, MedCom (who operate sundhed.dk), General Practitioners' Organization, and the Danish Regions. This taskforce oversaw the technical, legal, and operative aspects of the expansion.

4.3.2 Fifth functionality: Commodification of data.

From the perspective of research and innovation, the mandatory participation of GPs in DAMD constituted “one of the world’s most valuable research infrastructures” (Melbye, 2014) and commercial potential, since “Denmark should lead the way in using health data and making them commercially applicable” (Vækstteamet for sundheds- og velfærdsløsninger, 2013, p. 15). As such, the growing DA project attracted a new group of stakeholders with commercial interests. The State’s Serum Institute had an interest in the DAMD data and strived to gain direct and continuous access to the database. By gaining access to DAMD, the State’s Serum Institute would be able to pass on information to, for instance, pharmaceutical companies and researchers, like they were already doing with information from the medicine, laboratory, and health insurance registers. The commodification of DAMD data was never fully realized, but in December 2013 it was determined by the Quality and Research Committee of DAMD that data could potentially be delivered to research projects associated with private companies if the primary data processor and researcher was connected to a university and the project had the required legal permissions. A summary of a meeting held by the Quality and Research Committee read that “data may have great value to private firms and the first real applications with commercial interest have been received” (DAK-E, 2013, p. 2).

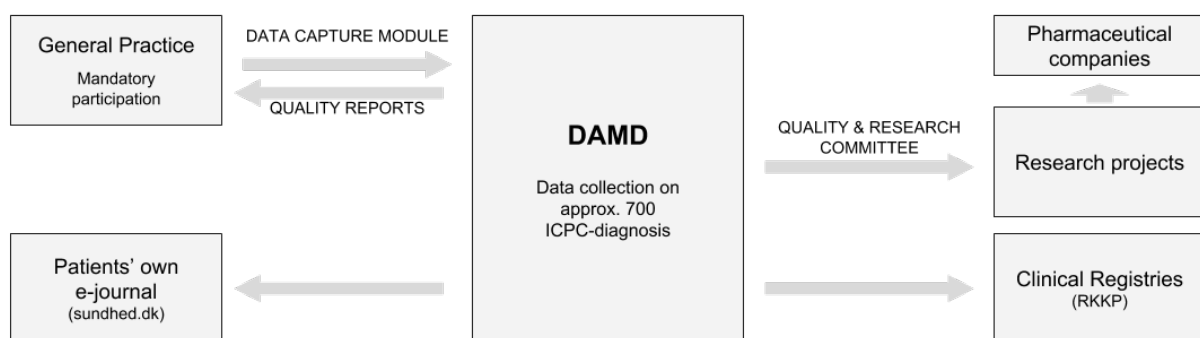


Figure 4. Dataflow with function 4 (deliver patient's own data) and function 5 (commodification of data) added.

4.4 Phase 4: Collapse (2014 – 2015)

In September 2014, the Danish Broadcasting Corporation harshly critiqued the DAMD project in regards to the comprehensive collection and (re)use of data without patient consent (Tynell & Fischer, 2014). This initiated widespread public debate and media coverage where doctor-patient-confidentiality, commodification of data, and unresolved conflicts about two additional functionalities were the main topics of discussion. This is described next.

4.4.1 Sixth functionality: Deliver data to other healthcare professionals

The data from DAMD was to be accessible by other professionals within and across the healthcare sector to support a more coherent healthcare system and improve continuity of patient care (P. Hansen, 2014). This functionality was only partially implemented. Numerous GPs feared that more comprehensive data sharing could undermine doctor-patient relationships and risk that patients might withhold sensitive information if the doctor-patient confidentiality was perceived to be broken (M. Hansen, 2014). In contrast, actors from the Danish Patients stated that “no one has died from the misuse of data. But it is possible that you can die if data sharing does not occur” (Lauridsen & Røndbjerg-Christensen, 2014).

4.4.2 Seventh functionality: Deliver data to healthcare management

DAMD was initially established to support local decision-making in general practices (function 1). However, it was also considered whether the data could also support managerial decision-making on a

regional or national level? This question resulted in a major ongoing conflict between key stakeholders. As the data controlling authority, the regions saw new opportunities to use the DAMD data themselves to ensure more effective and coordinated healthcare efforts (for instance through “pay for performance”). The Healthcare Act of 2014 gave the regions access to DAMD for planning, quality assurance, and financial control. Both DAMD and General Practitioners’ Organization (PLO) were strongly against this and argued that the risk of being sanctioned could affect the impartiality of doctors in terms of patient treatment and diagnostics. From this perspective, such a function would risk undermining the validity of data, as described by a manager from the DAMD project organization: “GPs of course wouldn’t report data that could risk compromising their own position.” The conflict remained unresolved and it was never settled whether authorities should have data access on an individual level. These dynamics may result in a kind of “red team-dynamics” (Lazer et al., 2014), where the data provider (here GPs) “attempt to manipulate the data-generating process to meet their own goals, such as economic or political gain” (ibid.: 4).

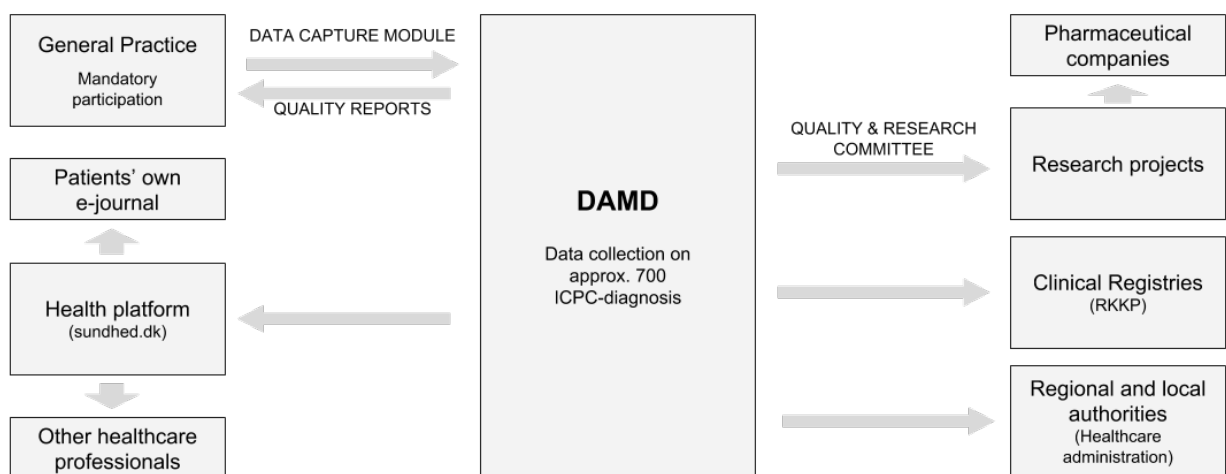


Figure 5. Final dataflow before collapse with function 6 (deliver data to other healthcare professionals) and function 7 (deliver data to healthcare management) added.

4.4.3 Deletion of DAMD

Data collection for DAMD was suspended in Autumn 2014 due to an investigation conducted by The State’s Serum Institute to determine the legality of DAMD (SSI, 2014). The inquiry concluded that the comprehensive use of the database lacked legal basis. However, the inquiry also (indirectly) revealed how intertwined stakeholder interests and the many functionalities complicated the accountability of DAMD, as described by the chairman of The General Practitioners’ Organization:

“The problem is that authorities, such as the State’s Serum Institute and the Danish Health Authority, who should have maintained a high standard and safeguarded the patients, are also subject to political control by politicians who have other intentions – that is, the intention to use information for control and resale. This means that even the authorities that should be protecting the patients in this case are, in reality, subject to other people’s interests. The Danish Regions, who want to develop the quality of the healthcare system and general practices, are simultaneously interested in being able to control and, if necessary, sanction the general practitioners, if they do not feel that doctors are doing well enough” (Dibbern, 2014)

In Table 2, we have summarized our analysis and illustrated the different functions and stakeholders involved in the project.

Phase	Functions	Stakeholders	Expectations	Preoccupations
<i>Emergence (2003-2005)</i> . Bottom-up experimentation and development of new data project system to handle patient information in general practices.	<ul style="list-style-type: none"> - Patient data capture module based on four ICPC-diagnoses - SQL-database 	<ul style="list-style-type: none"> - GPs (voluntary participation) - Danish Regions (interest organization) - GPs' professional organization (PLO) 	<ul style="list-style-type: none"> - Patient overview and quality support for individual GPs - From paper-based patient journal system to digital solution. 	None
<i>Formalization (2006 – 2009)</i> . The data project was transformed into a formal organization and categorized as a Clinical Quality Database.	<p><i>Additional:</i></p> <ul style="list-style-type: none"> - Export data to clinical registries - Deliver data to research projects - Expanded data collection; from four diseases to approx. 700 ICPC-diagnoses 	<p><i>Additional:</i></p> <ul style="list-style-type: none"> - GPs (increased participation) - Steering Committee for the data project - Quality and Research Committee - RKKP (national database operators) 	<p><i>Additional:</i></p> <ul style="list-style-type: none"> - New research opportunities in primary care and medical history. - New quality improvement opportunities due to broader data collection - Cross-sectorial linkage of data 	None
<i>Expansion (2010–2013)</i> . Mandatory participation among GPs resulting in significantly increased data volume.	<p><i>Additional:</i></p> <ul style="list-style-type: none"> - Export patients' own data to personal profile on Danish national e-health platform (sundhed.dk). - Commodification of data 	<p><i>Additional:</i></p> <ul style="list-style-type: none"> - Mandated participation of GPs - Patients - The State's Serum Institute - Medical industry 	<p><i>Additional:</i></p> <ul style="list-style-type: none"> - Better self-care opportunities for patients through self-monitoring of own health data - Denmark as a frontrunner for using health data and making them commercially applicable 	<ul style="list-style-type: none"> - Privacy and patient-doctor confidentiality
<i>Collapse (2014-2015)</i> . Increased public debate and media coverage. Closure of the data project.	<p><i>Additional:</i></p> <ul style="list-style-type: none"> - Export data to other healthcare professionals - Export data to regions and municipalities for management purposes 	<p><i>Additional:</i></p> <ul style="list-style-type: none"> - Healthcare professionals - National, regional, local government 	<p><i>Additional:</i></p> <ul style="list-style-type: none"> - Unified healthcare sector - Improved administrative control 	<p><i>Additional:</i></p> <ul style="list-style-type: none"> - Insufficient legal basis for the data project. - New incitements in data input may harm data validity - Inequality in access to healthcare

Table 2. Summary of DAMD analysis

5 DISCUSSION

Using evidence from a large-scale Danish DA initiative (DAMD), we have shown how the reuse of health data for multiple purposes to serve different stakeholder interests carries a dark side – not only for technical or ethical reasons but also because it may result in unclear organizational boundaries and insufficient liability arrangements, rendering a system less transparent or even illegal. A central discussion about DAMD concerns the shift from the purposes that the data were originally generated for and the contexts that the data were subsequently used in. As the use of data moved away from quality development for the individual patient, a landscape of dilemmas emerged, reflecting the many objectives

that different actors ascribed to DAMD. While patients' diagnoses and medication understandably should be shared between general practices and the wider healthcare system, it is less obvious that the data should be used by the medical industry and by regions for monitoring and planning issues. As such, our research challenges the technical and optimistic parts of big data-literature, which denotes data as "a magical diamond mine that keeps on giving long after its principal value has been tapped" (Mayer-Schönberger & Cukier, 2013: p. 104).

5.1 Theorizing function creep and stakeholder creep

While the traditional iron triangle in project management hints at the fact that more is better, data intensive (mega)projects like DAMD show that this theory needs a more sophisticated approach. The analysis of DAMD reveals how a data project that had all the characteristics of a successful project in the long-term actually ended up failing. More and more actors joined the project and they requested more and more functions. We call these phenomena function creep and stakeholder creep. We view these two concepts as closely intertwined and driven by the continual pressure of stakeholder interests that resulted in an expansion of the data project that was not originally envisaged. This indicates that there exists a "sweet spot of success" in large-scale data projects. The sweet spot of success happens just before the project success spins out of control. This is not easy to identify because, as Dahl & Sætnan (2009) note, the term 'creep' carries a hint of "sneakiness," which indicates that these expansions are unanticipated and happen gradually and stealthily. It can be hypothesized that stakeholder creep and function creep are more pronounced in large data projects – like DAMD – because of the complexity and variety of deliverables, the multiplicity of actors involved, and the long timespan which all dilute the possibility to effectively control the development.

Larson & Larson (2009) argue that function creep can happen due to lack of stakeholder involvement in project management since this may lead to autonomous project groups spinning out of control. Based on the analysis of DAMD, we argue that *too* many stakeholder interests can lead to function creep. Furthermore, we have introduced the notion of stakeholder creep as an active force in widening the scope of data projects. Budzier & Flyvbjerg (2012) show that in the context of IT projects, continual expansion of project scope can have proportionate effects; the larger the scope of a project, the higher the number of changes are requested. Following this logic, we conceptualize stakeholder creep and function creep as an ongoing and reinforcing interaction in the sense that new functionalities create new data analytics opportunities with potential social effects. This attracts more stakeholder interests, which in turn requires more functionalities and so on. It is a self-reinforcing cycle that is difficult to stop once it starts.

Function creep refers to the multiplication of functions within and beyond the scope of the data project itself. A core element in data projects is reuse of data based on the assumption that data is permanent and does not expire or deteriorate when used (Mayer-Schönberger & Cukier, 2013; Schmidt & Cohen, 2014; Sloan & Warner, 2014; Custers & Uršič, 2016). This implies an inherent driving force of the continual expansion of functions that exploit data beyond what was originally envisioned. These functions afford new ways to create value from data and are constantly developed, as demonstrated in the analysis of DAMD. The effect of additional functions in data intensive project is ambivalent. On the one hand, the introduction of new functionalities to meet emerging stakeholders' needs is the cornerstone of participative development and value creation. On the other hand, it is suggested that a libertarian addition of features and functionalities to a product bears the risk of budget and schedule overrun (Kearns, 1998; Flyvbjerg & Budzier, 2011). For large projects, function creep mechanisms may be an almost unavoidable part of project dynamics driven by a stakeholders' growing "wish list" or by developers' wish to improve the product or cope with unforeseen challenges (Norman, 1988; Shmueli & Ronen, 2017). However, as the DAMD analysis revealed, this ongoing expansion of functions or widening of scope in data intensive projects is not just an issue related to budgets and schedules, but also a matter of privacy, legality, and performance since the purpose or the use of technology changes (Wisman, 2013; Galloway, 2017). When considering the use of massive amounts of personal data, function creep can implicate an "invasion of privacy beyond what was originally understood and considered socially, ethically, and legally acceptable" (Dahl & Sætnan, 2009: 83).

Stakeholder creep refers to the surges of stakeholders in large-scale data project that involve the continuous addition of stakeholders beyond those initially involved in a specific data project. The DAMD case shows that in addition to multiple functions in a data project complicating the transparency and legality of the dataflow, the continual involvement of multiple stakeholders can impede accountability and division of responsibility in the project. Such issues are particularly relevant in data-intensive projects. As Manovich (2011) notes, data ecosystems create a new digital divide among those who create data (both consciously and by leaving digital footprints), those who have the means to collect it, and those who have the expertise to analyze it. Yet, these digital divides become more complicated with the involvement of more and more stakeholders with different and sometimes competing interests. In the DAMD case, we saw how the “polarization” of interests” (Boonstra et al., (2017) occurred between general practitioners and regional authorities (healthcare management) when dealing with data governance dilemmas concerning purpose, access, and incentives/sanctions embedded in the data infrastructure.

5.2 Practical implication, limitations, and future research

The practical implication of our study is for managers of DA projects to understand the dangers of function creep and stakeholder creep. Even though introducing new functionalities based on an existing data project can satisfy new stakeholder interests with little extra costs, we argue that for data projects such installments also bear the risk of slowly building up a dark side that may ultimately backfire. Elaborating on Boonstra et al.’s (2017) notion of the impact of stakeholder interests on IT projects, we suggest that data intensive projects and their functionalities should be managed and regarded as heavily intertwined with stakeholder interests, seeking to define the data generating process, data collection, data flow, and data usage (data-analytics).

Although (public) healthcare provides a particularly important research area to explore potential unintended consequences of DA due to highly sensitive data, it only represents one example. It would be naïve to suggest that function creep and stakeholder creep could be found, in a similar form, in all projects and circumstances. Challenges in data intensive project may differ and vary from one context to another and, therefore, we suggest that future research should strive to identify and further develop the notion of function and stakeholder creep in a variety of settings. In particular, what influences function and stakeholder creep should be explored, since issues of reusing personal data in new contexts continues to be controversial, as the current Facebook-Cambridge Analytica data scandal reminds us.

6 CONCLUDING REMARKS

DA and the associated fields of BI and big data have been of increasing interest among IS scholars and has become an integral part of many organizations. In contrast to the unprecedented focus on the opportunities and benefits of DA, we have concentrated on a problematic implication of the (re)use of health data in long-term data projects. Grounded in the case analysis of DAMD, we identified and theorized the existence of two mechanisms and their interaction – function creep and stakeholder creep – and posited that these ultimately resulted in the failure of the DAMD project. Because of the social implications of data projects like DAMD, surges of stakeholders and functions result in a creeping dark side. As the dark side brews and grows in a stealthy fashion, its nature and consequences are not clear and the few early warning flags are easily dismissed as scarecrows. After all, most patients and general practitioners would see the advantages of keeping a personal medical record. However, our study shows that large-scale data projects have a “sweet spot of success.” As stakeholders and functions get out of control, the project turns into a failure uncovering a dark side that becomes real and dangerous even for stakeholders that were once supporters.

References

- Ahangama S., and Poo, D. (2015). "Improving Health Analytic Process through Project, Communication and Knowledge Management", In: *Proceedings of the International Conference on Information Systems - Exploring the Information Frontier, ICIS 2015*. Ed. By T. Carte, A. Heinzl, & C. Urquhart. Association for Information Systems. Texas, USA, p. 418-427.
- Boonstra, A., Eseryel, U.Y., and Offenbeek, M.G. (2017). "Stakeholders' enactment of competing logics in IT governance: polarization, compromise or synthesis?", *European Journal of Information Systems* 17(2), 100-111.
- Brynjolfsson, E., Hitt, L., and Kim, H. (2011). "Strength in Numbers: How does data-driven decision-making affect firm performance?". Unpublished. URL: <http://dx.doi.org/10.2139/ssrn.1819486>
- Budzier and Flyvbjerg (2012). "Double Whammy – How ICT Projects are Fooled by Randomness and Screwed by Political Intent". Saïd Business School Working Paper. URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2238057
- Chen, H., Chiang, R. H. L., and Storey, V. C. (2012). "Business Intelligence and Analytics: From Big Data to Big Impact". *MIS Quarterly* 36 (4), 1165-1188.
- Constantiou, I.D. and Kallinikos, J. (2015). "New games, new rules: big data and the changing context of strategy". *Journal of Information Technology* 30 (1), 44–57.
- Custers, B. and Uršič, H. (2016). "Big data and data reuse: a taxonomy of data reuse for balancing big data benefits and personal data protection". *International Data Privacy Law* 6(1), 4–15.
- Dahl, J. Y. and Sætnan, A. R. (2009). "It all happened so slowly": On controlling function creep in forensic DNA databases. *International Journal of Law, Crime and Justice* 37 (3), 83-103.
- DAK-E. (2013). Referat af møde i Kvalitets- og Forskningsudvalget 12. december 2013. URL: http://www.dak-e.dk/files/46/20131212_damd_kvalitets_og_forskningsudvalget_referat.pdf
- Dibbern, H. (2014). Radio interview in DR 'P1 Orientering' 14-10-2014, 16:07 - 18:05. URL: <http://www.dr.dk/radio/ondemand/p1/orientering-695/#/> (visited on 04/20/2018).
- Fischer, A., and Tynell, J. (2014). "OVERBLIK Det handler data-sagen om." URL: <http://www.dr.dk/Nyheder/Indland/2014/12/04/144653.htm> (visited on 04/20/2018).
- Flyvbjerg, B. and Budzier, A. (2011) "Why Your IT Project May Be Riskier than You Think," *Harvard Business Review* 89(9), 23-25.
- Galloway, K (2017). "Big Data: A case study of disruption and government power". *Alternative Law Journal* 42(2), 89 - 95
- Greenwald, G. (2014). *Overvåget - En insiderberetning om Edward Snowden, NSA og den amerikanske overvågningsstat*. Informations Forlag.
- Günther, W.A, Mehrizi, M.H.R., Huysman, M., and Feldberg, F.. (2017). "Debating big data: A literature review on realizing value from big data". *The Journal of Strategic Information Systems* 26(3), 191-209
- Hansen, M. K. (2014). "Vi må ikke gamble med patienternes tillid". *Ugeskrift for Læger*, 29/9/2014
- Hansen, P. E. (2014). "Big Data - Poul Erik Hansen. Fra Oplæg ved Lægedage 12. november 2014". URL: https://www.youtube.com/watch?v=CzPgRfeU_rE (visited on 04/20/2018).
- Kearns, D. (1998). "Microsoft's case of feature creep". *Network World* 15(37), 22.
- Khan, Z., and Vorley, T. (2017). "Big data text analytics: an enabler of knowledge management". *Journal of Knowledge Management* 21(1), 18-34.
- Langley, A. (1999). "Strategies for theorizing from process data". *Academy of Management Review* 24 (4), 691-710.
- Larson, R. and Larson, E. (2009). "Top five causes of scope creep ... and what to do about them". Paper presented at PMI® Global Congress 2009—North America, Orlando, FL. Newtown Square, PA: Project Management Institute.

- Lassen, J. A. (2013). "Dansk databank skal tiltrække udenlandsk sundhedsforskning". *Berlingske* (04-06), p. Business 4.
- Lauridsen, J.B., and Røndbjerg-Christensen, H. (2014). "Patientforening: Ingen er død af datamisbrug". *BT - Berlingske Nyhedsbureau*, May 12th 2014.
- Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). "The Parable of Google Flu: Traps in Big Data Analysis". *Science* 343(6176), 1203-1205.
- Locke, K. (2001). *Grounded Theory in Management Research*. Great Britain: SAGE Publications.
- Manovich, L. (2011) "Trending: the promises and the challenges of big social data". In *Debates in the Digital Humanities*. Ed. M. K. Gold. The University of Minnesota Press, Minneapolis.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H. 2011. "Big Data: The Next Frontier for Innovation, Competition, and Productivity," McKinsey Global Institute. URL: <http://www.mckinsey.com/business-functions/business-technology/ourinsights/big-data-the-next-frontier-for-innovation>
- Margolis, R., Derr, L., Dunn, M., Huerta, M., Larkin, J., Sheehan, J., Guyer, M., & Green, E. (2014): "The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data". *Journal of the American Medical Informatics Association* 21(6), 957-958.
- Mayer-Schönberger, V., and Cukier, K. (2013). *Big Data - A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt Publishing Company. Boston, New York.
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., and Barton, D. (2012). "Big data. The management revolution". *Harvard Business Review* 90(10), 61-67.
- Melbye, M. (2014). "Mads Melbye: Fordele ved og udfordringer for styrket dansk sundhedsdataforskning". URL: <http://www.etiskraad.dk/Nyhedsarkiv/2014/december/Video-fra-debatdag-om-sundhedsdata.aspx>
- Miles, M.B., and Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks: Sage Publications.
- Miloslavskaya, N. & Tolstoy, A. (2016). "Big Data, Fast Data and Data Lake Concepts." *Procedia Computer Science* 88, 300-305.
- Norman, D. (1998). *The Invisible Computer*. MIT press, Cambridge, MA.
- Paulsen, M. S., and Thomsen, J. L. (2013). "Data til kvalitetsudviklings- og forskningsprojekter fra Dansk Almen Medicinsk Database (DAMD)". *Practicus* 216, p. 8.
- Raghupathi, W., and Raghupathi, V. (2014). "Big data analytics in healthcare: promise and potential". *Health Information Science and Systems* 2(3). <https://doi-org.zorac.aub.aau.dk/10.1186/2047-2501-2-3>
- J. S. Saltz and I. Shamshurin, "Big data team process methodologies: A literature review and the identification of key factors for a project's success". *2016 IEEE International Conference on Big Data (Big Data)*, 2872-2879.
- Schmidt, E., and Cohen, J. (2014). *The New Digital Age - Reshaping the Future of People, Nations and Business*. John Murray.
- Schroll, H. (no date). "DAK-E's historie - historien, der har ført til kvalitetsudvikling og forskning i almen praksis via bl.a. ICPC og Sentinel Datafangst". URL: https://www.dak-e.dk/resources/files/dak_es_historie.pdf (visited on 04/20/2018).
- Shmueli, O., and Ronen, B. (2017). "Excessive software development: Practices and penalties". *International Journal of Project Management* 35(1), 13-27,
- Sivarajah, U., Kamal, M.M., Irani, Z., and Weerakkody, V. (2017). "Critical analysis of Big Data challenges and analytical methods". *Journal of Business Research* 70, 263-286.
- Sloan, R. H., and Warner, R. (2014). "Beyond Notice and Choice: Privacy, Norms, and Consent". *Suffolk University Journal of High Technology* 14 (2), 370-412.

- Srinivasan, U., & Arunasalam, B. (2013). "Leveraging Big Data Analytics to Reduce Healthcare Costs". *IT Professional* 15 (6), 21-28
- Stonebraker, M (2015). "Chapter 12: A Biased Take on a Moving Target: Data Integration". In *Readings in Database Systems, 5th Edition*. <http://www.redbook.io/pdf/ch12-dataintegration.pdf>
- SSI - Statens Serum Institut. (2014). "Udredning vedr. ansøgninger om godkendelse af DAMD som klinisk kvalitetsdatabase fra 2007 og frem". SSI J. nr. 14/23193.
- Thomsen, J. L. (2014). "Big Data - Janus Laust Thomsen. Fra Oplæg ved Lægedage 12. november 2014". URL: <https://www.youtube.com/watch?v=gMsgKWEixU0> (visited on 04/20/2018).
- Tynell & Fischer (2014). "Fortrolige sundhedsdata samles ulovligt i database". *Danmarks Radio*. URL: <http://www.dr.dk/Nyheder/Indland/2014/12/04/144653.htm> (visited on 04/20/2018).
- Van den Broek, T., and Van Veenstra, A.F. (2015). "Modes of governance in inter-organizational data collaborations". *ECIS 2015 Completed Research Papers*. Paper 188.
- Vestergaard, M., Vedsted, P., Obel, C., Sandbæk, A., Christensen, B.; Olesen, F., Lauritzen, T., Bro, F., Andersen, J. S., Søndergaard, J., Halling, A.; Bentzen, N.; Mäkelä, M.; Bjerrum, L., Olivarius, N.F.; Kragstrup, J., and Reventlow, S. (2014). "Den uundværlige datafangst". *Jyllands-Posten*, Oct. 22th 2014.
- Vækstteamet for sundheds- og velfærdsløsninger. (2013). *Vækstteamet for sundheds- og velfærdsløsninger: Anbefalinger*. Regeringen, DK.
- Wang, Yichuan; Hajli, Nick (2017). "Exploring the path to big data analytics success in healthcare". *Journal of Business Research*, Vol. 70, p. 287-299
- Wisman, T.H.A., "Purpose and function creep by design: Transforming the face of surveillance through the Internet of Things". *European Journal of Law and Technology* 4(2). <http://ejlt.org/article/view/192>
- Yi, Liu, Liu, and Jin (2014). "Building a network highway for big data: architecture and challenges". *IEEE Network*. 28(4), 5-13.