Towards a Co-design Perspective on Data

Foregrounding Data in the Design and Innovation of Data-based Services

By
Cathrine Seidelin
IT University of Copenhagen

A dissertation submitted in compliance with the requirements for the degree of
DOCTOR OF PHILOSOPHY (PhD)

Academic supervisors
Yvonne Dittrich, IT University of Copenhagen
Erik Grönvall, IT University of Copenhagen

Industry supervisors
Henrik Amdi Madsen, Industriens Uddannelser
Louise Sonne Dyreborg, Industriens Uddannelser

Committee
Lene Nielsen, IT University of Copenhagen
Jeanette Blomberg, IBM Research
Vincenzo D’Andrea, University of Trento

To my first audience:
my parents and my sister.

PART I
INTRODUCTION COVER

Table of Contents

PART II
RESEARCH PUBLICATIONS

Acknowledgements

The personal and professional support and encouragement of so many wonderful people made this dissertation possible.

First and foremost, I would like to thank Industriens Uddannelser, Dansk Industri, Dansk Metal and 3F for having the vision to initiate and carry out this Industrial PhD project. I would also like to thank Innovation Fund Denmark and the Industry's Foundation for Education and Collaboration who financed this project. I am very thankful to all my colleagues, who showed great curiosity and participated in research activities – usually without knowing what the outcome would be. This dissertation simply would not have been possible without you! Also, a heartfelt thank you to my dear colleague and co-author Stine Moeslund Sivertsen. Your endless curiosity, creativity, and optimism are invaluable. Moreover, I would like to thank my industry supervisors: Henrik Amdi Madsen – thank you for your visionary outlook, constant encouragement and inspiring discussions; Louise Sonne Dyreborg – thank you for always looking after me, and for your insightful advice.

I would like to thank my primary academic supervisor, Prof. Yvonne Dittrich. Throughout this process, you have always gone above and beyond to support me and my work. Thank you for your our many rewarding discussions, your helpful guidance, and for challenging me. I would also like to thank my academic co-supervisor, Assoc. Prof. Erik Grönvall, for our inspiring discussions and valuable feedback. I am also much indebted to Prof. Irina Shklovski, Asst. Prof. Christian Østergaard Madsen, and Prof. Minna Isomursu for their thoughtful comments and suggestions during my midway evaluation. You feedback has contributed greatly to the development of my arguments.

Next, I would like to thank the people all over the world who have contributed and helped refine this work. I would like to thank the Oxford Internet Institute, Oxford University, for hosting me during my research abroad stay. I would especially like to thank Prof. Jonathan Bright, Bharath Ganesh, and Thomas Vogl for bringing me on board the Data Science for Local Government study, and for your excellent collaboration – also after I left Oxford. Furthermore, I would like to thank the Human Centred Design and Engineering Department, University of Washington, for hosting me during my second research stay abroad. A special thanks to Assoc. Prof. Charlotte P. Lee and the CSC Lab for fruitful discussions, which contributed to the refinement of my arguments. Thanks to all the people who provided feedback and engaged in interesting discussions during courses, doctoral consortiums, and research visits. I would also like to thank my fellow doctoral candidates at ITU and abroad – our support for one another on our individual but somewhat parallel journeys has been a great help to me.

Finally, I would like to extend my deepest thanks to my great friends, who have listened and have kept me sane during this process, and to my dear family for their endless support and encouragement. Last but not least, a heartfelt thanks to Jakob – meeting you in Oxford has turned this journey into an adventure.

Summary

Data supports an increasing number of services in society. This has created a growing need for organizations to consider data a key facilitator of service innovation and development. However, research reveals that organizations lack the tools to support creative and innovative work with data in ways that help to promote data-driven innovation. To address this problem, this dissertation examines how organizations can design and innovate their data-based services. Specifically, it explores how domain experts who are not IT professionals may participate in designing and innovating the data and data structures that underpin the digital services they use and provide, as part of their work practices. The dissertation demonstrates how it is possible to enable domain experts to design with data when data is carefully foregrounded. It also demonstrates that domain experts may collaboratively design with data in a way that takes into account that many organizations are connected to external stakeholders and organizations through shared practices, systems, and, indeed, data. The dissertation is based on a three-year action research study with Industriens Uddannelser (IU), an organization that works to maintain and develop vocational and continuing education programmes related to Denmark's industrial sector. This dissertation takes a practice perspective to explicitly focus on day-to-day data practices as a way to investigate how IU may work with data, to innovate and design their data-based services.

The long-term action research project with IU forms the core of this dissertation's six principal contributions. First, the dissertation discusses how data is used and handled today by local government and organizations in the public sector. Second, the dissertation presents a diagram that reveals the complex network of stakeholders which frame how an organization may provide and innovate essential data-based services. Third, the dissertation investigates how an organization can develop and establish a culture of design and innovation to foster data-driven innovation. Fourth, the dissertation elaborates on the tools developed to enable domain experts who are not IT professional, to participate in the design of the data-based services they use and provide as part of their work practices. Fifth, based on the tools developed, the dissertation proposes a Data Mode Map, which is an instrument that supports reflection on the design of data notation for co-design. Finally, the dissertation's principal theoretical contribution is the proposal to develop a co-design perspective on data. This perspective aims to support organizations in developing their existing as well as new data-based services using data. Additionally, this perspective promotes collaborative methods that reveal and take into account varying data practices in the design process – not only within the individual organization, but across the network of stakeholders who are more or less involved or influenced by new data-driven initiatives. The last contribution, in particular, offers suggestions for future research.

Sammenfatning

Data understøtter et stigende antal digitale tjenester (også kaldet ”digitale services”) i samfundet. Det har skabt et stort behov for, at organisationer i højere grad kan anvende data i forbindelse med udvikling og innovation af deres services. Forskning viser dog, at organisationer mangler værktøjer til at arbejde kreativt og innovativt med data, for på denne måde at fremme datadrevet service innovation. Denne afhandling undersøger derfor, hvordan organisationer kan designe og innovere deres databaserede services. Mere specifikt undersøger afhandlingen, hvordan domæneeksperter, som ikke er IT-fagfolk, kan inddrages i design og innovation af data og datastrukturer, der understøtter de digitale services, som disse domæneeksperter anvender og leverer som en del af deres arbejdspraksis. Afhandlingens resultater viser, at domæneeksperter kan designe med data, når repræsentationer af data fremhæves med stor omhu i designprocessen. Derudover viser resultaterne, at domæneeksperter ved brug af co-design metoder kan designe med data således, at der tages højde for forskellige organisationers varierende praksis, systemer og databehov. Denne afhandling er baseret på et treårigt aktionsforskningsstudie i samarbejde med Industriens Uddannelser, som er en organisation, der arbejder for at vedligeholde og udvikle erhvervsuddannelser og efteruddannelsesprogrammer til den industrielle sektor i Danmark. Afhandlingens teoretiske ramme bygger på et praksisperspektiv og undersøger på denne måde hvordan organisationer såsom Industriens Uddannelser kan gøre brug af data som led i udviklingen og innovationen af deres databaserede services.

Afhandlingen præsenterer seks centrale biddrag fra det langvarige aktionsforskningsprojekt med Industriens Uddannelser. Det første biddrag omhandler en diskussion om, hvordan data bliver anvendt i kommuner og organisationer inden for den offentlige sektor. Denne diskussion leder til en bedre forståelse af datapraksis i denne kontekst. Afhandlingens andet biddrag præsenterer en figur, der synliggør det komplekse netværk af aktører, som danner en ramme, der er både understøttende og begrænsende for, hvorledes en organisation kan levere og udvikle databaserede services. Dernæst biddrager afhandlingen med dybdegående empirisk indsigt i, hvordan en organisation kan udvikle og forankre en kultur, der fremmer design og innovation. Afhandlingens fjerde biddrag omhandler værktøjer, som er udviklet i løbet af projektet. Værktøjernes formål er at gøre det muligt for domæneeksperter, der ikke er IT-fagfolk, at deltage i designet af de databaserede tjenester, de bruger og leverer som en del af deres arbejdspraksis. Det femte biddrag bygger på disse værktøjer, og afhandlingen præsenterer et ”datatilstandskort”, som er et redskab, der understøtter refleksion i forhold til videreudvikling af data repræsentationer til co-design. Det sjette og sidste biddrag udgør også afhandlingens primære teoretiske biddrag, som består i den spæde udvikling af et co-designperspektiv på data. Formålet med dette perspektiv er at hjælpe organisationer til at kunne udvikle deres eksisterende såvel som nye databaserede services ved brug af data. Dertil fremmer dette perspektiv samarbejdsmetoder, der synliggør og tager højde for varierende datapraksis – ikke blot i den enkelte organisation, men på tværs af det netværk af interessenter som i større eller mindre grad er involveret i eller influeres af nye data-drevne initiativer. Særligt dette biddrag præsenterer både teoretiske og praktiske forslag til videre forskning.

Part I
Introduction Cover

… just because we have big (or very big, or massive) data does not mean that our databases are not theoretically structured in ways that enable certain perspectives and disable others. (Bowker 2014, 1797)

Chapter 1: Introduction

This dissertation explores how domain experts who are not IT professionals can participate in designing and innovating the data and data structures that underpin the data-based services they use and provide as part of their work practices. In discussions of how organizations and society at large may become more ‘data-driven’, researchers have paid great attention to the development of new technological tools for data collection, analysis, and application. Moreover, data and the possibility of analysing large amounts of data are increasingly regarded as a key enabler for service innovation in society (Antons and Breidbach 2018; OECD 2019). However, many organizations struggle to understand how they can effectively collect and implement data, and how to use data in the context of designing and innovating services (Ostrom et al. 2015). A leading cause of this struggle may be that working creatively and innovatively with data is currently restricted to those with a computer background or skills (Boyd and Crawford 2012; D’Ignazio 2017; D’Ignazio and Bhargava 2015; Gray, Gerlitz, and Bounegru 2018).

Using data aggregation and data analytics tools for innovative purposes often requires highly specialized skills and knowledge. The need for specialized skills means that core stakeholders and users are often left out of the design decisions that determine which kinds of data are included in, and excluded from the technical infrastructures that support data-based services (Boyd and Crawford 2012; King, Churchill, and Tan 2017). From a practice perspective, this is problematic because the underlying technical infrastructure form many aspects of how domain experts (can) act and interact in their everyday work practices. Thus, if only highly specialized people can engage in the discussions that eventually shape the development of common technical infrastructures, this will promote certain perspectives over others (Bowker 2014). This dissertation addresses this concern by questioning how we may include additional domain experts in the design of data and data structures. This is not to say that computer scientists and data scientists should be regarded as unnecessary. On the contrary, technological development and its effect on society make it clear that these forms of expertise are (and will continue to be) in high demand. However, to empower non-technical audiences, we need to bridge the gap between people who work effectively with data, and people who cannot (D’Ignazio 2017). This dissertation contributes to this line of argumentation by showing how we may begin to develop approaches that enable other forms of expertise (and thereby other perspectives) to be involved in the design of data and data structures to a greater extent. Below, I elaborate on how I have approached this objective.

1.1 Research perspective

This dissertation takes a practice perspective on understanding the organizational practices and processes that play off each other to produce the data work that underpin data-based services. Working with data involves certain practices. Knowing how to find, collect, and analyse data, or ‘use data in innovative ways’ implies *doing* something with data. Specifically, it involves everyday activities that collectively constitute data-science practices, for example. To foreground the importance of these particular data-related work activities, this dissertation takes a practice perspective, which is an approach that draws attention to how practical knowledge enables people to do things in the world. Moreover, a practice perspective sheds light on how this knowledge is reproduced in peoples' everyday activities (Blomberg and Darrah 2015). From a practice perspective, the meaning of data, and thus, what constitutes data, is established through the ways data is embedded in everyday work. This perspective enables an explicit focus on day-to-day data practices as a way of exploring how an organization may work with data to innovate and design its data-based services.

Providing and developing services involves certain practices. In line with the previous paragraph, the practice perspective also influences the dissertation’s view on the concept of service. Thus, I define services as abstract propositions of socio-material configurations that ‘are embedded with practice and are animated through practices’ (Blomberg and Darrah 2015, 74). This work focuses on *data-based services*, which refer to services that are supported by digital technologies, and where data is a core component of the service provisioning and delivery. Thus, I aim to understand services and their related practices through the data that are needed to provide and develop services.

Design also involves practices. Researchers and practitioners have increasingly made use of design practices to address complex problems that have been created by technological advancements and the accumulating number of ‘datafied’ processes in society, among other things (Wulf et al. 2018). Working in ‘designerly’ ways involves a set of conscious practices, which build on shared understandings and values that establish what it means to design (Julier et al. 2019). In recent years, the value of participatory, human-centred, and holistic design has significantly influenced design practices in the context of information technology (Simonsen and Robertson 2012). These participatory and collaborative approaches to design have enabled various professional practices to come together and influence design processes. Throughout this dissertation, I use the term ‘co-design’ when describing such design approaches. Specifically, I use co-design to describe a general concept for collaboration amongst people that ‘come together to conceptually develop and create things that respond to certain matters of concern and create a (better) future reality' (Zamenopoulos and Alexiou 2018, 12). Thus, co-design both constitutes a 'field of practice' itself, but also embeds the practice of bringing together other and varying fields of practices. I use co-design to identify a set of design practices that constitutes one approach to examining how organizations may design and innovate their data-based services. However, I argue that for co-design practices to be useful when designing and innovating data-based services, it is necessary to make data an explicit part of co-design. Specifically, I argue that it is necessary to foreground data in ways that support domain experts' understanding of data as an object of design, meaning that domain experts perceive data and data structures as malleable entities that are, and may be, designed (Feinberg 2017).

A practice perspective makes it apparent that data, service, and design constitute different 'fields of practices' (Blomberg and Darrah 2015). I argue that in the context of designing data-based services, these fields of practice overlap to some degree. For example, the practices related to service provision may be connected to certain data practices that are carried out to provide a given data-based service. Therefore, in this work, I consider how these different fields of practice may complement each other, with respect to involving domain experts, who are not IT professionals, in the design of data and data structures that underpin data-based services.

1.2 The industrial setting

This dissertation emerged from examining and intervening in day-to-day data practices in an organizational context. Specifically, this research is based on an Industrial PhD project (Innovation Fund Denmark 2020). This PhD programme comprises a three-year, industry-oriented research project and a PhD programme that was collaboratively carried out by the author, the IT University of Copenhagen, and Industriens Uddannelser (in English, 'The Education Secretariat for Industry', henceforth, 'IU'). The research was carried out at IU, which is an organization in Denmark's public sector. IU works to maintain and develop vocational and continuing education programmes related to the industrial sector. The research project originated in the quest to address the societal challenge of advancing small and medium-sized organizations' capacity to develop and support ways of innovating and designing services by using data more intelligently.

1.3 Research Questions

To address the overall objective of how to support small and medium-sized organizations to develop ways of designing services by using data, the guiding research question for this dissertation is, *How can organizations innovate and design their data-based services?*

To structure the research process, I developed three practical sub-questions that shed light on various aspects of the overarching research question:

- What are the common data practices of organizations?
- How may organizations design concrete, data-based services?
- How may organizations explore new data sources and experiment with their usefulness?

I considered these research questions from a practice perspective and by conducting a long-term action research project at IU. The research questions were prompted by the project's industrial setting. They were developed in collaboration with members of the organization and formulated in this specific manner to make sure they resonated with the organization. However, when addressing these questions, it became evident that the project warranted a broader discussion of how we design with data - specifically, how to do so in a co-design manner that involves various actors in the organization as well as external stakeholders.

1.4 Contributions

This dissertation makes six primary contributions based on a comprehensive action research project, the first part of the thesis, and eight research publications.

- First, the dissertation discusses how data is used and handled today by local governments and organizations in the public sector
- Second, the dissertation presents a diagram that reveals the complex network of stakeholders which frames how an organization may provide and innovate essential data-based services.
- Third, this dissertation investigates how an organization may develop and establish a culture of design and innovation to foster data-driven innovation.
- Fourth, this dissertation elaborates on the tools developed to enable domain experts who are not IT professional to participate in the design of the data-based services.
- Fifth, based on the tools developed, this dissertation proposes a Data Mode Map, which is an instrument that supports reflection on the design of data notation for co-design.

- Sixth, this dissertation's principal theoretical contribution is the proposal to develop a co-design perspective on data.

1.5 Reading guide

The remainder of this thesis cover is structured as follows. In chapter 2, I situate the research project by explaining the field site, the work's focus on data-based services, and the connection between this research project and broader societal trends. Chapter 3 presents the theoretical framework of the dissertation by explaining the practice perspective underpinning the key themes in this research: data, data work, and design. Chapter 4 elaborates on the applied action research methodology and presents the research process and research activities in detail. Chapter 5 introduces the publications included. At this point, I advise the reader to read the publications in full before continuing with the general discussion in the following chapters. The dissertation then turns into a broader discussion of how we may design with data, and how we may do so collaboratively: Chapter 6 presents the Data Mode Map, and discusses how it may be beneficial for researchers and practitioners to consider two prominent dimensions when foregrounding data in a design context. Chapter 7 discusses some of the necessary arrangements that are crucial for organizations to be able to design and innovate their data-based services. Finally, chapter 8 develops the proposal to establish a co-design perspective on data, to support the design and innovation of data-based services.

Chapter 2: Contextualizing the research project

This chapter has three objectives: First, it aims to provide an overview and understanding of the field site in which this research primarily took place. Second, it aims to establish what constitutes a data-based service in this context. The third aim is to establish the relevance of this research by relating the project to societal tendencies that increase the datafication of society. I begin this chapter by presenting 'Industriens Uddannelser' (IU), the organization that is the focal point of this research, and the context in which this organization exists. I provide a brief account of the notion of service, to establish what constitutes a data-based service in this context. Finally, I contextualize the research project in relation to growing societal trends, such as big data and open data, which advance organizations' need to consider data a strategic asset for the innovation and design of services.

2.1 The field site

This section begins with a presentation of IU, which is the main setting of this research project. The presentation is followed by a brief account of the context in which IU exists. The context is a particular area of Denmark's public sector, which focuses on vocational and continuing education that targets the Industrial sector. In this connection, I introduce one of the main contributions of this dissertation, specifically, a diagram that depicts the complex setting in which IU provides a number of essential data-based services to the larger network of stakeholders in this particular area of the Danish public sector.

2.1.1 Industriens Uddannelser

IU is a medium-sized service organization in Denmark's public sector. The organization is one of 19 education secretariats in Denmark, each of which works to maintain and develop vocational and continuing educational programmes in different areas of industry. IU focuses on the educational programmes and courses targeted at Denmark's industrial sector. For example, these include educational programmes for auto mechanics or Computer Numerical Control (CNC) technician. IU alone handles 39 vocational education programmes, and more than 1000 continuing education programmes related to the industrial sector. The organization primarily maintains and develops these many educational programmes through highly-organized committee work, on which I elaborate below.

IU's internal organization comprises five main departments that include 63 employees and 6 managers (see organization diagram below, figure 1). Most of the organization members are either employed in an administrative position or as educational consultants. The latter category of organization member works in either of two departments: The Industrial Sector's

Joint Committee (IF) and the Committee for the Metalworking Industry (MI). These departments are named for the two principal councils to which IU provides services. IU administers 12 Sector Skills Councils, which are authorities that are responsible for making sure that the vocational education programmes and continuing education are developed according to the needs of the labour market. A sector skills council is made up of multiple stakeholders: representatives from both employer associations and unions, and an education consultant from IU, who handles and supports the council and its members.

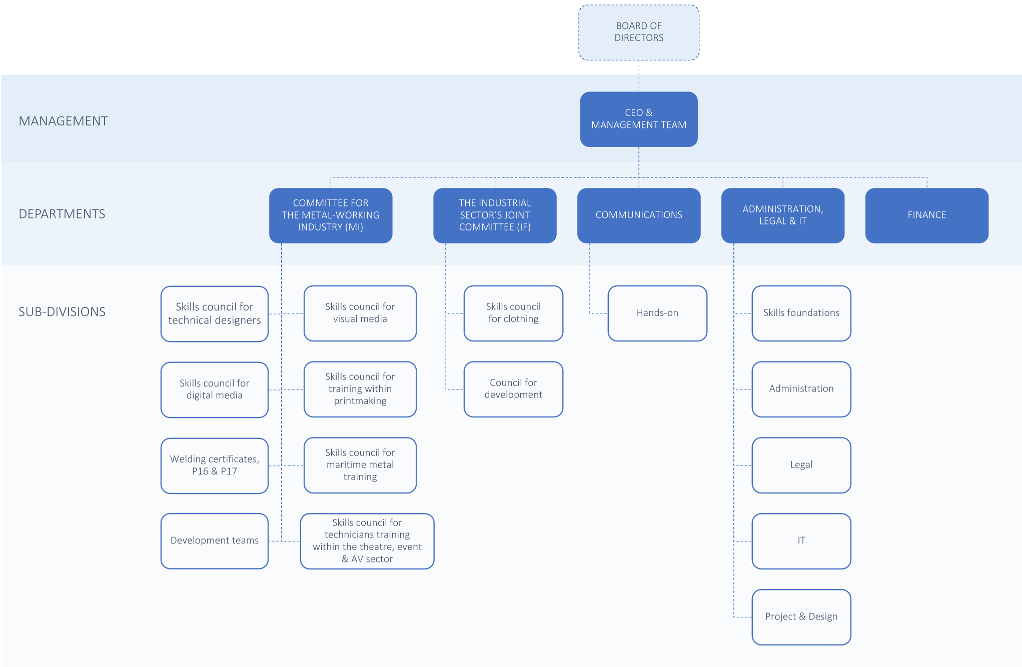


Figure 1. Organizational diagram of IU

IU has six overall tasks that emphasise the work the organization performs in this public sector arena. These overall tasks include: 1) Education development, 2) The operations of educations, 3) Events, 4) Communications, 5) Policy-support, and 6) Administration. Table 1 below shows an overview of the clustering of tasks and describes what the work is about, and which stakeholders are involved. Thus, in different ways IUs overall tasks aim to support various actors (e.g. schools, students, and industrial companies) in the network. This emphasizes a key characteristic of IU as an organization: most of the tasks are done in collaboration with other stakeholders in a large network that works to provide and support vocational education and continuing education for Denmark's industrial sector. I elaborate on this broader context in the following section.

	Task	The work is concerned with	Stakeholders involved
1	Education Development	Education development is governed/steered by the Vocational Education Act, IU's by-laws, and the Sector Skills Councils individual rules of procedure. The work is concerned with developing educational schemes and courses that can help to refine the given trade. The task also includes quality assurance at individual vocational colleges.	IU consultant(s), Sector Skills Councils, Trade associations, Industry companies, vocational colleges
2	The Operations of Educations	The many processes related to apprenticeships, for example, when companies apply for approval to train one or more apprentices; negotiation of special terms in apprenticeship contracts; and the comprehensive administration related to apprenticeship tests.	Administrators and education consultants at IU, vocational colleges (incl. examiners), Industry companies (incl. industrial examiners), students
3	Events	IU is responsible for organizing and coordinating several events yearly which aim to increase the prestige of IU's vocational educations as well as raise awareness about the educations. These events include, for instance, the "Metal Industry's Apprentice Award" and participation in the national skills championship.	Communications department, education consultants, administrators at IU, representatives from trade associations, students, industry companies
4	Communications	Attracting more "activity" (an increase in the number of students). Concrete tasks involve press work, press and marketing materials, maintaining various websites and portals, and campaigns. The tasks typically stem from the Sector Skill Councils' decisions and are resolved in close dialogue with them.	Primarily the communications department and relevant stakeholders (e.g. students, trade associations etc.)
5	Policy support	providing and supporting the trade associations with facts and knowledge, e.g. through producing statistics and data, creating inputs to the Committee for Vocational Education, e.g. suggestions to the "skills assessment procedure", digital competencies.	Education consultants, Sector Skills Councils, employer associations, unions.
6	Administration	Administrative operations related to amongst other things personnel management, IT operations and security, finance, running two foundations, providing legal services, and reporting.	Administrative and finance employees, IT consultants, and jurists at IU

Table 1. IU's core tasks (IU 2019a)

2.1.2 The Danish vocational education and continuing education system

The following brief description illustrates the course of a vocational education programme from a student's perspective. If, for example, you want to work as a Computer Numerical Control (CNC) technician in Denmark, you have to complete the CNC engineering degree (a vocational education). Such a degree provides basic knowledge about computer-controlled metal processing, which qualifies you to work at a manufacturing company, for example. It takes three years to complete the CNC technician programme, during which time the student studies at a vocational college and serve an apprenticeship at a local company that is certified to take on an apprentice. The degree is completed when the student passes her apprenticeship test. This example presents the main steps towards acquiring a vocational degree. However, what is not apparent to the student (and many others) is the comprehensive collaborative work that takes place among many different stakeholders, for these steps and progress to occur.

This research project reveals that it requires a lot of collaborative work by many stakeholders to provide and develop vocational education and continuing education. However, to make sense of the various stakeholders and their collaborative ways of working, it is necessary to explain a key, underlying governing structure that forms this organization: The Danish vocational education and continuing education system is governed by the Danish Labour Market Model. This model defines the organization of the Danish labour market and its partners (the state and the social partners, namely, employer associations and trade unions). The model is composed of 3 elements: Collective Agreements, Tripartite Cooperation, and a High Degree of Organization. Collective Agreements refer to one of the two predominant ways pay and working hours are regulated. The other way constitutes individual employment contracts. In Denmark, there is no statutory minimum wage. It is assumed that the social partners are accountable to the agreements being made. However, the state is a part of the negotiations when more general topics, such as "work environment" or "education", are being discussed. This constitutes the Tripartite Cooperation. The third element of the model is a High Degree of Organization, which mean that a large number of Danish workers are members of a trade union. Approximately 67% of Danish workers are members of a trade union, and the majority of the Danish companies are members of an employers' association (Danish Business Authority 2019)' (See also Publications 6).
This tripartite collaboration particularly influences IU and the broader network of stakeholders. The actors in the network are unified by their joint goals of attracting, educating, and graduating students who become skilled workers, and help to secure the current and future workforce for Danish industry. To accomplish this joint mission, the network has made arrangements to embed this governing framework in its everyday work

practices and organizational structures. Thus, the Danish Labour Market Model constitutes an essential set of guidelines for, and constraints on how this particular area of the public sector – and IU as an organization in it – (may) work, collaborate, and innovate.

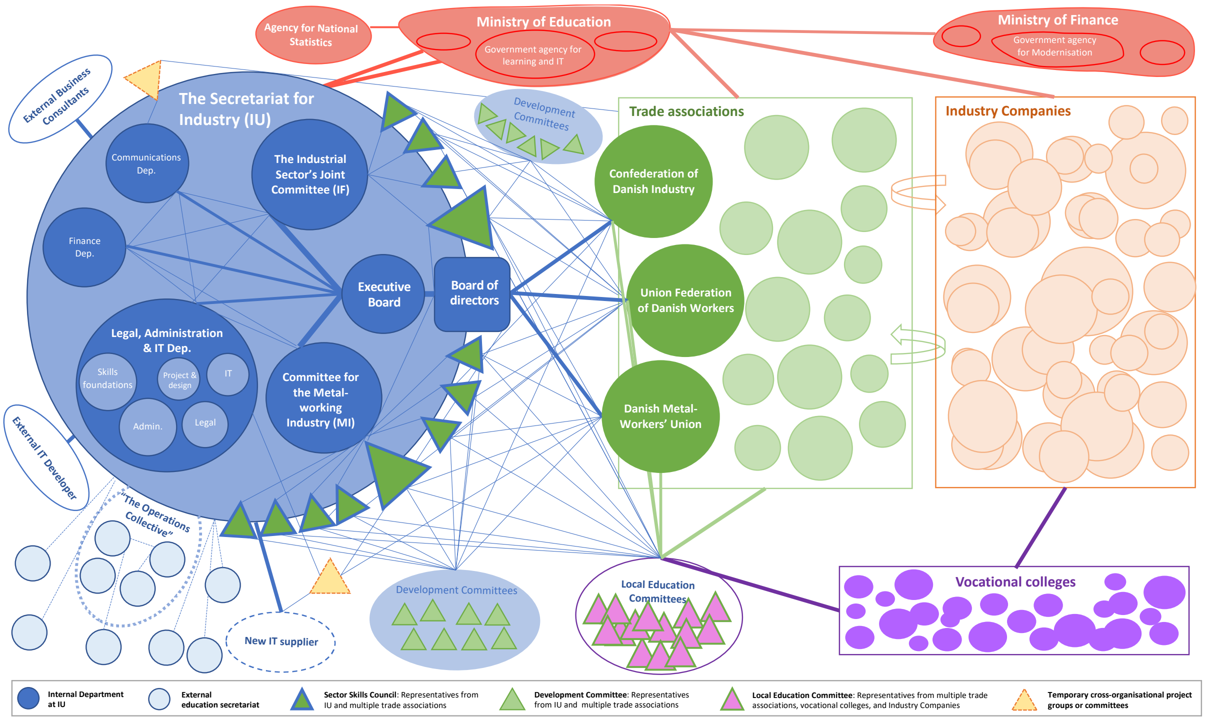


Figure 2. The public sector arena of vocational education and training in Denmark[1] (Publications 6, 9)

Figure 2 depicts the broad network of stakeholders that work and collaborate to organize, provide, and develop vocational and continuing education related to the industrial sector. From IU's perspective, this diagram illustrates how many external stakeholders the members of the organizations – in different ways and for different purposes – collaborate to provide the services necessary, for example, to support a CNC technician student from matriculation to receiving their degree (and eventually upskill through continuing education). I have provided a detailed description of this extensive network of stakeholders in Publication 6, where we examine the role data plays in this network. In this publication, we draw on the notion of 'social arenas' (Balka, Bjorn, and Wagner 2008) and conceptualize this network as a public sector arena, to frame the many stakeholders that include ministries, governmental agencies, vocational colleges, trade unions, employer associations, companies, and education secretariats. They all continuously collaborate to realize their shared or overlapping projects and concerns related to vocational education and continuing education courses. As mentioned, IU has been established in the midst of this public sector arena, to

1 The size of the figure does not indicate the actual size of the organizations. Owing to the situatedness of the research project, the figure highlights IU's perspective. This means that the figure might have been depicted differently if another stakeholder in the arena had been the focal point of this project (Publication 6).

support and facilitate much of the cross-organization collaboration that takes place in various councils and committees. This research project also reveals that much of this cross-organization collaboration is supported by data (Publication 1, Publication 2, Publication 3). Thus, data is a key component of many of the services that IU provides. I elaborate on this observation in the next section.

2.2 A data focus on services

Data is crucial for IU's service provision. As mentioned above, IU collaborates with multiple stakeholders to execute key tasks and manage central meeting structures that ensure collaboration and development work in the public sector arena. In most cases, data is an essential component the IU members need to be able to provide the organization's data-based. To further contextualize this research project, this section elaborates on the concept of service, and with this as a basis, elaborates on the data focus used to examine some of the services IU maintains and develops.

Traditional definitions of 'service' focus on its differences from 'product' (Sangiorgi and Prendiville 2017). Such attempts to define what constitutes a service often emerge from service marketing and management fields (Cowell 1980; Hipp and Grupp 2005; Lovelock 1983), and emphasize four characteristics: intangibility, heterogeneity, the inseparability of production and consumption, and perishability (Blomberg and Darrah 2015). Considering services as intangible market offerings has prompted a growing interest in understanding the notion of service as an approach to value creation (Sangiorgi and Prendiville 2017). This has caused a shift from viewing value as an embedded part of products, to 'value as co-created with users in their own context of use and in interaction with a wider array of other resources' (Sangiorgi and Prendiville 2017, 4). To articulate this shift, Vargo and Lusch (2004) introduced the notion of service-dominant logic, which provided an alternative to understanding services in itself, rather than as secondary to tangible products. Blomberg and Darrah (2015) emphasize that understanding services as co-produced by stakeholders (e.g. service providers and service 'receivers') promotes a focus on the necessary interactions among these stakeholders. Thus, the value of a service is determined by particular relationships among actors.
I use Blomberg and Darrah's (2015) conceptualization of services. They define services as 'fundamentally abstract propositions or transformations [that] are replaced with socio-material configurations of people and their know-how, artifacts and spaces' (Blomberg and Darrah 2015, 74).Their definition emphasizes that services depend on the 'doing' of relevant stakeholders, which implies that services are 'entrenched in practices and animated by practices' (Blomberg and Darrah 2015, 74).

Just as there are various perspectives on the notion of service, there exist many forms of services, including self-service, health services, public services. Blomberg and Darrah (2015, 12) highlight ‘the particular perspective taken to understand services and service worlds undoubtedly influences our ability to manage their impact and shape their design’. In other words, it is relevant to critically consider how one frames the notion of service in relation to a particular situation or context. This dissertation considers services that are in the public domain, supported by digital technologies, and of which data is a core component of the service provision and delivery. I refer to this type of service as a *data-based service*. For example, in the context of IU, a data-based service might be when the established ‘Statistics Team’ prepares and shares statistics and infographics that are a central part of the committee’s work to facilitate the meetings and align the stakeholders involved in the network (Publication 1). This specific focus on services is based on the growing service economy, which is, to a great extent, enabled by digital technologies (Blomberg and Darrah 2015). At the same time, digital technologies are to some extent shaped by the growing data economy, which influences governments’ and organizations’ opportunities to provide and innovate services (OECD 2019). However, despite the growing influence of digital technologies, data is rarely emphasized as a key element of these technical infrastructures in the context of service innovation and service design. By focusing on data-based services, this dissertation explicitly aims to understand services and service worlds, in part through the data that are needed to provide and/or develop services. Given that many of the services that IU provides and maintains are underpinned by heterogeneous data sources, innovation of these services (and the design of new services) might involve changes to the data that is needed to provide a data-based service. In the next section, I elaborate on why, in the context of IU, a data focus on service has relevance beyond the boundaries of this particular organization.

2.3 Service innovation in the age of datafication

To define the societal relevance of this dissertation, this section relates the research project to trends that increasingly transform social and human action into data. Researchers have conceptualized this trend as the ‘datafication’ of society, in order to describe the growing possibility of rendering many aspects of the world as data (Cukier and Mayer-Schönberger 2013). It is estimated that the amount of digital data worldwide will grow from 33 zettabytes in 2018 to an expected 175 zettabytes in 2025 (European Commission 2020). As the volume of data produced in the world continues to grow rapidly, the number of ‘datafied’ processes are accumulating in society (Gray, Gerlitz, and Bounegru 2018). The ever-increasing amount of digital data is being put to new uses with assistance of new technologies, for example, artificial intelligence and big data analytics. These developments are changing societies, with great implications for how daily operations (can be) run, in the public sector, among others things (OECD 2019).

The increasing focus on data in public sectors (and society at large) is important to this research, because it has a significant effect on the possibility of organizations improving their services and developing new services by using various data sources. For example, the city of San Francisco has improved service delivery to disadvantaged youth by creating an integrated data system, which allows for better case coordination among numerous agencies. Establishing this system also enabled a team to eventually provide a new service that generated insights that support caseworkers. For instance, the integrated system's aggregated data showed that 51% of San Franciscans that were registered in multiple systems were convicted of a serious crime. Thus, the interplay among these data sources created a way for caseworkers to anticipate and plan efforts that could identify high-risk youth, to divert them from getting into troubles in the future (OECD 2015b). Another example of data being used as a key component in service innovation in the public sector is the 'Cycling Infrastructure Database'. This database was created by Transport for London, to address the problem that people feel uncomfortable about bicycling. The database contains 'the location of more than 240,000 elements of the cycling infrastructure in London, including places to park and the location of cycle lanes' (Transport for London 2019), and provides a basis for understanding how to expand the walking and cycling networks in London. Although these examples may primarily emphasize the benefits of 'datafied' public services, it has been shown that it also presents challenges, including citizens' trust in government, data ethics, and data security. At an organizational level, the accumulating processes of datafication also present the challenge of a lack of skills and competence, and indicate a need to understand how more people at organizations may make sense of data (OECD 2019).

Previous research has addressed the emergence of data-driven public sectors from different perspectives. Examples include studies that underline how new technologies may be beneficial, but also involve socio-technical challenges (López-Quiles and Rodríguez Bolívar 2018; OECD 2019), emphasizing ethical issues related to the increased automatization of public services (Eubanks 2017), and broader political and economic concerns about using data as a fundamental facilitator of decision-making processes, or as an agent of capitalist interests (Kitchin 2014). These examples illustrate how the increased datafication of society is being addressed in various ways. Many of the aforementioned studies focus on the implications of machine learning and artificial intelligence that partially shapes the technical infrastructures that underpin (public) digital services. This dissertation adds to this research by using a practice perspective as a theoretical position for understanding how to we may develop tools that enable domain experts to design with data and data structures. I elaborate on this theoretical position in the next chapter.

Chapter 3: Theoretical Position

This dissertation takes a practice approach, and revolves around three topics: data, practices, and design. This chapter raises three questions based on these guiding topics and presents literature relevant to answering these questions, which positions the dissertation as a whole. First, I explain why this dissertation applies a practice perspective on data, and how this influences my understanding of data. This is followed by a brief account that questions how previous research has examined data practices in the workplace. Finally, I question what it means to design with data from a practice perspective, and use this to further elaborate on the two design approaches – participatory design and service design – that underlie my understanding and application of the notion of co-design in this dissertation.

3.1 A practice-based perspective on data

This section aims to establish my – and thus, the dissertation's – understanding of data in this research project. As indicated above, I take a practice perspective on data. Such a statement implies some fundamental questions: What are 'practices'? What constitutes a practice perspective? What are data? These questions are all broad. Therefore, I do not claim to provide definitive answers to the preceding questions. However, as part of the theoretical framing of this work, addressing these questions and bringing together my brief answers will guide us to what constitutes a practice perspective on data in this context.

We all engage in taken-for-granted, everyday practices. A well-known example of such a practice is teaching, which unfolds when a teacher and students come together to perform certain roles. Everybody (or at least those who follow social norms) acts according to their 'roles': for example, the students listen while the teacher presents a given topic that is part of a curriculum. However, at the time of writing, the world is challenged by the implications of the COVID-19 virus, and thus, the practice of teaching is modified, as students and teachers are at home, participating in classes in virtual classrooms. This example illustrates how practices are configurations of interconnected social, cultural, and material elements. Reckwitz proposes a more specific definition, which is that practices are 'a routinized way in which bodies are moved, objects are handled, subjects are treated, things are described and the world is understood' (2002, 250). when proposing this definition, Reckwitz (2002) distinguishes between 'practice' and 'practices'. Thus, he emphasizes that a 'practice' is performed individually at a particular place in a particular context; however, central elements of 'practices' are constituted collectively at a structural level (e.g. how students and teachers ought to behave while 'teaching' takes place) (Reckwitz 2002). Using this perspective to examine the world creates a view that draws attention to how practical knowledge that enables people to get things done in the world is transmitted, and how this knowledge is 'reproduced' in peoples' everyday activities (Blomberg and Darrah 2015). In other words, a practice perspective 'foregrounds the importance of activity, performance, and work in the creation and perpetuation of all aspects of social life' (Nicolini 2013, 3).

Although there are various ways to encapsulate what constitutes a practice approach, (Nicolini 2013), I choose to use Blomberg and Darrah's (2015) definition:

> A practice approach views social order as emanating from the repetition of routines over time and is thus grounded in social reproduction. Change then occurs as everyday routines evolve in response to shifts occurring in social and material conditions and is sustained through knowing actors, interpreting and responding to the situation at hand. (Blomberg and Darrah 2015, 2)

In the context of this research project, this definition of a practice approach is useful, in particular, because it emphasizes that 'routinized ways' (can) change if social and material conditions are altered. Moreover, Blomberg and Darrah (2015) examine their practice approach in the context of designing services. This leads them to argue that 'services are embedded within practice and are animated through practices' (Blomberg and Darrah 2015, 74). Thus, a practice approach enables me to explicitly consider the day-to-day practices at IU, to explore how an organization may work with data as a way to innovate and design their data-based services. Having established how I understand practice and what constitutes a practice perspective in this dissertation, I turn to consider the question 'what is data?' in the following section.

In computing, data is often regarded as a set of values of quantitative or qualitative variables concerning one or more objects or persons. In this perspective, data refers to a set of values that has been translated into a format, and is represented or coded in a way that is efficient and suitable for transmission, processing, and usage (Ramakrishnan and Gehrke 2003). The ever-increasing ability to collect, store, and analyse data has generated technical definitions of 'big data', early versions of which particularly emphasize the vast amount of digital data (Gandomi and Haider 2015; Laney 2001). However, others have suggested referring to big data in a technical sense, as a collection of processes that is needed to make data available for analysis (Berman 2013, 230). This project takes a different approach to understanding what constitutes data. From a practice perspective, the meaning of data and what data is in a social context is defined by the ways in which data is embedded in everyday work. For example, the way in which a computer scientist structure a database define what will determine data (Bowker 2014). Likewise, the meaning of data in a specific use context is depending on both the structure of the data developed by the work practices of software engineers and the work practice of the users, in which the data is used for a specific purpose. This means that data is dependent on everyday practices. Such an understanding of data is consistent with critical data studies that have problematized how data-intensive technologies and approaches mediate most aspects of life (at least in computationally advanced societies). This field of research has in various ways questioned what constitutes data (Kitchin 2014; Rosenberg 2013). Addressing this central question has led to a growing acknowledgement of considering data as socially constructed, rather than neutral, and a practice perspective is one way of looking at the social construction. Based on this perspective, scholars have argued that data is not inherently objective or raw (Gitelman 2013). One example is offered by Sumarjoto et al. (2016, 39), who suggest understanding data 'as a "lively", rich and emergent aspect of human experience that constitutes part of how we continue to make sense of the world'. Moreover, critical data scholars argue that data is not merely collected, but also produced, and this production is steered by epistemic acts of categorizing and prioritizing that otherwise have no boundaries or shared experience (Bowker and Star 2000). The same applies when data is collected, used, and reused. Thus, in this perspective, 'big data' refers to 'the data phenomena of that very moment' (Beer 2016) – rather than the volume, velocity, or variety of data. As Kitchin (2014, 2) emphasizes, 'data do not exist independently of the ideas, instruments, practices, contexts, and knowledge used to generate, process and analyse them'. This emphasizes that data is inextricably linked with the assemblages in which they are embedded. Consequently, a practice perspective on data is a view that emphasizes that data production is based on social and material (here, in particular, technical) elements, and emphasizes the need to make sense of the work practices related to the creation and subsequent 'liveliness' of data.

3.2 Studying data practices in the workplace

This dissertation's practice perspective on data emphasizes the need to examine how people in organizations actually use data. This is consistent with usage in the field of computer-supported cooperative work (CSCW), which focuses on mundane work and how to understand the role information technology plays in workplace settings, for example. This section questions how previous research has examined data practices in the workplace to position this dissertation's work with IU.

Prompted by the increasing use of, and focus on data at organizations, previous CSCW research examined data-related practices, for instance, the handling and sharing of data within and among organizations (Jackson and Baker 2004; Passi and Jackson 2018). Some have studied the emergence of data and information infrastructures (Leonelli 2016). Thus, data work and digital data practices have been studied in various fields, including e-Science,

library science, Information science, and Ocean Informatics (Futrelle et al. 2009; Karasti and Baker 2008; Koesten et al. 2017; Paine and Lee 2020; Scroggins et al. 2019).
These studies highlight the importance of the dynamic human shaping of data (Pink et al. 2018), and emphasize the inescapable social aspects of data generation, analysis, and usage.

To articulate work practices related to data, scholars have conceptualized the notion of *data work* to address 'any human activity related to creating, collecting, managing, curating, analysing, interpreting, and communicating data' (Bossen et al. 2019, 466). This broad definition demonstrates that, to various degrees, data work is included in many forms of work, and thus it is difficult to demarcate data work. In line with Holten, Møller and Bossen (2019), I argue that this underlines a point about data work. Specifically, data work is often entangled in other practices, and thus may easily become invisible work because it is categorized as a task that removes the focus from the data-related aspects of the work, for example. Moreover, the concept of data work is a useful part of the theoretical framework of this dissertation, because it suggests that data-related practices must be articulated (Strauss 1988), in order to make sense. Through my data collection (see detailed description in chapter 4), I have attempted to encourage this articulation and representation of data-related practices. Another reason that underlines the usefulness of the notion of data work in this context is found in the concept's capacity to highlight the collaborative aspects of producing, collecting, and using data. For instance, Fischer et al. (2016) report on a co-design project that explored how the data work of professional energy advisors could be augmented by environmental data from sensors set up in clients' homes. Their findings suggest that data work revolves around the interpretation of data. In other words, '*that the meaning of the data cannot simply be 'read off' the representations of it (e.g. graphs and charts). Rather what the data means, what it refers to, what it reveals is, without remedy, wrapped up in the situated interaction between parties to its use*' (Fischer et al. 2016, 5933). This point is further supported by Bossen et al. (2019), who considered data work in healthcare. They stress that '*data work is interdependent with – and has implications for – data work at other sites*' (Bossen et al. 2019, 468). Thus, data work practices seem to be characterized by being complex, distributed, and often dependent on multiple stakeholders (Bossen et al. 2019; Fischer et al. 2017). The next section questions how this practice perspective on data and data work may emerge in the context of design.

3.3 Designing with data from a practice perspective

The third and final topic used to position this dissertation revolves around design. A central part of the general objective of this research is exploring how organizations may design with data to innovate for their existing data-based services, or develop new ones. The underpinning practice perspective of this work also influences the understanding of what it means to design with data. From a practice perspective, designing with data implies a need to involve the people who are producing, using, and making sense of the data in the design process. For this reason, the dissertation works towards a *co-design perspective on data*, to show that designing the data and data structures that underpin data-based services has to take place in collaboration with the people involved in a (future) service's embedded practices. The notion of *co*-design suggests 'the collaborative, cooperative and collective or connective nature of this engagement in design' (Zamenopoulos and Alexiou 2018, 12). Therefore, in this context, co-design is understood in a broad sense that refers to the general concept of a number of people collaborating on design. Thus, in the context of this dissertation,

> 'Co-design means that people come together to conceptually develop and create things/Things that respond to certain matters of concern and create a (better) future reality. People come together despite, or because of, their different agendas, needs, knowledge and skills'. (Zamenopoulos and Alexiou 2018, 12)

There exists a number of design approaches that have developed methods and techniques, specifically to involve people – users and domain experts – during the process of designing. To explore how we may design with data from a practice perspective, I have made use of Participatory Design and Service Design, which in various ways offer inclusive design disciplines and toolboxes. I have chosen to use participatory design because it is a design approach with a strong emphasis on participation and participatory design processes. I draw on service design because this approach includes several notations that support constructs such as stakeholder mapping and user journeys (Stickdorn and Schneider 2011). Specifically, I used service design to explore how we can develop ways to relate the technical and the social in the participatory design process. In the following paragraphs, I briefly describe both design approaches, and argue why it made sense to combine the two in the context of this research project.

3.3.1 Participatory design

Participatory design is relevant to this research project because it is a design approach that explicitly addresses the democratic aspects of design (Simonsen and Robertson 2012). Participatory design employs direct interaction with users to articulate, create, and develop users' ideas and visions. Thus, shared experimentation and reflection are central aspects of a participatory design process (Kensing and Blomberg 1998; Simonsen and Robertson 2012). Participatory design (or cooperative design) emerged in Scandinavia in the 1970s, where research projects that addressed user participation in systems development were established. Key examples include the NJMF, Demos, DUE, and UTOPIA projects (Simonsen and Robertson 2012). These projects resulted in the development of strategies and techniques that enabled workers to influence, design, and use computer applications in their workplaces. Principles and practices for participatory design evolved around ideas of tools and processes to facilitate participation and joint negotiation, and thereby elicit respect for varied knowledge, opportunities to learn about others' domains of knowledge, and collective learning (Blomberg 2009; Greenbaum and Kyng 1991; Schuler and Namioka 1993). Owing to the workplace settings in which many of the early participatory design projects took place, the design ideals primarily focused on 'democracy at work' (Björgvinsson, Ehn, and Hillgren 2010). This focus is particularly fitting in the context of this research project, however, over time, participatory design has spread, and been applied to many other settings with various groups of people involved (Manzini 2015). This includes projects that aim to empower citizens, patients, and healthcare workers or vulnerable groups in the context of IT development and implementation (Bossen and Grönvall 2015; Ddamba and Dittrich 2015; Malmborg et al. 2015). Thus, participatory design research has developed and continues to question what constitutes participation, and how learning may be included in design processes.

3.3.2 Service design

Technological development has led to increased digitization at organizations, and digital transformations of society at large (Gray, Gerlitz, and Bounegru 2018). Especially during the past three decades, this has led to a growing focus on digital services, however, in ways where the supporting IT infrastructures often remain invisible. The focus on services has made room for the service design discipline, which aims to explicitly design and innovate socio-material configurations that constitute the abstract propositions known as services (Blomberg and Darrah 2015). As such, service design is about the process and act of designing services (Kimbell 2011). This discipline originates in the hybridization of business and management, service science, and other earlier design approaches, including participatory design. Service design has been characterized as a human-centred, holistic, creative, and iterative approach to creating new, or improving existing, services (Blomkvist, Holmlid, and Segelström 2010; Meroni and Sangiorgi 2011). Service design has adopted participatory and co-design approaches to involve stakeholders in the design process. In practice, this means that service design draws on co-design methods and participatory design techniques. However, the service focus has expanded the toolkit to include notation that emphasizes service systems and user journeys (Stickdorn and Schneider 2011).

3.3.3 Combining fields of practice

This dissertation is infused with a practice perspective, which also involves an understanding and application of design. I consider participatory design and service design fields of practice (Blomberg and Darrah 2015). Previous research has indicated that these fields of practice overlap to some degree, given service design's partial participatory approach (Blomberg 2009). Specifically, Saad-Sulonen et al. argue that there is 'an opportunity to combine existing participatory and service design approaches to participation in the way they weave connections between design, IT, digitization and democracy, focusing on the context of the public sector' (2020). An argument for using both disciplines when designing includes service design's increasing popularity and implementation as an approach to innovation in industry. This may benefit participatory design that 'has remained academic', and thus, the participatory design 'approach to democracy and IT has not yet gone mainstream' (Saad-Sulonen et al. 2020). Another reason to combine these fields of practice includes the opportunity to increase the scale of design, and thereby facilitate the design of democratic infrastructures and governance (Saad-Sulonen et al. 2020). These arguments emphasize the relevance of drawing on these two fields of practice in the context of this research project. Therefore, this dissertation draws on both participatory design and service design, and thus combines these fields of practice to explore how organizations can innovate and design their data-based services.

Chapter 4: Methodology, project description, and research activities

The main aims of this chapter are to present the project's applied research methodology and describe the project and its research activities in greater detail. As this is a paper-based dissertation, it comprises 8 publications (presented in part 2), each focused on different parts and aspects of the whole research project. However, when woven together, they collectively constitute a proposal to develop a co-design perspective on data. To outline the recurring methodological considerations throughout this research project, this chapter is divided into four parts. First, it presents the project's overall action research methodology. Second, I describe the project, and elaborate on the research activities throughout the process. Third, I present the Data Science for Local Government project in which I participated during my research-abroad stay at the Oxford Internet Institute, and show how this work provides a triangulation study for this dissertation. Finally, I elaborate on the steps taken to ensure scientific rigour throughout the research process.

4.1 Action Research

This section describes my understanding of action research, and explains how an ethnographic stance and research-through-design inspired and informed my way of conducting this action research project.

Action research is a methodology that is based on explicit democratic, participative, and interdisciplinary values, which aim to support collective action and (social) innovation (Gaventa and Cornwall 2008; Hayes 2011). A key characteristic of action research is that it aims to induce change, to improve certain aspects of the targeted research domain. To do so, action research often involves participants (e.g. members of an organization) in the preparation and implementation of the research. Therefore, when doing action research, the focus is on making research efforts *with* people who are experiencing real challenges in the research domain, rather than to doing research *for* or *about* the people involved (Hayes 2011).

Historically, Action research builds on practice-oriented currents such as the work of the early pragmatists, including John Dewey, who were interested in everyday practices and concerns related to the public (Hayes 2011; Robson 2002; Stringer 2007). Hayes (2018) emphasizes that Dewey in particular developed the idea that thought and action, or practice, are inseparable. Therefore, the practice has been the core of action research from the very beginning. However, Kurt Lewin first made use of the notion of *action research* (Lewin 1946). He regarded action research as a way to learn about organizations by attempting to change them (Robson 2002). Thus, Lewin's push for change made intervening in research settings an acceptable approach to conducting scholarly inquiry (Hayes 2018). On this basis, action research has continued to encourage organizational change and development. Over time, action research has been established as an approach with a strong and explicit concern for emancipation. Lewin, whose research and publications emerged right after the Second World War, treated action research as an approach to advancing democracy (Robson 2002). This was eventually interpreted and taken on as 'an embodiment of democratic principles in research' (Robson 2002, 200). To emphasize this emancipatory aspect of this research approach, Stringer (2007) refers to 'community-based action research'. This is to underline the fundamental premise of action research, which is to empower groups of people in various settings by enabling the participation of those involved in a given problem in the research process. This dissertation draws on this viewpoint, in that it also works with the belief 'that all stakeholders – whose lives are affected by the problem under study – should be engaged in the processes of investigation' (Stringer 2007, 10).

In contrast to the positivist research tradition, where the ideal is for the researcher to have an external and objective relationship to the field of study, in action research the ideal is for the researcher to actively contribute to democratic development and change in the field (Bradbury 2015). Action research emphasizes an understanding of the world and a change/transformation in the world. Thus, action research differs from other research approaches in that these beliefs put the researcher and his or her relations with the research participants at the centre of the research process. Moreover, action research explicitly recognizes that this constellation influences all aspects of data collection and analysis, how the research is communicated, and how change is implemented (Hayes 2011).

Action research ascribes to ontological and epistemological commitments which differs from other research approaches (McNiff and Whitehead 2006). In research, ontological commitments refer to how we, as researchers, consider ourselves in relation to our work and to other people, such as research participants. In action research, it is crucial for people to be aware of these commitments, owing to the action researcher's deep engagement with the research domain, and the partnerships between the researcher and the research participants (Hayes 2011). This high level of engagement means that action research cannot be value neutral, 'because researchers bring their own values with them into the field. Researchers inherently act in relation to the field site, the research literature, and the available resources' (Hayes 2011 p 3). Moreover, an action research approach would claim that it is not only the researcher who influences the research, it is simultaneously influenced by other actors involved the project (e.g. the research participants, collaborators, and the broader community), who also bring their own values to the process.

Ontological commitments influence the underlying epistemological commitments that action research ascribe to (Hayes 2011). The role of the action researcher is to be a co-creator of knowledge through trusting and equal relations with research participants. This co-creation of knowledge includes both examining and documenting existing situations, and experimenting with causing change, which is meant to improve the situation while maintaining a democratic perspective throughout the process (Aagaard Nielsen and Svensson 2006). Thus, an action research approach argues that knowledge is generated through collective research processes. This means that knowledge generation implies that action researchers are committed to the idea that knowledge is co-constructed and evolves (Hayes 2011). In other words, an action research approach implies that knowledge is generated *through* action. Hayes emphasizes that 'a practice perspective provides action researchers with a way to engage and learn about the world by focusing on everyday practices. In this view, doing and knowing are more important than what is done and what is known, meaning that the practice perspective engages with the world in its becoming rather than the idea that it 'is' at any given point in time' (2018, 303–4). This again identifies action research's inherent focus on practices, which connects with this dissertation's underlying practice perspective.

Action research is a perspective that employs an array of methods, and thus is not itself *a* method. Hayes (2018) suggests considering action research a 'meta-practice', to shed light on how action research and practice theory's shared academic traditions, the organization members application of scientific thinking, and an emphasis on details from a day-to-day practice provide a compelling approach to transformative technological interventions and creation of critical knowledge. To encourage these aspects of action research, throughout this research project I took an ethnographic stance. Moreover, owing to the project's focus on design, a significant part of the research revolved around designing as a form of action to create change in the research domain. In the following paragraphs, I elaborate on the roles of these subordinate but supporting methodological approaches.

4.1.1 An ethnographic stance

Stringer (2007) argues the action researcher's task is to enable different stakeholder groups to formulate 'jointly constructed descriptive accounts of the situation at hand' (p. 67).

As a way of understanding the existing situation, I have taken an ethnographic stance, meaning that I draw on characteristics of ethnographic fieldwork. Blomberg et al. (1993, 139) describe ethnography as 'a way of developing a descriptive understanding of human activities', and emphasize guiding principles for doing ethnography. These include conducting the field work in a field setting, considering how the activities studied relate to a broader social context, developing a descriptive understanding of people's actual behaviour, and understanding the world from the participants' point of view (Blomberg et al. 1993, 125–27). This use of an ethnographic approach aims to develop a description and interpretation of human activity in its everyday settings, where the activity takes place (Robson 2002). Although this is an action research project, I wanted to design interventions and cause change based on a rich understanding of the existing situation and data practices from the organization members' point of view. Thus, taking an ethnographic stance enabled me to better understand the relationalities of the practices that might be involved in, and affected by the action research.

4.1.2 Research-through-Design as a critical inquiry process

The second supporting methodological approach that inspired this action research project is research-through-design (Frayling 1993), which I used as a critical inquiry process. Research-through-design has been defined as '*a research approach that employs methods and processes from design practice as a legitimate method of inquiry*' (Zimmerman 2010). Moreover, it is known as a research approach that acknowledges how design actions play a formative role in the generation of knowledge (Stappers and Giaccardi 2014). The notion of research-through-design originates in Christopher Frayling's influential distinction among three design-research approaches: research into art and design, research through art and design, and research for art and design (Frayling 1993). Since then, research-through-design has established itself as a maturing research discipline, and has been applied in a growing number of studies in the field of human–computer interaction (HCI) (Hansen and Halskov 2018; Vaughan 2017). Zimmerman et al. (2010) argue that the increased interest of the HCI community relates to the growing engagement with 'wicked problems', which demand more complex design practices. Furthermore, Zimmerman et al. (2010) emphasize three main reasons for using research-through-design as an approach to scientific inquiry. (1) A research-through-design approach allows the researcher to rely on designerly activities as a way to address complex situations with vague or conflicting agendas. (2) A research-through-design approach prompts the researcher to focus on research for the future, rather than that of the past or the present. Finally, (3) the focus on the future that a research-through-design approach embeds enables the researcher to be an active and intentional producer of the change desired by the participants in the research domain. As Koskinen

(Hansen and Halskov 2018; Vaughan 2017) has also pointed out, these three reasons provide useful overlaps between action research and research-through design.

To summarize, this research uses an action research approach supported by an ethnographic stance and research-through-design as critical process of inquiry. The interplay of these methodologies has enabled me, as a researcher, to engage with members of IU to stimulate organizational change that is based on a rich understanding of the existing situation and, in part, takes place through design activities. In the next section, I elaborate on the action research interventions and the related research activities.

4.2. Project description and research activities

This section presents the research project in greater detail. First, I provide an overview of the whole project. This is followed by an overview of the research activities. Finally, the section ends with a description of the project's four general research phases.

4.2.1 Overview of the research project

A widely used representation of action research presents the approach as a spiral or cycle, where each intervention involves three general stages: (1) planning a change, (2) acting and observing what happens following the action(s), and (3) reflecting on the processes and the observed changes, to plan for further change and the continuation of the cyclical process (Robson 2002). This view of action research also influenced the design and implementation of this action research project. Figure 3, below, illustrates the general design of the research project.

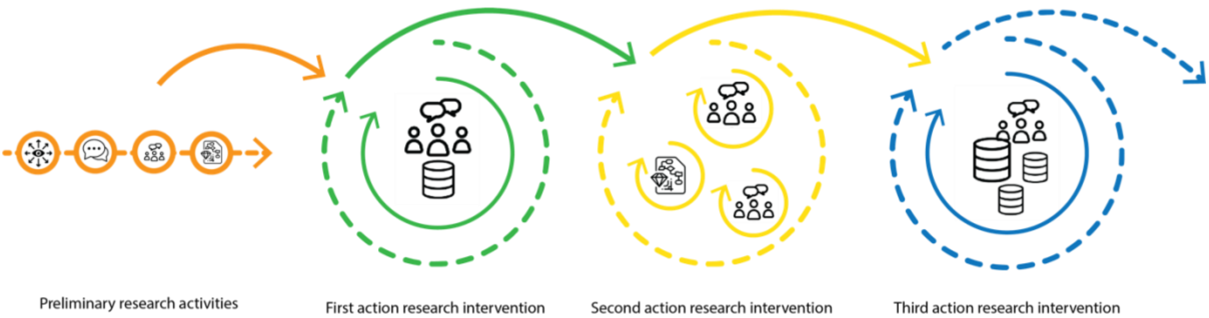


Figure 3. General structure of the action research project. The colour codes (orange, green, yellow, and blue) are related to the overview of the research activities and the related data collection presented in figure 3.

As the diagram shows, the project began with preliminary research activities, followed by three action research interventions. These four general phases are explained in greater detail in section 4.2.3. At a general level, the three action research interventions were designed to have different purposes that addressed various aspects of the broader research objective, which is to understand how organizations may innovate and design their data-based services. However, it is important to note that the interventions were highly interrelated, in the sense that the learning, methodological explorations, and feedback from the first intervention fuelled the second intervention, which fed the third. Moreover, each action research intervention was designed to create two parallel learning cycles, and thus produce knowledge in two different ways. First, it aimed to provide insights and improve the situation of the research problem of each intervention (this is depicted as the central cycle in each of the interventions in figure 2). Second, the action research interventions also aimed to extract new organizational knowledge from each of the processes (this is depicted by the encircling broken lines in figure 2). In the next section, I elaborate on how this general structure of the research process manifested through the research activities.

4.2.2 Overview of the research activities

This section provides an overview of the research activities undertaken and various methods used during this project. The idea of applying several methods to carry out various forms of inquiry is well-known in action research. Employing several forms of inquiry allows for various perspectives to be represented in an action research intervention, and thus in the change it intends to create (Bradbury 2015).

This dissertation is based on a large body of empirical work, which was conducted between September 2016 and December 2019. As stated in Publication 5, the general fieldwork comprised more than 250 units of observation, including (1) design, facilitation, and documentation of 22 workshops, (2) participation and observation of 51 meetings, (3) 12 in-depth interviews, (4) approximately 70 documents (email, reports, presentations), (5) images, and (6) ongoing field notes to document informal conversations, observations, and reflections throughout the project period. Figure 3, below, illustrates the (sometimes concurrent) research activities that formed the action research interventions. Building on figure 2, this diagram depicts the four general data-collection processes that informed the action research. The orange process includes the preliminary activities, and what I term ‘general activities’, that is, research activities that not did not specifically relate to one of the action research interventions. However, these general activities supported my ethnographic stance by enriching my ongoing understanding of the existing situation. In contrast, the green, yellow, and blue data-collection processes illustrate the inquiries that were associated the individual action research interventions. Finally, figure 3 also presents my research-abroad stays at the Oxford Internet Institute, (OII) at Oxford University, and more recently, at the Computer Supported Collaboration Lab (CSC Lab) at the University of Washington. However, during my stays abroad I maintained close contact with IU members and management, and as the diagram shows, I conducted interviews and participated in meetings virtually.

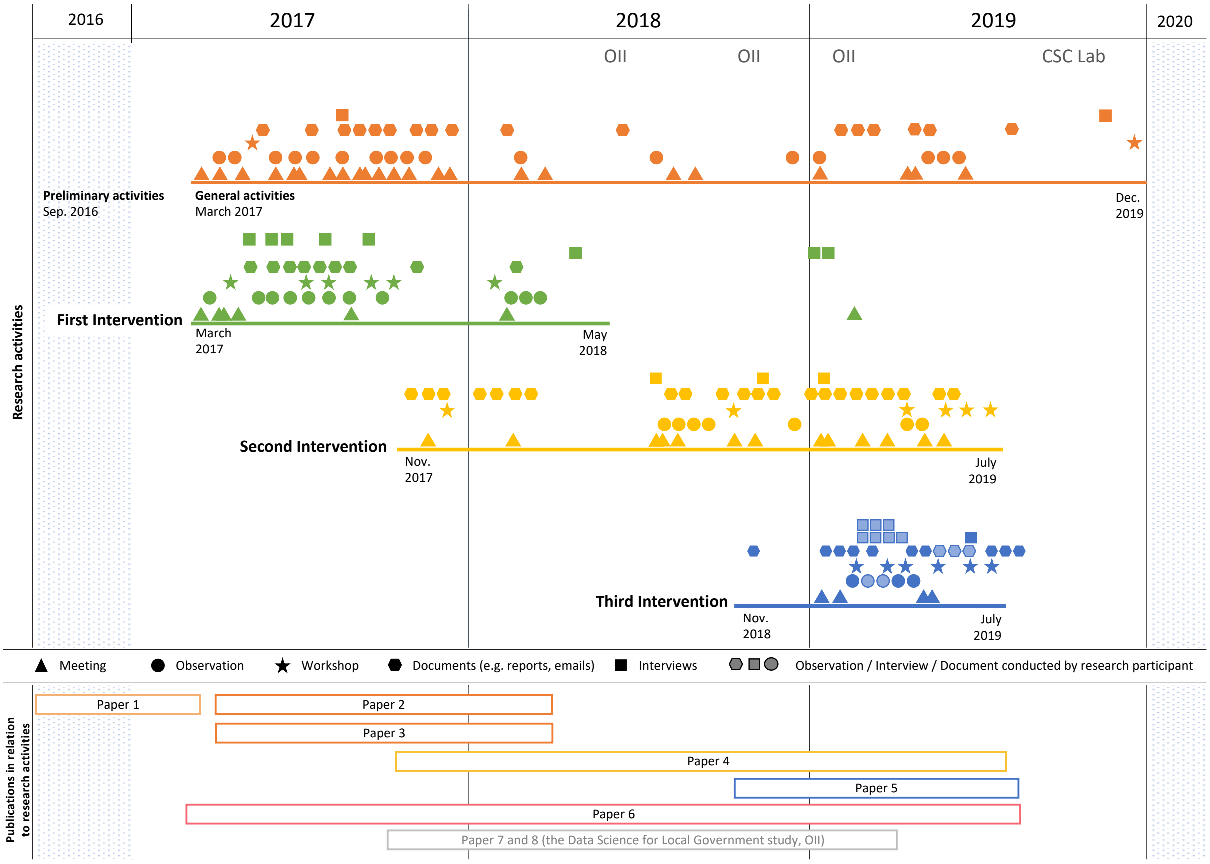


Figure 4. Overview of the action research interventions and research activities, and how the publications relate to the field work.

Throughout the project, I used various methods, and thus conducted (participant) observations, semi-structured interviews, and read relevant documents. I did this to continuously develop my understanding of people's activities and the changing contexts of the project. Using these methods enabled me to develop a rich understanding of the field site. Furthermore, these means of inquiry generated insights that were used to inform the design of workshops and tools. In the following section, I elaborate on how these various methods came into play during the four data collection processes.

4.2.3 The project's four data collection processes

Preliminary activities

In this section, I elaborate on two principal preliminary activities I conducted with IU prior to or during the initial phase of my doctoral project. Stringer (2013) argues that preliminary activities are an important part of action research, because they support interaction amongst stakeholders, and helps to establish the researcher(s) presence in the research setting.

The first preliminary activity was an exploration of existing data practices at IU. The exploration included desk research, four semi-structured interviews, 12 hours of observation, and a future workshop with IU members. During this period (September to December 2016) I worked at IU 2 to 3 days a week. This engagement with the organization provided initial insights into the organization, the key tasks for which it was responsible, with respect to the broader network, and how these tasks related to existing data practices. For example, during this time it became clear that members of the organization often used data as either *evidence* to ensure accountability on the part of the stakeholders involved, or as a *basis for decisions* that would help to steer negotiation and decision-making processes. I also identified four categories of tools that supported data work at the organization. These categories were Scripts, System Interfaces, Tables, and Infographics. My initial engagement with members of IU showed that the currently-available tools did not support the employees' desire for future data-based service provision (see Publication 1 for a more elaborate description of this engagement and the findings).

The second preliminary activity revolved around the development of a map of IU's IT infrastructure. The aim was to expand the understanding of existing data practices by visualizing the central IT systems and data sources that IU employees used. Through a collaboration with IU's (at the time) only external IT developer and members of the organization, we developed a rough visual overview of the organization's IT infrastructure (see image 1, below). The organization had never done this, therefore this map (see figure 4 below) became a tool for understanding how internal systems were integrated, and how and to what extent the organization's IT infrastructure – and by extension, data practices – depended on external stakeholders' IT systems and web services.

Images 1. and 2. Workshop with IU's external IT developer

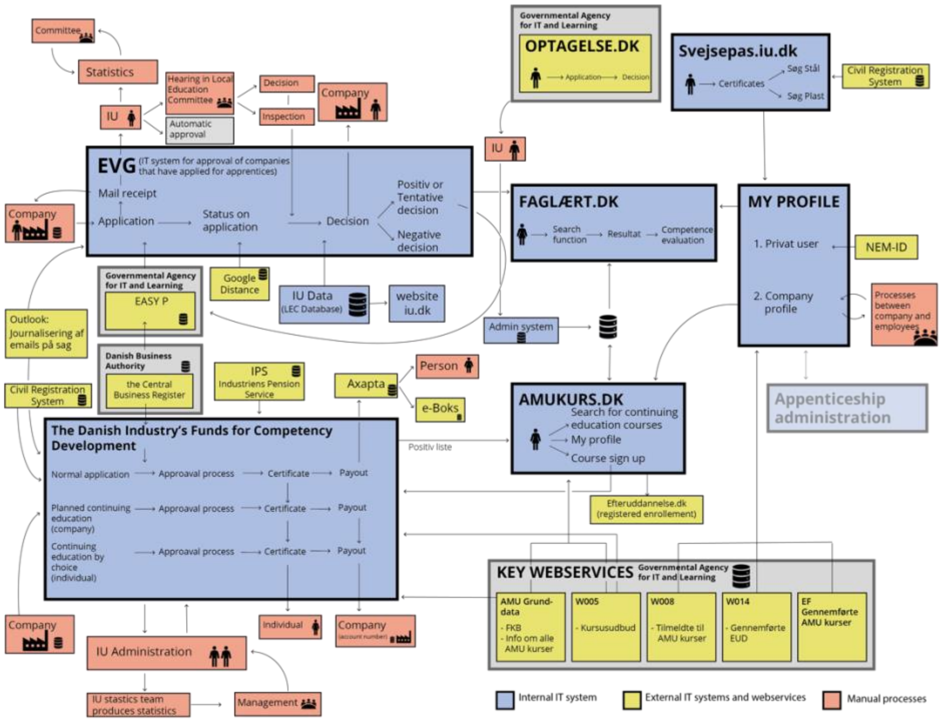


Figure 5. Map of IU's IT infrastructure (March 2017)

Overall, the preliminary activities were very useful, as they helped me to develop a rich understanding the field site. Specifically, they helped me to establish contacts within the organization, to identify stakeholder groups, to better understand the complex world of vocational education, and to learn the language of these particular domain experts, each of which provided clues to the social dynamics involved in the data practices related to the organization's many service-provision and development processes.

The first intervention

The first intervention was aimed at understanding IU's existing data practices in greater depth. The preliminary activities indicated that IU was highly dependent on other stakeholders and external IT expertise, IT systems, and data sources, to provide their data-based services. However, it was not yet clear how the large network of external stakeholders influenced the data work performed by members of IU. Based on the map of the IT infrastructure, I, in collaboration with IU management, decided to focus the first intervention on one IT system and its related database, known as 'The LEC Database' (see figure 5). This particular database contains crucial personal and organizational data about all the individual Local Education Committee (LEC) members of the 165 LECs that fall under IU's service provision.

Administrative employees at IU had been requesting a new IT system that could better support the data work related to the LEC. However, they did not know how to address the question. To unpack this problem space, I interviewed and observed the administrative workers at IU who were responsible for maintaining and updating the data in the LEC database. This helped me to identify key stakeholders from external organizations. This led to subsequent interviews and participant observation with representatives from four organizations that are IUs key stakeholders. In this way, cross-organizational collaboration became apparent as a critical aspect of the existing data practices related to maintaining and updating the LEC data (also see Publication 2 for a detailed description). From this I argued that IU should invite the central stakeholders to a couple of collaborative design workshops, to make sense of the various data needs before proceeding with the development of a new IT system and LEC database. Together with administrative employees at IU, I organized two collaborative workshops with the external stakeholders, who had been previously involved in this process. The insights from these workshops led to a third internal design workshop, which aimed to concretize the improved LEC service concept. These three workshops are described and discussed in detail in Publication 3, and therefore I will not elaborate here. During this intervention I evaluated my observations and insights with IU members and the external workshop participants. The evaluation was part of the workshop and included follow-up interviews.

The initial action research intervention was meant to be an approximately 6-month long study that would identify central (data) needs and requirements for an improved LEC database and IT system. However, the revealed cross-organizational data interdependence resulted in a very complex process that is (at the time of writing) still ongoing at IU. The unexpected complexity clarified the need for IU to acquire internal IT expertise to manage this (and future) development processes. Although this action research intervention has not (yet) resulted in an improved LEC database and IT system, it led to organizational changes. For one thing, the observed effects of this action research intervention manifested as the establishment of an internal IT department at IU. This changed the way questions related to data and IT were organized and addressed. This action research intervention also introduced the organization to co-design, which demystified 'designerly' ways of working, to guide innovation work and projects. The introduction of these creative practices also became the foundation for the second intervention.

The second intervention

The second intervention aimed to develop design competence at IU. This objective was prompted by the initial project setup, which included the goal of empowering the organization in ways that would enable IU members to use project insights themselves, after the research project ended. Thus, the second intervention aimed to provide the organization with tools to design with data. When I entered IU in the autumn of 2016, the organization had very limited experience of 'designerly' ways of working and innovating. At that time, IU could be characterized as an organization with very limited design capacity (Malmberg 2017). For example, when I attempted to organize the Future Workshop as part of the preliminary activities, an employee at IU asked, *'Do we really have to call it a workshop – it scares people. I think we should stick to calling it a meeting'* (IU employee. November 2016). This reflected the attitude – at that time – to creative methods as part of general project work. Therefore, to enable IU to design and innovate data-based services after the research project ended, we decided to focus on developing (co-)design capabilities at the organization.

The second intervention was initially designed in a way that aimed to establish a formal service design group consisting of 4 or 5 IU employees from various departments. The idea was that the group would be taught about design thinking, tools, and techniques, to further support other projects and groups. This approach built on learning from other large organizations in Denmark and abroad, which had implemented design thinking and service design teams in this way. Moreover, previously IU had successfully established a cross-departmental Statistics Team. However, it was not possible to establish a service design group. Owing to the organization's limited resources and limited knowledge about the benefits of design thinking, management was reluctant to allocate people and resources to a particular service design group. The proposal was turned down, despite meetings and a workshop with carefully designed activities that aimed to convince IU management of the value of a service design group (see Publication 5 for a more elaborate description of this process). It has been suggested that as action research projects increase in scope and complexity, management's resources and tools become increasingly relevant (Stringer 2013). This was also the case here, and this development forced me to reconsider the design of the second intervention in a way that considered management resources to a greater extent. This resulted in the design of a more fragmented action research intervention. Rather than having one specific service design group, my new proposal was to infuse (co-)design thinking and service design tools into already-planned projects. The idea was to limit concerns about 'additional tasks' and 'lack of time', while building participatory design capabilities at the organization. The management at IU approved the proposal of the so-called ‘Service Design Micro Cases‘ (see Publication 5 for a description of the cases).

Observing the intervention’s effects showed that the diffused approach to building design capabilities generated an increasing number of autonomous co-design initiatives at IU. The fieldwork documented a greater appreciation of design methods and creative problem solving. Over time, this appreciation was formalized by establishing a ‘Project and Design’ subdivision as part of IU’s organizational structure. Also, my close colleague and co-author was promoted and designated ‘Service Designer’. My research abroad stays allowed me to observe which initiatives worked well or less well when I, an agent of change, was not present in the organization. Through ongoing evaluative discussions with the management, the Service Designer, and the IT department, it became apparent that the success of an initiative was largely due to the incorporation of tools and techniques. For example, visualization became a key tool for the members of the IT department, when they realized how to make the technique ‘their own’. Another example is that of the education consultants, who incorporated a design process model to promote innovation work during committee meetings. However, being creative and innovative *with data* was still a challenge for IU members. Therefore, this question became the focal point of the third intervention.

The third intervention

Developing learning by examining data practices and building design capabilities at IU led to the third action research intervention, which focused on how IU members could work creatively and innovatively with data. Thus, the aim of the third intervention was to combine the learning from the previous interventions that addressed existing data practices and building design capabilities at the organization, to further explore how IU could explore new data sources, and experiment with their usefulness.

The second intervention’s fragmented approach proved to be an effective way to establish sustainable co-design capabilities at IU. However, it also resulted in the intervention being prolonged. Therefore, owing to the scope of the research project, I decided to together with the management at IU form a temporary project group to carry out the last action research intervention. The group consisted of 5 education consultants who were appointed by their manager, and agreed to participate.

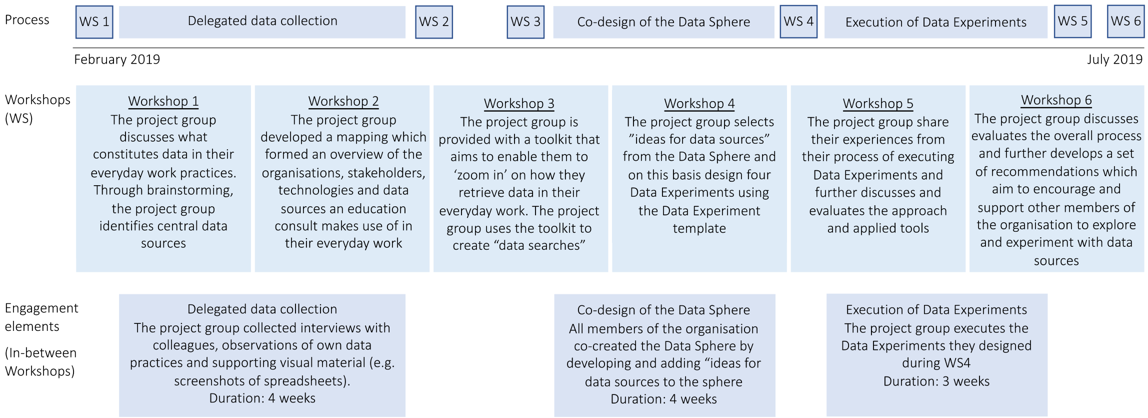
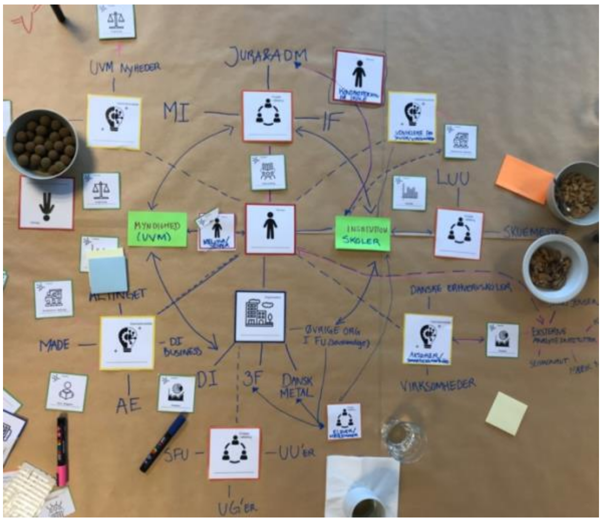
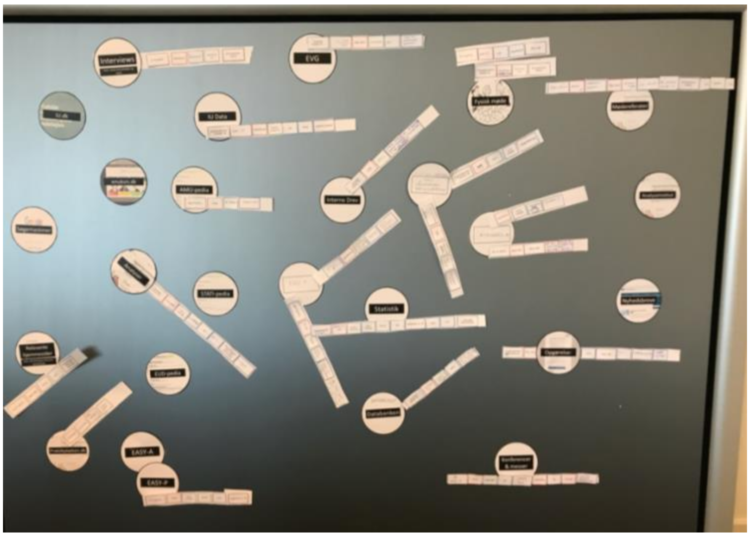


Figure 6. Process model of the third intervention (Publication 5)

The intervention was designed as an end-to-end process for examining how the structure of research activities might be a useful tool. The process revolved primarily around a series of 6 workshops, each of which included different objectives and activities (figure 6). The series of workshops began by exploring what constitutes data for an education consultant at IU (WS 1). This was followed by an exploratory phase, when the members of the project group were asked to participate in a data collection process (delegated data collection). They were asked to conduct interviews with colleagues, make observations, and use visual material. This was intended to make the group actively contribute to our 'co-constructed knowledge production' (Hayes 2011). Working from these inquiries and insights, the group developed a map of 'the education consultants' landscape', which included a representation of the current use of technologies and data in their everyday work (WS 2). This was followed by a workshop in which, in contrast to the previous activity, the group would zoom in and create 'data searches' as a way to explore how they look for data (WS 3). Together, these workshops explored ways of 'zooming out and zooming in', an approach (Nicolini 2013) to collectively understand existing data practices. This combination of 'macro- and micro-levels' made demonstrated that the education consultants worked primarily with data sources that were 'ready at hand', and made only limited use of data in new and innovative ways.



Images 3. and 4. Mapping exercise during the second workshop in the third action research intervention.



Images 5. and 6. During the third workshop of the third action research intervention, the project group carried out data searches by putting together 'search statements' they would use in the context of a specific data source.

The exploration of combining 'macro- and micro-levels' was followed by the implementation and co-design of IU's Data Sphere, which is a tool that aimed to encourage all members of the organization to consider and generate ideas for new data sources that the project group could explore and experiment with. The project group processed the input from the Data Sphere, and based on this, designed so-called Data Experiments (WS 4). During the final stage of the process, the project group evaluated their experiences and presented their learning in a guide to support them and their colleagues in working more innovatively with data in a participatory manner (WS 5 and 6) (for a more detailed description of the Data Sphere, the Data Experiments, and their effects, see Publication 5). During this final intervention, evaluation took place on an ongoing basis, and was documented as part of each workshop. Observed effects of the third intervention included increased acknowledgement and appreciation of co-design, with regard to working creatively and innovatively with data in a highly-connected organization such as IU. Another observed effect of this process was that its inherent focus on data helped the project group and the organization to consider data something that may be designed.

These three action research interventions explored various aspects of the general research question, specifically, how organizations may design and innovate their data-based services. In the next chapter, I elaborate on how this manifested in the publications. However, before turning to the written research output, I describe how I ensured scientific rigour throughout this research process.

4.3 Reliability of the empirical research

To ensure scientific rigour, action research puts the notion of *trustworthiness* at the centre of the research process, to establish a reliable alternative to generalizability (Hayes 2011; Robson 2002). Stringer (2013) highlights four concepts that support the trustworthiness of scientific inquiry: credibility, transferability, dependability, and confirmability.

In this project, action research was used to ensure credibility with respect to my long-term involvement with the research setting, which allowed me to collect data in situ. My situatedness allowed for informal accounts and the development of an extensive understanding of particular practices and relations among the actors and in the broader context. My close engagement also enabled me to incorporate project descriptions, reports, presentations, and design activities by using ‘the words of the participants themselves’ (Stringer 2007, 99). Moreover, my action research approach allowed me to include multiple perspectives by including research interventions with different purposes, activities, and various IU members and the external stakeholders involved. Furthermore, this project facilitated various ways to triangulate the results: the use of more than one method of data collection helped to support data triangulation (Robson 2002). I also made use of member checking and debriefing to ensure the rigour of the empirical research (Hayes 2011; Robson 2002). Member checking often took place during workshops where the participants were asked to voice their concerns and comment on the activities in the situation, and afterwards (e.g. in a following workshop or a written workshop summary), when the participants were asked to verify the data collected about them.
Observer triangulation (Stringer 2013) was carried out in this project by using more than one observer during workshops. This took place with help of my main supervisor, Yvonne Dittrich, who observed several of the workshops, and my colleague at IU, Stine Moeslund

Sivertsen, who not only took part in the workshops, but also took field notes for the duration of the research project. These additional records allowed us to compare findings, in order to confirm our observations. Finally, during my stay at the Oxford Internet Institute I participated in the ‘Data Science for Local Government’ (DSLG) study. The project aimed to better understand the spread and impact of data science in the context of local government in the United Kingdom. The DSLG study ran from May 2017 to August 2018, and was led by Professor Jonathan Bright and his research group. I joined the research group in April 2018. At that point, the study had been designed, and therefore I participated primarily in the late stage of the data collection, analysis, and writing. The study was based on a mixed-methods approach, and included an extensive documentary review, a nationwide survey of local authorities in the United Kingdom, and 34 in-depth interviews with practitioners working with data science initiatives in the public sector. Although the study applied various methods, to a certain extent they were used to examine the same phenomenon, namely, data practices at organizations in the public domain (in the United Kingdom). The project’s key findings included the identification of key challenges to developing new data practices to establish data science projects with local authorities (Publication 7). Specifically, the challenges met in the public sector included a lack of skills and knowledge that would enable people to work creatively with data. Thus, some of the insights from the DSLG project help to triangulate my findings at IU.

I have established transferability by documenting and describing the ways in which the findings from this research project emerged, and therefore may be applied and evaluated in other contexts (Stringer 2007). Moreover, the concept of a co-design perspective on data that I develop in this dissertation may be transferable, and may be used as a theoretical lens for other projects, to generate insights. Another example is the description of the tools I developed throughout this process. These may also be used and evaluated in various contexts.

Stringer (2007) emphasizes that dependability and confirmability in action research are ensured by an audit trail that explains the ways in which data is collected and analysed. I have documented the research activities throughout the research process. For example, I wrote fieldnotes in a systematic way that clearly divided my observations and reflections on the observations by using cloud-based software called Evernote, to secure the empirical data (see figure 7).

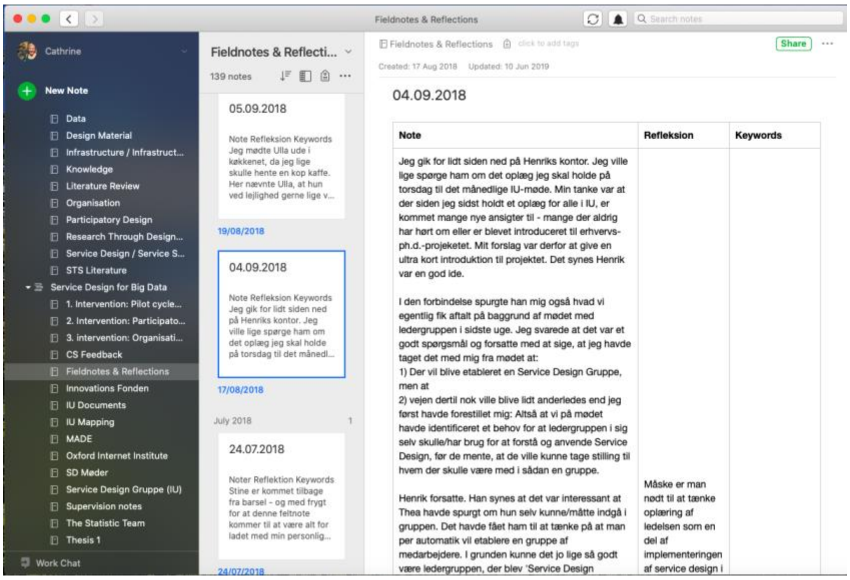


Figure 7. Field notes. This screenshot exemplifies how I structured my notes by date, and further divided the observations by 'note', 'reflection' and 'keywords'.

I also audio- and video-recorded most of the meetings and workshops in which I participated and facilitated. Taken together, these steps support the reliability of the empirical research, which is the foundation of this dissertation.

Chapter 5: Summaries of the publications

This chapter introduces the 8 original research publication that are included in the dissertation. Generally, this research output addresses the question of how organizations can design and innovate data-based services. The eight publications are included in their entirety in part 2 of this dissertation. The aim of this chapter is to present an overview of the content and main findings of each publication. I begin by summarizing each paper and its key contributions to this dissertation. This is followed by a model and description, which outline how each publication addresses one or more of the three research sub-questions: What are common data practices at organizations? How can organizations develop concrete, data-based services? How can organizations explore data sources and experiment with their usefulness?

Publications 1 to 6 build on empirical data from the action research project with IU. These papers form the core of this dissertation. Publications 7 and 8 are based on the Data Science for Local Government study, in which I participated during my research-abroad stay at the Oxford Internet Institute (also see figure 4 for an overview of the connection between the papers, action research interventions, and research activities). Table 2, below, illustrates the how the various publications address the research question through the sub-questions.

Research question	**How can organizations design and innovate their data-based services?**		
Sub-questions Publications	What are common data practices at organizations?	How can organizations design concrete data-based services?	How can organizations explore data sources and experiment with their usefulness?
1	X		
2	X	X	
3		X	
4		X	X
5			X
6	X		
7	X		
8	X		

Table 2 Relations between research questions and the publications.

Publication 1

Seidelin, C., Dittrich, Y., Grönvall, E. (2017) *Identification of data representation needs in Service Design.* Selected papers at IRIS, Issue Nr. 8. (2017) 10.

The first paper of this dissertation reports on the preliminary research activities. This paper presents the motivation for considering and integrating data as a central component in service design. This study examines the tools used to work with data and analytics at IU. Moreover, the study identifies the organization's (at that time) applied ways of representing data and data analytics. This paper discusses whether, and if so, how, these representations of data support data-driven innovation, which we also compare with current service design representations. This comparison suggests that service design representations lack ways to include data as a central component of the design of data-based services. This study proposes to make use go the notion of expansiveness as a way to evaluate future data representations for the design of data-based services.

This publication contributes to this dissertation by analysing some of the common data practices and tools used at the field site. Finally, it also defines the need to represent data to a greater extent designing data-based services.

Publication 2

Seidelin, C., Dittrich, Y., Grönvall, E. (2018) *Data work in a knowledge-broker organisation: how cross-organisational data maintenance shapes human data interactions*. Proceedings of the 32nd International BCS Human Computer Interaction Conference 32. http://dx.doi.org/10.14236/ewic/HCI2018.14

This paper explores how data is used across organizational boundaries for multiple stakeholders in the network to provide essential services to other actors in the network. The study is based on data from the preliminary activities and the first action research intervention. Specifically, the study focuses on the redesign of an old database and its related IT system, which is centralized, and maintained by IU. However, IU is not the owner of this data; instead, IU acts as a knowledge-broker that provides a number of stakeholders with relevant data and information in this database. In this way, the network has established IU as a broker that ensures data accountability. This paper examines the notion of Human-Data Interaction (HDI) as a lens through which to consider data as a central part of Human-Computer Interaction, as proposed by previous research (Crabtree and

Mortier 2015; Haddadi et al. 2013). Based on our analysis, we suggest extending the notion of HDI to include the increased complexity that exists when multiple stakeholders interact with the same data. This contrasts with previous work, which focused primarily on the interaction between the individual user and their personal data. Most importantly, this publication contributes to the dissertation by showing that in this cross-organizational context, data is collaboratively evolved, maintained, and used.

Publication 3

Seidelin, C. Dittrich, Y., Grönvall, E. - *Foregrounding Data in Co-design: An Exploration of Data as a Design Object*. [Resubmitted to International Journal of Human Computer Studies]

The third publication explores how data may be foregrounded in co-design in ways that enable domain experts to contribute their expertise in the design of data-based services and the services' underlying data structures. The study revolves around three collaborative workshops, which took place during the first action research intervention as part of the redesign process of the old database and related IT system. During these workshops, I introduced specific data notation, and employed service design notation as an experimental way to make data an explicit element of the process of co-designing a data-based service. We use Feinberg's (2017) design perspective on data (Feinberg 2017) as a lens to guide our analysis. This publication contributes to the dissertation by showing that using carefully designed data notation may enable domain experts who are not IT professionals to engage in the design of the data that underpins their data-based service provision.

Publication 4

Seidelin, C., Sivertsen, S., Dittrich, Y. (2020) *Designing an organisation's design culture: How appropriation of service design tools and methods cultivates sustainable design capabilities in SMEs*. Proceedings of the 6th ServDes Conference. Melbourne. 6th-9th of July 2020.

The fourth publication aims to help understand how organizations can overcome the barriers that prevent them from building internal design capabilities, and to develop a sustainable design culture (Julier 2006). This publication elaborates on the second intervention, which was intended to build service design capabilities at IU. The paper briefly presents how I addressed this through seven so-called 'service design micro-cases'. The analysis of these learning activities prompted autonomous service design initiatives at IU, which eventually fostered a sustainable service design culture at the organization. The paper emphasizes that in this case, the adaption of service design tools and methods was essential for the successful development of a service design culture. This publication contributes to the knowledge of how an organization with very limited design capacity can develop and establish a culture of design and innovation, to foster data-driven innovation.

Publication 5

Seidelin, C., Dittrich, Y., Grönvall, E. – *Co-creating Data Experiments: Exploration and Experimentation with Data Sources*. [Submitted to Designing Interactive Systems (DIS) Conference 2020]

The fifth publication examines how domain experts may be supported in their exploration of self-selected data sources, and experimentation with their usefulness. The paper is based on empirical data from the third action research intervention. The paper presents two tools, The Data Sphere and The Data Experiment Template, which I designed and implemented to prompt data-driven innovation at IU. The findings indicate that the proposed tools' tangible character and concreteness support domain experts' understanding of how to identify, explore, and experiment with various data sources. This paper contributes to the dissertation by emphasizing co-design as a useful approach for fostering data-driven innovation in an organizational context.

Publication 6

Seidelin, C., Lee, CP., Dittrich, Y. – *Exploring the role data play in a public sector arena*. [Resubmitted to the European Conference on Computer-Supported Cooperative Work 2020]

The sixth publication examines how data work takes place and the role data that play in a large and highly connected network of stakeholders in the public sector. This paper cuts across the three action research interventions, and builds on empirical data gathered throughout the project. This paper presents a diagram of the complex setting in which IU provides a number of crucial, data-based services to the larger network of stakeholders in Denmark's public sector. This diagram is a key contribution of this dissertation, because it depicts how many different stakeholders collaborate in various ways, in this case to make vocational and continuing education for the industrial sector work. By further asking what role data plays in this space, the paper finds that data work in this context rarely occurs in one organization, but that data produced, maintained and used through collaborative efforts of multiple stakeholders in the network. Thus, this publication contributes to the dissertation by showing the complexity in which data-driven innovation needs to take place, and underlines the necessity of considering the cooperative aspect of data and data practices.

Publication 7

Bright, J., Ganesh, B., Seidelin, C., Vogl, T (2019) *Data Science for Local Government*. Oxford Internet Institute. Oxford University. Available at: https://smartcities.oii.ox.ac.uk/wp-content/uploads/sites/64/2019/04/Data-Science-for-Local-Government.pdf

The seventh publication is an industry report, which describes the Data Science for Local Government study. The aim of the report is to examine the existing data practices in local government in the United Kingdom, especially with regard to how the growth of 'data science' affects daily operations in this context. The study is based on a documentary review, a nationwide survey of local authorities, and 34 in-depth interviews with practitioners. The report provides a guide to the various types of data science implemented in the United Kingdom, identifies related opportunities and challenges, and helps to understand how some of these challenges are being addressed. The report contributes to this dissertation by triangulating observations of common data practices and implications for new forms of data work at organizations in the public sector.

Publication 8

Vogl, T., Seidelin, C., Bright, J. – *Smart Technology and the Emergence of Algorithmic Bureaucracy: Artificial Intelligence in UK Local Authorities*. [Resubmitted to Public Administration Review – Special Issue on Transformation in Government]

The final publication included in this dissertation examines how local authorities in the United Kingdom have begun to use 'smart technologies' to support service delivery. Based on the empirical data from the Data Science for Local Government study, the paper describes key implications of smart technologies in this context, and emphasizes that public administrators and technology overlap in their delivery of public services. We conceptualize this overlap as 'algorithmic bureaucracy', to describe the increasing number of interactions between public servants and computational algorithms that are part of the everyday work environment. We propose a framework that explores how smart technologies transform the socio-technical relationships between people working at local authorities and their tools, and how this work is organized. This paper contributes to this dissertation by examining the implications of changing data practices.

To summarize, the publications show that data practices at organizations are characterized by their being essential to people's ability to do their work. However, data practices are often hidden in other work practices, and thus data work becomes 'invisible work', unless it is labelled, for instance, as 'doing data science' or 'creating statistics' (Wolf 2016). The research also shows that data work in one place and data work taking place at other sites are often interdependent (Publications 1, 2, 6, 7, and 8). With regard to how organizations can design concrete, data-based services, the research underlines that it is necessary to develop and establish design capabilities at the organization, as a foundation for the way the organization designs and innovates for their data-based services. Moreover, the research shows how carefully designed data notation can support domain experts in designing with data in ways that enable them to engage in the design decisions that shape the data and the data structures that underpin data-based services. The research also finds that co-design is a useful approach for designing concrete, data-based services, because it supports multiple stakeholders when they consider cross-organizational data practices (Publications 2, 3 and 4). Finally, the research presents tools developed to foster exploration and experimentation of data sources. This publication shows that the adaption of tools and methods is central to establishing design capabilities that prompt exploration and experimentation with data (Publications 4 and 5).

As mentioned in the introductory chapter, examining the concrete research questions developed additional questions that suggested broader discussions of how organizations can design with data, and how they may do so collaboratively. Therefore, I advise the reader to read the publications at this point. Based on the preceding chapters in this cover and the publications, the following chapters address this broader discussion. Chapter 6 begins to elaborate on the ways data has been foregrounded throughout this research project, and discusses how it may benefit researchers and practitioners in future investigations to consider two prominent dimensions of these explorations. Chapter 7 discusses some of the necessary arrangements that need to be present for organizations to be able to design and innovate data-based services. Finally, chapter 8 develops the proposal to establish a co-design perspective on data, to support the design and innovation of data-based services.

Chapter 6: Making data an explicit element in the co-design of data-based services

A key contribution of this dissertation is the exploration and development of ways in which data and its schemata may be considered integrated parts of co-design practice. The chapter on methodology (chapter 4) and the publications describe the proposed data notation, the ways in which I have explored data representation in existing service design representations, and developed new data representations for exploring and experimenting with data sources. This chapter elaborates on these design explorations, to further discuss how to make data an explicit element of the co-design of data-based services. Specifically, this chapter discusses two dimensions of 'data modes', which are called 'concrete and abstract data design' and 'routine and emergent data needs'. These dimensions were identified as important aspects that greatly influenced both the intended inclusion of data as an explicit element of co-design, and how the data notation was eventually applied in the co-design situation. Together, these dimensions constitute a map, which I call 'the data mode map'. The map makes up an instrument that supports reflection on the process of designing data notation for co-design. This chapter ends with a discussion about the potential use and implications of this map.

6.1 Concrete and abstract data design

This research shows that the level of abstraction with which data is represented is an important dimension in terms of how domain experts relate to, and make use of data notation. To explain the varying levels of abstraction of the data representations I have used in this research project, I refer to a concrete–abstract data design continuum. Here, 'concrete' or 'abstract' refer to the extent to which data is concretely structured in an IT system. Thus, I refer to concrete data design, when describing design situations where the data source(s) and data entities are formalized, and to some extent known. In contrast, I refer to abstract data design when describing design situations where data or data sources are (as yet) unknown or undetermined, and thus, where the representation of data is more general. For example, the data icons I developed and implemented during the first action research intervention, represented data rather concretely (Publications 2 and 3). This data notation resembled concrete data entities from an existing database, and aimed to make the domain experts aware of the data that underpins the practices and collaborative work they performed in relation to a specific field of work. Another example is from the third action research intervention, where I represented data more abstractly, with another the set of icons. In this design situation, the icons were used by the project group to produce a map of the education consultant's collaboration with other people and organizations, and the use of technologies and data sources that support their (data) work practices (see images 3 and 4). In this case, however, the icons themselves did not illustrate specific data entities in a specific IT system, but acted as an ‘open’ category, where the data(sources) were undetermined prior to the mapping exercise. The expansiveness of the icons prompted the domain experts to discuss and negotiate what constitutes a data source in this context. This emphasizes that although their representations may be similar, the design of the representation and its intended (and afterwards actual) design and use may differ significantly.

This research shows that the concrete–abstract data design continuum is an important dimension to consider when designing data notation for co-design. Building on the examples above, it seems that abstract representations of data work better when the design task at hand is rather open-ended and undetermined, as was the case when the project group worked with data icons for the mapping. In contrast, during the first intervention, where the project was much more predefined, the use of more concrete data notation seemed to help the multiple stakeholders to relate the data to concrete data practices.

6.2 Routine and emergent data needs

The second dimension that became apparent as important for the work of developing and implementing data representations in co-design was routine and emergent data needs. The field work showed that at IU, there exist both routine and emergent data needs. Specifically, our analysis of the roles data plays emphasized how the established structures generate rather well-known and predictable data needs on an ongoing basis. These routine data need included, for example, statistics to support committee work. However, the analysis also showed that new forms of coordinated actions involved new data work, which further created emerging data needs (Publication 6). An example of emerging data needs was revealed by the education consultant, who changed the way stakeholders in the network cooperated around Elective Specialization Courses, by including a new data source. However, eventually, this emergent data need turned into a more routine data need, as many education consultants in the organization and in the external committees began to use – and thus establish practices based on – this data source. Another instance where routine and emergent data needs became visible was during the first intervention, where domain experts from IU and external organizations (representing key stakeholders) worked with the data icons. This design activity took as its starting point the consideration of existing data entities in the IT system, and through discussions the participants were able to reach a mutual understanding concerning routine data needs related to their current work practices. However, as the discussions developed over the course of the three workshops, we observed how emerging data needs also became visible (Publication 3). For example, workshop participants stated that the redesigned IT system should also be able to document the many meeting minutes that Local Education Committees are required to produce and share after each meeting. In this case, the emerging data need was expressed as a vision of making better use of this data source. Based on this, I refer to a 'routine data' need when data is identified as essential to supporting a data-based service that is provided regularly. In contrast, an 'emergent data need' describes a situation where data that has not been previously used as part of a data-based service is introduced and identified as a useful service innovation or service provision. This research shows that it is important to consider the routine–emergent data needs continuum when designing data representations for co-design, because it supports reflection on the 'state' and 'familiarity' of the data: is this data critical for the organization to provide a specific data-based service? Is the data 'well-known' and already implemented in the organization somehow, or does the data involve an increased level of complexity and/or uncertainty? In other words, routine and emergent data needs require different toolkits.

Current data science practices propose more sophisticated support for routine work, for example, how to use artificial intelligence to assist workers in local government to provide services to citizens (Publications 7 and 8). Through such data science practices, complex algorithms are embedded in established information systems; however, these algorithms mainly support routine data needs. An example is the increasing number of autonomous agents ('chatbots') in local government services, which aim to decrease the pressure of face-to-face and telephone services by allowing citizens to conduct transactions online (Publications 7 and 8). Autonomous agents are often established and further developed based on routine data needs, for instance, frequently asked questions. Thus, to design with data in ways that support routine data needs requires organizations to be able to identify more specific and recurrent data needs. However, I argue that it is also necessary to consider how emergent data needs may be supported so organizations may take these evolving needs into account in the process of developing data-based services. Taking care of emergent data needs requires a different toolkit, because it implies that domain experts should be enabled to flexibly explore and analyse data and data sources. In co-design practice, this calls for tools and representations of data that promote exploration and experimentation with data, to further recognize emergent data needs.

6.3 A data mode map for co-design

I propose combining the two ‘data dimensions’ presented above (concrete and abstract data design, and routine and emergent data needs) to create a map that illustrates various ‘data modes’ in design. The map was inspired by Manzini (2015), who developed a ‘design mode map’ that illustrates the various ways design capabilities are enacted. By considering various modes of ‘designing’ and ‘being designers’, he suggests that the design mode map may support an understanding of who the design experts are, and what they do in various situations (Manzini 2015). The data mode map presented below (figure 8) has a different objective. It suggests an outline for how we may contemplate including data as an explicit element of the co-design of data-based services. I have populated the data mode map below with the main data representations included in this work, to illustrate the use of this tool. The data representations are mapped according to both their *intended* inclusion of data as an explicit element of co-design, and the use of their representation, meaning, how they were applied in a co-design situation. For example, the Data Sphere was intended to prompt users to think about new data sources that might be interesting in the context of service innovation in the organization. Moreover, The Data Sphere was used in an open-ended manner at IU, where all its members were invited to contribute their ‘data ideas’. Based on this, I have placed the Data Sphere in the upper right corner of the data mode map. I have done so to illustrate the tool’s the abstract configuration of data and simultaneously its aim to identify emergent data needs. It is important to note that this placement is based on a specific instance in a specific context. That is to say, if this tool for representing data was applied to a new context, it might generate a different effect, and thereby position the Data Sphere differently on the data mode map. Thus, the data representations do not prescribe how the map is used. I suggest that the data mode map is a tool that can support researchers’ or practitioners’ reflection on our (more or less articulated) intention to include data in co-design processes, and to reflect on the later use of a proposed data representation. In this way, the map is a tool that attempt to address the difficulties of reconciling an understanding of data as interpretive, flexible, and situational when designing (Feinberg 2017).

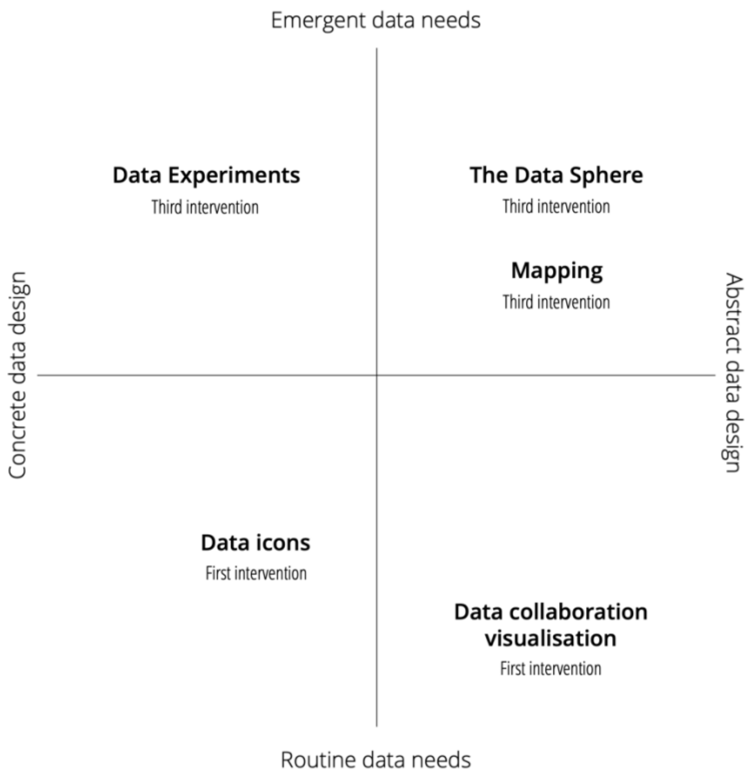


Figure 7. The Data Mode Map

Maps such as the Data Mode Map have certain limitations. From a critical point of view, this map might promote a more reflective inclusion of data in co-design processes; nevertheless, it is still a way of framing data. Thus, this map might support certain 'data constructions' rather than others. It is important to be aware of the ways in which it might generate certain constructions of the world that feed into the design of data-based services. Therefore, it is vital to critically consider – on an ongoing basis – whether the proposed dimensions are relevant, or whether dimensions are missing.

Another limitation is the map's embedded assumption that organizations know *how* to go about developing data representations and apply these in practice. To be able to benefit from the idea of these different data modes implies that an organization has certain established design capabilities, in order to enact proposals to work innovatively with data. The next chapter elaborates on this challenge by discussing how an organization with very limited design capacity such as IU can establish a culture of design and innovation.

Chapter 7: Establishing a culture of design and innovation

This chapter emphasizes and discusses two aspects that emerged as essential for creating a foundation that enables organizations in the public sector to innovate and design their data-based services. This chapter begins by discussing the first aspect, that is, the need to cultivate user-driven innovation to foster data-driven innovation in an organization. This is followed by a discussion about the second aspect, that is, the need to expand the organization's 'innovation toolkit' to support innovation with data.

7.1 Democratizing data-driven innovation in the organization

Organizations are constantly challenged by the need to innovate products, processes, and services to stay competitive (Kline and Rosenberg 2009). Using data and analytics as a means to innovate is highly complex work (OECD 2015a). This complexity is reflected in studies that show that few organizations are successful in their attempts at data-driven innovation, despite this being a commonly stated objective during the past decade (Bean and Davenport 2019). Drawing on Dittrich et al. (2017), this suggests a need for organizations to both sustain and develop capabilities for data-driven innovation. Previous research emphasized user-driven innovation as a way for organizations to promote innovation (Von Hippel 2005). In this section, I argue that user-driven innovation may be a stepping stone to organizations' ability to foster data-driven innovation.

The notion of innovation is popular, and is used by both industry and research (often) to describe a new product, process, or service, or it may be viewed as the application of improved solutions that meet (un)articulated needs. Many contemporary definitions are based on Schumpeter's (1947) conceptualization of innovations as 'new combinations' of production factors, including the production of new products or the introduction of new processes (Fagerberg and Verspagen 2009). However, innovation is a complex concept that is difficult to measure, and this has resulted in multiple and varied perspectives on what constitutes innovation. Fagerberg and Verspagen (2009) show that some communities in the field of innovation studies continue to build on Schumpeter's definition, whereas others focus on management, geographic, or economic aspects of innovation. This shows that research on innovation is becoming widespread and comprehensive. To delimit the focus, this dissertation concentrates on a specific strand of research, which is known as user-driven innovation.

Von Hippel (2005) argues there is an ongoing trend that enhances the democratization of innovation. This trend is based on opportunities (e.g. the Internet) that enable users to modify products, which thereby enables them to take the first step towards innovations. This differs from more traditional perspectives that regard innovation as something due to designers, developers, or research and development departments (Kline and Rosenberg 2009). To some extent, these perspectives also acknowledge the role of users and their needs, however, these are channelled into design, where specialists develop the solutions (Von Hippel 2005). Von Hippel's (2005) alternative perspective uses the concept of 'democratizing innovation' to discuss the development of user-driven innovation processes. He argues these user-driven innovation processes not only have great potential to enable users to develop what they need, they also make 'the learning associated with creativity and membership in creative communities […] more widely available as innovation is democratized' (Von Hippel 2005, 123). Von Hippel's (2005) perspective on innovation is relevant in the context of this research for two principal reasons. First, it emphasizes that innovation can occur everywhere, and thus is not restricted to certain people. Second, this perspective supports users being creative and taking the initiative to make changes that improve the existing situation, which applies to this project's general action research approach and use of co-design. I discuss democratized innovation from the perspective of users, who in this case are IU members. Moreover, in this context, democratizing innovation focuses on the need for the organization to take advantage of its members' capabilities, and thus allows the domain experts to be (more) creative with data.

As stated in chapter 3, this dissertation takes a practice perspective, and argues that data is embedded in practice. This further implies that *data*-driven innovation needs to be rooted in practice. From a practice perspective, user-driven innovation becomes a necessary aspect of fostering data-driven innovation in an organizational context. This work has aimed to empower the organization in ways that would enable IU members to make use of their insights, and thus to innovate and design with data, even after the research project ended. In line with Dittrich et al. (2017, 168), I argue that

> 'organisations need to make use of and cultivate the capabilities of their members, the communities that they are part of, and the networks they have access to – inside and outside the organisation. This is a process that involves both user-driven innovation and organisational learning of how to make use of innovations to add organisational value'.

This research has focused on developing (co-)design capabilities at IU, to foster user-driven innovation and organizational learning that could further lead to data-driven innovation. Publication 4 presents how 'Service Design Micro Cases' were introduced as an approach to teach IU members about design practices based on their own work practices. This brought about user-driven innovation in the organization. One example is that of the education consultants that struggled to understand how they could encourage a greater focus on innovation work in various committees. In this case, the education consultants learned about design processes and a particular design process model, which we subsequently adapted in order for the tool to work in their specific context by taking existing ‘committee work practices’ into account. Eventually, the organization members developed their approach, which emerged and became the new established way to do committee work one of the two largest departments at IU (see IF in the organizational diagram, figure 1). Therefore, I argue that advancing user-driven innovation is an essential stepping stone to fostering data-driven innovation in an organizational context. To further establish a culture of democratized, data-driven innovation processes in an organizational context entails to enable members of the organization to innovate with data. I elaborate on this point in the next section.

7.2 Expanding the organization’s ‘innovation toolkit’ through co-design

This research project has explored co-design as an approach to expanding an organization’s innovation toolkit in a way that promotes democratized, data-driven innovation. In this section, I discuss why co-design is a useful approach to advancing design and innovation with and through data in an organizational context.

During the long-term action research engagement with IU, it became clear that in this context, data is collaboratively evolved, maintained, and used through cross-organizational practices. For example, as discussed in Publication 2, data about members of local education committees is created, collected, maintained, used and updated through a number of practices that occur in various contexts. This means that if – or rather when – these data-related practices are changed in one place, it influences practices in other places, and by extension, the data. In this case, it became apparent as ‘data discrepancies’, meaning that the data would be wrong or missing. These discrepancies were often rooted in changed or missing practices. This is consistent with Bossen et al. (2019), who also find that data work is highly interdependent, and has implications for data work taking place in other contexts. Therefore, at organizations such as IU, where data is collaboratively produced and used among a broad network of stakeholders, user-driven innovation alone is not enough to make design and innovation with data meaningful. In this situation, it is necessary to explicitly consider the cooperative aspect of innovating with data, as it will most likely affect the ‘common’ data as well as data work at different sites.

Manzini argues that ‘when confronted with new problems, human beings tend to use their innate creativity and design capacity to invent and realize something new: they innovate’ (2015, 9). Although members of an organization are able to be creative in ways that may lead to innovation, previous research also emphasizes that users generally work with 'local' information, which refers to 'information already in their possession or generated by themselves – both to determine the need for and to develop the solutions for their innovations' (Lüthje, Herstatt, and von Hippel 2005, 962). Thus, when innovating, organization members are likely to use the tools they have ready at hand. This suggests that in order for organizations such as IU to be innovative with data, it is necessary to expand the organization's 'innovation toolkit'. At IU, co-design provided a new set of practices to approach innovation collaboratively. Moreover, to make co-design useful as an approach to promoting data-driven innovation, the organization was also introduced to, and experimented with, data notation and adapted service design representations that aimed to make data an explicit part of the co-design. We (the organization and I) experimented in this way with creating co-design practices that allowed domain experts to include a cooperative perspective on data and data work in the context of design and innovation. For example, the first action research intervention demonstrated how making data an explicit element of co-design enabled the domain experts to consider data as malleable entities that may be designed and innovated. Moreover, the third action research intervention showed how co-design in the form of Data experiments made the interconnectedness of data work visible to the members of the project group, and advanced their appreciation of co-creation (Publication 5).

I do not to claim that innovative work with data did not or cannot occur without co-design. The field work shows that some IU employees were already working innovatively with data to improve their own (and other's) data work before being introduced to co-design. For example, as illustrated in Publication 6, an education consultant at IU innovated with data in a way that changed the data work related to how he and the Sector Skills Councils could evaluate and develop Elective Specialization Courses. The initiative was much appreciated by colleagues and management at IU, who also adopted these new data practices. However, the changes were not presented as data-driven innovation at the organization. Instead, the initiative was framed as a way to work more effectively. Dittrich et al. (2017) have shown that sometimes, user innovations are not recognized. By including the education consultant's initiative in the third intervention, the education consultant reframed the initiative, which in this way became visible as data-driven innovation at IU. The structured co-design approach made these new practices visible as data work in the organization. This suggests that co-design is also a useful approach for framing and making democratic, data-driven innovation visible in the organization.

Overall, this research indicates that co-design is a useful set of practices for supporting organization members in including a cooperative perspective on data and data work, and for making data-driven innovation visible in the organization. However, it is important to note that in this case, the benefits of co-design as an approach to democratized data-driven innovation could not emerge without significant adaption of tools and methods (Publication 4). Julier et al. (2019) argue that developing a culture of design often involves a set of practices that establishes shared understandings and values, which may include ‘the identification and establishment of specific infrastructural support, common linguistic tropes, key personalities and support systems’ (Julier et al. 2019, 227–28). These processes of negotiation, presented as part of establishing a shared understanding of what constitutes data-driven innovation in the context of IU and its network of stakeholders, are still emerging. However, the increased focus on data and collaborative design as means of innovation has become a guiding principle that is stated in the latest IU 2020–2025 strategy document (IU 2019b).

Chapter 8: Towards a co-design perspective on data

The preceding chapters and publications have prepared the ground for this final chapter. They have shown the relevance of making data an explicit element of co-designing data-based services, and discussed central organizational arrangements and practices that are necessary to collaboratively foster design with data. This chapter builds on these contributions, and carves out the main theoretical contribution, which is the proposal to develop a co-design perspective on data. The chapter begins by emphasizing the societal relevance of adopting a cooperative approach to design with data and data structures. This is followed by briefly restating the key points of Feinberg's (2017) design perspective on data. The chapter ends with a discussion of the importance of extending this perspective, and shows how this dissertation has done so in two respects.

Through its examination of data practices, this research project has shown how IU is very connected to, and dependent on external stakeholders to provide and develop the organization's data-based services. This connectivity becomes particularly apparent in figure 2, which shows how collaboration in this public sector takes place among various stakeholders within and across organizational boundaries. This resonates with what has been stated by Manzini (2015), who argues that this high and increasing level of connectivity reduces the solidity of organizations. Here, the understanding of connectivity covers both connectivity in the sense of human interactions, and it also comprises the ever-growing number of digital technologies that depend on advanced connectivity and differentiated networking. The growing level of connectivity makes organizations more and more interdependent with external organizations and stakeholders through shared systems, practices, and collaboration (Manzini 2015). The extreme level of connectivity influences how organizations design and innovate their data-based services. This dissertation has shown two central aspects of how this influence manifests in practice. On the one hand, IU is highly dependent on external stakeholders and organizations to be able to maintain and develop a number of essential services in the network; on the other hand, this extreme connectedness with other stakeholders also presents an obstacle for IU when the organization wishes to try new ideas or initiatives, because these changes in practice immediately affect others in the network (see Publications 2, 5 and 6 for examples). Therefore, one might argue that the high level of connectivity is both a barrier to and a cohesion for the organization's ability to provide and develop their data-based services. Thus, the question is, how can we take this inherent condition of many organizations into account, for example, when designing and innovating data-based services?

Feinberg (2017) introduced a design perspective on data that provides a lens through which to consider data as something that is and may be designed *through* practices. (For an in-depth account of Feinberg's work, I refer the reader to Publication 3, section 2). This is an important contribution when questioning how we can design with data, because it enables us to consider the 'design in use' (Henderson and Kyng 1991) of data and data structures. Previous research described the concept of 'design in use' as the 'practices of interpretation, appropriation, assembly, tailoring and further development of computer support in what is normally regarded as deployment or use' (Dittrich, Eriksen, and Hansson 2002, 124). By highlighting the creativity that emerges through the use of technological artefacts over time, the notion of 'design in use' helps to make visible the ongoing collaboration on design practices in everyday use (Dittrich, Eriksen, and Hansson 2002). In her work, Feinberg (2017) focuses specifically on data, and examines how users are able to adapt a data infrastructure in concrete use situations. For example, by generating and collecting data by using a digital service, such as online dating, the user is 'manipulating' the data infrastructure. Although this data generation or collection may seem banal, it does involve creative decision-making because this data input helps to create the data infrastructure, and thus, the service, on an ongoing basis. Thus, Feinberg emphasizes that a design perspective on data supports 'the empirical realities of practice and enables innovative reconceptualizations of data creation and use' (2017, 2957). Furthermore, she suggests,

> 'we might purposefully design data infrastructure to function more directly as design material – to support a range of possibilities for data creation, just like we design computer interfaces to function as material for new ways of working and living with devices' (Feinberg 2017, 2959).

Overall, this perspective constitutes a substantial foundation for ways to theorize about how we (can) design with data. However, I argue that when designing with data in a connected world, we need to extend this perspective, for it to be useful in a cross-organizational context.

This dissertation has extended Feinberg's (2017) design perspective on data in two respects. The first aspect concerns whether the design with data happens in a more or less conscious manner. Feinberg (2017) draws on examples from online dating apps, as a way to illustrate how 'people collecting data interpret data infrastructures creatively, flexibly, and situationally', shows that a data infrastructure 'does not determine data; it provides conditions under which people create data'. However, these examples illustrate unwitting design with data, meaning that the users of an online dating app are not consciously designing the use of data. This dissertation explores a different aspect because it aims to promote conscious design of both the use of the data and also the underlying data structures (data schemata). The second way in which this work extends Feinberg's (2017) design perspective on data is by presenting ways to collaboratively practice design with and of data and data structures. This dissertation's empirical work reveals that in an organizational context where multiple stakeholders are co-dependent on data and data infrastructure, it is necessary to take into account the collaborative aspects of data, data work, and design with data. As Manzini emphasizes, 'in a connected world, all designing processes are in fact co-designing processes, unless special barriers are set up to isolate the work of the design team from its context' (2015, 48). From this, I argue that we should not only purposefully design data and data schemata, we should do so collaboratively. This is particularly relevant during the process of designing and innovating data-based services, because the service provision and the underpinning data structures are so interconnected and interdependent that changes to one imply changes to the other.

To summarize, I propose extending the design perspective on data. The extension entails more conscious design with and of data, where data is foregrounded as an explicit element of cooperative design, to account for the high level of connectivity in organizations. Thus, this dissertation presents the first steps towards a *co*-design perspective on data. In this way, it contributes to the emerging debate on how researchers as well as practitioners may articulate the design work, which revolves around the data and data structures that underpin data-based services.

Chapter 9: Conclusions and future work

This dissertation has argued for the development of a co-design perspective on data as an approach to addressing how organizations can design and innovate their data-based services. I have drawn on a three-year action research study to consider how an organization may develop design capabilities that enable organization members to undertake this form of data-driven innovation. Specifically, this dissertation has explored how domain experts at organizations may participate in the design of data and data structures that underpin data-based services.

The research for this dissertation has drawn on a practice perspective to investigate how the foregoing may be done. In this perspective, working with data, providing services, and designing all revolve around the everyday practices in a specific context. This forms the basis for considering the practices related to data work and the development and provision of data-based services in an organizational context. The empirical work has involved a long-term action research study which comprised preliminary research activities and three action research interventions. The action research builds on a variety of methods, from observation and interviews to co-design workshops.

Through my ongoing engagement with, and intervention at the field site, this dissertation makes six main contributions.

First, the dissertation contributes to **a better understanding of data work**. The research examines how data is used and handled today in various organizational contexts. Together, the action research study with IU and the Data Science for Local Government study emphasize, for one thing, the high level of connectivity and interdependence of the many stakeholders and organizations, which is manifested through shared practices, systems, and, indeed, data.

Second, the refined understanding of data work in this context further supported **a better understanding of cross-organizational data work**. This work presents several diagrams that visualize the ongoing collaboration that occurs across organizational boundaries through shared practices, systems, and data (see figure 4 in Publication 2, and figures 2 and 5 in this cover). In particular, this research presents a diagram of the complex setting in which IU exists and provides a number of essential data-based services for the broad network of stakeholders (figure 2). However, this diagram not only acknowledges the existing complexity, it is also a tool for accurately identifying the site(s) of the intervention, and thus, it helps to situate the ‘local accountabilities‘ (Suchman 2002) of data practices in relation to the broad network of stakeholders.

The third main contribution of this dissertation is **the promotion of user-driven innovation as means of enabling data-driven innovation** in an organizational context. The dissertation explores how an organization may develop a culture of design and innovation by using action research to foster data-driven innovation in the organization. This research highlights two essential aspects that address an organization’s need to cultivate user-driven innovation and to expand the organization’s ‘innovation toolkit’. This dissertation explores co-design as set of practices that may develop organization members’ ability to engage with data-driven innovation. The research suggests that co-design is a useful approach to supporting domain expert’s consideration of a cooperative perspective on data and data work and is a framework for making data-driven innovation visible in the organization. However, the research also emphasizes that to apply co-design as a set of practices that foster data-driven innovation requires a significant adaption of tools and methods. The research shows that adopting appropriated tools and methods is most successful when the appropriation is undertaken through ongoing collaboration among the organization members.

This dissertation’s fourth contribution is **a toolkit that comprises several ways of foregrounding data in co-design**. Specifically, the tools developed in this work explore how domain experts who are not IT professionals can take part in the design of the data-based services they use and provide as part of their work practices. This dissertation presents a specific data notation and adapted service design notation that aims to foreground data in co-design processes. Moreover, this dissertation presents tools developed to foster the exploration of, and experimentation with, data sources. Together, these tools and notations facilitate the initial steps towards making data an explicit element of collaborative design practices. The research shows that by foregrounding data in this context, domain experts are able to consider data an object of design. This indicates that domain experts can participate in designing data and data structures when the tools and notation are carefully designed. Thus, foregrounding data enables inclusion of more perspectives in the design discussions and decisions that eventually shape our databases, and by extension, data structures.

This dissertation’s fifth contribution is **the Data Mode Map**, which may be used to support future development and investigations of similar tools and notations. The map is a tool for reflecting on the process of designing data notation for co-design.

Finally, this dissertation proposes and initiates the development of **a co-design perspective on data**. This research extends Feinberg’s (2017) design perspective on data in two respects: it calls for more conscious design with and of data, where data is foregrounded as an explicit element that allows domain experts to take part in this form of design, and it calls for collaborative ways of undertaking this conscious design with and of data, in order to account for cross-organizational data practices.

This dissertation presents several possibilities for future work. One aspect centres on the ongoing collaboration around data in a large network of stakeholders: how do we provide better tools or approaches that enable researchers and practitioners to support these forms of often invisible corporation when undertaking an intervention? Another aspect is the need to develop new notation that can support co-design for data-driven innovation. Future studies based on this research might consider the usefulness of the proposed notation forms and tools in other contexts as a way of identifying characteristics of new notation forms. A third aspect concerns the development of the theoretical proposal of a co-design perspective on data. For instance, how might considering data as 'design things' help us understand the agency that data, and thus data practices, involves, and how might this influence the process of co-design? Thus, although this dissertation addresses some questions, it evidently also raises additional ones. I hope this work has inspired the reader to ask new questions that challenge and advance the idea of making data an explicit element of co-design.

References

Aagaard Nielsen, Kurt, and Lennart Svensson. 2006. *Action and Interactive Research: Beyond Practice and Theory*. Maastricht: Shaker.

Balka, Ellen, Pernille Bjorn, and Ina Wagner. 2008. 'Steps toward a Typology for Health Informatics'. In *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work - CSCW '08*, San Diego, CA, USA: ACM Press, 515. http://portal.acm.org/citation.cfm?doid=1460563.1460645 (December 17, 2019).

Bean, Randy, and Thomas H. Davenport. 2019. 'Companies Are Failing in Their Efforts to Become Data-Driven'. *Harvard Business Review*. https://hbr.org/2019/02/companies-are-failing-in-their-efforts-to-become-data-driven (March 26, 2020).

Beer, David. 2016. 'How Should We Do the History of Big Data?' *Big Data & Society* 3(1): 205395171664613.

Berman, Jules J. 2013. *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information*. Amsterdam: Elsevier, Morgan Kaufmann.

Björgvinsson, Erling, Pelle Ehn, and Per-Anders Hillgren. 2010. 'Participatory Design and "Democratizing Innovation"'. In *Proceedings of the 11th Biennial Participatory Design Conference on - PDC '10*, Sydney, Australia: ACM Press, 41. http://portal.acm.org/citation.cfm?doid=1900441.1900448 (February 25, 2019).

Blomberg, Jeanette. 2009. 'On Participation and Service Innovation'. In *(Re)Searching the Digital Bauhaus*, Human-Computer Interaction Series, eds. Thomas Binder, Johan Löwgren, and Lone Malmborg. Springer-Verlag London Limited, 121–44.

Blomberg, Jeanette, and Chuck Darrah. 2015. 8 *An Anthropology of Services: Toward a Practice Approach to Designing Services*. Morgan & Claypool Publishers.

Blomberg, Jeanette, Jean Giacomi, Andrea Mosher, and Pat Swenton-Wall. 1993. 'Ethnographic Field Methods and Their Relation to Design'. In *Participatory Design - Principles and Practices*, eds. Douglas Schuler and Aki Namioka. , 17.

Blomkvist, Johan, Stefan Holmlid, and Fabian Segelström. 2010. 'Service Design Research: Yesterday, Today and Tomorrow'. In , 306–13.

Bossen, Claus et al. 2019. 'Data Work in Healthcare: An Introduction'. *Health Informatics Journal* 25(3): 465–74.

Bossen, Claus, and Erik Grönvall. 2015. 'Collaboration In-between: The Care Hotel and Designing for Flexible Use'. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*, Vancouver, BC, Canada: ACM Press, 1289–1301. http://dl.acm.org/citation.cfm?doid=2675133.2675243 (February 18, 2019).

Bowker, Geoffrey C. 2014. 'The Theory/Data Thing Commentary'. *International Journal of Communication* 8(5): 5.

Bowker, Geoffrey C, and Susan Leigh Star. 2000. *Sorting Things out: Classification and Its Consequences*. Place of publication not identified: publisher not identified. http://cognet.mit.edu/book/sorting-things-out (April 19, 2020).

Bradbury, Hilary, ed. 2015. *The SAGE Handbook of Action Research: Edited by Hilary Bradbury*. Third edition. Los Angeles: SAGE Publications.

Cowell, D.W. 1980. 'The Marketing of Services'. *Managerial Finance* 5(3): 223–31.

Crabtree, Andy, and Richard Mortier. 2015. 'Human Data Interaction: Historical Lessons from Social Studies and CSCW'. In *ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work, 19-23 September 2015, Oslo, Norway*, eds. Nina Boulus-Rødje et al. Cham: Springer International Publishing, 3–21. http://link.springer.com/10.1007/978-3-319-20499-4_1 (February 18, 2019).

Cukier, Kenneth, and Viktor Mayer-Schönberger. 2013. 'The Rise of Big Data: How It's Changing the Way We Think about the World'. In *The Best Writing on Mathematics 2014*, ed. Mircea Pitici. Princeton: Princeton University Press, 20–32. http://www.degruyter.com/view/books/9781400865307/9781400865307-003/9781400865307-003.xml (March 2, 2020).

Danish Business Authority. 2019. 'The Danish Labour Market Model'. *Danish Business Authority*. https://danishbusinessauthority.dk/danish-labour-market-model (January 20, 2020).

Ddamba, Joshua, and Yvonne Dittrich. 2015. 'Data for Participation and Participation as Data: Supporting Incremental Participatory Decision-Making in Urban Planning'. *Selected Papers of the IRIS, Issue Nr 6 (2015)*. https://aisel.aisnet.org/iris2015/5.

D'Ignazio, Catherine. 2017. 'Creative Data Literacy: Bridging the Gap between the Data-Haves and Data-Have Nots'. *Information Design Journal* 23(1): 6–18.

Dittrich, Yvonne, Johan Bolmsten, and Jeanette Eriksson. 2017. 'End User Development and Infrastructuring – Sustaining Organizational Innovation Capabilities'. In *New Perspectives in End-User Development*, eds. Fabio Paternò and Volker Wulf. Cham: Springer International Publishing, 165–206. http://link.springer.com/10.1007/978-3-319-60291-2_8 (March 24, 2020).

Dittrich, Yvonne, Sara Eriksen, and Christina Hansson. 2002. 'PD in the Wild; Evolving Practices of Design in Use'. In *The Participatory Design Conference*, , 12.

Eubanks, Virginia. 2017. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. First Edition. New York, NY: St. Martin's Press.

European Commission. 2020. *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (March 22, 2020).

Fagerberg, Jan, and Bart Verspagen. 2009. 'Innovation Studies—The Emerging Structure of a New Scientific Field'. *Research Policy* 38(2): 218–33.

Feinberg, Melanie. 2017. 'A Design Perspective on Data'. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, Denver, Colorado, USA: ACM Press, 2952–63. http://dl.acm.org/citation.cfm?doid=3025453.3025837 (February 18, 2019).

Fischer, Joel E. et al. 2016. '"Just Whack It on until It Gets Hot": Working with IoT Data in the Home'. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing*

Systems, CHI '16, San Jose, California, USA: Association for Computing Machinery, 5933–5944. http://doi.org/10.1145/2858036.2858518 (March 19, 2020).

———. 2017. 'Data Work: How Energy Advisors and Clients Make IoT Data Accountable'. *Computer Supported Cooperative Work (CSCW)* 26(4–6): 597–626.

Frayling, Christopher. 1993. 'Research in Art and Design'. *Royal College of Art* 1(1): 1–9.

Futrelle, Joe et al. 2009. 'Semantic Middleware for E-Science Knowledge Spaces'. In *Proceedings of the 7th International Workshop on Middleware for Grids, Clouds and e-Science - MGC '09*, Urbana Champaign, Illinois: ACM Press, 1–6. http://portal.acm.org/citation.cfm?doid=1657120.1657124 (January 18, 2020).

Gandomi, Amir, and Murtaza Haider. 2015. 'Beyond the Hype: Big Data Concepts, Methods, and Analytics'. *International Journal of Information Management* 35(2): 137–44.

Gaventa, John, and Andrea Cornwall. 2008. 'Power and Knowledge'. In *The SAGE Handbook of Action Research*, 1 Oliver's Yard, 55 City Road, London England EC1Y 1SP United Kingdom: SAGE Publications Ltd, 172–89. http://methods.sagepub.com/book/the-sage-handbook-of-action-research/d17.xml (March 29, 2020).

Gitelman, Lisa. 2013. *'Raw Data' Is an Oxymoron*. Cambridge, Massachusetts: The MIT Press. http://ieeexplore.ieee.org/xpl/bkabstractplus.jsp?bkn=6451327 (February 18, 2019).

Gray, Jonathan, Carolin Gerlitz, and Liliana Bounegru. 2018. 'Data Infrastructure Literacy'. *Big Data & Society* 5(2): 205395171878631.

Greenbaum, Joan M., and Morten Kyng, eds. 1991. *Design at Work: Cooperative Design of Computer Systems*. Hillsdale, N.J: L. Erlbaum Associates.

Haddadi, Hamed, Richard Mortier, Derek McAuley, and Jon Crowcroft. 2013. 'Human-Data Interaction'. : 9.

Hansen, Nicolai Brodersen, and Kim Halskov. 2018. 'Teaching Interaction Design by Research through Design'. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction - OzCHI '18*, Melbourne, Australia: ACM Press, 421–31. http://dl.acm.org/citation.cfm?doid=3292147.3292159 (January 19, 2020).

Hayes, Gillian R. 2011. 'The Relationship of Action Research to Human-Computer Interaction'. *ACM Transactions on Computer-Human Interaction* 18(3): 1–20.

———. 2018. 'Design, Action, and Practice'. In *Socio-Informatics: A Practice-Based Perspective on the Design and Use of IT Artifacts*, , 302–18. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1724959 (April 11, 2020).

Henderson, Austin, and Morten Kyng. 1991. 'There Is No Place like Home - Continuing Design in Use'. In *Design at Work: Cooperative Design of Computer Systems*, eds. Joan M. Greenbaum and Morten Kyng. Hillsdale, N.J: L. Erlbaum Associates, 219–40.

Hipp, Christiane, and Hariolf Grupp. 2005. ‘Innovation in the Service Sector: The Demand for Service-Specific Innovation Measurement Concepts and Typologies’. *Research Policy* 34(4): 517–35.

Holten Møller, Naja, and Claus Bossen. 2019. *Perspektiver På Dataarbejde*. https://issuu.com/diku/docs/perspektiver_paa_dataarbejde (January 16, 2020).

Innovation Fund Denmark. 2020. ‘Industrial PhD: All Areas (Private Sector)’. https://innovationsfonden.dk/en/programmes/industrial-researcher/industrial-phd-all-areas-private-sector (April 13, 2020).

IU, Industriens Uddannelser. 2019a. ‘Core tasks (In Danish: “Kerneopgaver”)’. https://iu.dk/om-os/iu-organisation/hvad-arbejder-vi-med/kerneopgaver/ (January 7, 2020).

———. 2019b. *The Strategy for the Education Secretariat for Industry 2020-2025*. https://iu.dk/media/4317/strategi-2020-2025-med-forside.pdf.

Jackson, Steven, and Karen S. Baker. 2004. ‘Ecological Design, Collaborative Care, and Ocean Informatics’. In *Proceedings of the Eighth Participatory Design Conference*, Toronto, Canada: Association for Computing Machinery.

Julier, Guy. 2006. ‘From Visual Culture to Design Culture’. *Design Issues* 22(1): 64–76.

———. 2019. *Design Culture: Objects and Approaches*.

Karasti, Helena, and Karen S. Baker. 2008. ‘Digital Data Practices and the Long Term Ecological Research Program Growing Global’. *International Journal of Digital Curation* 3(2): 42–58.

Kensing, Finn, and Jeanette Blomberg. 1998. ‘Participatory Design: Issues and Concerns’. *Computer Supported Cooperative Work (CSCW)* 7(3–4): 167–85.

Kimbell, Lucy. 2011. ‘Designing for Service as One Way of Designing Services’. *International Journal of Dsign* 5(2). http://www.ijdesign.org/index.php/IJDesign/article/view/938/345 (March 4, 2019).

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Los Angeles, California: SAGE Publications.

Kline, Stephen J., and Nathan Rosenberg. 2009. ‘An Overview of Innovation’. In *Studies on Science and the Innovation Process*, WORLD SCIENTIFIC, 173–203. http://www.worldscientific.com/doi/abs/10.1142/9789814273596_0009 (March 24, 2020).

Koesten, Laura M., Emilia Kacprzak, Jenifer F. A. Tennison, and Elena Simperl. 2017. ‘The Trials and Tribulations of Working with Structured Data: -A Study on Information Seeking Behaviour’. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI ’17*, Denver, Colorado, USA: ACM Press, 1277–89. http://dl.acm.org/citation.cfm?doid=3025453.3025838 (January 18, 2020).

Koskinen, Ilpo Kalevi. 2012. *Design Research through Practice from the Lab, Field, and Showroom*. Waltham, MA: Morgan Kaufmann/Elsevier.

Laney, Doug. 2001. *3D Data Management Controlling Volume Velocity and Veriety*. https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (March 29, 2020).

Leonelli, Sabina. 2016. *Data-Centric Biology: A Philosophical Study*. Chicago ; London: The University of Chicago Press.

Lewin, Kurt. 1946. 'Action Research and Minority Problems'. *Journal of Social Issues* 2(4): 34–46.

López-Quiles, José Miguel, and Manuel Pedro Rodríguez Bolívar. 2018. 'Smart Technologies for Smart Governments: A Review of Technological Tools in Smart Cities'. In *Smart Technologies for Smart Governments*, Public Administration and Information Technology, ed. Manuel Pedro Rodríguez Bolívar. Cham: Springer International Publishing, 1–18. http://link.springer.com/10.1007/978-3-319-58577-2_1 (April 19, 2020).

Lovelock, Christopher H. 1983. 'Classifying Services to Gain Strategic Marketing Insights'. *Journal of Marketing* 47(3): 9–20.

Lüthje, Christian, Cornelius Herstatt, and Eric von Hippel. 2005. 'User-Innovators and "Local" Information: The Case of Mountain Biking'. *Research Policy* 34(6): 951–65.

Malmberg, Lisa. 2017. *Building Design Capability in the Public Sector: Expanding the Horizons of Development*. Linköping University Electronic Press.

Malmborg, Lone et al. 2015. 'Mobilizing Senior Citizens in Co-Design Work'. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '15*, Copenhagen, Denmark: ACM Press, 916–19. http://dl.acm.org/citation.cfm?doid=2786567.2794304 (January 19, 2020).

Manzini, Ezio. 2015. *Design, When Everybody Designs: An Introduction to Design for Social Innovation*. Cambridge, Massachusetts: The MIT Press.

Meroni, Anna, and Daniela Sangiorgi. 2011. *Design for Services*. Routledge. https://www.taylorfrancis.com/books/9781315576657 (September 23, 2019).

Nicolini, Davide. 2013. *Practice Theory, Work, and Organization: An Introduction*. First Edition. Oxford: Oxford University Press.

OECD. 2015a. *Data-Driven Innovation: Big Data for Growth and Well-Being*. OECD. https://www.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en (March 26, 2020).

———. 2015b. *Rebooting Public Service Delivery - How Can Open Government Data Help to Drive Innovation*. http://www.oecd.org/gov/Rebooting-Public-Service-Delivery-How-can-Open-Government-Data-help-to-drive-Innovation.pdf (April 8, 2020).

———. 2019. *The Path to Becoming a Data-Driven Public Sector*. OECD. https://www.oecd-ilibrary.org/governance/the-path-to-becoming-a-data-driven-public-sector_059814a7-en (December 30, 2019).

Ostrom, Amy L. et al. 2015. 'Service Research Priorities in a Rapidly Changing Context'. *Journal of Service Research* 18(2): 127–59.

Paine, Drew, and Charlotte P. Lee. 2020. ‘Coordinative Entities: Forms of Organizing in Data Intensive Science’. *Computer Supported Cooperative Work (CSCW)*. https://doi.org/10.1007/s10606-020-09372-2 (April 11, 2020).

Passi, Samir, and Steven J. Jackson. 2018. ‘Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects’. *Proceedings of the ACM on Human-Computer Interaction* 2(CSCW): 1–28.

Pink, Sarah, Minna Ruckenstein, Robert Willim, and Melisa Duque. 2018. ‘Broken Data: Conceptualising Data in an Emerging World’. *Big Data & Society* 5(1): 205395171775322.

Ramakrishnan, Raghu, and Johannes Gehrke. 2003. *Database Management Systems*. 3rd ed. Boston: McGraw-Hill.

Reckwitz, Andreas. 2002. ‘Toward a Theory of Social Practices: A Development in Culturalist Theorizing’. *European Journal of Social Theory* 5(2): 243–63.

Robson, Colin. 2002. *Real World Research: A Resource for Social Scientists and Practitioner-Researchers*. 2nd ed. Oxford, UK ; Madden, Mass: Blackwell Publishers.

Rosenberg, Daniel. 2013. ‘Data Before the Fact’. In *‘Raw Data’ Is an Oxymoron*, Cambridge, Massachusetts: The MIT Press, 16–40.

Saad-Sulonen, Joanna, Amalia de Götzen, Nicola Morelli, and Luca Simone. 2020. ‘Service Design and Participatory Design: Time to Join Forces?’ In *The 16th Participatory Design Conference 2020- Participation(s) Otherwise*, Manizales, Columbia: ACM.

Sangiorgi, Daniela, and Alison Prendiville, eds. 2017. *Designing for Service: Key Issues and New Directions*. London ; New York, NY: Bloomsbury Academic, an imprint of Bloomsburgy Publishing Plc.

Schuler, Douglas, and Aki Namioka, eds. 1993. *Participatory Design: Principles and Practices*. Hillsdale, N.J: L. Erlbaum Associates.

Schumpeter, Joseph A. 1947. ‘The Creative Response in Economic History’. *The Journal of Economic History* 7(2): 149–59.

Scroggins, Michael J et al. 2019. ‘Thorny Problems in Data (-Intensive) Science’. : 10.

Simonsen, Jesper, and Toni Robertson. 2012. *Routledge International Handbook of Participatory Design*. London: Routledge. https://www.taylorfrancis.com/books/9780203108543 (February 18, 2019).

Stappers, Pieter, and Elisa Giaccardi. 2014. ‘Research through Design’. https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/research-through-design (April 11, 2020).

Stickdorn, Marc, and Jakob Schneider, eds. 2011. *This Is Service Design Thinking: Basics - Tools - Cases*. 2. pr. Amsterdam: BIS Publ.

Strauss, Anselm. 1988. ‘The Articulation of Project Work: An Organizational Process’. *The Sociological Quarterly* 29(2): 163–78.

Stringer, Ernest T. 2007. *Action Research*. 3rd ed. Los Angeles: Sage Publications.

———. 2013. *Action Research*. Fourth edition. Los Angeles, California: SAGE.

Suchman, Lucy. 2002. ‘Located Accountabilities in Technology Production’. 14: 16.

Sumartojo, Shanti, Sarah Pink, Deborah Lupton, and Christine Heyes LaBond. 2016. ‘The Affective Intensities of Datafied Space’. *Emotion, Space and Society* 21: 33–40.

Transport for London. 2019. ‘World’s Largest Cycling Database Set to Make Cycling in the Capital Easier’. *Transport for London*. https://tfl.gov.uk/info-for/media/press-releases/2019/august/world-s-largest-cycling-database-set-to-make-cycling-in-the-capital-easier (April 8, 2020).

Vargo, Stephen L., and Robert F. Lusch. 2004. ‘Evolving to a New Dominant Logic for Marketing’. *Journal of Marketing* 68(1): 1–17.

Vaughan, Laurene, ed. 2017. *Practice Based Design Research*. London New York: Bloomsbury Academic.

Von Hippel, Eric. 2005. *Democratizing Innovation*. Cambridge, Mass: MIT Press.

Wolf, Christine T. 2016. ‘Seeing Work: Constructing Visions of Work in and Through Data’. In *Proceedings of the 19th International Conference on Supporting Group Work*, GROUP ’16, New York, NY, USA: ACM, 509–512. http://doi.acm.org/10.1145/2957276.2997028 (October 21, 2019).

Wulf, Volker et al., eds. 2018. *Socio-Informatics: A Practice-Based Perspective on the Design and Use of IT Artifacts*. First edition. Oxford, United Kingdom: Oxford University Press.

Zamenopoulos, Theodore, and Katerina Alexiou. 2018. *Co-Design as Collaborative Research*.

Zimmerman, John, Erik Stolterman, and Jodi Forlizzi. 2010. ‘An Analysis and Critique of Research through Design: Towards a Formalization of a Research Approach’. : 10.

Part II

Research publications

Publication 1

Seidelin, C., Dittrich, Y., Grönvall, E. (2017) ***Identification of data representation needs in Service Design****.* Selected papers at IRIS, Issue Nr. 8. (2017) 10.
http://aisel.aisnet.org/iris2017/10

Winter 12-31-2017

Identification of Data Representation Needs in Service Design

Cathrine Seidelin
IT University of Copenhagen, cfre@itu.dk

Yvonne Dittrich
IT University of Copenhagen, ydi@itu.dk

Erik Grönvall
IT University of Copenhagen, erig@itu.dk

Identification of Data Representation Needs in Service Design

Cathrine Seidelin[1,2], Yvonne Dittrich[1] and Erik Grönvall[1]

1 IT University of Copenhagen, Rued Langgaards Vej 7, 2300 Copenhagen, Denmark
2 Industriens Uddannelser, Vesterbrogade 6D, 1620 Copenhagen, Denmark
{cfre, ydi, erig}@itu.dk

Abstract. Organisations are looking for new service offers through innovative use of data, often through a Service Design approach. However, current Service Design tools conceal technological aspects of service development like data and datasets. Data can support the design of future services but is often not represented or rendered as a readily workable design material. This paper reports on an early qualitative study of the tools used to work with data and analytics in a medium-sized organisation. The findings identify the current representations of data and data analytics used in the case organisation. We discuss to which extend the available representations of data and data analytics support data-driven service innovation. A comparison of our findings and current Service Design representations show that Service Design lack to represent data as design material. We propose the notion of expansiveness as a criterion to evaluate future data representations for data-driven Service Design.

Keywords: Service Design, Big Data, Design Artefacts, Organisations, Service Innovation.

1 Introduction

Data and the potential to analyse huge amounts of heterogeneous data has become a key asset for our society. More and more organisations turn towards Big Data to seek new or higher profit, new business possibilities or to improve existing work tasks [1]. The popular notion of Big Data often refers to the vast amount of data that may be analysed to reveal complex patterns and behaviours, allowing an organisation to for example discover trends and consumer patterns as large amounts of data are processed by computers [2]. However, while many organizations talk about applying Big Data, many actually work with Data rather than Big Data (e.g. organizations working with datasets containing a few gigabytes of data and not hundreds of terabyte). Working with (Big) Data encounters many different and often complex processes in order to make the heterogeneous data sources available for analysis and application. In line with [3:230], we propose to rather than referring to Big Data in terms of a particular amount of Data, we refer to Big Data as the collection of processes that is needed in order to make Data available for analysis.

In parallel with the increased awareness of Big Data's many possibilities, Service Design is becoming a recognized design discipline within industry. Service Design is as a design discipline facilitating a move from a product-oriented mind-set towards services. The use of Service Design tools allows organisations to better grasp and consider the intangible aspects of service development and innovation. An increasing number of organisations and companies attempt to make use of Service Design as a way to understand the role Big Data may have in the organization and how to design and implement services around Big Data [4]. As put forward by Ostrom et. al [5], Big Data has fundamentally changed how organisations can provide and innovate services. The evolving Service Design discipline aims to provide an explorative and holistic approach to the development and enhancement of services [6,7]. However, Service Design has not yet targeted specifically the design of data-driven services that require the design of analysis and integration of data from heterogeneous sources and across different organisations as part of the participatory design process. Thus, on the one hand organisations have difficulties with understanding how to make use of Big Data for service innovation and development. On the other hand, Service Design as a design discipline lacks methods and tools that enable organisations to design with Big Data and thus explore innovative possibilities for developing smart services [4,5].

To address this issue, the article questions how currently available and applied data and analytics tools confine data-driven service innovation in a Danish medium-sized service organisation. The article explores this question by juxtaposing the core representations in Service Design with empirical results from our early study on the use of representations of data and data analytics in the context of service provisioning.

The remainder of this paper is structured as follows: Section 2 discusses the paper's related work. We draw on the notions of design artefacts and expansive visibilization as theoretical underpinning to discuss the role of representations in design and present the core representations in Service Design. Section 3 presents a description of research methods, and Section 4 introduces the research setting. In Section 5, we present the findings from our early study. Then, in Section 6, we relate our findings to the literature discussed. Finally, in Section 6, we conclude by asserting a lack of representations of data in Service Design and argue it is necessary to consider and develop "a data dimension" in new tools and methods to support design of data-driven services in small and medium-sized organisations.

2 Related work

Why do representations of data in Service Design matter? This is a significant question to ask in our attempt to connect the field of Big Data and the Service Design discipline. To elaborate on the question, we first include the discussion on the role of representation in design. Here, we refer to the discussion on Design Artefacts in Participatory Design, Co-design and developmental work research. We elaborate on what constitutes Service Design and furthermore look at the kind of representations Service Design is offering, especially for facilitating the participatory design of data driven services.

2.1 Why Design Artefacts are important

Most design disciplines work with haptic or semiotic representations of different aspects of the design in progress: representations serve to communicate knowledge of the current situation and the design challenge as well as anticipating future work practices and technologies [8]. Especially, in Participatory Design (PD) the representation of the future system and software have received special attention. PD aims at involving domain experts, that is, for example the future users in the design of their future work and tools. Design Artefacts here have the role to support communication and cooperation across professional disciplines. Already in 1995, Morten Kyng discussed in his article 'Making representations work' the need to choose representations well to support the open ended cooperative design [9]. Building on this work, Bertelsen discusses in depth the role of representations as Design Artefacts mediating and facilitating the design in three dimensions, being Construction, Communication and Conception [10]. A specific Design Artefact would support all three dimensions: The Construction dimension describes, how a design artefact supports the concrete implementation of design; a mock-up for example provides the instruction of the overall layout of the application. The Communication dimension describes the how an artefact can support the communication between different stakeholders. To use mock-ups again as an example: the mock-up provides a deictic space for users and designers to relate to functionality and data by pointing to interface elements. The Conception dimension that is facilitated by design artefacts is the conception of new ideas, the creativity that is part of all design. The mock-up supports conceptualisation when it allows the participants of the design session to follow ideas and take apart and reassemble a paper mock-up in line with innovative functionality. The mediating quality of Design Artefacts though does depend on the design constituency it is used with: Design Artefacts that serve well the cooperation with non-IT professionals might not be suitable to mediate a design discussion between the software architect and the development team.

Engeström's article 'Expansive Visibilization' from 1999 [11] can be read as an elaboration of the last of Bertelsen's facilitation dimensions: He compares different ways of representing work processes, and argues for the need that the representations support not only the communication of workflows and the social and spatial arrangements that implement it, but also the learning and change of the arrangements.

As further discussed below, in the context of data driven services, data is not any longer only an enabler of services, but becomes part of the material that can be used to improve or design new services. In order to be subject to cooperation between service designers, domain experts and software developers, data has to be represented in a way that supports not only the construction of computer support, e.g. in form of integration of heterogeneous data sources, but also needs to represent data as design material. Applying these discussions on the representation of data sources in the context of Service Design, which prompt a list of questions: How do different representations relate to the design of the existing data infrastructure? Can the design be related to concrete future technical functionality? Do representations of data support communication and cooperation between designers, domain experts and IT experts? Last but not least, the representations of data need to support conception or the innovation of services. With other words, how can we create representations of heterogeneous data (re-)sources that support continuous improvement of data-driven services? These questions are relevant for future work in order to develop tools and methods for data-driven Service Design. However, to explore and answer these questions goes beyond the scope of this article. To build a foundation for future work, this paper focuses on the current situation and thus how available and applied data and analytics tools confine data-driven service innovation.

2.2 Service Design

Service Design has emerged from the needs, and perceived possibilities, of companies and other organisations to provide services to their customers [5]. A main objective for Service Design is to establish a holistic, user-centred perspective throughout the design process. (The term 'user' refers here to the service user, not necessarily an IT user.) A traditional product-centric business model focus on selling products such as a computer or coffee mugs. Here the company-customer relationship constitutes very few encounters, for instance at the time of purchase, and the value is exchange-based as the customer receives the product in exchange for money [12]. In the case of services, a company would not sell for example a computer once, but rather sell the service of on-demand computational power. That also means that a service per se does not have any value by itself, but that value is created through service use [6]. Some even go as far as stating that a service only exists, when it is used [13]. Designing a service is hence something else than designing a product, and being a service provider is different from being a product manufacturer or retailer. As pointed out by Polaine et al, Service Design as a consequence is different than other design practices such as Industrial, interaction or experience design [6].

Service Design uses methods and tools developed for a wider purpose, for example Personas [14] and Storyboards [15], but has also as a field developed its 'own' tools which specifically targets the design of services. Examples of these tools includes Customer Journeys [16], Service Blueprints [17], and Service Ecology Maps [6].

Customer Journeys, Service Blueprints, and Service Ecology Maps are all examples of tools strongly related to the Service Design community. These tools often have two functions in the design process: they work both as analytic tools to document a given situation or as a representation depicting the anticipated future (e.g. before or after a service has been (re-)designed).

A Service Blueprint is a tool that facilitates the process to map out and understand how a service will look like, unfold from a user or customer perspective, actions needed at specific locations and infrastructural needs. A blueprint is divided into two sections by 'a line of visibility'; a front stage part that the user 'see', and a backstage part that contains important elements for the service but that is not noticeable by the user.

A Service Ecology Map represents actors and their relationships. It can take different graphical expressions, but is often a circular shape where the further away a representation of an object (like an actor, a technology, or infrastructure) is placed from the centre, the further away from the core service it is. The circle can be split up into different sections, representing for example different aspects and actors of the service, like how is something performed, who performs it, when is it performed, where is it performed, what enables it, and why is it performed.

A Customer Journey is a (graphical) representation of a scenario, visualizing how one or more actors interact with a service. The customer journey may for example visualize a trip to the hospital, being based on a particular patient persona and his or her envisioned use of a healthcare provisioning service. A customer journey can help the design team to foresee, plan and discuss possible user behaviour and service interactions based on for example a persona.

The above examples are representative tools used within the field of Service Design. While they allow a quick overview of both the current or envisioned future situation, with embedded possibilities and shortcomings, these tools are less optimal to use by themselves and in isolation for service design work where complex and high quantities of data are the main service enabler. To work with precise data flow analysis and design that can prepare a service for implementation, Unified Modeling Language or other tools must often be used to 'engineer' the technical side of the service, preparing for it to be programmed by a software developer. When designing services around Big Data, these tools do allow service designers to open up and explore aspects of data in these tools. They though are not meaningful to facilitate design together with domain experts.

3 Methodology

This research constitutes an early study for a subsequent action research PhD-project. Due to this linkage and because of the case organisation's underlying wish to create change through these research activities, we likewise adopted an action research approach for this study [18, 19, 20]. This section elaborates on the research setting, the applied methods and the analysis.

3.1 Research Setting

The empirical research took place at The Educational Secretariat for Industry, Industriens Uddannelser (IU), which is a medium-sized service organisation based in Copenhagen, Denmark. IU's main services and service provisioning are centred around the development of educational programmes for vocational training and adult vocational training in the industrial sector in Denmark. The organisation integrates heterogeneous data sources including government data, personal data, and data generated through their service provisioning. Referring to the literature on Big data [18], IU's work with data resemble Big Data in terms of high variety, velocity and veracity though the volume is not comparable to that of data generated e.g. through social media platforms. As mentioned in the introduction, we thus refer to Big Data in this context. Besides being a service organisation, IU can also be seen as a knowledge broker organisation, that for example utilizes heterogeneous data to answer to knowledge and information needs [21]. Being a knowledge broker organisation, IU cooperates with a large number of organisations in order to generate and provide data for their key stakeholders, who have different needs for information and data analysis. IU and its cooperating organisations share the interest of many organizations, being how they can utilize data in more innovative ways, e.g. to improve their services. However, unlocking the data potential can be challenging for organizations, to transform data into a viable and reliable resource that can inform databased services and ideally create a competitive advantage and fuel growth [1, 4, 19]. For many organisations, like for example IU, it is also challenging to build and implement the necessary organisational structures to support data-based service provisioning as there is no “one size fits all” solution for how to create and implement data-based strategies [19, 20]. This study was initiated as a way to investigate IU’s current work with data and data analytics tools in order to further understand how the organisation’s current “data practices” can be changes and developed.

3.2 Methods

The data collection focused on how currently available data analytics tools mediate different ways of working with and exploring data in relation to service innovation. The primary data sources thus consisted of 4 semi-structured interviews, observations, participatory observations, a workshop, and studies of the tools used for data-related activities in the organisation (see table 1). The fieldwork was conducted at IU, and specifically focused on the work of the Statistic Team, a group (four people in total) in the organisation that was responsible to create periodical statistical reports and support other members of the organisation with data analytics. Members of the Statistic Team were interviewed about their organisational role and data-related tasks. The members of the Statistic Team were also observed as they performed individual data-related activities in their offices. Ambiguities which emerged were investigated by follow-up questions.

Table 1.

Empirical Data	**Amount**	**Total Length**
Individual interviews with the members of the Statistic Team	4	6 hours
Observations of the individual members of the Statistic team and team meetings	12	8 hours
Participatory observation of the statistic team’s Statistic Seminar for	1	4 hours
Workshop	1	3 hours

The study lasted 3 months. During this time, the first author worked at the organisation and thus became of the everyday life at IU. Moreover, she immersed herself into data-related activities and initiatives taking place in the organisation to collect data from ongoing work concerned with data analytics tools. The data collection was documented through audio recordings, field notes, photos and documents distributed at participatory events. To prepare the analysis, audio recordings were transcribed word by word.

3.3 Analysis

A thematic analysis was used to identify and understand the employee's use of the currently available data analytics tools [19]. The analysis started in parallel with the field work through ongoing status meetings amongst the authors. The themes that emerged focused on barriers for data exploration, statistical data representations, the work processes of the statistic team, technical infrastructures and 'silo IT-systems'. The themes emerged based through two coding iterations; open coding and coding which focused specifically on data-related actions. In this article, we focus on four categories of tools used for data-related activities and on the purpose of their usage: Scripts, System Interfaces, Tables, and Infographics. This paper explores these categories in depth and questions how these tools confine data-driven service innovation.

The study included several ways to assure the trustworthiness of our results. Throughout the research and the analysis, the second and third authors took part in debriefing sessions supporting the reflection and direction of the research. We used multiple data sources to triangulate the findings. The statistic team was invited to comment on the developing themes and in a workshop the results were presented and discussed by a wider group of members of IU.

4 Findings

IU works to develop educational programmes for vocational training and adult vocational training in the industrial sector in Denmark. It is responsible for 45 vocational training programs and more than 1000 adult vocational training courses. Moreover, IU acts as a knowledge broker in a network of more than 20 cooperation organisations, which all work together to future-proof the Industry by creating the conditions that can provide the necessary, skilled labour. More specifically, they do so by aiming to get more people to choose (and complete) vocational educations and to get more unskilled workers to become skilled through attending adult vocational training. At the current state, the usage of data at IU primarily serves two overall objectives: Data is both used as proof to subsidise argumentation and as a foundation for decision-making.

Data constitute central elements in terms of how employees at IU deliver and improve services. The empirical data shows that data and analytics used to be applied in particular cases to support specific decision-making processes. To make the use of data and data analytics less time-consuming and more valuable for the organisation, the management decided to appoint a statistic team. The team consists of four employees from different departments in the organisation, who are responsible for staying updated on topics such as a data access and data security. Moreover, they are responsible for creating and publishing statistics about the development of all vocational training programmes and adult vocational training programmes on a regular basis. The tasks within the team are divided so that two of the members mainly provide the periodic statistical reports, while the other two members to a greater extend communicate with external stakeholders on topics such as data access. By observing the team members' work practices, we identified the tools, which are used to perform the various data-related tasks.

In presenting this study, we elaborate on the four identified categories of tools, which emerged in the analysis; Scripts, System Interfaces, Tables, and Infographics. Together, these categories of tools make up the IU's present approach to working with data, analytics and representations of data. Figure 1 illustrates how the tools are connected in relation to IU itself, external data providers, the public and stakeholders. The four categories embody representations of data and data analytics in different ways. The categories were divided into two groups based on the tools' and thus the categories' overall objective: while Tables and Infographics are representations of data; Scripts and System Interfaces are representations meant for producing data analytics. These categories of tools are relevant in the context of Service Design, as they might provide representations that allow for exploration of the potentials in data that need to be extracted through analysis. By elaborating on the four identified categories of tools, this paper aims to create a foundation for further research that can further support the creation of representations of data in Service Design, which will allow non-IT experts to explore and design with (Big) data.

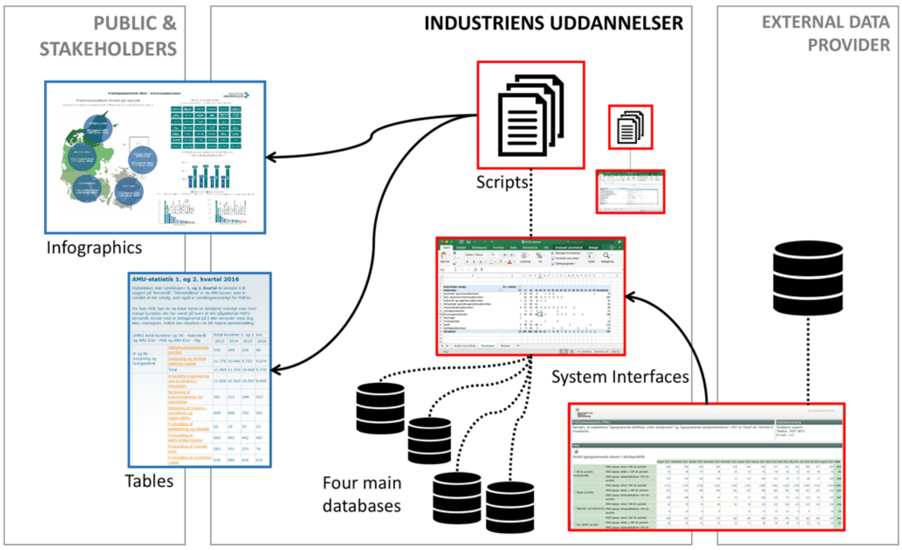


Fig. 1. Categories of tools currently used at Industriens Uddannelser for data-related activities

4.1 Scripts

As mentioned, the statistic team is responsible for creating and publishing statistics and statistical representations for internal and external use on a regular basis. The data and the representations of the data are produced in a manner, which makes them comparable to previous statistical statements. During the process of extracting, analysing and representing the data, the appointed employee makes use of detailed instructions in form of documents to solve the data-related task. These documents functions as scripts to make up visualisations, as guidelines that enable an employee to do specific data analysis activities. It explains systematically how a human actor can complete a specific and predefined activity. In this way, scripts serve as tools with the objective of making data analytics. However, the script in itself does not invite its users to go beyond and explore the data. The empirical data shows two forms of scripts: One form are standard word documents containing screenshots and detailed descriptions of various procedures. The second form constitutes a Wikipedia-inspired page known within the organisation as “Stati-pedia”. Located within Microsoft OneNote, this page is a part of a larger knowledge-sharing initiative across the organisation. The page contains specific information and links to websites that are often used to access system interfaces where government data can be accessed in order to generate relevant datasets. “Stati-pedia” was introduced by the Statistic Team as a way to make their way of working with data-related task transparent and accessible for the whole organisation.

4.2 System Interfaces

This tool category refers to an interface as a point of interaction, which enables people to engage with a computer-based system. System Interfaces supports the transfer of data between a user and a computer system. By facilitating this transfer, the System Interface becomes a visualisation of data analytics. This category emerged from observations of the members of the Statistic Team and their use of various System Interfaces during the process of producing data analytics. More specifically, the empirical material shows the usage of three different System Interfaces, which includes Excel, an Excel Macro customized to produce statistical representations, and “The Databank” (a system interface provided by the Danish Ministry of Education). Systems Interfaces are more open for free exploration compared to Scripts, in that they do not ‘dictate’ certain actions. System Interfaces add a layer to data analytics, which aims to enable users to interact with data in a less predefined way. However, it is necessary for a user to know and understand the interfaces in order for him or her to use them for exploratory purposes. This means, that System Interfaces embody increased possibilities for data exploration, but they are at the same time difficult to access for users, who do not have comprehensive knowledge about the interface’s expansiveness.

4.3 Tables

At its core, a table is a data arrangement that consist of columns and rows, which furthermore enables a relation between these. In contrast to the two preceding tool categories, tables are schemes with the objective of representing data. At IU, tables are often used to show selected data. An example is when the organisation represents statistics about adult vocational training on the organisation's website. Representing data in this way depicts data in a linear manner; from one data point to the next and often in relation to time. Moreover, these tables only depict data from the past and thus do not include any databased extrapolation or prediction. This form of data representation excludes all aspects of the data analysis process, and only shows the result of that process. In sum, tables are structured schemes that do not invite users to explore the data further.

4.4 Infographics

The final tool category constitutes Infographics, which are graphical visualisations of selected data that intends to present data analysis quickly and clearly. Infographics was initiated by the Statistic Team as an attempt to represent data in a new way. The deviation from the standardized data representations in forms of tables generated positive feedback from external stakeholders. This new way of representing data also created a new demand in form of additional requests for visualising other types of data in the form of Infographics. Compared to tables and graphical representations of statistics, the Infographics enabled a new and different way of representing data, which resulted in positive reactions and new demands. However, in order to visualize data in this 'easy to read' manner, there are two central prerequisite steps: First, the employee has to generate a specific dataset, which is then further prepared using a separate system. In other words, the employee is required to make use of both System Interfaces and Scripts to be able to create Infographics about new data. This makes Infographics less accessible.

In a workshop with both the statistic team and representatives of the consultants and the administrative staff they support with data analytics, both the need and wish to work with the data in different and innovative ways in the future was discussed. There was a widespread awareness about problems with the current way of making use of data: issues about privacy, veracity and lack of overview were discussed. Likewise, the potential of new ways of working with data were clearly articulated. As a first steppingstone, knowledge sharing workshops and the above mentioned Statipedia have been initiated by the statistic team. However, the challenge remains to make data accessible and understandable for non-IT professionals. These challenges are further detailed in the following section.

5 Discussion

The empirical research described above indicates that data in organisations like IU is an intrinsic part to the services provided. Data is in this case not a technical commodity underpinning the service delivery that can be black-boxed, but it needs to be made visible and accessible to design as core ingredients when designing the service. Below we further discuss the qualities and limitations of the representations used in IU respectively available as part of the service design toolbox.

5.1 The expansiveness of data and data analytics

In the related work in Section 2, we emphasised that design artefacts need to represent relevant aspects of the design in an expansive way, that means in a way which invites creative ideas and new conceptualisations. How would expansive visibilizations of data look like? This has yet to be investigated. However, we can see how expansive the representations of data and data analysis are today. Above, we have identified four categories of tools that have been used for data-related activities at IU. Despite having different functionalities and objectives, these tool categories also vary in terms of possibilities for data exploration. The majority of the organisation's currently available tools for data and analytics related activities reflect what Engeström calls the linear dimension [11]. Scripts communicate how to extract one specific set of data. Tables represent data without allowing to explore correlations. Though infographics make more dimensions of data accessible, they do not allow the reader to explore additional relations. They only make the outcome of the employee's data analysis and processes visible and they represent static data. System Interfaces, though, are tools made to enable users to engage with data analytics. However, a system interface's ability to support data exploration is closely related to the skills of the employee, who uses it.

During the workshop described above the members of the statistic team and the representatives of the consultants supported by the team clearly stated that the current data analytics tools are insufficient to support the improvement of the quality of the current services. It also became clear that potentials for future developments, e.g. designing a set of data that would be indicative to the health of an educational program or the ambition for prediction of educational needs based on past and current data, were hardly accessible based on the current tools. For this, more exploratory tools were needed that allowed the domain experts to connect to the possibilities the rich data sources IU has at its hands.

Last but not least, the discussion here indicates that expansiveness is a relevant criterion to evaluate data representations as design artefacts. However, it also indicates that expansiveness might be dependent on whether the representation is understandable and accessible to the user. As database-level system interfaces allow exploring the bare bone data model, such an interface is only accessible for people with at least a basic understanding of databases and data analytics.

5.2 Making (Big) data meaningful in service innovation

Like IU, more and more organisations are looking to develop smarter, data-driven services that e.g. can automate processes and in this way, improve an organisation's service provisioning. However, little attention has been given to the challenges that the process of designing with data encounters. As the empirical research reported above shows, members of organisations like IU who try to develop new ways of working with data do not have adequate tools. This research analysed the present data analytic tools used at IU in order to understand how data-related work and innovation is supported – or limited – by them. The findings show that the currently available tools facilitate very limited possibilities for exploring data, unless the user has developed advanced knowledge about the tool and the data available through it. This exemplifies how data analytics tools, which are difficult to access, and thus implement in work practices, affect employees' ability to make sense of and innovate with data. Moreover, the tools' limited possibilities for data exploration make it difficult for the user to make sense of what data is available/accessible beyond the scope of a particular, pre-defined data-related task, which arguably restricts data-driven service innovation. The research indicates the necessity of implementing data analytics in the organisation's future service innovation. It is essential for organisations, such as IU, to be able to discuss data as part of their service innovation as a malleable material to design with. Therefore, a question for future work remains; how do we facilitate this discussion?

5.3 Service Design as a sensemaking activity

Weick first introduced the concept of sensemaking to describe how we structure the unknown in order to navigate and thereby be able to act in it [24]. The empirical material shows how IU's cooperation organisations increasingly requests data, which means that IU needs to allocate an increasing amount of time and resources to be able to understand how to go about these new incoming data-related tasks.

Prendiville, Gwilt and Val propose that Service Design can be developed into an approach for organisations in the pursuit of turning "the abstract and intangible nature of Big Data into human-centred services with social and economic value; thus transforming highly technical forms into something that can be understood and consumed by broader communities" [4: 225]. They argue the use of tools to facilitate visualisation, mapping and co-design in Service Design offers sensemaking activities that can function as a foundation to establish the necessary organisational structures to bring together relevant stakeholders required to enable data-driven service innovation. However, at the same time, they underline that the currently available tools need to adapt and evolve in order to enable organisations to act in the unknown world of data. It is thus in this context that the identified gap of Service Design's incapacity of data representations manifest itself as an issue. As mentioned in the related work, most design disciplines work with haptic or semiotic representing different aspects of design in progress. Kyng argues that well-established representational artefacts often continue to be used "not because they mirror that which is represented, but because they do not, that is, the representation captures a few intentionally selected qualities of that which is represented and nothing more" [9: 46]. This quote underlines an inevitable contradiction, which all representations embody. On the one hand, the simple configuration of Service Design tools makes them accessible for people who are unfamiliar with practicing Service Design. In part, the simple configuration is a key enabler for organisations, who draw on Service Design to innovate services despite of internal competences [12]. On the other hand, Service Design tools conceal technological aspects of service development and especially the data, which can support the design of smart, data-driven services. The identified gap between Service Design and (Big) data calls for new or improved tools that includes representations of data, in order to support stakeholders to collaboratively explore, make sense of and design with data. This prompt the final question in this paper, which is how the identified necessity for data exploration in service innovation changes the requirements for Service Design tools? We discuss this question in the following section.

5.4 Data exploration changes requirements for Service Design tools

As described in the Related Work section, a number of Service Design tools exist to facilitate the design process. These tools constitute simple representational artefacts to support a defined design activity [10]. At the current state, it is only the Blueprint of the aforementioned tools, which to a limited extend considers technical integrations of service development. As a tool, the Blueprint [17], represents the phases of a service experience from start to end including points of interactions between users and service, and the support processes which occur throughout the service journey. Through “the line of visibility” the Blueprint facilitates the considerations of actions and processes that might occur even though they are not visible to the user of the service. However, the Blueprint does not represent data in a way that allows to design with data.

A first step towards tools for data-driven service design seems to be to avoid black-boxing data. Next, the research indicates two additional requirements for the design of new Service Design tools: First, data needs to be represented in “expansive” ways that enable exploration. Second, the representation and exploratory tools need to cater to non-IT experts and thus need to abstract from unnecessary complexity.

6 Conclusion

We started out with the aim to explore the representations useful for Service Design with Big Data. We did so by juxtapositioning related representations from Service Design and practices in a broker organisation to give an understanding of the needs such representations have to fulfil.

First of all, we can state that there is a need for representations that allow non-IT professional to explore and work with data when improving their data dominant services. Second, we also can conclude that the representations used are not very expansive, as they do not support exploration of and learning with and about data for the normal domain expert. And third, we need to admit that Service Design does not provide adequate representation, as data and its analysis are normally not subject to service design but black-boxed as technical commodities provided by software developers.

So the main result of the article is the identification of a gap: a need for representations that at the same time are expressive and expansive enough to make data and data analysis accessible as ingredients for services design but abstract from technical aspects not necessary for the design. The gap between Service Design and (Big) data calls for new or improved tools that includes representations of data, in order to support stakeholders to collaboratively explore, make sense of and design with data. We propose to evaluate the expansiveness and accessibility of such visibilizations as criteria to evaluate such representations. This directly points to the future research that we have recently started.

Acknowledgments. We are grateful for the collaboration with Industriens Uddannelser, who made this research possible, in particular the members of the Statistic Team and workshop participants. We thank the reviewers for their valuable contributions that have improved the paper. This research was supported by Innovation Fund Denmark and The Danish Industry's Education and Cooperation Fund.

References

1. European Commission. (2016, February). Towards a thriving data-driven economy. Retrieved October 2016, from Digital Single Market - Digital Economy and Society: https://ec.europa.eu/digital-single-market/en/towards-thriving-data-driven-economy
2. Berman, J., (2013). Principles of Big Data – Preparing, Sharing and analysing complex information. MA: Morgan Kaufman Publishers.
3. De Mauro, A., Greco, M., Grimaldi, M. (2016). A formal definition of Big Data based on its essential features, Library Review, Vol. 65 Issue: 3, pp.122-13
4. Prendiville, A., Gwilt, I., & Val, M. (2017). Making sense of data through service design opportunities and reflections. I D. Sangiorgi, & A. Prendiville, Designing Services (s. 225-235). London: Bloomsbury Academic.
5. Ostrom, A. L., Parasuraman, A., Bowen, D. E., Patrício, L., & Voss, C. A. (2015). Service Research Priorities in a Rapidly Changing Context. Journal of Service Research, Vol. 18(2), pp. 127-159.
6. Polaine, A., Løvlie, L., & Reason, B. (2013). Service Design - From Insight to Implementation. Brooklyn, NY: Rosenfeld Media.
7. Kimbell, L. (2011). Designing for Service as One Way of Designing Services. International Journal of Design, 5(2), pp. 41-52.
8. Kensing, F., Munk-Madsen, K. H. (1991). Generating visions: Future workshops and metaphorical design. In J. Greenbaum and M. Kyng (Eds.), Design at work: Cooperative design of computer systems (pp. 155-168). Hillsdale, NJ: Erlbaum
9. Kyng, M. (1995, September). Making Representations Work. Communications of the ACM, Vol. 38(9), pp. 46-55.
10. Bertelsen, O. W. (2000). Design Artefacts: Towards a designoriented epistemology. Scandinavian Journal of Information Systems, 12(1), pp. 15-2
11. Engeström, Y. (1999). Expansive Visibilization of Work: An Activity-Theoretical Perspective. Computer Supported Cooperative Work(8), pp. 63-93.
12. Boztepe, S. (2007). User value: Competing theories and models. International Journal of Design, (2), p. 55-63.
13. Nielsen, L. (2005). Personas In: Encyclopedia of Human-Computer Interaction. p. 1-37.

14.van der Lelie, C. (2006). The value of storyboards in the product design process. Personal and Ubiquitous Computing, 10 (2). p. 159-162
15.Morelli, N. (2002). Designing product/service systems. A methodological exploration." Design Issues 18(3). p. 3–17.
16.Richardson, A. (2010, November). Using Customer Journey Maps to Improve Customer Experience. Retrieved September 2016, from Harvard Business Review's website: https://hbr.org/2010/11/using-customer-journey-maps-to
17.Shostack, G. L. (1984). Designing Services That Deliver. Retrieved September 2016, from Harvard Business Review: https://hbr.org/1984/01/designing-services-that-deliver
18.Kitchin, R. (2014). The data revolution, big data, open data, data infrastructures & their consequences. In R. Kitchin, Conceptualising Data (p. Chapter 1). London: Sage Publications Ltd20.
19.Chevalier, J.M., Buckles, D.J (2013). Participatory Action Research - Theory and Methods for Engaged Inquiry. London: Routledge 18. Robson, C. (2002). Real World Research - A resource for Social Scientists and Practitioner-Researchers (Second Edition ed.). Oxford: Blackwell Publishing.
20. Robson, C. (2002). Real World Research - A resource for Social Scientists and Practitioner-Researchers (Second Edition ed.). Oxford: Blackwell Publishing.
21.Mayer-Schönberger, V., Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think. London: John Murray.
22. Meyer, M. (2010). The Rise of the Knowledge Broker. Science Communication. 32(1) p. 118-127.
23.Ramakrishnan, R., Shahabi, C. (2014, July). Big Data and Its Technical Challenges – Exploring the inherent technical challenges in realizing the potential of Big Data. Communications of ACM, 57(7), pp. 86-94.
24.Weick, K. (1995). Sensemaking in organizations. Thousand Oaks, CA: Sage.

Publication 2

Seidelin, C., Dittrich, Y., Grönvall, E. (2018) ***Data work in a knowledge-broker organisation: how cross-organisational data maintenance shapes human data interactions***. Proceedings of the 32nd International BCS Human Computer Interaction Conference 32. DOI: http://dx.doi.org/10.14236/ewic/HCI2018.14

http://dx.doi.org/10.14236/ewic/HCI2018.14

Data Work in a Knowledge-Broker Organisation: How Cross-Organisational Data Maintenance shapes Human Data Interactions

Cathrine Seidelin
IT University of Copenhagen
Copenhagen, Denmark
Cfre@itu.dk

Yvonne Dittrich
IT University of Copenhagen
Copenhagen, Denmark
Ydi@itu.dk

Erik Grönvall
IT University of Copenhagen
Copenhagen, Denmark
Erig@itu.dk

The term Human-Data Interaction (HDI) conceptualizes the growing importance of understanding how people need and desire to use and interact with data. Previous HDI cases have mainly focused on the interface between personal health data and the healthcare sector. This paper argues that it is relevant to consider HDI at an organisational level and examines how HDI can look in such a context, where data and data maintenance are core assets and activities. We report on initial findings of a study of a knowledge-broker organisation, where we follow how data are produced, shared, and maintained in a cross-organisational context. We discuss similarities and differences of HDI around personal health data and cross-organisational data maintenance. We propose to extend the notion of HDI to include the complexity of cross-organisational data work.

Human-data Interaction. Data Work. Data Maintenance. Collaboration. Knowledge-Broker Organisation.

1. INTRODUCTION

In general, society becomes more and more populated with technology, sensors and other data gathering and processing entities. While much attention is given to the physical design of technology and its digital interfaces and user interactions, it is not only the physical and digital surfaces we interact with and how they are designed that is important, but also the very data and how we interact with data. Researchers have argued that more research is needed to further understand the processes affecting new forms of data work and data-driven accountability (Blomkvist et al., 2015, Bossen et al., 2014, Hogan et al., 2017). This study is one response to this need. In this paper, we explore how data is maintained in a cross-organisational or otherwise distributed contexts where many use, or ‘interact’ with the same data, and what is required to improve or support this cross-organisational data work.

The study reported here relates to data and people's interaction with data, in particular in an inter- and cross-organisational context to inform the initial stages of the design and development of a new interactive system. Our study will inform a final database design, but regards peoples use and interaction with data in a distributed context rather than technical database challenges or solutions.

Our study is situated at the medium-sized service organisation Industriens Uddannelser (English: Education secretariat for industry, hereafter the acronym IU is used), an education secretariat based in Copenhagen (Denmark). IU facilitates the collaboration between diverse labour market partners to develop educational programs for vocational training and adult vocational training in the industrial sector in Denmark. Data is at the core of this collaboration; involving data collection, processing, analysis, and intra- and cross-organisational data exchange. The need to collaborate around data makes IU and other organisations more and more interconnected through shared objectives, policies, IT-systems, interfaces, and indeed *data*. This results in complex networks of data flows, including data production, maintenance, processing, sharing and usage. This “data interconnectedness” generates a joint, cross-organisational responsibility for data maintenance. The complexity of inter- and cross-organisational data management, where data updates can origin from different organisations and stakeholders, has led to the establishment of IU as the knowledge-broker organisation within this complex network of stakeholders with different knowledge interests

(Meyer, 2010, Jackson and Baker, 2004). Through this setup, IU becomes a central actor and facilitator for data quality and transparency in data management. While these two dimensions of data management may seem straight forward in the context of data work, they constitute central aspects of the new General Data Protection Legislation in Europe, which underline that they should not be overlooked as a means of supporting and improving cross-organisational collaboration (eugdpr.org, 2018).

Due to the central role of data in our case, the study turns towards recent work in the field of Human-Data Interaction (Crabtree and Mortier, 2015, Mortier et al., 2014, Haddadi et al., 2013, Wilke and Portmann, 2016) to find a suitable analytic perspective. The emerging research field of HDI proposes to place "*the human in the center of the flows of data and providing mechanisms for citizens to interact with these systems and data explicitly"* (Mortier et al., 2014, p. 1). The increased attention to, and use of data, in society makes data and understanding how we use and interact with data increasingly important. Thus far, the field of HDI has mainly been used in healthcare contexts and have discussed the interface between personal data (e.g. health data) and an organisational entity (e.g. the healthcare sector) (Cabitza and Locoro, 2017, Crabtree and Mortier, 2015). However, data and data interaction (e.g. data maintenance) become increasingly core assets supporting central databased services that thereby goes beyond the interaction between the individual user and his or her personal health data (Karasti and Baker, 2008). Furthermore, given the growing and wider use of "Big Data", these aspects are relevant to consider from an organisational perspective. HDI is a first step to consider data as a central part of HCI. However, the focus on health data (e.g. the relation between a patient and patient data management) leaves out the cross-organisational dimension. We therefore argue it is beneficial to study cross-organisational data work and organisational data from a HDI perspective also in non-healthcare contexts. Through our study, we explore different kinds of data (being personal, public, administrational or organisational data entities) as boundary objects for the collaboration at IU and with the organisation's key stakeholders. Thus, this study contributes to existing work by further exploring the concept of HDI and what constitutes HDI in a cross-organisational context.

The paper proceeds as follows: In the next section, we discuss the related work, which focuses on the concepts of Data work and Collaborative Care, Human Data Interaction Studies and Data as Boundary Objects. Then follows a case description and the research methods are presented. The paper then proceeds to our analysis and discussion, which focuses on the social practices and collaborative management related to data use and maintenance at IU. In particular, the analysis investigates the needed ongoing coordination of data production, potential data discrepancies, IU's responsibilities as a knowledge-broker. Following a discussion, the paper concludes by proposing a wider notion of HDI.

2. RELATED WORK

In the following, we briefly touch upon the concepts of Data work and Collaborative Care to frame our study. We then review HDI-related studies to support our argument that interactions with data is at the core of cross-organisational data maintenance. Afterwards, we build on the existing work, as we elaborate on the concept of boundary objects in order to underpin our discussion of what constitutes HDI in a cross-organisational context.

2.1 Data Work and Collaborative Care

The concept of Data work has been coined to address the "*the social practices in and through which data is accountably collected, used, and acted upon*" (Fischer et al., 2014, p. 1). As such, the notion of Data work is relevant to our case in trying to understand people and organisations' interaction with data. Related studies have indicated how emerging technologies demand new practices in order to make visible, anticipate and perform work that have data at its core (Fischer et al., 2014, Bossen et al., 2016, Elsden et al., 2016). With an increased data collection and new possibilities for data-driven innovation through for example Big Data, organisations need and desire the ability to understand, explore and thus interact with their data (Kitchin, 2014). Such Data work is complex, distributed and often interdependent of external stakeholders, organisations and third parties (Fischer et al., 2014, Bossen et al., 2016). Previous examples of Data work and studies of digital data practices and infrastructure in cross-organisational and multi-stakeholder contexts do exist, for example within e-Science, library science, Information science and Ocean Informatics (Fearon, 2017, Futrelle et al., 2011, Koesten et al., 2017, Jackson and Baker, 2004, Karasti and Baker, 2008, Bowker, 2000). In these studies, data is an acknowledged entity and Data work is a recognized activity, but we are not aware of Data work-studies that take on a knowledge-broker perspective for crafting multi-stakeholder and cross-organisational system designs. In this paper, we add to the existing body of work on Data work by exploring how the role and presence of a knowledge-broker organisation affects collaborative Data work across organisational boundaries, not only in initial systems design work but also in system use and everyday work. As such, when we talk about data, we perceive it as a malleable entity, both in initial design work and in later use of for example a system and its data (see similarities with infrastructuring (Karasti, 2014, Seravalli, 2012, Pipek and Wulf, 2009)).

To further frame our study, we also draw on the notion of Collaborative Care proposed by Jackson and Baker (2004). The concept has emerged from a study on collaborative tensions, which occur as a result of collaborative undertakings that aimed to join and construct information infrastructures within the fragmented fields of Ocean Science (Jackson and Baker, 2004). Based on this study, Jackson and Baker (2004) propose the concept of Collaborative Care as a means to embrace, bridge and preserve heterogeneity in collaborative interaction. We apply Collaborative Care perspective in order to examine how trust and compromise is established in a cross-organisational context with a knowledge-broker organisation at the centre of a complex network of actors with different knowledge interests.

2.2 Human-Data Interaction Studies

The concept of HDI was coined by Haddadi et al. (2013), in order to conceptualize the increasing ethical and practical challenges concerning collection, analysis and trading of personal health related data. Haddadi et al. further propose that HDI does not consider explicit interactions, but rather passive scenarios which allow one to consider how people interact with "*apparently mundane infrastructure, which they generally do not understand or would rather ignore*" (2013, p. 5). Haddadi et al. (2013) emphasize that HDI further differs from HCI by focusing on aspects or dimensions of people's interaction with computer systems that is usually not in the center of attention within the existing body of HCI work: First of all, HDI focuses on the social interaction with data *itself*. Secondly, HDI differs in terms of scale, in that dealing with infrastructures for sharing data takes a bigger part than what is usually considered in interaction studies (Mortier et al., 2014). While this paper applies HDI as a theoretical framing, we do argue that the concept of HDI has shortcomings, which we will elaborate on in the following paragraphs.

In one of the earlier works on HDI, Mortier et al. (2014) presents a model (Figure 1) that illustrate the concept of HDI. The model makes visible how personal data feeds into more or less invisible data-ecosystems, in which the individual has little or no control over his or her personal data. On this basis, and as pointed out in the introduction, Mortier et al. stress the need for placing "*the human at the center of the flows of data, and providing mechanisms for citizens to interact with these systems and data explicitly"* (2014, p. 1). They further highlight three challenges that HDI raises: First, they argue there is a need for data to be more legible, in order for people to understand it. Secondly, they argue that it requires giving people agency so they are able to act within complex data ecosystems. The third challenge they emphasize focus on the current data ecosystems favour of data aggregators over the individual user, which create an imbalance of power between these actors. These are all challenges that resonate with the later developed and adopted European GDPR (eugdpr.org, 2018), and thus reflects a growing societal need for research that explores the areas which the field of HDI addresses.

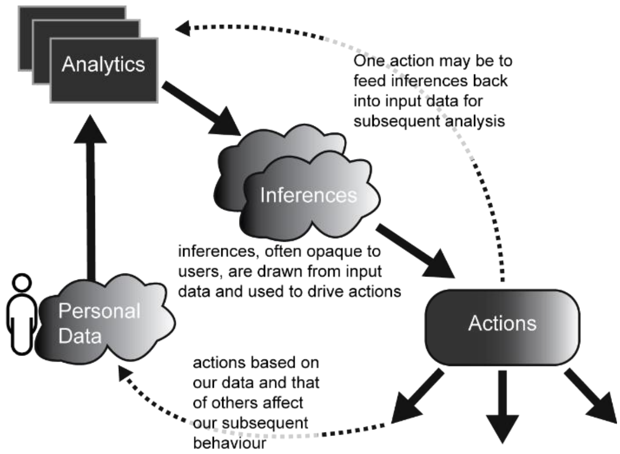


Figure 1. *Human Data Interaction (redrawn from Mortier et al., 2014)*

To address these challenges, some subsequent HDI-related studies have taken a more solution-oriented approach. Building on studies about collaboration tools for visual and data analytics (McAuley et al., 2011, Mashhadi et al., 2014), Crabtree et al. (2016) propose "The Databox Model" to discuss core research challenges in HDI. They identify issues revolving around personal data discovery, data ownership, data legibility, and data tracking. Even if our case does not concern personal data as applied in the healthcare examples, but rather data about individuals and organisational data, the above concerns are indeed present issues also in our study. Cabitza and Locoro (2017) analyse how HDI can be applied in healthcare and propose a tripartite perspective to personal health data in order to ensure data quality. By distinguishing between primary, secondary and tertiary health data, they argue it could mitigate issues regarding reuse of data and thus differentiate agendas. Koesten et al. (2017) stray from the healthcare domain and analyses people's information seeking behaviour, when searching for new sources of structured data. They propose a framework for human structured data interaction. They identify challenges that occur when people try to find and access data in the context of their daily work activities. Finally, Wilke and Portmann (2016) proposes granular computing as a theoretical, formal and methodological basis for HDI, in order for new systems to support data legibility to a greater extent. They propose information granules as a prerequisite for data legibility.

So far most of the cases discussed focus on the interaction between the individual user and his or her personal data (i.e. health data in the reported-on studies), and how to further improve user-centric ways in which these interactions can take place. In our opinion, it also stresses one of the limitations with the current explanation of HDI: The previous studies make complex and entangled data infrastructures visible, and thus to some degree indicate the complexity of HDI. However, the perspective does not encounter data interactions beyond the individual and his/her personal data. We argue that HDI at this point conceals an often-present level of complexity, as data are often produced, conducted, analysed and used by others than the individual himself in order to maintain and develop services for instance in organisations or governmental agencies. Moreover, considering the adopted and soon to be enforced GDPR, any organisation that controls personal data processing (including collecting, using, storing and disclosing it) is required to demonstrate compliance with the Accountability Principle that aims to ensure that *what* is done with the data and by *whom* is made visible (Regulation, 2016). For this reason, we argue, it is relevant to consider HDI from an organisational perspective in order to better understand the "passive scenarios" that come about when people interact with data. As such, it may be that HDI should not be studied as an isolated discipline, but rather be perceived as an extension to the fields of HCI and CSCW.

2.3 Data as Boundary Objects

HDI-studies have suggested to apply the notion of boundary objects as a means to view and understand how data as an object is embedded in human interactions (Elmqvist, 2011, Crabtree and Mortier, 2015). Building on this idea, we argue it might also be useful to consider data as a boundary object to extend the concept of HDI at an organisational level. According to Star and Grisemer, boundary objects are "*both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites*" (1989, p. 393). The notion of boundary objects has extensively been used within the HCI and CSCW literature to analyse, understand, design and support collaboration (Blomkvist et al., 2015, Lee, 2005, Bødker and Grönvall, 2013). Drawing on Star and Grisemer's ((1989) early insight, we understand boundary objects as artefacts that (to varying degrees) cohere amongst different communities of practice and thus support communication and collaboration across organisational boundaries. In this sense, boundary objects derive from action and are thus objects that people can act with and upon (Star, 2010). Boundary objects are often artefacts, being health data records (Bossen et al., 2014) or a shared web-interface for collaboration (Borchorst et al., 2009). In our case, the collaboration is crafted around data as the boundary object. While data is intangible by nature, the different stakeholders create their own views and extensions that render the data meaningful for them and allow them to interact with the data in a meaningful way. In doing so, data becomes malleable, a tool to work with and collaborate around; a boundary object for translating, or rendering understandable, the needs and situation in and between organisations and their employees.

3. CASE STUDY

This action research case study took place at Industriens Uddannelser, an education secretariat based in Copenhagen (Denmark). The research is part of a larger, on-going, 3-year collaborative action research project between the university and the case organisation. IU is a medium-sized organisation that works to develop educational programs for vocational training and adult vocational training for the industrial sector in Denmark. IU is a self-governing institution but is owned by both employer associations and unions, which means that IU needs to consult and consider the interests of both sides. IU can be seen as a knowledge-broker organisation (Mashhadi et al., 2014), in that the organisation navigates within a large network of cooperation organisations and government agencies. IU makes use of heterogeneous data sources to answer to internal and external knowledge and information needs. A sub-section of IU's data sources constitute the focal point of our study, which we describe below.

To understand our case, it is essential to be introduced to the context in which it exists: Tripartite negotiations between the state and the social partners form the basis of the Danish labour market. This negotiation procedure affects and regulates amongst other things the vocational educational system and the adult vocational training system. IU is responsible for facilitating the collaboration between the social partners around the vocational education and training programmes of the industrial sector. Each vocational education in Denmark is controlled by a Skill Sector Council that consists of representatives from both employer and employee associations. The Skill Sector Council determines the educational framework in collaboration with the Danish Ministry of Education. The framework is then locally implemented at the vocational colleges through governing bodies known as Local Education Committees (LEC). The LEC members consists of representatives from both employer and employee associations. LEC members come from the local industry and have been appointed by their affiliated employer or employee association. It is among IU's administrative tasks to produce, maintain and communicate all relevant information about each individual LEC member and their affiliation as needed. This paper reports on the initial findings from studying data work and interactions related to the collaborative maintenance processes of the LEC member database. This work was undertaken to further inform the development of a new system that efficiently incorporates the current needs for supporting IU and its main stakeholders with the administrative tasks related to LEC members. The current database system uses outdated technology and does not support well the work and role of IU in the related data ecosystem. This result in a number of problematic work-arounds, including the communication of data through excel sheets or e-mail rather than through the intended interfaces.

4. METHOD AND STUDY ACTIVITIES

A medium-sized knowledge-broker organisation such as IU, produces, maintains and uses many different data entities in order to provide their services. Given our action research approach, we involved the case organisation in this decision-making process (Chevalier and Buckles, 2013, Robson, 2002, Van de Ven, 2007). To create a common understanding of the organisation's IT-infrastructure, the first author created a map, which visualized IU's internal IT systems and data flows, as well as external web services and data sources that are necessary to run IU's internal IT systems. The map was developed in collaboration with IU's external IT-developer and provider and it was complemented with inputs from management and employees at IU. Based on this mapping, the management at IU and the authors jointly decided to use the LEC database and its associated interfaces and systems as a starting point for studying how data are produced, used and maintained.

The subsequent data collection took place over the cause of 9 months (March-October 2017). Throughout this period the first author spent approximately 3 days a week at IU "following the data" related to the LEC database. First, central actors in relation to the case were identified together with relevant employees and management at IU. On this basis, the first author conducted about 20 hours of field observations and 15 semi-structured interviews with administrative "data workers", who represented 12 organisations (including education secretariats, employer and employee associations, vocational colleges, and IU's external IT-provider). The observations and interviews focused on how "LEC data" were produced, conducted, analysed, shared, maintained and updated. Moreover, two workshops were conducted with representatives from IU and the three key data providing and receiving cooperation organisations. The workshops lasted approximately two hours each.

The first workshop focused on how the practices concerning data maintenance related to the LEC database could be improved. All representatives worked with LEC members and data about them and were thus central actors to the production, maintenance and updating of data relevant to this case. The workshop roughly followed a 'future workshop' scheme and thus included a problem, fantasy and implementation phase (Jungk and Müllert, 1987). Each phase lasted 30 minutes, leaving 30 minutes for a short introduction and a wrap-up at the end. It was necessary to limit the duration of the workshop in order to gather the relevant representatives from the external organisations at the same time. Indeed, it was crucial that both the employer associations and the unions were equally represented in matters of idea development and the initial decision-making.

Figure 2. *Organisation representatives organize and discuss the "data entity icons" of the current LEC database.*

The second workshop was of a more experimental character (see Figure 2). The purpose of this workshop was to gain insights about how the participants understand the data they produce, conduct, extract, analyse and apply in their everyday work in order to maintain and update the LEC database. To explore this, a set of simple graphical icons that each represented the data entities in the LEC database. All of the workshop participants had very limited knowledge about IT-systems and databases. Thus, the reason for representing the data entities in this way was to enhance the data literacy and thereby make it easier for the participants to relate and understand what a data entity meant in the context of the LEC database. The participants were first asked to remove and/or add data entities (icons) they thought were either redundant or (un)necessary. Next, they were asked to discuss how they thought the data entities were related. During the discussion, they collaboratively organized the icons and drew lines between them to visualize, how the data entities were connected

(Figure 3). The participants decided to draw lines with different colours as a way to represent the different organisations that were represented in the workshop and how these organisations related with each data entity.

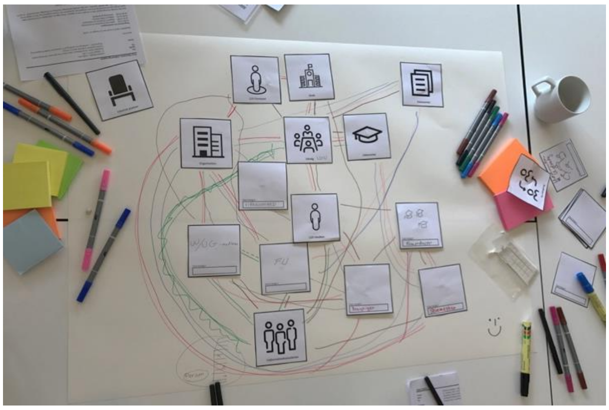


Figure 3. *Final visualisation of how the workshop participants percieve how they relate to each of the data entities in the current LEC database.*

To document the fieldwork, the interviews were audio recorded and transcribed in full. Field notes were conducted during all observations and the workshops were video recorded and later thematically analysed. The transcribed interviews and field notes were used to perform an open coding by the first author. On this basis, the authors collaboratively produced a thematic analysis (Robson, 2002) where our point of departure was IU and how people interact with the LEC data. We followed the flows of producing, maintaining, sharing and using the LEC data at IU's collaboration organisations, vocational colleges and LEC members. We also considered how the data work was articulated in a cross-organisational context in order to maintain the data, and thus joint services.

5. ANALYSIS: COLLABORATIVE MANAGEMENT OF THE LOCAL EDUCATION COMMITTEES DATA

While the daily activities and focus of the LECs are centred on providing advice to vocational colleges that offer vocational education and training, a number of actors are required to appoint the members and to organize the LECs' work. This organizing depends on various data about the LEC members distributed across different organisations. Interacting with data in order to collect, maintain, update and use the data in a cross-organisational context presents a number of collaborative challenges. We elaborate on the observed challenges below.

5.1 Continuous coordination of data production

There are 165 Local Education Committees alone in the industrial sector in Denmark (IU, 2017). The number of LEC members in each LEC vary depending on the size of the related vocational college and the number and size of vocational education programs the committee advices. On average, a LEC is made up of 4-8 committee members that represent both employer and employee associations, and two representatives from the local vocational college. A considerable proportion of the LEC members are active in more than one LEC. It requires careful organisation to keep track of the LECs' members and to make sure that each committee is equally staffed with members from both employer and employee associations. In this context, IU acts as a "neutral" part between the cooperating organisations, and has been trusted with the task to collect, store and maintain all relevant data in the so-called LEC database. However, in order for IU to be able to maintain the data, it is constantly necessary to collaborate with the external stakeholders. As illustrated in Figure 4 (see next page), the LEC database and its data is connected to a large network of internal and external collaborators that contribute and adjust to the process of producing and maintaining the data. In this case, the LEC data constitutes both individual member's data (e.g. name and Civil registration number) and organisational data (e.g. place of employment and which appointing association a member is affiliated with). When data in the LEC database needs to be updated, IU initiates an array of events that includes various actors across organisations. Often data maintenance is needed because a LEC member retires from a LEC, or because an employee/employer association decides to appoint a new LEC member. The processes concerning data maintenance in such cases differentiate slightly across the observed organisations due to organisational culture, constellation and internal IT-systems. Our data shows that these processes often occur as follows (please refer to Figure 4 for a description of what the numbers represents): A LEC member chooses to secede from a committee to which s/he has been appointed. S/he informs a contact person (administrative worker) at the association s/he is affiliated to (1), who initiates internal processes and updates - if existing - internal systems and database(s) (2). Thereafter, the contact person contacts the LEC's presidency at the related vocational college to confirm the decision/information (3). This is documented by filling out different templates (word documents) (4), which are communicated via email to the vocational college's representatives (5). Once this data is produced, it is forwarded to an administrative worker at IU (6), who adds the new data to the LEC database through an interface (7). After the data update, the administrative worker renews the information on IU's website (8) that publicly shows

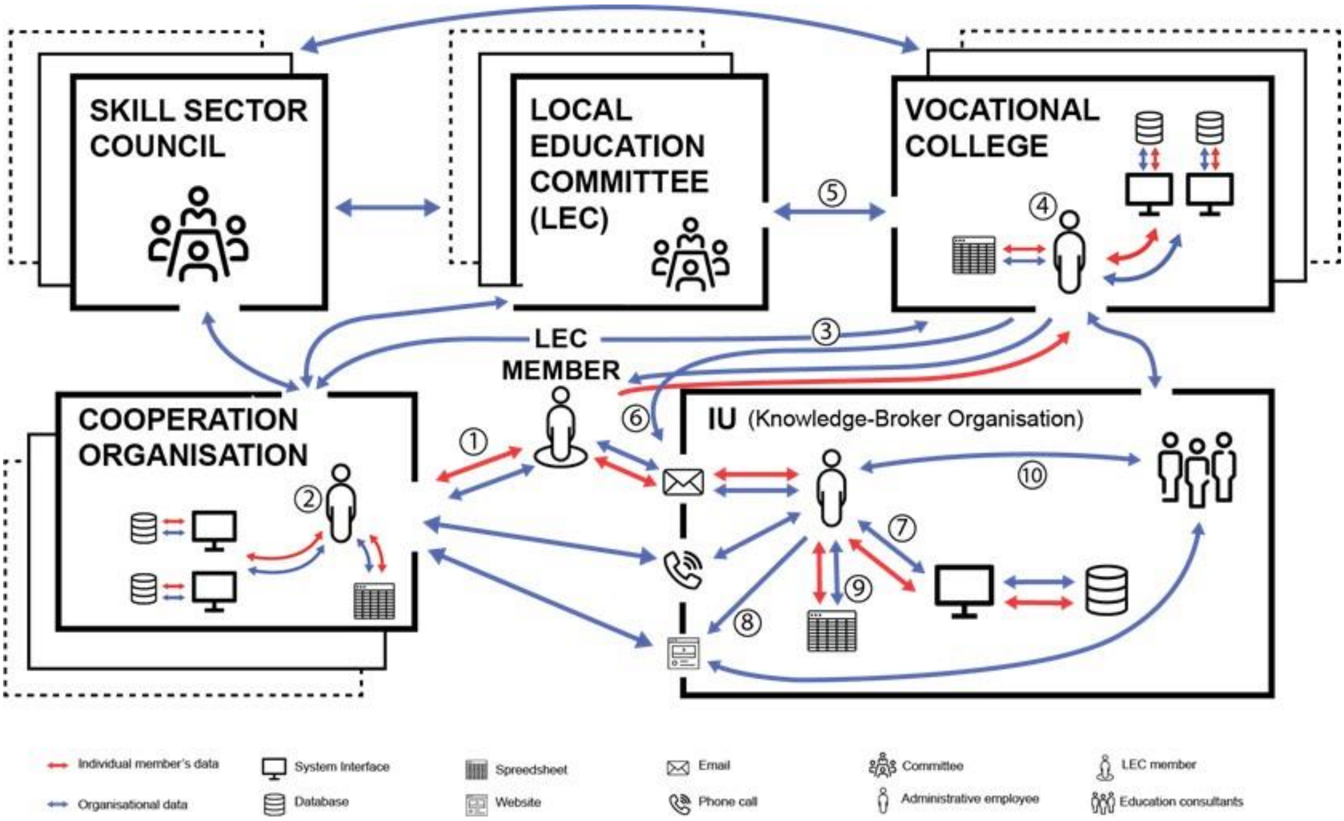


Figure 4. *Collaboration involving data about LECs and LEC members (Please refer to the text regarding the numbers.)*

which representatives are connected to which LEC. Furthermore, the administrative worker at IU also updates an internal spread sheet, which is used to keep track of the LEC members and vacant positions (9). The worker also informs the relevant education consultant at IU about the change (10). If, however, the LEC member chooses to contact IU directly, the flows of data production and maintenance take place in a slightly different order. In this case, there is also a need for even more communication and documentation between the LEC member, IU, the relevant collaboration organisation and the vocational college. The work practices described above might seem frictionless, however, in reality these processes encounter numerous breakdowns that makes the data work highly complex. The breakdowns include: the LEC member fails to notify anyone about him/her seceding from the committee; lack of updating the individual and internally shared spread sheets; the organisations forget to inform IU about new changes, which results in out-dated information, for instance on IU's website and different data sources out of sync, potentially existing within diverse organisations, and finally, in practice these breakdowns can create political imbalance in the LECs, which is required by law to uphold equal parity between the labour market partners.

5.2 Data discrepancies

The majority of the LEC member data stays "the same" for long periods because LEC members are, in general, active for several years. However, people might move and change address or get a new job. These seemingly small changes in the datasets generate continuous strings of actions across organisations in order to maintain and keep the LEC member data accurate and up to date. What became apparent from our fieldwork was how the "LEC data workers" independently had created spread sheets, which were stored on their personal computers and used as a means to keep track of the LEC data that was relevant to their other LEC-related tasks. During fieldwork the first author (Author) observed and recorded how an administrative worker (Admin) who works in the largest employer association and is in charge of managing and maintaining the LEC data, applied workarounds to ease some of her task related to LEC. To illustrate, an excerpt from the fieldwork follows here below:

Author: *How do you keep track of the data?*

Admin: *Well, I get lists from [IU], but I also have a long Excel sheet that I try to keep up to date… but there are only the names, social security number and Department… I don't need the email address or home address, so I have deleted that…*

The quotes exemplify how the LEC data is also tweaked according to the workers other LEC-related tasks. In doing so, the Admin worker creates additional maintenance tasks, as she has to examine and compare "their lists" – the personal spread sheets with LEC data – with the data-lists they receive from IU. The administrative workers across IU, collaboration organisations and vocational colleges are aware that these manual work-practices sometimes result in data discrepancies, meaning that the LEC data held at IU does not align with the data held by a given external collaborator. An administrative worker in one of the collaboration organisations explains how human errors and thus data discrepancies may occur: *"…but then I know the chairman of the LEC, because he is also from [our association], so I just use the opportunity to call to say 'Hi, how is it going and who is it you are going to appoint?'… and there is so much of 'now he is out and he is in instead', so sometimes it [data maintenance] just fails…*" [Administrative worker, union]. Our data indicate that the possibilities for data inconsistency have formed a common understanding across the network of organisations that constitutes IU as the governing body for ensuring data quality and transparency. However, with the current system and data infrastructure, IU is not able to complete this role, which is also a reason as to why a new system is needed. Below, this data responsibility is further elaborated.

5.3 IU's responsibilities as a knowledge-broker

As a result of the possibility for data discrepancies, IU constitutes the main reference point across organisations, vocational colleges, committees, and members. In particular, this is manifested in how the cooperation organisations depend on IU to keep track of the LEC data. An external education consultant from one of the largest employer associations explains his organisation's dependency: "*Well, we rely on IU – that [IU] have a system, a well-functioning system that is… We haven't established a large database for this purpose in-house. Obviously, we feed IU with data about existing and future LEC members, but once we've done that, we sit back safely and count on that [IU] are in control of the data. If we then need to communicate with our LEC representatives, [administrative worker] typically calls [IU's] LEC contact or sends an email, and then we'll get a list from [IU]…"* [Education consultant, union]. The consultant concluded that several tasks in his and his colleagues work are connected to the LEC data, in particular the processes of dismissing or appointing new LEC members. This data dependency is expressed by the majority of the interviewees. An administrative worker explains how she deems the public LEC information on IU's website better than the lists in her organisation's internal system: *"… I also use [IU's] website a lot if I need to see who is a member of a particular LEC… I often use it when in doubt, then I check IU's website because it is updated. I think I use it almost every day…*" [Administrative Worker, vocational college]. All in all, this data dependency establishes IU as a knowledge-broker (Meyer, 2010) that move knowledge (data) around and through this data create connections between e.g. cooperation organisations and LEC members. In other words, IU becomes "the care facilitator" that works (and is expected) to ensure trust and transparency in terms of how data is handled and maintained, and moreover, to facilitate ongoing compromise and collaboration amongst multiple stakeholders (Jackson and Baker, 2004). IU's role as a facilitator of mutual care between the stakeholders also became visible during the second workshop. Through the discussions it became clear for the various stakeholders that they in some cases ascribed value to different kinds of data according to their organisational knowledge interests. To illustrate, the quote below shows how these differences emerged during the discussion (quotes transcribed from the video-recordings of the second workshop):

IU employee: *"… and for you [refers to a specific employer organisation], shouldn't there be a piece that says "company"? I assume it is important for you that it is registered…"*

Representative from the employer organisation replies: "*Yes, indeed. It is very important for us, because we have to distinguish between so many companies. The name of the member [refer to another data entity icon] is not enough…"*

[The IU employee begins to create a new "company"-data icon]

A representative from one of the trade unions: *"… but that is not so important for us. We organize the LEC members based on their social security number in our system…"*

The discussion that was raised by representing data entities as icons at the workshop shows the important role of IU as a knowledge-broker. At the workshop, an IU employee made the other organisations aware of available data leading to further discussions on what data are available and how it can be used either in isolation or combined with other data sources. Through these discussions, the workshop participants became aware of their small, yet significant, differences in interacting with and interpreting data. This further indicate how including data as malleable entity in the workshop let to a mutual understanding of how the 'same' data is understood, used and valued in different ways across organisations.

6. DISCUSSION: HOW CROSS-ORGANISATIONAL DATA MAINTENANCE SHAPE HDI

The majority of the activities that IU and its collaborators perform to keep track of the 165 LECs have data at its core. Thus, in our presented case, data is essential for inter-organisational and cross-organisational collaboration to happen. This "data condition" shapes how people interact both with other people within their own and other's organisations, but also with the actual data.

6.1 The complexity of cross-organisational data maintenance

Based on the LEC case, cross-organisational data maintenance entails arrangements of data work that are dependent on updated and accurate data and, simultaneously, a lot of manual labour, that is people interacting with the data through the different stages of the maintenance process. As shown in the analysis (and illustrated in Figure 4), this constant involvement of various collaborators creates a complex data ecosystem including many potential sources of data updates and correspondent errors. This depicts a difference between the LEC case and previous HDI-studies, as in our case data is used and updated by different people in different contexts. A second noticeable difference is that in the LEC case, the data update has to be confirmed by specific actors within the network, and IU is responsible to assure this confirmation. Drawing on Crabtree and Mortier's (2015) acknowledgement, we argue that the individuated HDI model as proposed by Mortier et al. (2014) is not sufficient from an organisational perspective. Based on the LEC case, we argue that for the concept of HDI to be useful from an organisational perspective, it is necessary to look beyond one single data entity or transaction in isolation. In a cross-organisational context, it is necessary to understand how data are produced, maintained and updated by multiple actors. Thus, we argue it is necessary to expand the notion of HDI in order to consider the wider network of actors, and how they use distributed and shared data.

6.2 Data as boundary object and the role of the knowledge broker

In the following, we consider data as boundary object to further clarify why it is relevant to extend the notion of HDI so it becomes useful from an organisational perspective.

As outlined in the Related Work, previous HDI-studies have proposed to consider data as boundary objects (Crabtree and Mortier, 2015). More specifically, Crabtree and Mortier emphasizes how "*human data interaction turns upon 'a mutual modus operandi' involving 'communications' and 'translations' that order the 'flow' of information through 'networks' of participants'. This, in turn, creates an 'ecology' of collaboration in which data interaction becomes stable. As stable entities boundary objects inhabit 'several intersecting worlds'… and meet the information requirements of each.*" (2015, p. 8). This is also evident in our case, where the organisational data likewise constitutes boundary objects. Through our work with "data entity icons", it became visible that for example, "the affiliation information" of each LEC member is an essential data entity that is needed by all involved organisations to perform the majority of tasks related to the work of the LEC as well as to data maintenance of the LEC database.

The affiliation data about a LEC member might be seen as a boundary object, as it enables a given organisation to verify the individual member, while at IU it functions as a measurement to ensure that in each LEC employer and employee association are equally represented. Furthermore, for the individual member the membership data is a referral to the organisation to whom s/he belong. Finally, for the local vocational college it resembles the local Industry and a training location for student members. Thus, for IU and its collaborative partners, data becomes a boundary object that goes beyond 'several intersecting worlds' and facilitates the cross-organisational collaboration that is necessary for these actors to maintain and provide the jointly needed LEC-related services. When the focus expands to organisational data and data supporting cross-organisational collaboration, its use and management is not any longer the responsibility of an individual but rather shared organisational responsibility. In previous HDI studies (Crabtree et al., 2016, Crabtree and Mortier, 2015), the individual both is the origin of, and (ideally should) acts, as the broker of his or her personal (health) data. In the LEC case however, IU is assigned the role as the knowledge broker: IU keeps the reference version of the data, and IU is responsible to communicate the right information at the right time to the right persons and organisations.

Understanding data as a malleable entity makes visible how specific types of data are understood used and prioritized across organisational boundaries in different contexts. From an organisational perspective, it is therefore necessary to extend the notion of HDI for it to comprehend the complexity, which exists when people interact with data in a cross-organisational context. As the data management takes place in collaboration between organisations, not only the need to agree on responsibilities but also the requirements for data quality and transparency in data management becomes core issues of the distributed data management. These aspects will be further discussed below.

6.3 Data quality and transparency as important dimensions

The projects cited as related work mainly focused on a single source or a single consumer of data. Given the individuated focus, these cases do not render data quality and transparency of data management visible as relevant issues for the individual. They have however become visible as core aspects in the LEC case.

As presented, data quality becomes a requirement for cross-organisational collaboration: it is necessary for employees of the different organisations involved to be able to trust the accuracy of the data they receive, apply and reuse. What furthermore becomes visible from considering the wider data eco-system, is that in this case IU holds a dual role: On the one hand IU constitutes a knowledge-broker and is thus the main reference point in terms of receiving accurate LEC data. On the other hand, and given IU's role as a knowledge-broker, the data eco-system has also established IU as a "data accountability mechanism", which is expressed in the way all data-updates (ideally) have to be confirmed by a qualified worker at IU in order to be considered accountable.

A second dimension that becomes visible through the LEC case is the need for transparency. This need is expressed in two ways: First, transparency is necessary in order to understand how the data came about; who produced it (which LEC or individual LEC member), who documented it (which affiliation), and whether it has been confirmed by the knowledge-broker (i.e. IU). Second, transparency is a necessary quality in a cross-organisational context to visualize who has access to what data and for which purpose. As a knowledge-broker, IU navigates in a large network of actors that has to be treated equally so that neither the employee or employer associations are in the majority in the LECs. Here, data plays a central role in constantly supporting equality within the cross-organisational context, which further can promote continuous collaboration.

7. CONCLUSION

This paper shows that to 'interact' with data that is produced, collected, used, maintained and updated by many different stakeholders across organisations is not simply a question of providing infrastructure. On the contrary, this study shows that in this case, where no single stakeholder is the owner nor in control of the data, cross-organisational collaboration is necessary in order to perform the data work that support central data-based services.

What surprised us when conducting, and later analysing the empirical research is the large number of actors involved in the maintenance and usage of even one of the smaller databases at IU. It was also surprising how distributed the data actually was across different organisations. With this setup, the level of complexity increases in order to maintain the data. The reason why such complexities have not been broader discussed in a HDI context might have been that previous research focused mainly on the interaction between the individual user and his or her personal data. However, taking a cross-organisational perspective in other domains, such as healthcare (where HDI have previously been studied), might reveal complexities similar to what we have identified in our case. With this in mind, we have proposed to extend the notion of HDI as a way to include the level of complexity which exists when multiple stakeholders interact with the same data.

In our depict case, IU acted as a knowledge-broker taking care of the data that constituted a boundary object between organisations, stakeholders and tasks. Our analysis shows how such a knowledge-broker organisation interacts through and about the data with the different stakeholders in order to manage the update of data originating in different places in the network. Moreover, as the knowledge-broker organisation within this complex network, IU also becomes a central "care facilitator" that is expected to ensure mutual trust – through data quality and transparency – in order to nurture on-going, data-based cross-organisational collaboration. In such a complex collaborative network with partly adversary interests, data quality and transparency of data management thus become visible as core issues: data and data management need to be accountable for all actors of the collaboration.

The empirical work reported in this paper is part of the preparation of revising the IT support for managing the LECs and thereby the many members involved. By perceiving data as a malleable entity, we argue, designing with data becomes part of designing the future functionality. Our results suggest that when data is made visible, workshop participants can discuss with and through the data allowing them to consider and design data-related aspects of a future system; for instance, how data is prioritised and handed across organisational boundaries. Moreover, from an HDI perspective, our study suggests how users (beyond the individual, and not necessarily IT-experts) can be informed and involved in the design of a future system's data collection, processing and analysis of personal and organisational data, thereby adding a layer of transparency and accountability already in the initial design phase of a new IT-system.

REFERENCES

Blomkvist, J. K., Persson, J. & Åberg, J. 2015. Communication through Boundary Objects in Distributed Agile Teams. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Seoul, Republic of Korea: ACM.

Borchorst, N. G., Bødker, S. & Zander, P.-O. 2009. The boundaries of participatory citizenship. In: Wagner, I., Tellioğlu, H., Balka, E., Simone, C. & Ciolfi, L. (eds.) ECSCW 2009. Springer London.

Bossen, C., Jensen, L. G. & Udsen, F. W. 2014. Boundary-Object Trimming: On the Invisibility of Medical Secretaries' Care of Records in Healthcare Infrastructures. Comput. Supported Coop. Work, 23, 75-110.

Bossen, C., Pine, K., Elllingsen, G. & Cabitza, F. 2016. Data-work in Healthcare: The New Work Ecologies of Healthcare Infrastructures. Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion. San Francisco, California, USA: ACM.

Bowker, G. C. 2000. Biodiversity Datadiversity. Social Studies of Science, 30, 643-683.

Bødker, S. & Grönvall, E. 2013. Calendars: Time Coordination and Overview in Families and Beyond. In: Bertelsen, O. W., Ciolfi, L., Grasso, M. A. & Papadopoulos, G. A. (eds.) ECSCW 2013: Proceedings of the 13th European Conference on Computer Supported Cooperative Work, 21-25 September 2013, Paphos, Cyprus. Springer London.

Cabitza, F. & Locoro, A. 2017. Human-Data Interaction in Healthcare. Smart Technology Applications in Business Environments. IGI Global.

Chevalier, J. M. & Buckles, D. 2013. Participatory action research: Theory and methods for engaged inquiry, Routledge.

Crabtree, A., Lodge, T., Colley, J., Greenhalgh, C., Mortier, R. & Haddadi, H. 2016. Enabling the new economic actor: data protection, the digital economy, and the Databox. Personal Ubiquitous Comput., 20, 947-957.

Crabtree, A. & Mortier, R. Human data interaction: historical lessons from social studies and CSCW. ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work, 19-23 September 2015, Oslo, Norway, 2015. Springer, 3-21.

Elmqvist, N. Embodied human-data interaction. ACM CHI 2011 Workshop "Embodied Interaction: Theory and Practice in HCI, 2011. 104-107.

Elsden, C., Kirk, D. S. & Durrant, A. C. 2016. A Quantified Past: Toward Design for Remembering With Personal Informatics. Human–Computer Interaction, 31, 518-557.

Eugdpr.Org. 2018. GDPR Portal: Site Overview [Online]. Available: https://www.eugdpr.org/ [Accessed 29 March 2018].

Fearon, D. S. 2017. Expanding Data Management Service Boundaries at Johns Hopkins University. University of Massachusetts and New England Area Librarian e-Science Symposium.

Fischer, J. E., Costanza, E., Ramchurn, S. D., Colley, J. & Rodden, T. 2014. Energy advisors at work: charity work practices to support people in fuel poverty. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. Seattle, Washington: ACM.

Futrelle, J., Gaynor, J., Plutchak, J., Myers, J. D., Mcgrath, R. E., Bajcsy, P., Kastner, J., Kotwani, K., Lee, J. S., Marini, L., Kooper, R., Mclaren, T. & Liu, Y. 2011. Semantic middleware for e-Science knowledge spaces. Concurrency and Computation: Practice and Experience, 23, 2107-2117.

Haddadi, H., Mortier, R., Mcauley, D. & Crowcroft, J. 2013. Human-data interaction. Computer Laboratory

Hogan, M., Ojo, A., Harney, O., Ruijer, E., Meijer, A., Andriessen, J., Pardijs, M., Boscolo, P., Palmisano, E., Satta, M., Groff, J., Baker, M., Détienne, F., Porwol, L., Scarano, V. & Malandrino, D. 2017. Governance, Transparency and the Collaborative Design of Open Data Collaboration Platforms: Understanding Barriers, Options, and Needs. In: Ojo, A. & Millard, J. (eds.) Government 3.0 – Next Generation Government Technology Infrastructure and Services: Roadmaps, Enabling Technologies & Challenges. Cham: Springer International Publishing.

Iu. 2017. IU-Overview [Online]. Available: https://iu.dk/luu/alfabetiskoversigtoverudvalg/ [Accessed 18 Sept 2017].

Jackson, S. J. & Baker, K. S. Ecological design, collaborative care, and ocean informatics. PDC 04 Proceedings of the eighth conference on Participatory design: Artful integration: interweaving media, materials and practice, July 27 - 31 2004 Toronto, Ontario, Canada. 64-67.

Jungk, R. & Müllert, N. R. 1987. Future workshops How to create desirable futures. Institute for Social Inventions, London.

Karasti, H. 2014. Infrastructuring in participatory design. Proceedings of the 13th Participatory

Design Conference: Research Papers - Volume 1. Windhoek, Namibia: ACM.

Karasti, H. & Baker, K. S. 2008. Digital data practices and the long term ecological research program growing global. International Journal of Digital Curation, 3.

Kitchin, R. 2014. The data revolution: Big data, open data, data infrastructures and their consequences, Sage.

Koesten, L. M., Kacprzak, E., Tennison, J. F. A. & Simperl, E. 2017. The Trials and Tribulations of Working with Structured Data: -a Study on Information Seeking Behaviour. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. Denver, Colorado, USA: ACM.

Lee, C. P. 2005. Between Chaos and Routine: Boundary Negotiating Artifacts in Collaboration. In: Gellersen, H., Schmidt, K., Beaudouin-Lafon, M. & Mackay, W. (eds.) ECSCW 2005. Springer Netherlands.

Mashhadi, A., Kawsar, F. & Acer, U. G. Human Data Interaction in IoT: The ownership aspect. 2014 IEEE World Forum on Internet of Things (WF-IoT), 6-8 March 2014 2014. 159-162.

Mcauley, D., Mortier, R. & Goulding, J. The Dataware manifesto. 2011 Third International Conference on Communication Systems and Networks (COMSNETS 2011), 4-8 Jan. 2011 2011. 1-6.

Meyer, M. 2010. The Rise of the Knowledge Broker. Science Communication, 32, 118-127.

Mortier, R., Haddadi, H., Henderson, T., Mcauley, D. & Crowcroft, J. 2014. Human-Data Interaction: The Human Face of the Data-Driven Society. Social Science Research Network.

Pipek, V. & Wulf, V. 2009. Infrastructuring: Toward an integrated perspective on the design and use of information technology. Journal of the Association for Information Systems, 10, 1.

Regulation, G. D. P. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46. Official Journal of the European Union (OJ), 59, 1-88.

Robson, C. 2002. Real world research: A resource for social scientists and practitioner-researchers (Vol. 2): Blackwell Oxford.

Seravalli, A. 2012. Infrastructuring for opening production, from participatory design to participatory making? Proceedings of the 12th Participatory Design Conference: Exploratory Papers, Workshop Descriptions, Industry Cases - Volume 2. Roskilde, Denmark: ACM.

Star, S. L. 2010. This is Not a Boundary Object: Reflections on the Origin of a Concept. Science, Technology, & Human Values, 35, 601-617.

Star, S. L. & Griesemer, J. R. 1989. Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. Social Studies of Science, 19, 387-420.

Van De Ven, A. H. 2007. Engaged scholarship: A guide for organisational and social research, Oxford University Press on Demand.

Wilke, G. & Portmann, E. 2016. Granular computing as a basis of human–data interaction: a cognitive cities use case. Granular Computing, 1, 181-197.

Publication 3

Seidelin, C. Dittrich, Y., Grönvall, E. **Foregrounding Data in Co-design - An Exploration of how Data may Become an Object of Design**. [Resubmitted to the International Journal of Human Computer Studies]

Foregrounding Data in Co-design

– An Exploration of how Data may Become an Object of Design

Cathrine Seidelin,[1] Yvonne Dittrich,[1] Erik Grönvall[1]
[1]IT University of Copenhagen, Denmark

Abstract

The rise of Big Data and data science has prompted a focus on data as an essential component of making and innovating data-based services. Traditionally, however, digital data has not been object to co-design as have other physical or functional dimensions of IT application design. This is problematic, because it hinders domain experts who are not IT professionals from taking part in the discussions and design of the data-based services they use and provide. We argue that to address this challenge, it is necessary to empower such domain experts to be able to consider data as an object of design, so they may contribute their expertise to the design of data-based services and their underlying data structures. This paper describes how data may be foregrounded as an object of design for domain experts participating in collaborative design situations. We present a detailed interaction analysis of video recordings of three collaborative design workshops, in which we propose a form of data notation and two data representations. We find that data may become an object of design for domain experts when tangible and flexible representations of data are used. Based on our findings, we discuss five lessons learned for foregrounding data in co-design. Together, these provide practical insights for future work.

KEYWORDS: Data, Data as input for design, Data as an object of design, Co-design, Participatory Design

1. Introduction

This paper explores how data may be represented through co-design in ways that enable professionals outside the field of computer and data science to make sense of, and critically consider, data and data schemata in design. This focus stems from the growing need to make sense of the ever-increasing availability of data (Russell et al. 2018). Developments such as Big Data and Artificial Intelligence have made it necessary for organizations to become more ‘data-driven’, in order to utilize and benefit from ‘the data economy’ (OECD 2019; European Commission 2020). However, employing tools for data aggregation and analytics often requires highly specialized skills and knowledge. This means that core stakeholders and users are excluded from the design discussions that determine what kinds of data are necessary, which existing data sources should be included or excluded, and which data sources should be explored to innovate and design data-based services (Boyd and Crawford 2012; King, Churchill, and Tan 2017). The growing use of digital technologies and data are transforming societies with significant implications for how organizations provide and develop their services (OECD 2019; Bright et al. 2019). The increasing digitalization of services is expected to improve not only the efficiency of service production, but also service quality and service provision in organizations (Sangiorgi and Prendiville 2017; Malmberg 2017). Thus, as data becomes a defining basis for many (digital) services, it becomes more and more important to critically consider the underlying data and data schemata that partly shape these services, which affect our everyday lives to various degree. Therefore, it is necessary to develop pathways that enable domain experts who are not IT professionals to engage with data and data-structure design.

This study addresses the need for tools that can enable domain experts to participate in design with and of data and data-structures. We investigate ways to foreground data in collaborative design situations, which enable domain experts to consider data and data structures as something that may be designed, rather than as a ‘given’ that only informs the design of a service, for example. Previous research has argued that data is intangible in nature, but that it may be represented in a number of ways, which emphasize different aspects of the data (Kitchin 2014). The way data is represented affects how people make sense of, and engage with it. Therefore, the representation of data frames what we may say, how we may say it, and how we eventually act, based on this (Boyd and Crawford 2012; Gitelman 2013; Kitchin 2014). More recently, these critical perspectives on data have been brought into the design field to demonstrate how the production, collection, and ongoing use of data may also be considered design activities (Feinberg 2017; Speed and Oberlander 2016). This paper builds on the perspective that data is designed, and data collection itself is a design activity (Feinberg 2017). The aim of this paper is to expand this perspective by demonstrating *how* data may become an object of design that enables domain experts to think about data and data schemata in collaborative design situations. Thus, the paper contributes to the emerging body of work that explores how data may become more accessible to a broader audience, to bridge the growing gap between people who can work effectively with data, and people who cannot (D’Ignazio 2017)

This paper is based on a study at a medium-sized service organization based in Copenhagen, Denmark. The main purpose of the organization, Industriens Uddannelser (English: ‘The Education Secretariat for Industry’, hereafter ‘IU’), is to coordinate processes related to maintaining and developing vocational education programmes and courses for continuing education in the industrial sector. IU exists in and navigates a large network of stakeholders in order to provide its core services. As part of a larger, ongoing project, this study focuses on the redesign of an existing data-based service, which revolved around an existing IT system and its one connected database. This paper builds primarily on three collaborative design workshops from the initial stage of this redesign process. We do so by presenting a study in which we analyse interactions among various stakeholders (employees and managers from various small and medium-sized organizations) during three collaborative design workshops. During these collaborative workshops we also explored various ways of making data a more explicit part of the design process. We draw on the established field of participatory design (PD), which is understood as a design discipline that aims for direct user participation in design processes (Kensing and Blomberg 1998). Moreover, we use and adapt representations from service design, to guide the design of the services related to the existing IT system. Therefore, we understand co-design in a broad sense that refers to a general concept for collaboration surrounding design, involving a number of people (Zamenopoulos and Alexiou 2018).

The paper proceeds as follows. In section 2 we present our related work by questioning why it is relevant to foreground data in design, what it means to design with data, and how to create paths to foregrounding data in design. In section 3, we describe our research setting. This is followed by section 4, where we elaborate on our methodology, the three collaborative design workshops, and our analysis of the empirical material. In section 5, we present our three main findings. (1) Our proposed data notation makes it possible for the domain experts involved to discuss data and to establish a mutual understanding about the data among several stakeholders. (2) The domain experts are able to distinguish between data and the real-life situation it models through the use of notation. (3) Notation may support domain experts' understanding of data as an object of design. In section 6, we discuss our findings and present five lessons learned from our explorations of ways to foreground data in design. Finally, section 7 presents the conclusion.

2. Related work: Why design *with* data?

This section presents the theoretical foundation for our exploration of how data may become an object of design in collaborative design processes. First, we briefly address previous research that has focused on rethinking the role data plays, and how to represent data in cooperative and participatory design. We refer to this as using data as input for design. Then, we present Feinberg's (2017) design perspective on data, which we use as a lens to examine what it means to design with data. On this basis, we argue it is important to also be able to empower domain experts to consider data as an object of design in some collaborative design situations. Finally, by referring to literature on participatory design and service design, we question how to create paths to making data an object of design for domain experts.

2.1 Why should data be a foregrounded element of design?

We begin by questioning the relevance of foregrounding data in design to position our research in relation to prior work. For thousands of years, mankind has experimented with physical representations as means to make sense of abstract data (Jansen et al. 2015). For example, the Incas of ancient Peru developed a system for transmitting information, which was based on spatial arrays of coloured, knotted cords (Ascher and Ascher 1997). Today, designers and artists increasingly use digital data to express meaning *in* data, as well as to convey messages *through* data, for instance, by creating data sculptures and digital artworks (Zhao and Moere 2008; C. Jordan 2009). In research, several strands of work aim to better understand how to meaningfully visualize and represent data and information. For example, Information Visualization and Data Physicalization have developed as research areas that examine how to make appropriate conversions from digital data to human-readable, and even tangible representations (Jansen et al. 2015; Dove and Jones 2014b; Goodwin et al. 2013). In the context of design, researchers have argued for the relevance and usefulness of (big) data in Human Computer Interaction (HCI) research and user interface design, while emphasizing the subjective character of data collection and cleaning (Boyd and Crawford 2011; Fisher et al. 2012). Previous studies have explored how Information Visualization may be used in the front end of design projects as tools that make quantitative data more accessible and engaging for workshop participants (Dove and Jones 2014b; Dove 2015). Others have examined how designers may incorporate a data science practices in their design process (Kun, Mulder, and Kortuem 2018; Kun et al. 2019). Importantly, HCI research has begun to identify the creation of data as design activity per se (Feinberg et al. 2017; Feinberg 2017). Generally, this prior work suggests and demonstrates the relevance of representing data in design to a greater extent, because it can help a broader audience to consider and engage with ‘the data’ that underpins more and more of the IT services we use every day.

Another reason for making data an explicit object of IT design lies in a critical perspective on data, which has developed in the fields of Science and Technology Studies (STS) and Computer Supported Cooperative Work (CSCW), raising the awareness of data as constructed, fundamentally contextual, and interpretive (Gitelman 2013, Ribes 2013). These critical perspectives emphasize that data-production is not a neutral act (Rosenberg 2013; Markham 2013; Kitchin and Lauriault 2018). On the contrary, every decision made about data before the data is collected supports the possibility of certain types of actions, while simultaneously restricting others (Sanches and Brown 2018). Decisions about what should and should not be measured, and which aspects of the data should be made visible, are choices that affect the social world to various extents, and therefore should not be underestimated (Churchill 2012; Ribes and Jackson 2013; Gray, Gerlitz, and Bounegru 2018). These decisions result in the schemata that describe the structure of the data that underpins the computation and interaction. For example, Sanches and Brown (2018, 3) state that *even before data is inscribed in databases, the work of producing data can have profound effects in the social world.* Other researchers have addressed this topic by proposing the concept of Human-Data Interaction, exploring the role of data in Smart Cities, examining how accountability is established through the use of IoT data, and proposing a research agenda for Human-Centred Data Science (Crabtree and Mortier 2015; McMillan et al. 2016; Fischer et al. 2017; Aragon et al. 2016). As such, critical perspectives on data emphasize that one should not take data and data schemata for granted. Rather, it is important to recognise that the production and framing of data is influenced by preceding decisions and assumptions. By being more aware of this condition, data and data structures can be explored differently in design (Feinberg 2017).

This section sheds light on two important reasons that data should be a foregrounded element of design. Our review of previous research shows that there is a growing body of work that indicates the need to make data and data structures visible in ways that enable people to critically consider data in a design context, and to actually design with data (Dove 2015; Dove and Jones 2014b; Speed and Oberlander 2016; Kun, Mulder, and Kortuem 2018). However, previous research primarily reports on exploration of design with data when data is used as *input for design*. For example, Dove and Jones (2014b) included energy data in a service design workshop that aimed to *generate ideas for a new service that utilise data generated by smart meters* (p. 283). They report how the workshop participants were introduced to data that *represented seven days' energy use for one household, based on a selection of possible consumption patterns rather than particular demographic factors* (Dove and Jones 2014b, 283). Thus, in this case the focus is on the design of the service, not on the design of the data used for the service. In other words, in this case data is used as input to inform the service design, rather than to consider data and its schemata as an integrated part of the service design. This exemplifies how so far, research has primarily explored data as something that may inform the design process (data as input for design), for example, of a digital service. These efforts are valuable steps towards foregrounding data in design. However, to consider data and its schemata as an integrated part of design practices, we argue it is also necessary to realize whether a given method can also expose *data as an object of design*: data as something that may be itself be designed. To develop this discussion, we present Feinberg's (2017) design perspective on data in the next section.

2.2 What constitutes a design perspective on data?

We apply Feinberg's (2017) design perspective on data as a lens to examine the question of what it means to design with data in a collaborative design context. Feinberg (2017) emphasizes data as a core component of today's information technologies, and points out that it is important to consider data and its schemata as an integrated part of the design. Feinberg (2017) proposes considering '*data collection as a multi-layered set of interlocking design activities*', and argues that the subsequent use of data is a continuation of these activities. This perspective emphasizes that from the initial decisions of choosing what to measure and how as well as what not to measure may be viewed as design activities (Feinberg 2017). In her work, Feinberg demonstrates that data collection is a result of design decisions, which she divides into three stages: conceptual infrastructure, collection processes, and aggregation processes.

Below, we use Google maps to exemplify Feinberg's perspective on data. Google Maps is a well-known example of a data-based service used in Western societies (Dittrich, Seidelin, and Grönvall 2018). Google Maps constitutes a data infrastructure, where data is not only used to store and retrieve information about specific persons, facts, or situations, but also, the data input itself helps to *create* the service on an ongoing basis. Feinberg argues data infrastructures such as Google Maps generate 'a set of conditions under which design after design can occur' (2017, 2956). In other words, the pre-set values (e.g. how you describe an address) are 'data decisions' that generate Google Maps as a service, and constitute an infrastructure through which data is created and collected. Moreover, this data infrastructure is continuously improved – or 'redesigned' by the users, who use the service by describing their wishes/needs (e.g. to find a certain location) through attributes that are controlled by the 'data decisions'. Feinberg (2017) refers to the accumulation of these preceding data decisions as the service's *conceptual infrastructure* (from a computer science perspective, they are captured as data schemata). It is through this conceptual infrastructure that ongoing *collection* can occur. Feinberg (2017) emphasizes that the data infrastructures can restrict the ways in which data are collected, which may further constrain the intended use-design of the data. In the context of Google Maps, these collection processes are expressed in the ways the service calculates and times various routes for the individual user. Finally, Feinberg (2017) refers to *aggregation processes*, when describing the accumulating number of acts of independent collection that, combined, can determine an 'outcome' of a data-based service. For example, this occurs when Google Maps calculates and suggests the fastest or shortest route, depending on the means of transportation and recent traffic information.

Feinberg's (2017) perspective on data has a technical underpinning that may resemble extract–load–transform processes of traditional database programming. Nevertheless, this perspective is different in that it does not take the extraction of data for granted. Overall, Feinberg's (2017) perspective on data sheds a light on the necessity of considering the preceding data decisions that feed into the definition of a data infrastructure (data schemata) that will eventually underpin data-based services.

What may be gathered from Feinberg's (2017) work is that data may constitute an entity that has been designed, created, and exists in the world, and data may also be a flexible object that may be altered during its existence. Thus, her work underlines the importance of considering data a malleable construct that may be designed. In this sense, Feinberg's idea about data design goes beyond the research where data is understood and used as input for design. As an alternative, she proposes rethinking how we design data infrastructures, and suggests a more reflective and/or critical design *of* data.
However, Feinberg's work does not (yet) address how to make data object to design. Therefore, to develop this area of research, this study has explored how data may become object to design in a collaborative design context. In the next section, we elaborate on our considerations on how to create pathways that can support domain experts' participation in the design of data and data structures.

2.3 How does one create pathways that foreground data as an object of design?

Researchers have called for more data-conscious design that makes it possible to explicitly understand data and data practices in ways that are more attuned to design processes (Churchill 2012; Speed and Oberlander 2016; Feinberg 2017; Kun, Mulder, and Kortuem 2018). Therefore, this section considers how we might foreground data as an element of design that resonates with domain experts, who are not IT-professionals. D'Ignazio proposes the notion of creative data literacy, and suggests that we need to create 'pathways towards data which do not come from technical fields' (D'Ignazio 2017, 7). She emphasizes that data literacy is not simply about acquiring technical skills, but also the emancipation bestowed by literacy. This emphasizes that, for the world to become fairer, it is not enough to teach people how to read a table, for instance. It is necessary to teach people how to use the table (D'Ignazio 2017), and – as we will argue – discuss how the table should look and what kind of data it should be able to capture. The increasing use of data and the continuous growth of the field of data science has generated new opportunities for applying new tools to aggregate, combine, and visualize data (Cao 2017). However, many of these tools require specialized IT skill sets that may act as a barrier, and prevent people and organizations from designing with data. Moreover, the increasing production, collection and aggregation of digital data simultaneously entangles and makes organizations dependent on certain data sources (Seidelin, Grönvall, and Dittrich 2018). This is interesting when we consider how organizations work with data, and more or less explicitly use and collaborate around (the same) data (Seidelin, Grönvall, and Dittrich 2018). This indicates a need to not only foreground data in design, but to foreground it in ways that make sense to multiple stakeholders, and represent 'multiple voices' (Suchman 1993; 2002). On this basis, we propose participatory design as a pathway-generating approach to foregrounding data as an object of design in collaborative design contexts. We elaborate on our proposal in the following subsection.

2.3.1 Participatory design as a pathway-generating approach

Participatory design is a design discipline that aims to elicit direct user participation in project definition and design processes (Kensing and Blomberg 1998; Simonsen and Robertson 2012). This strand of research introduced approaches that involve users in the design process, and aims to develop ways for users to articulate their needs and influence decision-making, in an innovation project, for example (Buur and Larsen 2010). To some extent, participatory design is about facilitating a process where different stakeholders can determine, and articulate their needs, and through that process make the design fit their needs, rather than adjusting their needs to fit the design. Through these objectives, participatory design emphasizes the situated expertise of various stakeholders, and how to establish collaboration among these actors (Buur and Larsen 2010).

Participatory design applies a range of methods and techniques that support engagement and ways to articulate and consider multiple voices in a design process. Collaborative design workshops and methods often involve end-users – whether older adults, children, workers, or people with a specific interest such as bird-watching – in the design of novel interactive modalities and interfaces for interacting with technology (Bossen and Grönvall 2015; Guha et al. 2004; Hiniker, Sobel, and Lee 2017; Still et al. 2004; Lindsay et al. 2012). Many co-design tools take an Interaction Design stance, and are oriented to Interface design, to map out communication flows and to identify key stakeholders in a situation (Bossen and Grönvall 2015; Halse et al. 2010; Kristensen, Kyng, and Palen 2006). Traditional PD methods have addressed data and data schemata indirectly, for example, in interface mock-ups where data is represented through the elements that provide the functionalities to manipulate the data. This means that for participatory design to be a useful pathway that helps to foreground data as an object of design in collaborative design contexts, it is necessary to develop data notation that supports diverse stakeholders in articulating their needs and use of their shared data to guide the design. To do so, we need to both unpack and make tangible the existing data structures, in order to enable domain experts to critique and (re-)design them. To this end, we wanted to explore methods that might enable domain experts to explicitly design with data in a collaborative workshop setting in a particular context. Our design and evaluation of data notation were inspired and guided by Engeström's concept of expansive visibilization (Engeström 1999), which emphasizes the need for representations that establish new design spaces. We describe the notion of expansive visibilization and our use of the concept in the following subsection.

2.3.2 Expansive Visibilization as a tool for evaluating data notation

We adopt the concept of expansive visibilization (Engeström 1999) as an analytical lens for examining the extent to which data representations support domain experts in their critical evaluation of the current situation, and their ability to explore possibilities for change and its design implications. Engeström (1999) introduced the notion of expansive visibilization as a way to conceptualize a representation's capacity to give rise to discussion, questioning, and critical reflection of the status quo. Thus, Engeström (1999) aimed to incorporate a developmental aspect when analysing (primarily) transformation at work. He argued that such activities are often represented in a two-dimensional manner, which he refers to as the linear and socio-spatial dimensions. Engeström (1999) exemplifies the linear dimension of work by referencing Business Process Modelling, which aims to represent the processes of an organization as a temporal series of activities with the purpose of optimizing these processes. This approach often results in a rigorous description. Engeström (1999) emphasizes that linear representations of work are often very limited in terms of generating discussion and allowing people to question the processes. According to Engeström, the socio-spatial dimension goes beyond the linear dimension of representation. The socio-spatial dimension refers to a way of representing work that encounters discourse and situated actions with and without artefacts, for example. This dimension challenges the linear representations by including a greater level of complexity, and by taking into account how processes are actually enacted collaboratively in time and space. Engeström (1999) emphasizes that the socio-spatial dimension fails to uncover how processes and their transformations come about. Therefore, Engeström (1999) proposes a third, developmental dimension that aims to frame the linear and socio-spatial dimensions in a broader perspective, to expand representation of work, and in this way support people's discussion of, and critical reflection on, work activities. On this basis, we apply the notion of expansive visibilization as a lens to characterize and evaluate ways of foregrounding data in design. Specifically, we use the concept to consider how and in which way(s) data representations create opportunities for design.

2.4 Summary of related work

The research presented in this section suggests that data may be, and is used as input to inform collaborative design processes (Dove and Jones 2014b; 2014a; Dove 2015; Kun, Mulder, and Kortuem 2018). However, owing to organizations' increasing dependence on various data sources, the related work also reveals a need to foreground data in ways that allow multiple stakeholders to *co-design data*. Applying Feinberg's (2017) design perspective on data, we emphasize that data may constitute two 'things'; it may be an entity that has been designed, and an entity that may be designed. We argue that it is important to be sensitive to this, to understand how different forms of notation support design with or of data, either as input for design or as an object of design. This sensitivity is essential when designing data-based services used by a number of stakeholders in different contexts and with different data-related practices, because the design of such a service implies data design. As a first step in exploring how this may be done, we adopted a participatory design approach, to build on traditional co-design methods, in order to explore how notation may support domain experts' understanding of data as an object of design. We also propose using the concept of expansive visibilization as an analytical tool to characterize and evaluate the applied notation.

3. Research setting and research focus

This section presents the research setting in which our study took place, and elaborates on the emergence of the research focus. The research was carried out at Industriens Uddannelser (English: Education secretariat for industry, hereafter the acronym IU is used), which is an education secretariat based in Copenhagen, Denmark. IU develops and ensures the quality of vocational education and adult vocational training programmes in the industrial sector. IU is a main driver for ensuring the continuous development and upgrading of the qualifications of Danish industry (IU 2019). Thus, maintaining and developing these educational programmes constitutes IU's core services. In doing this, IU navigates a highly complex network of stakeholders, including ministries, governmental agencies, trade unions, employer associations, vocational colleges, various councils and committees, and other education secretariats. To navigate and eventually provide its services in this complex setting, IU is dependent on heterogeneous data sources, and therefore data is at the core of the internal work, and when IU engages in its many collaborative activities with external stakeholders. What constitutes data in this context varies. A few examples include quantitative data on students participating in vocational education (number of students in a specific course, grades, number of apprenticeship), relevant reports and analyses produced by government agencies and think tanks, and qualitative interviews with management and employees working in industrial enterprises.

This study originated in the examination of the work and cross-organizational collaborations required to maintain the data in a single database, and the redesign of an existing information system that supports the management of this particular database (Seidelin, Grönvall, and Dittrich 2018). In this case, IU collaborated with multiple stakeholders to facilitate the service provision of the data on the individual members of the more than 160 Local Education Committees (LECs). To understand why LECs exist and the role of these governing bodies, it is necessary to elaborate on the broader context of the vocational educational system in Denmark. The Danish labour market is based and builds on tripartite negotiations among the social partners of the labour market (the government, employers' associations, and unions). Among other things, this negotiation affects and regulates the vocational education system and the adult vocational training system. IU was established to facilitate the ongoing collaboration between the social partners, to ensure the maintenance and development of vocational education programmes, specifically for the industrial sector in Denmark. Furthermore, each vocational education programme is controlled by a Skill Sector Council that consists of representatives from unions and employers' associations. Together with the Danish Ministry of Education, these Skill Sector Councils determine the educational framework. Finally, these frameworks are locally implemented through governing bodies known as LECs, which are affiliated with specific vocational colleges. Generally, an LEC consists of four to eight members, including representatives from the vocational college, and relevant employer associations and unions. The LEC members include a student and a teacher from the local college, and representatives from the unions and employers' associations. The representatives from the employer association(s) and union(s) are appointed by the union or employer association with which they are affiliated, and usually work in some local industry. IU provides services to maintain this setup by managing and communicating all relevant data about each LEC member, as needed (findings from our study on data maintenance in this cross-organizational context appeared in Seidelin, Grönvall, and Dittrich 2018). This data work requires many supporting manual processes, which often result in 'data discrepancies' among IU's three main stakeholders, each of which depends on this data. Therefore, in collaboration with the key stakeholders, IU decided to launch a redesign of the LEC database and its related IT system. To support this development, we proposed three collaborative design workshops, which included a data notation that specifically represented the entities in the current LEC database's schemata, and several other forms of adapted service design notation intended to support the initial steps of designing a new IT system that can manage this cross-organizational work. In the next section, we present our methodology, elaborate on our workshop activities, and describe the applied notation to further analyse how these support design with data.

4. Methods

The research presented here is a part of a larger, ongoing three-year action research project involving the IT University of Copenhagen and IU. The action research project focuses on how domain experts (who are not IT professionals) in organizations can take part in the design of data and data structures that underpin the services used by and influencing the work practices of these same domain experts and the network of stakeholders with which they are collaborating. We understand action research as a methodology, which means that the research aims to induce change and improvement of certain aspects of a targeted research domain (Robson 2002; Stringer 2013). In this case, IU constitutes the research domain. This paper builds on the first of three interventions, and thus presents the initial phase of the action research project. The objective of the first action research intervention was three-fold: 1) to understand what constitutes data in the context of IU, and examine how people use and make sense of data, 2) to introduce collaborative design methods at IU as a new approach to collaboration that supports innovation, and 3) to explore ways to foreground data in collaborative design situations. This project conducted its inquiry through design methods and techniques, as a way to induce change. In line with Binder and Brandt (2017), we argue that not all design practices are research practices, as research practices must respond to a research question or concern beyond, such as the (design) brief of a project. Thus, this project applies Research-through-design (Frayling 1993) as an intervention process (Zimmerman, Stolterman, and Forlizzi 2010; Koskinen et al. 2013).

4.1 Data collection

We collected data from March 2017 to January 2018 (figure 1). During this period, the first author spent approximately 3 days a week at IU. We decided, in collaboration with the management at IU, to focus our action research intervention on the LEC database because it appeared to be the simplest and most "isolated" area of IU's IT infrastructure. The first author identified key stakeholders related to the LEC data work together with employees and managers at IU. Then, we "followed the data" (Leonelli 2016; Bates, Lin, and Goodale 2016) by first conducting 20 hours of field observation that encompassed 15 semi-structured interviews with administrative 'data workers', who made use of, and contributed to the maintenance of the LEC data. Together, they represented 12 different organizations (including education secretariats, employer and employee associations, vocational colleges, and IU's external IT provider). The observation and interviews focused on the practices that go into producing, collecting, applying, maintaining, and updating the LEC data across organizational boundaries. By exploring the production and use of LEC data, we developed a comprehensive understanding of current data practices at IU and the cross-organizational work, including collaborative data practices that are necessary for maintaining the database. The initial data collection also revealed the key stakeholders' very limited knowledge about how the LEC data was produced, maintained, and used in other organizational contexts (Seidelin, Grönvall, and Dittrich 2018). Therefore, we suggested designing and facilitating three collaborative workshops that aimed to support the redesign of the LEC database in a way that put the domain experts at the centre of the development work. We also aimed to establish mutual understanding amongst these key stakeholders. Finally, we wanted to use this opportunity to explore how data can become object to design in a collaborative design context. On this basis, the workshops were designed to explore in various ways how data may become an object of design. Each of the three workshops lasted approximately 2.5 hours, and all were documented as video recordings, images and workshop materials. In the following subsections, we elaborate in greater detail on the design of the workshops, and the workshop participants.

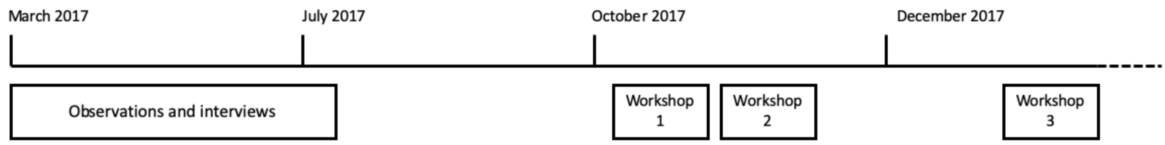
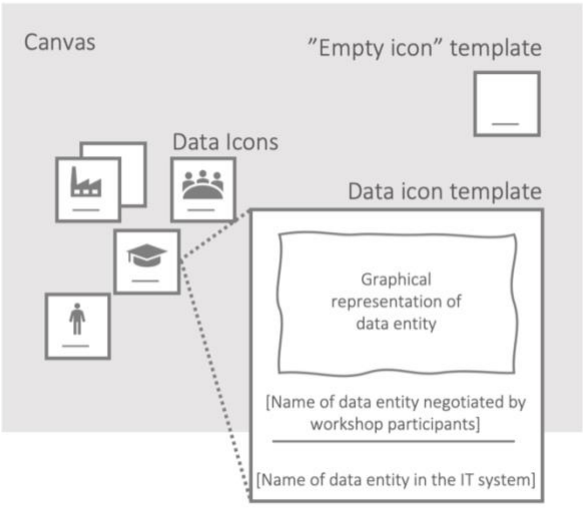


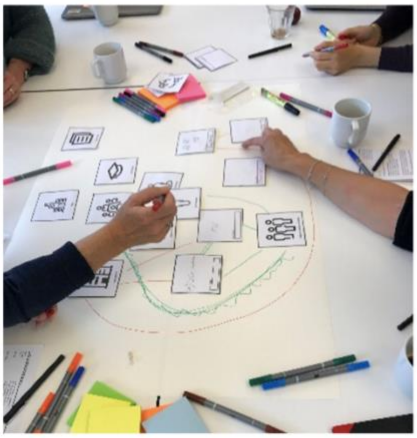
Figure 1. Timeline for the data collection.

4.1.1 The first workshop

The first participatory design workshop involved six participants, one facilitator (the first author) and one observer (the second author). The workshop had two general objectives. The first goal was member checking (Robson 2002), by verifying our initial analysis of the empirical material (observations and interviews) with the workshop participants, the administrative workers that were responsible for LEC-related tasks at IU, two unions and an employer association, which are the four key stakeholders in this case. The second goal was to explore notation for explicitly foregrounding data in a collaborative design context. The notation consisted of a set of simple graphic icons that represented the data entities in the current LEC database (see figure 2 below). We also included "empty icons" to allow the participants to create their own, e.g. if they felt a data entity was missing (figure 2). We chose to represent the data entities as icons, in an attempt to enhance the participants' data literacy in a non-technical manner. This is in line with previous studies that called for ways to foreground data so it supports learning and creative work (D'Ignazio 2017; Wolff et al. 2016). Examples of data entities are Schools, Education Programmes, and Education Consultants (which were represented as a building, a graduation cap, and the outline of three people, see Figures 2 and 3). To some, our notation may appear to be a physical, movable Entity-Relationship Diagram, a structural model that is often used in database design as a way to visualize relationships amongst the data entities in one or more databases (Ramakrishnan and Gehrke 2003). Therefore, it is important to emphasize that our notation is not to be confused with an entity-relationship diagram. First, the aim of our notation was not to produce a data model, but to support the workshop participants' understanding of data and database design as something that influences their work practices, and vice versa. Furthermore, the aim of our notation was to make visible to the participants how various stakeholders made use of data in similar and/or different ways, and in this way show how their underlying information and data needs determine the data schema that constitutes the service provision framework of the LECs. As such, our notation may be considered a useful tool and design activity when used prior to developing an Entity Relationship data model.

We experimented with the notation by using it in two different ways. First, we included in activities where the participants could relate to the notation on a blank canvas. Then, we used the data notation in a more structured format. To provide a structured format, we developed a matrix to prompt the participants to critically question aspects such as who generates the LEC data, how data is shared amongst various actors, and who uses the LEC data. This matrix constituted a 2×2-metre poster with a grid that emphasized stakeholders and activities. We refer to this structured format as the Data Matrix. In the next section, we explain the workshop activities and our use of the data notation.





Figures 2. and 3. On the left: A generic remake of the canvas, data icons and "empty icons". On the right: The workshop participants are negotiating the meaning of the data notation and their relationships. The data notation employs a blank canvas and a set of tangible data icons.

The workshop was designed as follows. First, we presented the insights we had collected from the field observation and interviews, in which all the participants had participated, and we invited them to comment on and discuss these (for a more in-depth description of the findings, please see Seidelin, Grönvall, and Dittrich 2018). On this basis, the first author introduced the notion of a database, and explained that the structure of a database depends on the data one needs, emphasizing that a database, and by extension, the data schema, may be altered and organized according to these needs. The workshop participants (apart from the participating authors) had very limited knowledge of IT systems and databases, despite their regular maintenance of the LEC database, and thus this perspective on IT and data was new to them. In the second phase of the workshop the participants were introduced to our data notation. The participants were asked to discuss the notation, to establish a shared understanding of what the various icons represented. This activity was carried out on the blank canvas (see figures 2 and 3). Then, they were asked to remove and/or add data entities (using “empty icons”), if they agreed that there were either redundant or missing entities. Finally, they were asked to discuss how they thought the data entities were related. This was intended to prompt discussion on the participants’ joint and varied data needs and practices. In the third phase of the workshop, the participants were asked to place the data notation on the data matrix (see figures 4 and 5). In this way, we observed how a more structured format influenced the workshop participants’ ability to talk about data. The workshops’ final phase and activity was a feedback session, where the participants were asked to reflect on the workshop, their experience with the data notation, and the implications of the two different formats (blank canvas and data matrix).

DATA MATRIX Activities / Stakeholders	Industriens Uddannelser	Key unions and employer associations	LEC Member	Vocational Colleges	Additional stakeholders?
Data generation					
Sharing of data					
“Data approval”					
Use data					
Delete/edit data					
Additional data activities?					

Figures 4. and 5. At the left: a translated version of the Data Matrix.

4.1.2 The second workshop

The second participatory workshop followed shortly after the first workshop. The purpose of the second workshop was also two-fold. First, it aimed to build on the insights from the first workshop, in order to generate ideas for the redesign of the LEC data and related IT system. Simultaneously, we wanted to examine whether foregrounding data in the first workshop would affect the ways in which the participants generated ideas. This workshop consisted of the same six domain experts, and three additional participants: one representative from the main employer’s association, one administrative worker from a vocational college, and one LEC member. Again, the first author was the moderator and the second author observed the workshop. This workshop was inspired by the ‘Future Workshop’ technique (Jungk and Müllert 1987), which is well-established in the PD community, because we wanted the participants to focus on identifying specific problems with the current situation, and to explore how to address these concerns.

The second workshop revolved around two design activities. The workshop started with an introduction of the participants, as additional participants had joined the second workshop. The first author also summarized the activities and discussions from the first workshop. The first design activity involved ‘the critique phase’, where the participants were asked to critique the current handling of LEC data and the related IT system. This was followed by a general discussion in which the participants shared the points of their critiques, which they subsequently categorised. We used this categorization as the basis for launching the second design activity, which was the ‘visionary phase’ in which the participants were asked to design visions of new and improved ways to manage and collaborate with and through the LEC data and IT system. To document this phase, we introduced the participants to Storyboarding (Tassi 2009), a method used in design processes to visualize concepts. The participants were unfamiliar with method, but nonetheless attempted to draw storyboards of their visions in pairs (see figure 6 for an example). This was followed by presentations of the storyboards and ideas. The final phase of the workshop was a feedback session, in which the participants were invited to reflect on their experiences and the use of tools for idea generation.

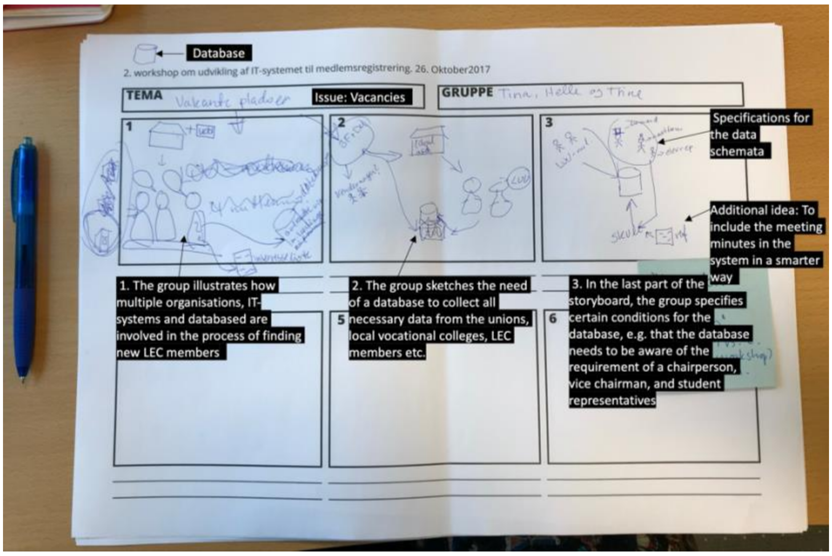


Figure 6. An example of a storyboard.

4.1.3 The third workshop

The third workshop differed from the previous two in that it involved only employees from various departments at IU. Thus, the workshop participants were three administrative employees, three managers, and one facilitator (the first author). This workshop aimed to bring together the insights and collected ideas, to create a foundation for the future development of a redesign of the LEC database and IT system. Moreover, we wanted to explore additional ways to foreground data at a slightly later stage of the design process. To do so, we complemented our participatory design approach with notation from service design (Blomberg and Darrah 2015), to also keep IU's service focus an explicit part of the design process. At its core, service design is about the process and act of designing services (Kimbell 2011). Service design draws on many disciplines (including participatory design), and may be understood as a design specialization, as it specifies a specific design focus. The emergent field of service design has developed tools and techniques that motivate a holistic perspective when designing and innovating a service and its context (Ostrom et al. 2015). We understand services as socio-material constellations that provide a functionality useful to a third party. Service design notation has been developed to support the design of these socio-material constellations (Blomberg and Darrah 2015). We chose to use two forms of service design notation. The first is known as the *service ecology map*, which is a tool used to map collaboration between service actors and stakeholders, in order to investigate the relationships amongst these parties (Polaine, Lovlie, and Reason 2013). Traditionally, this tool focuses primarily on human and nonhuman actors, which may include forms of technology used to support these relations (e.g. a computer or a phone). We wanted to explore whether a service ecology map could be incorporated to foreground data and to observe how it would play out in a collaborative design context. On this basis, we designed a 'service ecology map toolkit' that consisted of a canvas with a circular pattern, and octagonal icons in various colours to represent the following categories: 'Organization', 'Individual', 'Technology', 'Data Source', 'Questions' (to account for questions that might occur in the design situation), and 'New Ideas' (to facilitate the possibility of idea generation in the design situation). Although we initially identified central actors in collaboration with management and employees at IU, we chose to explore this tool as a means to 're-open' the design process and make room for new, additional actors relevant to the redesign of the LEC database.

The second notation we included in the third workshop is called the *service blueprint,* which is a tool that helps to visualize various service components (e.g. people, tools, and processes) and how they are linked throughout the service journey (Lynn Shostack 1982; Shostack 1984; Bitner, Ostrom, and Morgan 2008). A service blueprint is a matrix that allows users to identify each step of a service journey, and then visualize how each step affects service components above or below ‘the line of visibility’, which refers to the element that the user can or cannot see or experience while going through this process. In line with our inclusion of service ecology maps, we also wanted to examine how data could be foregrounded when using a service blueprint in a collaborative design situation (figure 7). Therefore, we altered the service blueprint to include two additional aspects below the line of visibility. We termed these aspects ‘Necessary data’ and ‘New potential data sources’. The aim was to observe whether this alteration of the notation would prompt the participants’ consideration of data that is needed to support the various steps of the data-based service and service provision, and to stimulate a consideration of data that could be explored in order to determine whether it could add value to the future service.

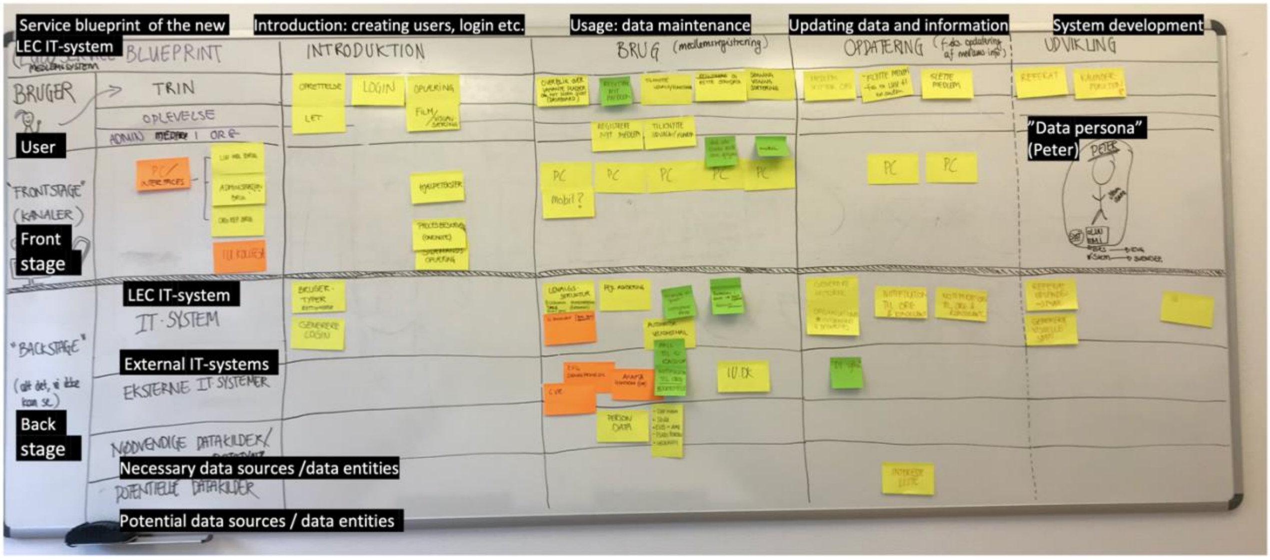


Figure 7. Extended service blueprint.

The workshop was designed as follows. First, the participants were divided into two groups, each of which got a ‘service ecology map toolkit’. Then, one group was asked to create a service ecology map from the perspective of an administrative worker at IU, and the other group was asked to create a service ecology map that put an administrative worker from an external organization at the centre of the map. This was followed by a joint discussion and comparison of the two service ecologies. Building on the mapping exercise, the next design activity revolved around the development of a service blueprint to support the redesign of the LEC database. We chose to make a ‘sketchy’ template for the service blueprint, to give the impression that it was not the final design. We did so by simply drawing the blueprint on a whiteboard, and asking the participants to collectively add Post-it notes with suggestions for the various stages and aspects of the service journey. We chose to include the service blueprint as a means to articulate and encapsulate the insights concerning the subsequent process of negotiating and collaborating with external IT developers in commencing the technical development of a redesigned LEC database and IT system. As we did in the first and second workshops, we concluded the third workshop with a feedback session in which the participants were asked to question and comment on the activities, and share their experiences.

4.1.4 Summary of the data collection

This article builds on a data collection consisting of field observation, semi-structured interviews, and three collaborative design workshops. The initial observations and interviews formed the basis for the three subsequent workshops. The general goal of the collaborative design workshops was to explore how various forms of notation facilitate explicit data design in collaborative design situations, and also to understand how such notation influences domain experts' ability to design with data. Despite the general goal of the workshops, each workshop had different objectives, and involved different forms of notation and activities (see table 1). In this article, we focus on the materials (video recordings, images and workshop materials) generated by the workshops. In the following section, we elaborate on our analysis, and how this led to our three main findings.

Workshop 1		**Workshop 2**		**Workshop 3**	
General objectives					
(1) To verify initial analysis of field observations and interviews with workshop participants, and (2) to explore our proposed notation for representing data in a collaborative design context.		(1) To build on the insights from Workshop 1, and (2) to examine whether the foregrounding of data in Workshop 1 would affect the subsequent design process.		(1) To bring together the insights and collected ideas collected in Workshops 1 and 2, and (2) to explore other ways of foregrounding data at a later stage of the design process.	
Number of participants					
6		9		6	
Workshop activities and design tools used					
Activity	Tool	Activity	Tool	Activity	Tool
Introduction		Introduction		Introduction	
Zooming in: Data exploration	Data notation	The Critique Phase	Clustering of brainstorming results	Combining user perspectives and data practices	Service Ecology Maps with a focus on technology and data
Zooming out: the context of the data	Data matrix	The Visionary Phase	Storyboards	Initiating the development of a service concept	Extended Service Blueprint
Feedback and wrap-up		Feedback and wrap-up		Feedback and wrap-up	

Table 1. Overview of the workshops, their main objectives and applied tools

4.2 Analysis

We chose to conduct a close interaction analysis (B. Jordan and Henderson 1995; Plonka et al. 2015) of the video recordings of the three workshops, in order to understand the interactions amongst the workshop participants when they worked with the various forms of notation. Our analysis unfolded as follows. The first author viewed all the video material with two questions in mind: 'when do people talk about data?' and 'how do they talk about data?'. Along with the initial analysis of the video material, the first author also reviewed the workshop output (e.g. the storyboards and images of the service blueprint). This step resulted in a selection of 54 video excerpts, each of which lasted between 17 seconds and 3 minutes. These excerpts were then categorized by the first author, which yielded 14 themes: current data practices, new data practices, data needs, data negotiation, data representation, imagining the future IT system, division of work, transparency, data priorities, real world relations, data relations, simplicity, complexity, and functionality. Bearing in mind that we were exploring how different forms of notation might support various aspects of designing with data (data as input for design, data as object of design, or both) and in light of the foregoing categories, the first author selected 10 video excerpts that best illustrated the 14 themes. These 10 video excerpts, including the participants' verbal and non-verbal interactions, were transcribed in detail. These transcriptions were later used in a workshop, where the three authors repeatedly viewed the selected video excerpts and read the transcripts. Based on this process, the authors analysed in detail how, for example, the workshop participants during the first workshop related the proposed data notation on the canvas to the data underpinning their common tasks. Our discussions of the observed and analysed interactions led to an understanding of two prerequisites that are necessary for domain experts to be able to consider data an object of design in a collaborative design context. First, domain experts need to be able to talk about data as data, meaning, for example, that they must recognize that they are talking about entities in a database. Domain experts also need to be able to distinguish between the data and the concrete situations it models. This enables the domain experts to collectively be aware of whether they are designing 'the service situation', for example, the user's interaction with a system, or whether they are designing the data schemata that underpin the service situation. Based on our empirical material, these prerequisites seem essential to domain experts' ability to consider data an object of design in a collaborative design situation. We consolidated this understanding in two steps. The first author reviewed all the video material to ensure that the excerpts were representative of the participants' discussions, critical considerations, and interactions. This led to the suggestion to include three additional excerpts. Therefore, we held another joint viewing of the three new excerpts. However, this discussion led to an agreement to stick to the selection of the 10 original video excerpts, as the three proposed video excerpts did not illustrate our understanding of the prerequisites more clearly. On this basis, we describe our three main findings, as follows. (1) The concrete data notation makes it possible for domain experts to talk about data, to further establish mutual understanding about the data among multiple stakeholders. (2) The domain experts are able to use the notation to distinguish between data and the real-life situation it models. (3) Notation may support domain experts' understanding of data as an object of design. In this article, we use 6 of the 10 selected video excerpts, as they best illustrate our findings. In the next subsection, we describe our efforts to ensure the reliability of the research. This is followed by a presentation of our three main findings in section 5.

4.3 Reliability

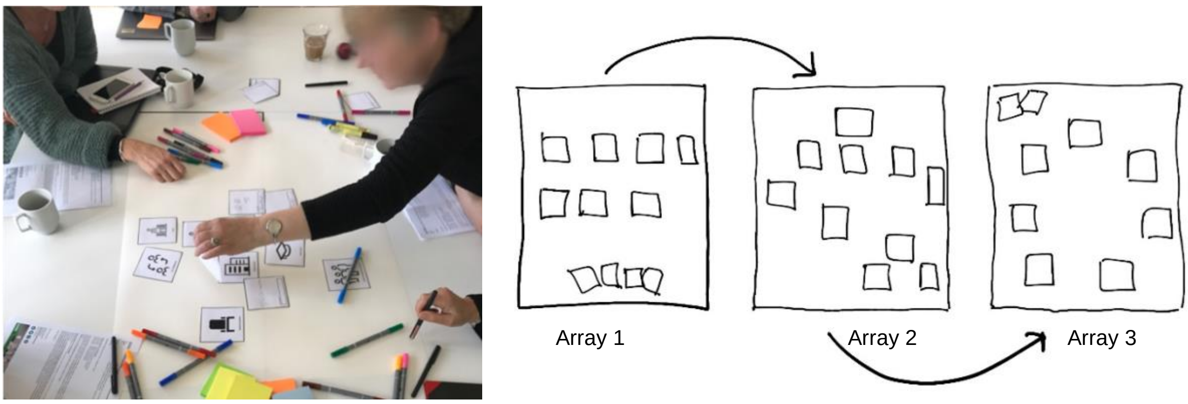
Our study ensured the reliability of our qualitative research (Robson 2002) in four principal ways. First, the fieldwork was carefully documented: interviews were audio recorded and transcribed, and participatory observation and workshops were video recorded. This generated an (1) audit trail and enabled thick descriptions, which enhanced the transparency of the field material and our findings. This also enables the reader to review our findings and insights. Secondly, (2) we triangulated in two ways. We triangulated data points, for instance, our initial participatory observations and interviews, with the participants' statements during the workshops. The first author collected photos, screen dumps, video and audio recordings, and documents, which are independent data points for triangulation. Furthermore, the fieldwork that forms the basis of earlier work presented by Seidelin, Grönvall, and Dittrich (2018), and this article, involved close collaboration among the authors, who compared the transcripts of the video excerpts with the video-recorded material, to ensure that the written excerpts reflected the recording. All three authors participated in (3) debriefing sessions, both following the initial research phase (interviews and participatory observations) and after each workshop, which supported the reflection on, and direction of the research. These debriefing sessions also influenced the first author's selection of the 10 video excerpts. Finally, (4) member checking was used as a technique to improve the accuracy and credibility of our study (Robson 2002). The interviewees, observed individuals, and workshop participants were invited to comment on the developing themes after each phase of the data collection. During the workshops, the participants were asked to evaluate the various forms of notation. This ongoing member checking helped to ensure that the participants' perceptions of the explorations were reflected in our findings (Robson 2002).

5. Findings

This section presents our three main findings. First, we elaborate on our proposed data notation: how it supported the workshop participants discussion of the data in the LEC database, and how it enabled them to explore the existing data schemata and their shortcomings. Second, we show how the data notation helped the domain experts to establish the difference and relationship between data and the concrete situation it models. Finally, we show how data may become an object of design for domain experts in a collaborative design situation.

5.1 Concrete data notation makes data discussable for domain experts

The first finding of our study focuses on how our proposed data notation supported the domain experts' ability to discuss the 'LEC data'. The data notation enabled the domain experts to explore the existing data and its shortcomings. Using data icons to conceptualize the data and to discuss the data through the notation was a first step towards the participants jointly transforming the data into an object of design. In this section, we elaborate on a situation that arose in the first workshop, where the domain experts worked to collaboratively to make sense of the data notation.



Figures 8. and 9. The image on the left shows how the workshop participants tried to make sense of the data notation by positioning the icons in different ways. The sketch (remake of the first author's note from the workshop) on the right shows the various formations the participants used: linear, hierarchical, circular.

The situation arose at the beginning of the workshop, when the participants were asked to talk about their understanding of the data icons and their position on the blank canvas. As mentioned, we chose to represent the existing data entities in the LEC database as simple graphic icons, and experiment with two different notation formats (a blank canvas and the data matrix). The data icons were labelled with the name of the data entity the icon represented (this was the same name that was used in the IT system). During the discussion of their understanding of the data notation, the domain experts placed the icons in various arrays on the blank canvas. The participants initially arranged the data icons in two separate rows, to overview of the individual icons and their relationships. This arrangement changed as the participants began to discuss the relevance, hierarchy, and priorities of the data icons. The configuration of the data icons transformed into a triangular array (see figures 8 and 9). Excerpt 1, below, illustrates how this process of group prioritization was initiated, and how it led to a shared understanding of the meaning of the data notation.

Excerpt 1. Understanding the data notation. (Workshop 1)	
Speaker	**Discussion**
Admin-X	The question is – what is the smartest thing to do? Is it to start with this one or that one? *Admin-X leans over the table and points to two icons of the canvas. The first icon is the 'LEC committee icon', the second is the 'LEC member' icon.*
Admin-Z	Yes… what should be the focus? I mean for me… I would always search for the committee [on IU's website] as a first step.
Admin-Æ	I do that, too. *Admin-X takes the committee icon and puts it in the middle of the canvas.*
IU-1	… and then we use that [the LEC committee icon] as a starting point and take it from there? *The other participants agree and nod.* Okay.
IU-2	And then we could put the member [LEC member icon] there, right? *The rest of the group agrees. Admin-Æ moves the LEC member icon beside the committee icon.*
Admin-Æ	And also, which school it [the LEC committee] is affiliated with, right? Like that? *Admin-Æ takes the school icon and put it close to the LEC committee icon. The group agrees.*
IU-2	Do you also look at which skills an LEC member has? *IU-2 points to the skill icon…*
Admin-Z	That skill [icon] and education [icon]… is that the same? *Admin-Z points to the icons.*
IU-2	No, I think 'education' covers areas of education that the committee addresses, whereas…
Admin-Z	The specialty? *Admin-Z looks at IU-2*
IU-2	That's how I understand it *The other participants nod and agree. Admin-Æ moves the 'Education' icon*
Admin-Æ	Then it [the education icon] needs to go here.
Admin-X	But that is important… but is it the same?
Author 2	The 'skill' [icon] is attached to the committee. *Admin-X takes the 'Education' icon.*
Admin-X	Ah, then we could say that this [the education icon] relates to the education at the school [vocational college].

Excerpt 1 comes from the initial phase of the first workshop. The excerpt illustrates in detail the use of data icons on a blank canvas to reconstruct the existing data set and its organization from the domain expert's point of view. The participants started by establishing their understanding of the data icons as stand-ins for the data, and decided how to begin the activity. They decided to use their own interaction with the system as the starting point. Initially, spatial proximity was established as indicating 'is related to' (see figure 9, array 1). The meaning of the notation and the meanings of the individual icons was (re-) established through discussion and by pointing to icons, holding them up, and moving them closer to each other. By pointing to and moving the data notation, they illustrated the data entities they were discussing with the other workshop participants. Thus, the participants also established a shared understanding of their practices, for example, when Admin X says, 'I would always search for…'.

The situation (Excerpt 1) showed how the participants used the icons to represent the data that the existing system administrates. Through these discussions, the domain experts also agreed that the triangular array of the data did not represent their understanding of the importance and connections among the various data entities. This led to a circular arrangement of the data notation (figure 9, array 3), which became the basis for the participants' subsequent discussion about the stakeholders' varying data needs. To guide this discussion, the participants were asked to draw lines on the canvas between the icons, using different colours to represent the relations between the various data entities, according to the various stakeholders' use of the LEC data. Although this array and these visible relationships may seem final, they did not stop ongoing discussions and negotiations, for example, about how the LEC data is prioritized differently by the stakeholders involved, and what it meant for the future redesign of the LEC database. Instead, these discussions continued throughout the three workshops.

The use of our proposed data notation also revealed that the tangibility of the data icons, and the flexibility of their placement appeared to be important characteristics of the notation. Specifically, they allowed the participants to explore connections and relations. We observed that drawing explicit connections on the canvas took longer. The participants stated that they found this form of activity more permanent. We also observed this aspect when the domain experts experimented with the data notation using the Data Matrix (a more structured format compared to the blank canvas) restricted the domain experts' ability to explore the relations among the various data entities, and to identify the stakeholders' varying data needs. This observation was supported by the participants when evaluating the design activities, and specifically, how they perceived the difference between the blank canvas and the data matrix. One of the workshop participants stated, '*I think it was difficult to start with a blank canvas, because then everything <u>is</u> possible, but it was interesting to discuss on such an open basis. Maybe that is also why – when we switched to the Data Matrix – I found it quite restrictive. However, I think the Data Matrix made it very visible to me that IU is at the centre of this. Just look at it: the majority of the icons are in IU's column*' (Admin-X, Workshop 1).

To summarize, our findings indicate that carefully designed data notation supports domain experts' ability to talk about data in ways that enabled them to collaboratively explore connections and relations among data entities. Our analysis also indicates that it is important to consider how expansive a data notation is in relation to the notation's intended aim and use.

5.2. Notation supports domain experts' ability to distinguish between data and the real-world situation it models

Our second finding focuses on the need for domain experts engaged in collaborative design to be able to distinguish between 'a real-world situation' and the data about the situation. Our domain experts needed to understand data as a purposeful representation of specific aspects of the real-world situation, to consider data and schemata as objects of design. In this section, we present two situations that show how the use of notation supports domain experts' ability to differentiate between data and 'reality', to negotiate the meaning of data.

The first situation we discuss comes from the first workshop, where the domain experts discussed whether the proposed data icons reflected the data they used when performing tasks related to maintaining the LEC IT system. Specifically, Excerpt 2 presents the participants' discussion of the meaning of an icon that represented the entity, 'Vacancies'. The participants discuss the difference between data and the situation it models through the use of the notation, by addressing both the larger question of filling vacancies in the LECs (reality), and the need for registered data that can help to solve the question at hand. Thus, the example illustrates how including the data notation in this particular collaborative design situation prompted the participants to critically consider what data is necessary to support their work practices and information needs.

Excerpt 2. Distinguishing between the data in the existing system and reality. (Workshop 1)

Speaker	Discussion
Admin-X	This is also interesting. I know it is annoying, but we all encounter this issue [refer to the 'vacancies' icon]. *Admin-X leans in over the table and takes the icon of the canvas. She shows the icons the group and puts it aside.*
IU-1	But you need that to have that data [vacancies in the LECs] registered right?
Admin-X	Yes! We think it is important that it is registered, but at the moment it doesn't show, it [the existing LEC system] just say 'vacancy' right? *The group nods and agrees*
Admin-Y	Exactly
Admin-X	Really, there are many more things [aspects] than this [the data icons] *Admin-X indicates it goes beyond the represented data entities. The participants nod.*

Excerpt 2 presents one of the occasions when the participants negotiated what was currently part of the data, and what additional information was needed to take action based on the data. The icon 'vacancy' is referred to as indicating a problematic situation. To resolve it, the administrators refer to knowledge about how LECs have to be assembled: a new member must be nominated or supported by a specific labour market organization. In the excerpt, Admin-X notes the inclusion of this constraint in the redesign of the system: 'We think it is important that it is registered, but at the moment it doesn't show, it [the existing LEC system] just says 'vacancy' right?' In the last round, Admin-X asks, at a more general level, which aspects of the situation should be modelled by the data. This question came up several times in the workshops, as the participants repeatedly discussed and negotiated the meaning of the data icons. When the participants used the data notation, it became clear that they understood data itself as an entity, and that the data model only reflected certain aspects that were relevant to their practices. Excerpt 2 shows how the participants discuss data as an independent part of an IT system.

The second situation we consider comes from the third workshop, where the participants had been asked to populate the incorporated service blueprint. As described in the methods section, we adapted this well-known service design notation to stimulate the domain experts to consider data when consolidating their insights from the first and second workshops into an initial description of future service provision related to the LEC IT system. Excerpt 3 shows a discussion in which the participants tried to establish "the boundaries of the future LEC IT system", as a way to understand what data is necessary to support the redesigned service concept.

Excerpt 3. People distinguish between data and reality: Peter and data about Peter. (Workshop 3)	
Speaker	**Discussion**
IU-3	…then we may as well consider 'where are we missing examiners?', 'which exams are being called for?' or… *IU-3 makes hands gestures to illustrate the various areas that a new system could support.*
IU-4	No! Because that is not what the system should handle. We are registering persons in here [in the system], because you are not going to use it… or then we need to expand this workshop, if we also need to consider how to handle and manage the call for apprenticeship tests… *IU-4 points her pen at the service blueprint.*
IU-3	But examiners are registered, right?
IU-4	Yes, but what I imagine… can I try to draw it, then we can erase it afterwards? *IU-4 grabs a pen and draws a stick figure.* If we imagine Peter. He is an LEC member, a member of [committee], then he is also an inspector [who approves whether or not a company is allowed to train and thus educate trainees] … and then he is an examiner. I don't think there are more things he can take part in… *IU-4 writes Peter's affiliations* We handle this data in [system A]. This [data] is handled in [system B], and this [part of the data] is handled in [system C] … It would be nice if it was just one system… But his master data [personal data] should be here in The System, for lack of a better name. *The other participants repeat* '*The System*'. Yes, so when we sign in and 'tick' his affiliations, then The System should recognize, 'ok this person is an examiner', this data is then sent to [system A], so [IU-employee] can access, for instance, a dashboard and then on this basis decide what types of examiners are needed, and see 'I got Peter'. [The IU-employee] doesn't have to think about updating data about Peter, because it's been updated in The System. This is how I envision The System – data-structure-wise – would support our practices. *IU-4 points her pen at the rest of the blueprint. The other participants nod.*

Excerpt 3 presents how some of the participants spoke about the future LEC IT system, in terms of the tasks the existing system supports, and which tasks the future system might support. The participants in the third workshop consisted entirely of IU members. This created a space, where the participants could discuss the data practices that were specifically related to maintenance of the LEC data. However, it also enabled the participants to discuss additional tasks that were peripheral to the LEC data work and currently supported by other internal IT systems. Specifically, IU-3 and IU-4 questioned how and where the data about a specific person with various roles should be stored in the future LEC database. IU-4 emphasizes a typical example of an experienced master craftsman who is a member of a specific LEC, acts as an examiner, and is a contact person for the practical education of apprentices in his company. They propose redesigning the data across various systems, and inventing a new one, 'The System'. This discussion about the boundaries of the future system created some confusion in the group, regarding which service(s) they aimed to improve and/or design. To resolve this uncertainty, IU-4 invented new notation that allows a representation of the various contexts in which the base data is used. The notation describes a 'data persona', a stick figure named Peter, with additional data attributes that represent how 'Peter' should be modelled in the system. IU-4 explained her idea, referring to the data persona and the common understanding of the work practices at the organization. This helped the group to jointly distinguish between the tasks that the LEC IT system can and should support, and the specific data that is needed to model the situation (see Excerpt 3, and figures 10 and 11).

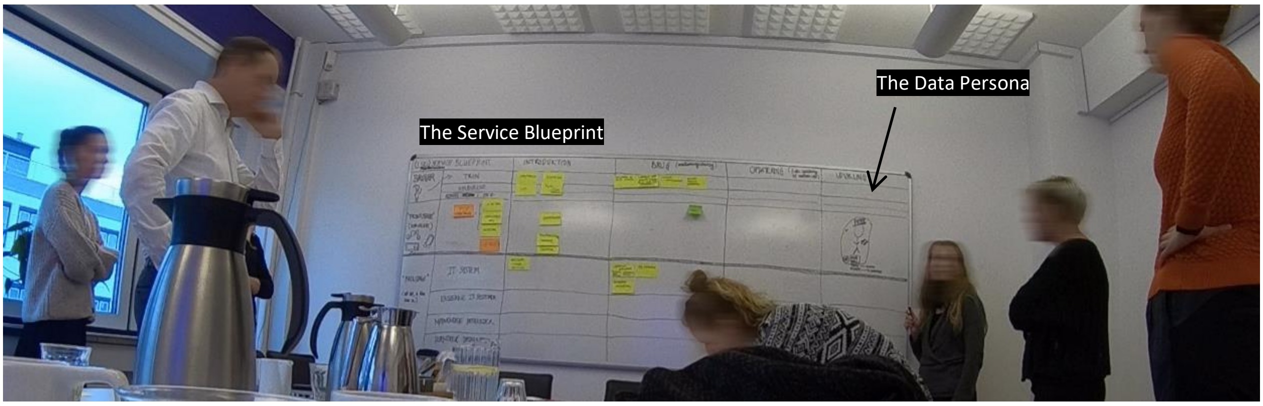


Figure 10. IU-4 explains her drawing of Peter; what constitutes the person and the data about the person.

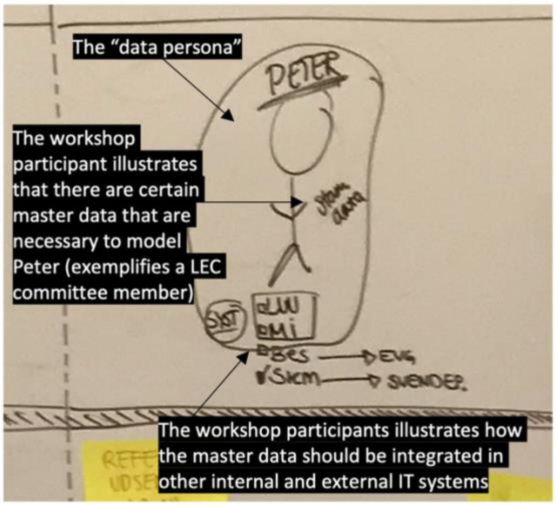


Figure 11. The 'data persona', Peter, drawn by IU-4.

In this section, we presented in detail two situations that in various ways show how notation, which aims to make data an explicit part of design, may support domain experts' ability to distinguish between the data and the real-world situation it models. This is an important finding, because the ability to distinguish between the two is a prerequisite for understanding data as an object of design.

5.3 Data may be an object of design for domain experts

This section presents three additional situations that suggest that data may become an object of design for domain experts in a collaborative design situation. Our third and final finding focuses on how the various forms of notation we used during the three workshops supported the participants' understanding of data as an object of design to varying degrees.

The first situation addressed in this section emphasizes yet another discussion that emerged during the first workshop when the participants were discussing whether there were redundant or missing data icons. Already by this point in the process, the participants began to think and talk about data as an object of design, as they started to discuss improvements to the existing system and data schema. The use of the proposed data notation prompted the participants to question that data schema. Excerpt 4 below emphasizes a discussion during the first workshop, where the participants considered whether certain data was missing from the existing IT system. Here, an IU employee questioned whether 'company' notation was not needed to support one organization's need to know in where their LEC members were employed/employers. This need was confirmed by the representative of the organization, and the data entity was added to the canvas (see Excerpt 4 below).

Excerpt 4. Data as something that may be designed (Workshop 1)

Speaker	**Discussion**
IU-1	… and for you [Administrative worker], I am thinking, a piece is missing – one that is called 'company', because it depends on whether the company is a member [of the association], right?…
Admin-Z	Yes, yes it does. *The IU employee grabs a blank icon template and creates a company icon, which she then hands to the administrative worker… Admin-3 adds the company icon to the blank canvas.*
Admin-X	… Then there also needs to be a 'vacancies' [icon], right?
Admin-Æ	Exactly.
IU-1	Exactly.
Admin-X	… It's a question of how you build it [the system] in a data-technical manner. If you know there need to be 3 [LEC members] from [organization 1] and 3 [LEC members] from [organization 2], or whatever… and then there needs to be – because it's not only our own representatives… there are also the schools and sometimes the student representatives, which could be added when it [the system] is constructed… So, there will be many more things than this… *Admin-X points and makes a circular movement with her pen to stress that she is referring to the icons on the table.*

Excerpt 4 illustrates how the data notation enabled the participants to critically evaluate the proposed data icons, in order to identify missing or redundant aspects of the existing data schema. Excerpt 2 already showed that the participants started to propose extensions of the model. Excerpt 4 is more explicit: IU-1 prompted the discussion referring to the constraints on manning LEC committees. To preserve the members' pre-negotiated parity, the company appointing a member representing the employers' side in the LEC needed to be a member of a specific trade organization. This led to further discussion of the details of how to model 'vacancies', as they are not simply the absence of members, but places that need to be filled by persons who fulfil specific requirements. This time, IU-1 raised the question of how far they should go with their redesign. The option to 'debug' the existing data model, for example, by adding a new data icon, enabled the participants to identify what kinds of data were missing from the existing system, and thus extend the data model with new objects (data entities). This indicates that the notation enabled the participants to consider data an object of design, and thus allowed them to design with data, so it could support the improved system functionality.

The second situation that we want to emphasize in this section supports the observation that the data notation supported the domain experts to consider data as an object of design. The second situation highlights how the domain experts in the first workshop discussed the limitations of the existing system. As shown in Excerpt 5 below, the participants note that the existing system is restricted to registering a certain category of LEC member, that is, members that have been appointed by the unions or employer associations. This has led IU employees to creatively (mis-)use the IT system to register additional members whose contact information may be useful for IU and other stakeholders at some point.

Excerpt 5. Creative misuse of the existing system. (Workshop 1)

Speaker	Discussion
IU-2	Well, we have never registered the representatives of the schools [vocational colleges].
Admin-X	No? *The IU employee grabs a blank icon template and creates a company icon, which she then hands to the administrative worker… Admin-3 adds the company icon to the canvas.*
IU-1	No, because – to say it frankly – it would be a nightmare. They [the schools] never tell us anything. *IU-2 looks at IU-1.*
Admin-X	Right
IU-2	Then [IU-1] would have to spend all her time getting in contact with the schools to ask whether there were any updates [to the data]… *IU-1 nods.* … which happens all the time.
IU-1	Yes, we only register them [school representatives] if they contact us, and then we register them as *the school* [in the system].

In Excerpt 5, the participants refer to a limitation of the existing system. So far, only LEC representatives appointed by the social partners have been registered systematically. However, for a number of reasons, it is important for IU to collect the contact information of the representatives of vocational colleges, when it is available. For example, this information may be useful for employees at IU, to serve representatives from unions and employer associations who request updated information about who, from a given vocational college, is affiliated with a particular LEC. As the existing system does not support this functionality, IU employees have registered the vocational college as a labour market organization in the LEC database, so the LEC system can show which representatives from a vocational college are associated with the LEC in question. These practices of creative misuse (excerpt 5) became visible to the domain experts when using the data icons and relating it to the use of the data. They were treated as indications of the limitations of the existing system. This suggests that our carefully designed data notation supported the domain experts in this particular collaborative design situation in transforming data into an object of design, which further enabled the design of additional system features through the data.

The third situation illustrates how the domain experts' experience with the data notation in the first workshop allowed them to more systematically relate to the data underpinning the new service in the second workshop. We introduced storyboards to prompt the participants to discuss and formulate initial service concepts that addressed the points of criticism that had been identified and categorized during the first phase of the second workshop. Excerpt 6, below, presents a situation where a group of participants, consisting of representatives from three different organizations, presents their service concepts by showing their storyboards to the other groups. Their idea addresses 'the vacancy problem', which refers to the challenge of recruiting new LEC members. This is a great challenge for the whole network of stakeholders, as it is required by law that all LECs have a certain number of representatives, and that each LEC has equal representation.

Excerpt 6: Redesign of the LEC database; finding potential new LEC members (Workshop 2)	
Speaker	**Discussion**
Admin-X	We chose to take the perspective of [employer association], because we [the various organizations working with LEC data] have different needs. We thought when [employer association] needs to appoint a new LEC member, they need a database similar to [existing database], with data about approved apprenticeships… *Admin-X points to her group's storyboard* … we talked about the necessity of a [new] database with approved training places… and then you need to be able to search by postal code, so you can look for approved companies in the local area. And then you can look and see if there are companies in the local area. *The representatives from the employer association nod.* … and then we discussed – now I move to the next field…. *Admin-X points with her pen to her group's storyboard to illustrate that she moves from one part of the story the next.* … then, if you consider a [new LEC] database; what needs to be in it for this to happen? It should contain and register LEC members, who is the chairman, who is the vice chairman, who is student representative, and meeting notes. *Admin-X makes hand gestures to show that she is counting the database attributes. The participants nod.* …
IU-2	Can I ask a question? Is it then supposed to be the vocational colleges that register this data? *The participants look at IU-2*
Admin-X	No, the schools should extract [the data]. *Admin-X points to the storyboard again.*

IU-2	Okay.
IU-1	So maybe it needs to be similar to when we extract data from [governmental agency], and then there should be a web service that enables us to say, ‘here we need to show X [geographical] area’. *The representative from the employer association nods. IU-1 grabs a Post-it note and starts to write.*

In Excerpt 6 the participants shared their ideas on how to solve one of the key problems: ‘the vacancy problem’ was already raised during the first workshop (excerpts 2 and 4), and was discussed again during the second workshop. Admin-X explained the storyboard developed by her subgroup. They proposed a ‘new database’ which should register companies with approved apprenticeship positions, because a company that has apprentices at a specific vocational college is more likely to be interested in the quality of the theoretical side of the education. The group suggested the new database should allow administrative workers across organizations to search for companies based on their postal code. This data entity (postal codes) would indicate the proximity from a company to local vocational colleges to which LECs belongs. Thus, they might be able to identify potential LEC members and in this way address the high number of vacancies in LECs. Furthermore, Admin-X explained that her working group proposed extending the existing data model so representatives from the vocational colleges that are part of a LEC may also be registered, without a workaround. Thus, during the presentations of the storyboards, it became evident that the participants (including Admin-X), who had worked with the data notation and their relationships, applied the negotiated meaning of data from the first workshop to further develop and substantiate their proposals for improving the LEC database (in contrast to those who did not participate in the first workshop).

5.4 Summarizing the findings

Overall, our findings show that data may be represented in collaborative design situations in ways that enable domain experts to regard data as an object of design. This indicates that domain experts can take an active part in designing data-based services’ technical infrastructure. The way the domain experts referred to data icons is an example of this. Our excerpts also show how foregrounding data enabled the domain experts to distinguish between data and the situation it represents, and to distinguish between, and connect the functionality of the future IT systems and the data underpinning it. The analysis above also shows that developing this ability to distinguish is not easy. Repeatedly, the domain experts discussed the relation between the data, their work practices surrounding it, and the situation is helps to administrate. These discussions led to an understanding of the role of data underpinning the new IT system’s functionality. Multiple representations helped the participants to distinguish and connect the various aspects. The series of workshops with their various related forms of notation for focusing on and connecting services, the functionality of the IT system, and design, supported the participants' ability to consider data as a design object. Our participants, without prior knowledge of database management and data attributes, became more 'data-aware' by applying data representations during the collaborative design workshops. This indicates that the use of relevant notation may help to better connect people who work effectively with data, and domain experts outside the field of computer science and data science (D'Ignazio 2017).

6. Discussion

We opened this article by asking why it is relevant to foreground data as an essential component of making and innovating data-based services. We presented related work that addressed why data should be foregrounded in design, what it means to design with data, and how to create paths to *co*-design with data. We explored participatory design in a service design framework as one approach to create such a path. Then, we elaborated on our interaction analysis of three collaborative design workshops, and presented the results of the analysis of our explorations with various forms of notation, each intended to make data an explicit part of the design process. Our study demonstrates how various forms of notation can mediate between design and data, and how data *can* become an object of design for domain experts in a collaborative design context. In line with Feinberg (2017), we challenge the understanding that data is simply input to design. Instead, we argue that data is designed – both when its structure is determined and when the structure is applied to collect data. We propose that domain experts should also be able to engage in the design of the data structures that underpin the data-based services these same domain experts use in and across organizational boundaries. For these domain experts, who are not IT professionals, to be able to this, we argue that data should also be an object of design during the process of co-designing data-based services. In the following sections, we discuss five lessons learned from our exploration of ways data may be foregrounded in a collaborative design context. We also address the limitations of this work. Taken together, we hope that the insights from this study will inform future investigations into how to support domain experts to co-design with data.

6.1 The data notation is an expansive representation

The first lesson learned from our exploration of how data may be foregrounded as an object of design for domain experts participating in collaborative design situations focuses on our proposed data notation. Specifically, it discusses how our proposed data notation supported the domain experts' discussions and critical reflections that created opportunities for design. The tangibility of the data icons allowed the participants to 'prototype' the mapping of existing and new data. The data icons were specifically designed notation that aimed to reflect the existing data model in the LEC database. The concreteness of the icons made it easier for the workshop participants to discuss which data was essential, less relevant or currently missing. In the workshop setting, the data icons were lifted, moved, pointed at, and placed side by side, while discussing their relation to the participants' tasks and to each other. The icons and the canvas provided a deictic space that anchored the discussion of data needs and usage. The participants were able to use the data notation to visualise and discuss how the domain experts had different data needs. Using Engeström's (1999) notion of expansive visibilization, we may say that the notation helped to anchor the discussion and the reformulation of the socio-spatial arrangements of the provision of the related services. Rather than limiting the possibility for discussion and development, this data notation allowed the participants to articulate their needs and question the present data design in this specific design situation. Even though the data notation represented the data entities of the LEC database (visualized through graphic icons) in a relatively linear manner, the way in which the data representation was introduced and used enabled the participants to critically evaluate and expand the existing data model. The possibility of being able to move the icons about on the canvas prompted exploration, for instance, when the participants made different arrays of the data icons, and additional icons were included. Excerpt 4 exemplified how the participants not only constructed the data, they also expressed how they understood the connection between a given kind of data and the functionality. In other words, data was not included and applied as a 'given', but as something that may be altered, based on shared decision-making. Also, the inclusion of 'blank icons' invited the participants to create additional icons, and thus expand the data model. We also observed how, when changing the framing of the notation (from the blank canvas to the data matrix) it prompted a different use of the notation. Although the data matrix had a more structured format, compared to the blank canvas, it supported the participants critical appraisal of the existing situation, in terms of how data is produced, maintained, and shared among actors in a specific context. It is important to emphasize that the case-specific preparation of the data icons involved allowed the domain experts in our study to relate the abstract notation to their reality, without limiting them to the status quo. For other cases, the representations, such as the data icons, would have to be adapted, to allow the domain experts to design with data in their specific contexts.

Our research indicates that it is possible to develop expansive representations that enable domain experts to relate to data as a design object; to critically evaluate the existing data schemata and the processes surrounding their maintenance, and to propose improvements to the data schemata. Being able to make sense of, and think critically about data is important, as it is used as an element that shapes the digital services that have come to play a significant role in how we experience everyday things, innovate at organizations, and structure society. Thus, such representations are essential to supporting people's ability to work with data as an object of design, and to allow for questioning and supporting the development of innovative ideas.

6.2. Data is a difficult design element to incorporate in co-design

The second lesson learned our study underlines that data is a difficult design element. As we showed in the findings section, it may be difficult to distinguish between what constitutes a concrete situation, and the data concerning the situation (see excerpt 5). Our explorations with various forms of notation indicate that it is necessary to consider the interplay between abstract data and concrete situations, for domain experts to be able to design with data. This is in line with the related work, where, for instance, Dove (2015) argues '*designing interfaces that visualise domain-relevant data with an intentionally ambiguous visual encoding appears to have a negative impact on co-designers' sensemaking, and reduces the appropriateness of their subsequent design ideas*' (p. 278). Although not intentional, the challenge of ambiguous visual encoding also became apparent when the workshop participants used the data icons. We had produced one set of icons, which represented the data entities in the existing system's interface. This resulted in situations where the participants were uncertain whether they were referring to an actor in the network (e.g. a person) or the data about the actor (e.g. the person's name, listed in the database). In the discussion, the participants managed to establish and agree on the relation between data (schemata) and the concrete situation.

Our study indicates that, although data is a difficult design element to incorporate to collaborative design situations, domain experts are able to design both the 'functionality level' and the 'data level' of the data-based services. This suggests that in collaborative design situations, the need to distinguish between data and the concrete situation it models should be reflected in data notation.

6.3 Using multiple forms of notation may help to foreground data in co-design

The third lesson learned from our explorative work addresses the use of multiple data representations and their complementarity throughout the design process. Although several insights were extracted from the work with the data icons, this notation did not stand alone. The participants were also prompted to consider data when using storyboards, and developing service ecology maps and a service blueprint (Shostack 1984; Bitner, Ostrom, and Morgan 2008). The adapted service ecology map helped the workshop participants to get an overview and identify the different needs of a potential user from IU, and a potential external user of the future LEC IT system. This was a useful way to remind the domain experts about varying data needs, as the third workshop included only members from IU. The expanded service design blueprint also offered useful notation. We adapted this notation to prompt the participants' consideration of data, when consolidating an initial redesign of the service provision related to the LEC IT system. Our analysis showed that the expanded service design blueprint helped the domain experts to make sense of how data underpins the service to be designed.

Our analysis suggests that it was helpful to introduce the most concrete notation for the data (data icons) during the initial part of the process. The shared sense-making related to the data, and enhanced understanding of data as an element of design, supported the inclusion of data in the other forms of design notation. The concrete and tangible introduction of data early in the process enabled the participants to bring the data 'forward' in the later workshops workshop (excerpt 6). This became apparent when the participants built on their knowledge of the LEC data, and drew databases on their storyboards and articulated what data was necessary to support their innovative service concept.

A limitation of our notation became evident when the participants designed their own notation to explain their design ideas. Even though we included several data representations, other/additional representations could have been included. When working with the expanded service blueprint, it became necessary for the group to augment one form of data notation (the service blueprint, extended with a space to define data sources in the form of data icons) with another ('data persona') (excerpt 3). The use of multiple forms of data notation, each with a different design objective, allowed the participants to address both the data design and how it underpins the service provision (e.g. the combination of data icons and storyboards). Again, this suggests that it is important to consider which notation to use, at which point of the design process, and how one data representation may be complemented by one or more other representations.

6.4 Foregrounding data helps to make the technical infrastructure of data-based services visible in co-design

The fourth lesson learned of our exploration of foregrounding data emphasizes and reiterates our emphasis on the relevance of considering both social *and* technical aspects of the process of designing data-based services. The growing number of digital services emphasizes the importance of considering and representing the technical infrastructures (data and data infrastructures and their interconnectedness) that underlie the ways in which we are able to provide and use services. Thus, the technical infrastructure of a service also constitutes essential parts of the context in which the service exists (or will exist). Our study revolves around the process of redesigning a database and its related IT system, which is maintained and provided by collaboration across many different organizations in a large network (Seidelin, Grönvall, and Dittrich 2018). Our analysis reveals that this form of service innovation does not simply influence the social aspects of how people interact and collaborate with the data, and how these practices are most likely to change because of the development and implementation of a new IT system. This form of service innovation also influences the technical infrastructure that underpins the data-based service. Our analysis shows that changing data practices in one setting – for example, how data is handled at IU – will affect how external stakeholders are required to work, for example, register and update data, for the service improvement to be successful. Thus, this will influence how data may be integrated with IT systems in external organizations. This illustrates how practices and services and technical infrastructures are interconnected (Blomberg and Darrah 2015). This emphasizes the importance of acknowledging that when we (re-)design data-based services, we are often building on and/or including established IT systems and practices. Therefore, when foregrounding data in co-design, it is important to acknowledge that we are always 'designing from somewhere' (Suchman 2002).

So far, design and engineering disciplines have focused mainly on either technical or social aspects, when developing data-based service (Seidelin, Dittrich, and Grönvall 2017). Building on Feinberg's work (2017), we show how data may become an object of design that enables exploration and innovation in collaborative, participatory workshops. In other words, we show that data, as a key component of the technical infrastructure, may be integrated into collaborative design situations. In this way, foregrounding data in concrete and tangible ways makes the technical infrastructure of data-based services visible for domain experts during the design process. We argue this may be a way to bridge the growing gap between people who can work effectively with data, and people who cannot (D'Ignazio 2017). In this regard, it is also relevant to consider various ways in which data notation may foreground social or technical (or both) aspects. In our related work, we argued that it is important to be to better understand how different forms of data notation support design with or of data (data as input for design and data as object of design, respectively). Based on our review of the related work (Dove 2015; Dove and Jones 2014b), we suggest that design situations where data is used as input for design may encourage a focus on the social aspects of a data-based service, whereas our intent to promote data as an object of design may advance more technical aspects of designing data-based services. However, future work is needed to examine this observation.

6.5 Participatory design is a useful approach for foregrounding data in collaborative design situations

In this final part of the discussion, we reflect on Participatory Design as a potential pathway-generating approach to foregrounding data in design in ways that do not originate from technical fields (D'Ignazio 2017). The increasing production and use of digital data make organizations dependent on certain data sources (Seidelin, Grönvall, and Dittrich 2018). This results in a growing need for multiple stakeholders to be able to make sense of the shared data, in order to design with data in ways that represent 'multiple voices', and thus address various needs and uses. This constitutes a second argument for why we need to unpack, represent, and make the existing data infrastructures tangible, to design with them.

As outlined in the related work, to the best of our knowledge, existing workshop design materials focus mainly on functionality or interaction interfaces. Well-known tools and methods in participatory design have included data in an indirect manner. For instance, the use of mock-ups to design an interface may include the data represented in the interface underpinning the functionality. We argue that by designing the functionalities of a system, the users simultaneously define which data will be represented in the system. The 'data decisions' are normally invisible to the users (Churchill 2012). One might question whether it is necessary to 'trouble' the users with making these underlying structures visible. In our case, an alternative design strategy could be to rely more on data experts, such as database designers, to design systems. However, even if these data experts were to interview the intended users, this would constitute design *for*, rather than *with* the IU personnel. Our ambition was to enable domain experts to be involved in the design of the data and data schemata that underpin the services used by these same domain experts, and in this way give the projected users a voice in the design. Our analysis shows that such involvement does provide the domain experts with new professional skills. Moreover, it also allowed IU to better understand how they work with data across organizations, how their data may be organized, and how it involves their services. Therefore, we argue that data-related decisions embed the prioritization of service needs and ways of provisioning the service. Data decisions become important to consider in cases where people and organizations depend on, and collaborate on data. These insights contribute to knowledge of the organization, and support the data experts. Domain experts and IT professionals need a way to make sense of and co-design with data. Our study compliments the existing participatory design toolbox by proposing notation that foregrounds data, to enable people to explicitly discuss data. Echoing previous research, we note that a participatory design approach helps various stakeholders to become aware of, and articulate their needs, in a way that makes the design fit their needs, instead of adjusting their needs to fit the design (Buur and Larsen 2010). Owing to the service focus of our study, we chose to include tools and techniques known from service design, to support the participants' awareness of what constituted the service, and the context in question. The use of service design tools helped the participants to make sense of the intangible data-based service for maintaining the LEC database. Furthermore, altering the service design tools helped to represent data in ways that prompted workshop participants to consider data throughout the design process. Based on this, we suggest that in the context of service innovation, it is useful to supplement a participatory approach with service design tools, to create a pathway to foregrounding data in design.

7. Conclusion

This paper explores how data may be explicitly foregrounded in ways that enable domain experts outside the field of computer and data science to make sense of, and critically consider data and data schemata in collaborative design situations. A growing body of research is exploring ways to include data as input for design in collaborative design contexts. This study is different, in that we explored how data may become an object of design for domain experts, meaning that domain experts may consider and recognise data as something that may be designed. We show that it is possible to do so, by using carefully designed data notation, applying a close interaction analysis of three collaborative design workshops. We also present five lessons learned from our explorations. These include that (1) the tangibility of our proposed data notation supported the domain experts' discussions and critical reflection, which led to collaborative design workshops of which data was a more explicit part of the design. (2) Our explorations also showed that data is a difficult element to incorporate in co-design, because of the observed difficulty in distinguishing between data and the concrete situation it models. We suggest that this aspect should be considered when designing data notation. (3) Moreover, in our case, it was helpful to use multiple data representations throughout the design process. Our findings indicate that multiple forms of notation can complement each other when used together. (4) We find that supporting the domain experts' understanding of data as an object of design enables them to critically consider a data-based service's technical infrastructure (data and data schemata). (5) Finally, we find that participatory design is a useful approach for foregrounding data in a co-design context. Together, these lessons offer practical insights that may inform future work. We acknowledge that our study includes a number of limitations, and we conclude that more research is needed, to further develop the insights from this work.

8. Acknowledgements

We are grateful for the collaboration of Industriens Uddannelser, who made this research possible. Especially, we thank the workshop participants for their efforts. We thank the reviewers for their valuable contributions, which have improved this paper significantly. This research was supported by Innovation Fund Denmark and The Danish Industry's Education and Collaboration Foundation.

9. References

Aragon, Cecilia, Clayton Hutto, Andy Echenique, Brittany Fiore-Gartland, Yun Huang, Jinyoung Kim, Gina Neff, Wanli Xing, and Joseph Bayer. 2016. 'Developing a Research Agenda for Human-Centered Data Science'. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion - CSCW '16 Companion*, 529–35. San Francisco, California, USA: ACM Press. https://doi.org/10.1145/2818052.2855518.

Ascher, Marcia, and Robert Ascher. 1997. *Mathematics of the Incas: Code of the Quipu*. Mineola, New York: Dover Publications.

Bates, Jo, Yu-Wei Lin, and Paula Goodale. 2016. 'Data Journeys: Capturing the Socio-Material Constitution of Data Objects and Flows'. *Big Data & Society* 3 (2): 2053951716654502. https://doi.org/10.1177/2053951716654502.

Binder, Thomas, and Eva Brandt. 2017. 'Design (Research) Practice'. In *Practice Based Design Research*, edited by Laurene Vaughan. London New York: Bloomsbury Academic.

Bitner, Mary Jo, Amy L. Ostrom, and Felicia N. Morgan. 2008. 'Service Blueprinting: A Practical Technique for Service Innovation'. *California Management Review* 50 (3): 66–94.

Blomberg, Jeanette, and Chuck Darrah. 2015. *An Anthropology of Services: Toward a Practice Approach to Designing Services*. Vol. 8. Morgan & Claypool Publishers.

Bossen, Claus, and Erik Grönvall. 2015. 'Collaboration In-between: The Care Hotel and Designing for Flexible Use'. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*, 1289–1301. Vancouver, BC, Canada: ACM Press. https://doi.org/10.1145/2675133.2675243.

Boyd, Danah, and Kate Crawford. 2011. 'Six Provocations for Big Data'. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1926431.

———. 2012. 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon'. *Information, Communication & Society* 15 (5): 662–79. https://doi.org/10.1080/1369118X.2012.678878.

Bright, Jonathan, Bharath Ganesh, Cathrine Seidelin, and Thomas M. Vogl. 2019. 'Data Science for Local Government'. SSRN Scholarly Paper ID 3370217. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=3370217.

Buur, Jacob, and Henry Larsen. 2010. 'The Quality of Conversations in Participatory Innovation'. *CoDesign* 6 (3): 121–38. https://doi.org/10.1080/15710882.2010.533185.

Cao, Longbing. 2017. 'Data Science: A Comprehensive Overview'. *ACM Computing Surveys* 50 (3): 1–42. https://doi.org/10.1145/3076253.

Churchill, Elizabeth F. 2012. 'From Data Divination to Data-Aware Design'. *Interactions* 19 (5): 10. https://doi.org/10.1145/2334184.2334188.

Crabtree, Andy, and Richard Mortier. 2015. 'Human Data Interaction: Historical Lessons from Social Studies and CSCW'. In *ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work, 19-23 September 2015,* *Oslo, Norway*, edited by Nina Boulus-Rødje, Gunnar Ellingsen, Tone Bratteteig, Margunn Aanestad, and Pernille Bjørn, 3–21. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-20499-4_1.

D'Ignazio, Catherine. 2017. 'Creative Data Literacy: Bridging the Gap between the Data-Haves and Data-Have Nots'. *Information Design Journal* 23 (1): 6–18. https://doi.org/10.1075/idj.23.1.03dig.

Dittrich, Yvonne, Cathrine Seidelin, and Erik Grönvall. 2018. 'Data as Infrastructure – Infrastructuring for Data Analytics'. In *Proceedings of the Participatory Design Conference*. Hasselt & Genk, Belgium. https://pure.itu.dk/portal/files/83630235/Data_as_Infrastructure_1.0.pdf.

Dove, G. 2015. 'CoDesign with Data'. https://core.ac.uk/reader/29017176.

Dove, G., and S. Jones. 2014a. 'Using Data to Stimulate Creative Thinking in the Design of New Products and Services'. https://core.ac.uk/reader/29017095.

———. 2014b. 'Using Information Visualization to Support Creativity in Service Design Workshops'. https://core.ac.uk/reader/76980229.

Engeström, Yrjö. 1999. 'Expansive Visibilization of Work: An Activity-Theoretical Perspective'. *Computer Supported Cooperative Work (CSCW)* 8 (1–2): 63–93. https://doi.org/10.1023/A:1008648532192.

European Commission. 2020. 'White Paper on Artificial Intelligence: A European Approach to Excellence and Trust'. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

Feinberg, Melanie. 2017. 'A Design Perspective on Data'. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 2952–63. Denver, Colorado, USA: ACM Press. https://doi.org/10.1145/3025453.3025837.

Feinberg, Melanie, Daniel Carter, Julia Bullard, and Ayse Gursoy. 2017. 'Translating Texture: Design as Integration'. In *Proceedings of the 2017 Conference on Designing Interactive Systems - DIS '17*, 297–307. Edinburgh, United Kingdom: ACM Press. https://doi.org/10.1145/3064663.3064730.

Fischer, Joel E., Andy Crabtree, James A. Colley, Tom Rodden, and Enrico Costanza. 2017. 'Data Work: How Energy Advisors and Clients Make IoT Data Accountable'. *Computer Supported Cooperative Work (CSCW)* 26 (4–6): 597–626. https://doi.org/10.1007/s10606-017-9293-x.

Fisher, Danyel, Rob DeLine, Mary Czerwinski, and Steven Drucker. 2012. 'Interactions with Big Data Analytics'. *Interactions* 19 (3): 50. https://doi.org/10.1145/2168931.2168943.

Frayling, Christopher. 1993. 'Research in Art and Design'. *Royal College of Art* 1 (1): 1–9.

Gitelman, Lisa. 2013. *'Raw Data' Is an Oxymoron*. Cambridge, Massachusetts: The MIT Press. http://ieeexplore.ieee.org/xpl/bkabstractplus.jsp?bkn=6451327.

Goodwin, S., J. Dykes, S. Jones, I. Dillingham, G. Dove, A. Duffy, A. Kachkaev, A. Slingsby, and J. Wood. 2013. 'Creative User-Centered Visualization Design for Energy Analysts and Modelers'. *IEEE Transactions on Visualization and Computer Graphics* 19 (12): 2516–25. https://doi.org/10.1109/TVCG.2013.145.

Gray, Jonathan, Carolin Gerlitz, and Liliana Bounegru. 2018. 'Data Infrastructure Literacy'. *Big Data & Society* 5 (2): 205395171878631. https://doi.org/10.1177/2053951718786316.

Guha, Mona Leigh, Allison Druin, Gene Chipman, Jerry Alan Fails, Sante Simms, and Allison Farber. 2004. 'Mixing Ideas: A New Technique for Working with Young Children as Design Partners'. In *Proceeding of the 2004 Conference on Interaction Design and Children Building a Community - IDC '04*, 35–42. Maryland: ACM Press. https://doi.org/10.1145/1017833.1017838.

Halse, Joachim, Eva Brandt, Brendon Clark, and Thomas Binder. 2010. 'Rehearsing the Future'. https://adk.elsevierpure.com/en/publications/rehearsing-the-future.

Hiniker, Alexis, Kiley Sobel, and Bongshin Lee. 2017. 'Co-Designing with Preschoolers Using Fictional Inquiry and Comicboarding'. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 5767–72. Denver, Colorado, USA: ACM Press. https://doi.org/10.1145/3025453.3025588.

IU, Industriens Uddannelser. 2019. ‘Mission og vision’. Mission and Vision. 2019. https://iu.dk/om-os/iu-organisation/mission-og-vision/.

Jansen, Yvonne, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper Hornbæk. 2015. ‘Opportunities and Challenges for Data Physicalization’. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI ’15*, 3227–36. Seoul, Republic of Korea: ACM Press. https://doi.org/10.1145/2702123.2702180.

Jordan, Brigitte, and Austin Henderson. 1995. ‘Interaction Analysis: Foundations and Practice’. *Journal of the Learning Sciences* 4 (1): 39–103. https://doi.org/10.1207/s15327809jls0401_2.

Jordan, Chris. 2009. *Running the Numbers: An American Self-Portrait*. Pullman : Munich ; New York, NY: Museum of Art, Washington State University ; Distributed by Prestel Verlag.

Jungk, Robert, and Norbert Müllert. 1987. *Future Workshops: How to Create Desirable Futures*. London: Institute for Social Inventions.

Kensing, Finn, and Jeanette Blomberg. 1998. ‘Participatory Design: Issues and Concerns’. *Computer Supported Cooperative Work (CSCW)* 7 (3–4): 167–85. https://doi.org/10.1023/A:1008689307411.

Kimbell, Lucy. 2011. ‘Designing for Service as One Way of Designing Services’. *International Journal of Dsign* 5 (2). http://www.ijdesign.org/index.php/IJDesign/article/view/938/345.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Los Angeles, California: SAGE Publications.

Kitchin, Rob, and Tracey Lauriault. 2018. ‘Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work’. In .

Koskinen, Ilpo, John Zimmerman, Thomas Binder, Johan Redstrom, and Stephan Wensveen. 2013. ‘Design Research Through Practice: From the Lab, Field, and Showroom’. *IEEE Transactions on Professional Communication* 56 (3): 262–63. https://doi.org/10.1109/TPC.2013.2274109.

Kristensen, Margit, Morten Kyng, and Leysia Palen. 2006. ‘Participatory Design in Emergency Medical Service: Designing for Future Practice’. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’06*, 161. Montreal, Québec, Canada: ACM Press. https://doi.org/10.1145/1124772.1124798.

Kun, Peter, Ingrid Mulder, Amalia de Götzen, and Gerd Kortuem. 2019. ‘Creative Data Work in the Design Process’. In *Proceedings of the 2019 on Creativity and Cognition - C&C ’19*, 346–58. San Diego, CA, USA: ACM Press. https://doi.org/10.1145/3325480.3325500.

Kun, Peter, Ingrid Mulder, and Gerd Kortuem. 2018. ‘Design Enquiry Through Data: Appropriating a Data Science Workflow for the Design Process’. In , 1–12. BCS Learning and Development Ltd. https://doi.org/10.14236/ewic/HCI2018.32.

Leonelli, Sabina. 2016. *Data-Centric Biology: A Philosophical Study*. Chicago ; London: The University of Chicago Press.

Lindsay, Stephen, Daniel Jackson, Guy Schofield, and Patrick Olivier. 2012. ‘Engaging Older People Using Participatory Design’. In , 1199–1208. ACM. https://doi.org/10.1145/2207676.2208570.

Lynn Shostack, G. 1982. ‘How to Design a Service’. *European Journal of Marketing* 16 (1): 49–63. https://doi.org/10.1108/EUM0000000004799.

Malmberg, Lisa. 2017. *Building Design Capability in the Public Sector: Expanding the Horizons of Development*. Linköping University Electronic Press.

Markham, Annette N. 2013. ‘Undermining “Data”: A Critical Examination of a Core Term in Scientific Inquiry’. *First Monday* 18 (10). https://doi.org/10.5210/fm.v18i10.4868.

McMillan, Donald, Arvid Engström, Airi Lampinen, and Barry Brown. 2016. ‘Data and the City’. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI ’16*, 2933–44. Santa Clara, California, USA: ACM Press. https://doi.org/10.1145/2858036.2858434.

OECD. 2019. 'The Path to Becoming a Data-Driven Public Sector'. OECD. https://doi.org/10.1787/059814a7-en.
Ostrom, Amy L., A. Parasuraman, David E. Bowen, Lia Patrício, and Christopher A. Voss. 2015. 'Service Research Priorities in a Rapidly Changing Context'. *Journal of Service Research* 18 (2): 127–59. https://doi.org/10.1177/1094670515576315.
Plonka, Laura, Helen Sharp, Janet van der Linden, and Yvonne Dittrich. 2015. 'Knowledge Transfer in Pair Programming: An in-Depth Analysis'. *International Journal of Human-Computer Studies* 73 (January): 66–78. https://doi.org/10.1016/j.ijhcs.2014.09.001.
Polaine, Andrew, Lavrans Lovlie, and Ben Reason. 2013. *Service Design: From Insight to Implementation*. Brooklyn, NY: Rosenfeld Media.
Ramakrishnan, Raghu, and Johannes Gehrke. 2003. *Database Management Systems*. 3rd ed. Boston: McGraw-Hill.
Ribes, David, and Steven J Jackson. 2013. 'Data Bite Man: The Work of Sustaining a Long-Term Study'. In *'Raw Data' Is an Oxymoron*, 147–66. Cambridge, Massachusetts: The MIT Press.
Robson, Colin. 2002. *Real World Research: A Resource for Social Scientists and Practitioner-Researchers*. 2nd ed. Oxford, UK ; Madden, Mass: Blackwell Publishers.
Rosenberg, Daniel. 2013. 'Data Before the Fact'. In *'Raw Data' Is an Oxymoron*, 16–40. Cambridge, Massachusetts: The MIT Press.
Russell, Daniel M., Gregorio Convertino, Aniket Kittur, Peter Pirolli, and Elizabeth Anne Watkins. 2018. 'Sensemaking in a Senseless World: 2018 Workshop Abstract'. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–7. Montreal QC, Canada: ACM Press. https://doi.org/10.1145/3170427.3170636.
Sanches, Pedro, and Barry Brown. 2018. 'Data Bites Man: The Production of Malaria by Technology'. *Proceedings of the ACM on Human-Computer Interaction* 2 (CSCW): 1–19. https://doi.org/10.1145/3274422.
Sangiorgi, Daniela, and Alison Prendiville, eds. 2017. *Designing for Service: Key Issues and New Directions*. London ; New York, NY: Bloomsbury Academic, an imprint of Bloomsburgy Publishing Plc.
Seidelin, Cathrine, Yvonne Dittrich, and Erik Grönvall. 2017. 'Identification of Data Representation Needs in Service Design'. *Association for Information SystemsAIS Electronic Library (AISeL)*, Selected Papers of the IRIS, , no. 8: 16.
Seidelin, Cathrine, Erik Grönvall, and Yvonne Dittrich. 2018. 'Data Work in a Knowledge-Broker Organisation: How Cross-Organisational Data Maintenance Shapes Human Data Interactions'. In *HCI '18 Proceedings of the 32nd International BCS Human Computer Interaction Conference*, 1–12. Belfast, United Kingdom. https://doi.org/10.14236/ewic/HCI2018.14.
Shostack, G. 1984. 'Designing Services That Deliver'. *Harvard Business Review* 62 (1): 133.
Simonsen, Jesper, and Toni Robertson. 2012. *Routledge International Handbook of Participatory Design*. London: Routledge. https://doi.org/10.4324/9780203108543.
Speed, Chris, and Jon Oberlander. 2016. 'Designing from, with and by Data: Introducing the Ablative Framework.' In *DRS 2016 International Conference: Future–Focused Thinking*, 1:2991–3004. Brighton, United Kingdom. https://www.research.ed.ac.uk/portal/files/25134011/drs_2016.pdf.
Still, Kaisa, Minna Isomursu, Johanna Still, and Pekka Isomursu. 2004. 'The Role of Technology in Interest-Based Communities: Interactivity, Immersion and Connectivity'. In *Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*, 681–685. WI '04. Washington, DC, USA: IEEE Computer Society. https://doi.org/10.1109/WI.2004.137.
Stringer, Ernest T. 2013. *Action Research*. Fourth edition. Los Angeles, California: SAGE.
Suchman, Lucy. 1993. 'Working Relations of Technology Production and Use'. *Computer Supported Cooperative Work* 2 (1–2): 21–39. https://doi.org/10.1007/BF00749282.
———. 2002. 'Located Accountabilities in Technology Production' 14: 16.

Tassi, Roberta. 2009. ‘Storyboard | Service Design Tools’. Service Design Tools. 2009. http://www.servicedesigntools.org/tools/13.
Wolff, Annika, Daniel Gooch, Jose J Cavero Montaner, Umar Rashid, and Gerd Kortuem. 2016. ‘Creating an Understanding of Data Literacy for a Data-Driven Society | The Journal of Community Informatics’. http://ci-journal.net/index.php/ciej/article/view/1286.
Zamenopoulos, Theodore, and Katerina Alexiou. 2018. *Co-Design as Collaborative Research*.
Zhao, Jack, and Andrew Vande Moere. 2008. ‘Embodiment in Data Sculpture: A Model of the Physical Visualization of Information’. In *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts - DIMEA ’08*, 343. Athens, Greece: ACM Press. https://doi.org/10.1145/1413634.1413696.
Zimmerman, John, Erik Stolterman, and Jodi Forlizzi. 2010. ‘An Analysis and Critique of Research through Design: Towards a Formalization of a Research Approach’, 10.

Publication 4

Seidelin, C., Sivertsen, S., Dittrich, Y. (2020) **Designing an organisation's design culture: How appropriation of service design tools and methods cultivates sustainable design capabilities in SMEs**. Proceedings of the 6th ServDes Conference. Melbourne. 6th-9th of July 2020.

ServDes.2020 – Tensions, Paradoxes, Plurality
RMIT University, Melbourne Australia
6-9th July 2020

Designing an organisation's design culture: How appropriation of service design tools and methods cultivates sustainable design capabilities in SMEs

Cathrine Seidelin[1], Stine Moeslund Sivertsen[2], Yvonne Dittrich[1]
cfre@itu.dk, stms@iu.dk, ydi@itu.dk
[1]IT University of Copenhagen, Denmark
[2]Industriens Uddannelser, Denmark

Abstract

Service design (SD) is acknowledged as an approach that can help organisations to address service innovation. However, organisations are struggling to build design capabilities and develop sustainable SD cultures within the organisations. This paper focuses on this central challenge by exploring how a small and medium-sized, "non-design-intensive organisation" can integrate SD both as a way to develop internal design capabilities and as an approach to service innovation. We report on an action research study in which we initiated seven SD micro cases. The findings show how our designed SD learning activities developed autonomous SD initiatives within the organisation, and thus over time fostered a sustainable SD culture in this context. Based on our findings, we conclude that organisational appropriation of SD tools and methods is crucial for an organisation's ability to build and sustain capabilities which can foster a SD culture.

KEYWORDS: service design, service innovation, design capabilities, organisational change, design culture

Introduction

It is becoming increasingly difficult for organisations to ignore the need to hold inherent capabilities for continuous improvement and development work (Wetter-Edman & Malmberg, 2016). Therefore, more and more organisations are investing in design-enhancing initiatives as a way to become more innovative and competitive (Lima & Sangiorgi, 2018; Wetter-Edman & Malmberg, 2016). SD has been recognised as a useful and beneficial approach to service innovation, i.e., due to its way of supporting the generation of innovative ideas through a user-centric and holistic perspective (Meroni & Sangiorgi, 2011). However, research shows that many organisations experience difficulty in developing a sustainable SD culture from within the organisation (Holmlid & Malmberg, 2018; Lima & Sangiorgi, 2018). So far, there has been little discussion about how small and medium-sized organisations (SMEs) can overcome the critical challenge of not only of integrating SD tools and methods but doing so in ways that foster a sustainable SD culture in the organisation. This paper explores how a medium-sized, non-design-intensive service organisation can integrate SD as an approach to build and sustain design capabilities and address service innovation.

SD constitutes a human-centred, holistic, creative, and iterative approach to creating new or improving existing services (Blomkvist, Holmlid, & Segelström, 2010; Meroni & Sangiorgi, 2011). While these definitions have proven useful in previous studies, this paper argues they are too limited when discussing the adoption of SD in SMEs. Instead, this paper makes use of an understanding of SD, as proposed by Blomberg and Darrah (2015). In this perspective 'designing' is understood as a bundle of activities rather than a single activity or process and 'services' constitute "fundamentally abstract propositions or transformations [that] are replaced with socio-material configurations of people and their know-how, artifacts and spaces" (ibid. p. 74). This means that services are deeply embedded within practices as well as enacted through practices (Blomberg & Darrah, 2015). This perspective embraces an understanding of SD that it can be practised beyond a single process and in between projects.

Based on our framing of SD, what does it mean, then, to build sustainable design capabilities (Malmberg, 2017)? The notion of design culture has emerged as a multifaceted concept which aims to shed light on the qualities by which design is practised, meaning how design is perceived, understood and enacted in everyday life (Julier, 2006). This means a design culture can exist at a very local level, for instance in a specific organisational context and is influenced by an organisation's design capabilities, as these make up how and to what extent design is practised within a given context (Malmberg, 2017). When adding 'sustainable' to the concept of design culture, it is essential to have the contextualization of the study in mind. SMEs are often incapable of simply hiring (service) designers and rarely have specific design departments that can drive change. Thus, if design should be part of a non-design intensive SME, it needs to be part of their DNA. Thus, a *sustainable* design culture for SMEs means integrating design in ways that are durable according to their size and resources in the long-term. On this basis, the term sustainable design culture will be used throughout this paper to refer to an organisation's ability to change dominant organisational cultures by making use of SD in ways that prompt continuous service innovation over time.

This study is situated at the medium-sized service organisation Industriens Uddannelser (in English: The Education Secretariat for Industry, hereafter the acronym "IU" is used), which is an education secretariat based Copenhagen, Denmark. "IU" facilitates the collaboration between multiple labour market partners to develop educational programs for vocational training and adult vocational training in the industrial sector in Denmark. Prior to this study, IU had minimal knowledge of and experience with creative problem solving and "design thinking" (Curedale, 2019). This paper presents an empirical study where the authors initiated SD initiatives, so-called "service design micro cases", to develop SD capabilities at IU. We show how these micro cases spurred additional initiatives and manifested an emergent design culture at IU. The paper takes a socio-cultural perspective to discuss how SMEs can initiate learning activities that help to overcome the challenges of integrating and maintaining SD as an approach to service innovation. Due to the study's organisational context, this paper makes use of the notion of service innovation as a new or improved process or service offering that is put into practice and adopted by an organisation to further create value to one or more actors in a service network (Patrício, Gustafsson, & Fisk, 2018).

This paper is organised as follows: In the next sections, we present the related work, which focuses on (a) the organisational challenge of adopting SD in organisations, (b) the current state of service innovation and SD literature, and (c) the concept of the Zone of Proximal Development (Cole, 1985), which constitutes our analytical lens. Then follows a description of our methodology. The paper proceeds to our analysis and discussions, which focuses on our proposed SD learning activities and their impact at IU. In particular, the analysis investigates how the learning activities transformed into three waves that in different ways brought about organisational- and cultural change. Following a discussion, the paper concludes by proposing three lessons learned for future practice that can support SMEs' integration of SD.

The challenge of adopting service design in organisations

In the past decade, SD has developed and established itself as a practice that enables Industry to innovate their services through a human-centred design approach (Miettinen, 2016). The prevalence of positive business cases has caused non-design intensive organisations to invest in initiatives that develop and enhance SD capabilities as a means to drive innovation and trigger organisational change (Brown, 2019; Lima & Sangiorgi, 2018; Malmberg, 2017; Sangiorgi & Prendiville, 2017). This tendency originates from a need "to build innovative organisations and organisations that inherently hold capacities for continuous improvement and development work" (Wetter-Edman & Malmberg, 2016, p. 516). However, this is easier said than done. Despite this growing interest, Holmlid and Malmberg (2018) find that few studies have been published on organisations' *successful* adoption of SD. They identify that it is a barrier for many organisations to disseminate design practices within their organisation, and thus develop a sustainable design culture. They argue that although individual members of an organisation participate in design-enhancing and capability-building initiatives, many of these projects do not diffuse SD knowledge or practice to other projects or additional members of the organisation (Holmlid & Malmberg, 2018). This means that while SD has proven to be a useful way for organisations in many different industries to approach innovation, they are struggling to expand and sustain their design capabilities.

There is a growing body of research that study organisations' introduction to and application of SD as an approach to innovation. These studies investigate both public and private organisations that have engaged in SD projects to address various issues. The areas of application range from innovating service offerings in the insurance and escalator industries (Miettinen, 2016; Polaine et al., 2013) to improvements of policymaking and healthcare (J. Bailey & Lloyd, 2016; Bailey, 2012). More recent evidence (Kurtmollaiev et al., 2018) shows that SD can be adopted successfully in order to improve an organisation's innovation capabilities. In their study of a large service organisation, Kurtmollaiev et al. (2018) find that top management can overcome the challenges of adopting SD in the organisation "by encouraging the creation of a service design based corporate language, by re-aligning KPIs with service design principles and objectives, and by providing room for experimentation" (ibid. p. 71). Other studies of large organisations' adoption of SD support these findings

(Madden, 2017; Miettinen, 2016). However, little is known about how SMEs' can successfully adopt SD as an approach to build inherent capabilities for continuous improvement and innovation work. This paper seeks to address this research gap by providing in-depth insights into the process of adopting SD in a medium-sized, service organisation.

The (missing) link between service innovation and service design

Service innovation and SD intuitively seems to be interconnected topics. However, it has been demonstrated that literature within these two research areas are still scattered and lack integration (Patrício et al., 2018). Studies have emphasised that service innovation is a priority in both service research and practice, due to the growing service economy, technological developments and increased globalization which challenges organisations' competitiveness (Ostrom et al., 2015; Patrício et al., 2018). Recent literature reviews have found that there are many different understandings and definitions of service innovation, which prevents knowledge development in the field (Snyder et al., 2016; Witell et al., 2016). In parallel, similar calls have been made to gain a better understanding of the service concept in order to advance knowledge of SD (Ostrom et al., 2015). From a research perspective, the gap between service innovation and SD is problematic because knowledge from both fields should be combined to develop the current discourse more holistically to establish further the research domains (Antons & Breidbach, 2018). While it is not the overall aim of this paper, this study contributes to strengthening the link between service innovation and SD research, by developing an understanding of how SD can support service innovation in SMEs.

The zone of proximal development as an analytical lens

This paper takes a socio-cultural perspective to discuss how SMEs can initiate learning activities that help to overcome the challenges of integrating SD as an approach to service innovation. In line with Holmlid & Malmberg (2018), the paper makes use of the concept of the Zone of Proximal Development (ZPD)(Cole, 1985) as an analytical lens to understand how members of an organisation develop knowledge through the participation in (practical) learning activities. The notion of the ZPD can be defined as the space between what a learner can do without help and where the learner needs support (Cole, 1985). In other words, the ZPD constitutes the edge where a learner (e.g. an organisational member) can succeed only with guidance from a mentor (e.g. a designer) or in collaboration with more capable peers (e.g. other organisational members with broader knowledge and skillset). These forms of mentoring are termed "scaffolding", which suggests flexible and temporary support that is enacted until the learning task is accomplished. At this point, the learner's ZPD has evolved, and scaffolding is moved to the edge of the now expanded ZPD (Cole, 1985). The underlying assumption behind the ZPD is that the development and instruction are socially embedded, which means that in order to understand these aspects it is necessary to analyse the context of the learning situation and its social relations. Thus, the notion of ZPD also shed light on the practice aspect in line with our understanding of SD. By considering the SD micro cases and surrounding activities at IU as learning activities, it is possible to analyse in which situations individual, organisational members reached their ZPD and further how adaption of SD tools and methods enabled them to overcome this challenge and thereby expand their ZPD.

Method

The study presented in this paper took place at IU. The research is part of a larger, 3-year action research project between the university and IU. We understand action research as a methodology, which implies that the research aims to induce change and improvement of certain aspects of the target research domain (Robson, 2002; Stringer, 2014). The overall project is comprised of three action research interventions. This study originates from the second intervention, which intended to build SD capabilities within the organisation as a way to address and advance service innovation.

The data collection happened over the cause of 17 months (February 2018 – July 2019). During this period, the first and second author spent 2-3 days a week at the case organisation, where they worked as an Industrial PhD student and part-time student worker, respectively. Both authors were familiar with the case organisation and trained in SD. This therefore created, what Holmlid and Malmberg (2018) describe as a rare setup in which the designers are also a part of the organisation where SD is being integrated. However, in this case, the authors were not hired as service designers per se, but rather as internal "motivators for SD". The authors' position allowed them to follow organisational processes from the inside, making continuous observations in situ, having formal and informal conversations with members of the organisation. Also, the overall frame of the Industrial PhD project provided a space for exploring and experimenting with the application of SD in this organisational context.

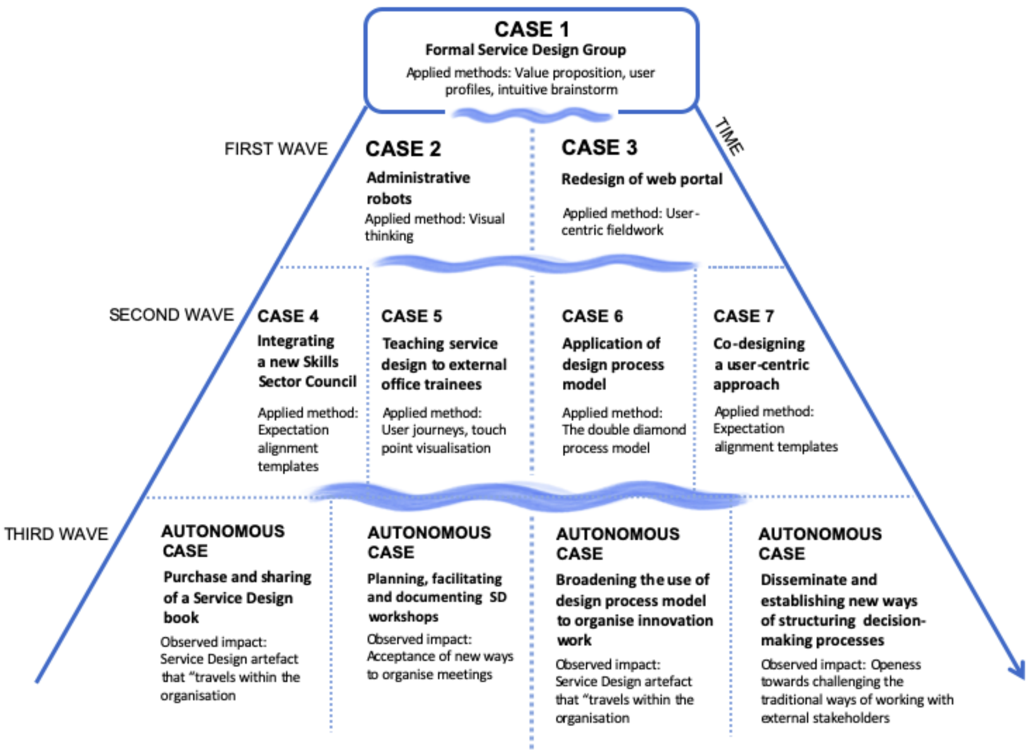


Figure 1. The evolution of SD micro cases at IU

SD was initiated in at IU through a proposal to establish a SD group (which will be explained in more detail in the Findings section). This proposal transformed into seven SD micro cases. The two first cases were selected by the authors based on (1) the perceived scope of the individual case, preferably as small as possible and (2) projects that were ongoing and at a nascent stage or planned to begin within the data collection period. Throughout the seven micro cases, the authors instructed 14 different learning activities, which included two customized SD compendiums, three 'miniature editions' (to offer organisational members short, condensed introductions to SD and with emphasis on specific elements or methods) and seven workshops, which each lasted 1-3 hours. During the workshops, the authors introduced the double diamond process model (UK Design Council, 2019), value propositions, empathy maps, user profiles, and ad-libs (Osterwalder et al., 2014), intuitive brainstorms (IDEO, 2015) visualisations tools such as user journeys (Kalbach, 2016) and service blueprints (Bitner et al., 2008) was introduced. As a way to introduce these tools and methods, 13 templates were appropriated or created to guide the practical learning activities. We documented the process by making audio or video recording the workshops, collecting the workshop outputs, conducting 12 meeting minutes, and taking 172 fieldnotes to capture informal chats, follow-ups, observations and reflections. We developed a shared system for precise and consistent record keeping, to ensure transparency and verifiability of data collected (Perecman & Curran, 2006). We emphasized a critically reflective practice, which closely related itself to the idea of learning from experience. This practice helped us to transform observations and reflections into subsequent actions, and we considered ourselves active participants in the organisational learning situations (Thompson & Pascal, 2012).

Our data analysis occurred in two main steps. First, we processed the data by categorizing the seven micro cases and detecting patterns by sorting the data based on twelve different case characteristics (e.g. the aim of the micro case, who and how many people were involved, and which methods and tools were introduced in each case). Based on this initial analysis, we identified 14 learning activities across the seven micro cases. We define a learning activity as actions that involves introduction to or collaborative use of SD tools and methods. An example of a learning activity is a SD workshop with management (this workshop is explained in more detail in the next section). By analysing these 14 learning activities resulted in the identification of 36 successful learning experiences, where organisational members expressed increased understanding or appreciation of SD. Due to the scope of the paper, we highlight four successful learning experiences to exemplify our analytical findings, which we present in the next section.

Findings

When trying to change cultural practices in an organisation, one can 'make waves' by challenging the status quo and initiate movements. In this section, we present our findings as three waves, which show how the developed SD learning activities built sustainable design capabilities within IU. First, we elaborate on our initial approach to integrate SD in the organisation and explain why this did not work out as planned. Then, we describe how our approach transformed into three waves of SD micro cases, which over time fostered a SD sustainable culture at IU.

Making waves: Initiating service design as an approach to innovation

We see a necessity to share our *adapted approach* to integrating SD at IU because this adaptation became our key catalyst to affect the organizational culture. We experienced this during the negotiation of what "form" SD should take within IU. A formalized group was not considered meaningful, as it would cause too big of a commitment and an additional load for a few selected employees. The learning gained from accepting a decentralized and informal approach to embedding SD as an approach to innovation was found in the fluidity of the approach. In this way, we could in the context of an SME, induce SD with a perception of less being at stake (especially in terms of committed resources), while reaching broader within the organization by exemplifying how SD could be contextualized to any given project and any given practice. We elaborate on this learning in the following paragraphs.

We initiated the study by proposing to establish a formalised SD team as a means to anchor and build SD capabilities within IU. This initiative was inspired by previous studies, which have reported on the use of internal resources as a beneficial way to anchor SD in an organisational setting (Lima & Sangiorgi, 2018). Moreover, establishing specialised teams were the most commonly used approach in IU to create cross-organisational collaboration to address overall issues. We proposed that this internal and cross-departmental team would get a crash-course in SD, allowing them to act as ambassadors with SD knowledge and practice. The proposal was that this "task force" should support other teams in the organisation by making use of SD tools and methods to address development work. The initiative was presented at a meeting with IU's management team. Despite our efforts to explain how this approach to SD could benefit the organisation, the six managers were reluctant and expressed concerns about their lack of resources. Also, one manager explained, "it is difficult to agree to this proposal, when you do not know what you are buying into" (manager, SD workshop, 20.09.2018). In this way, we identified a need to educate the managers about SD and showcase the use of SD through practice before they were able to decide whether to settle with a formalized SD team or not.

To broaden their horizon on SD, we designed a learning activity constituted a customized compendium with relevant resources, which took into account that the management team had little or no knowledge about SD. This aimed to function as a joint knowledge base. On this basis, the authors organised and facilitated a SD workshop, which took its point of departure in a project that was on the manager's agenda but had not yet been realised, due to limited resources. The project had the goal to develop an internal 'academy' to support the on-boarding process of new education consultants. The reason for making use of this project was to show the benefits of SD tools and methods through a use case that would simultaneously help the managers to progress with a stagnant project. In line with previous studies, we found that practising SD helped the management team to comprehend what SD is and how it could potentially help the organisation to become more innovative (Wetter-Edman & Malmberg, 2016). Making use of a concrete project as a way to mediate how SD can support service innovation helped the management to understand and internalise the benefits of SD. In this way, we found that contextualising SD is important. Working with SD tools and methods close to a relevant, concrete project was perceived very positively by the management team.

Despite the manager's positive experiences of using SD as an approach to development work, the decision about whether to establish a SD group was postponed. In the end, they proposed an alternative, which caused our approach to change. The initial idea of a formalised team transformed into decentralised "SD micro cases", which required less commitment and allocating of resources from a management perspective. This shift is central as it changed the perception of drawing (even harder) on existing resources, to the perception of allocation of additional resources (the researchers) to current projects. The decentralized micro cases aimed to incorporate SD in upcoming and ongoing development projects across the organisation. It became visible that there was a need to adapt the overall approach to the integration of SD. Our attempt to adopt formal structures did not work. Instead, we found that it was crucial to adapt our approach to account for the available resources, the current (lack of) design capabilities at IU and the context of the organisation. This was our opening to the everyday practices as well as the cultural and social context of the organisation. In the following, we describe how our new approach manifested as three waves of SD micro cases and elaborate on the impact of these waves.

The first wave of service design micro cases

The first wave exemplifies how it is possible to reinforce an emerging design culture by supporting the struggle that employees engage by approaching their daily practices in new ways. To overcome this struggle, employees need to be mentally prepared, for instance, through a "need to know" object, which can encourage them to go through the struggle. By ensuring alignment of expectations in a learning group, the learning environment is supportive and can stimulate collective, local learning experiences. By doing this, we learned that it is the motivation and positive experience of a ZPD expansion that feeds the "wave-making processes". We will elaborate on this learning here.

The first wave constituted two of the initially selected micro cases that were proposed by the management team. To illustrate this wave, we elaborate on one of the cases (2nd SD initiative, Figure 1), which focused on the exploration of possibilities for implementing administrative robots at IU. In this case, a project group aimed to identify potential work procedures that would benefit from automation. The group, which included two IT consultants and two members of the administrative department, were struggling with organising the identification and prioritization of the work procedures that potentially could be automated. As a way to incorporate SD tools and methods in this project, the authors suggested making use of visual thinking (Brown, 2019) their next project meeting. We proposed that they should visualise the processes in a manner inspired by "Customer Journey Mapping" (Stickdorn & Schneider, 2011). This is a well-known SD technique used to describe the service recipients as they operate and interact with touchpoints and service interfaces (Blomberg & Darrah, 2015). The project group agreed to approach the meeting in this way, which was new and different for all of them. Prior to the learning activity, the authors prepared a short document that explained what SD is and briefly introduced how visualisation tools can be used. We took into account that the project group had limited knowledge about SD and had different professions, and thus adjusted the document accordingly in order to prepare the participants mentally before making use of these new methods. As such, the document constituted the group's shared "need to know" object, which helped to align their expectations. When employees are motivated to make an effort to approach for instance a meeting situation differently, they engage in a struggle that goes beyond current cultural practices (in this context what it means to "have a meeting"). The moment in which this struggle immediately occurs can be described as the ZPD. This became visible during the meeting where one of the IT consultants was challenged, attempting to visualize a process on the whiteboard. He stated: "I do not know how to draw this, because I do not know this part of the process very well" (Meeting participant 15.10.2018). In this situation, the authors acted as mentors by suggesting that the IT consultant could draw a question mark (using signs) to express that there are steps in the process that needs further investigation. In this way, the IT consultant and the other group members extended their understanding of how they could make use of visual thinking in this context. For instance, one of the participants said "drawing the processes shows how many steps there are in each work procedure - how complex it is. It was good that you [the authors] suggested that we draw the processes" (Meeting participant 15.10.2018). Thus, by incorporating a SD learning activity as a part of a regular meeting allowed the participants to expand their ZPD. Moreover, by suggesting incorporation of SD elements in this way made the organisational members regard SD as a "generous offering" rather than a "bureaucratic burden", which leaves a positive impression of going through the struggle. These observations were further confirmed when one of the members of the project group showed how she had developed the visualisations from the meeting further. She did so by highlighting where value was created throughout their operational processes. This exemplifies the emerging interest for further exploration of the new tools and methods that occurred during this first wave. The group's knowledge about and positive experience with visual thinking was shared at the following "IU meeting" (a monthly meeting where management, departments, and employees share updates on projects and insightful experiences). We found that when learners share their positive learning experiences with their colleagues, they engage in "wave-making processes" which makes others curious to learn and expand their ZPD as well. They implicitly pass on the supportive environment they have experienced themselves, by ensuring others new to SD, that it is "safe" to welcome these new practices.

The second wave of service design micro cases

The first micro cases gave rise to an increased curiosity for SD thinking and induced organic growth of a new wave, constituting four additional micro cases. These micro cases differed from the first wave because firstly, they were put forward by organisational members rather than the authors or the management group. Secondly, they were put forward by members that all had been involved in one of the first micro cases from the first wave (see figure 2 below). The four identified micro cases all had different focus and objectives (see micro case 4-7 in figure 1), but all grew out of unforeseen changes or struggles experienced in daily procedures. Based on those changes or struggles, SD became an approach to gain a new perspective and a way forward. In the following, we elaborate on micro case #6 to illustrate the impact of the learning activities. This case generated learning in terms of stressing the importance of intentionally designing for repetitive participation of organizational members as well as a collaborative adaptation of methods and tools to ease the integration of new practices.

Micro case 6 constituted a SD initiative which aimed to understand how the inclusion of a design process model might support education consultants' wish to create room for and enhance innovation work when collaborating with appointed sector skills council. In this case, the structuring of the sector skills council had been rearranged, which offered a challenge for the consultants in terms of a mismatch of expectations to innovation processes, and the pace of concrete results being presented. The micro case was spurred by a department manager who had questioned whether the education consultants might be able to make use of a SD process model (UK Design Council, 2019) to (re)structure innovation work and redefine what was considered a result in the different phases of the innovation process. The department manager was inspired and had the idea from the SD compendium that the management team got prior to the initial SD workshop before the micro cases were initiated. However, the manager did not know how to apply the model in the context of an education consultant's everyday work practices and collaboration with the sector skill councils. In collaboration, the authors and two education consultants adapted the model to function in the context of their development work, which meant including a timeline to ensure external committee members that the development work would progress, while at the same time creating space for education consultants' creative problem solving and experimentation. These collaborative learning activities had a dual outcome: education consultants developed an understanding of SD and further expanded their knowledge by adapting the model together with the authors, who acted as mentors. Nevertheless, through this collaboration, the authors also gained a better understanding of how SD tools and methods could be adapted to the context of IU and its network of stakeholders and collaborators. This contextual understanding is vital to reach cultural changes, as it enables appropriation. The appropriation is a way to work around the challenge of integrating SD as an approach to service innovation. When the new ways of working fit with the daily context, it is easier to overcome the challenges of doing something new. When evaluating the adapted tool, one of the consultants stated that "because the tool was already adapted to them [external stakeholders] and their preferences [being a set timeframe] they thought it would be a useful way to address innovation work" (Education consultant 08.04.2019). This supports our finding of adapting tools and methods in collaboration with those that should be using it. This informal way of using SD and appropriating it is a way to include it in daily practices more efficiently. Another noteworthy observation is that all four emerging micro cases all included members of the organization that to a different extent, had been involved in one of the first micro cases, as mentioned above. This suggests that a decentralised approach to the integration of SD in SMEs is a good way to avoid "one-off projects", where the integration of SD will remain only on the introductory level. By designing for repetition, it is possible to disseminate knowledge and experience about SD across the organisation (Holmlid & Malmberg, 2018).

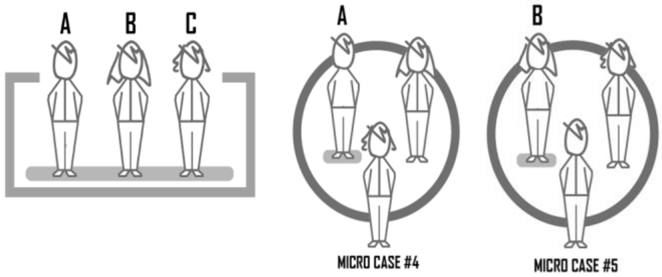


Figure 2. Dissemination of SD knowledge. SD micro cases support organisational members' repetitive participation in different SD learning activities.

The third wave: the impact of the service design micro cases

The third wave is a symbol of how cultural changes manifested itself at IU. This wave emerged without the involvement of the authors, new autonomous SD initiatives was observed, ranging from small initiatives such as a SD artefact circulating within the organisation, to the more comprehensive changes, where a department were all required to adopt the SD process model as their standard approach to innovation work. Those autonomous cases demonstrate how local learning experiences can evolve into more substantial structural changes, affecting the organisation and its culture at a broader level.

Over time we observed how the micro cases from the first and second wave surprisingly developed and generated new autonomous initiatives that were appropriated and incorporated in projects independently, without mentoring or guidance from the authors (See 3rd wave, figure 1). One example is based on observations of an administrative team, where several employees had participated in micro case #2. On their initiative, this team chose to expand their nascent SD knowledge and practice by buying and sharing a SD book. A team member shared that the aim was to develop their joint knowledge base and discussions on how they could make use of SD tools and methods to a greater extent as a way to innovate their internal procedures. Based on increased curiosity, the SD book started to travel from department to department, as a symbol of the increased interest of the new design capabilities that was starting to show locally in different departments. Another example of an autonomous SD initiative builds on our previous example of the education consultants who introduced the adapted SD process model in their sector skill council. They explained how their positive experiences of changing their development practices had created curiosity and awareness of SD in their department. Their dissemination of knowledge and use of the adapted model later resulted in an executive decision, which states that all educations consultants in the department should make use of this model as a tool to enhance innovation work in the skills sector councils. These autonomous SD initiatives support our findings stated above and suggest that this decentralised and informal approach is a valuable way for SMEs to overcome the challenges of successfully integrating SD as an approach to innovation, despite their inability to commit too many resources in doing so. Based on our understanding of design culture, these autonomous SD initiatives is a clear indication of a change in the dominant culture within the organisation. In other words, our study finds that our efforts to integrate SD through scaffolding and a range of learning activities have contributed to an emerging design culture at IU.

Discussion

So far, this paper has presented findings showing how SD capabilities can be built through learning activities with appropriated SD tools and methods to foster a sustainable design culture within a medium-sized, non-design-intensive service organisation. The paper has also shown that a decentralised and informal approach to adopting SD is useful in this organisational context, as it offers a fluidity that helps SD to reach broader with fewer resources allocated. With the designed and tested set of learning activities, the authors provide preliminary suggestions to how organisations can address the challenge of integrating SD as an approach to service innovation, and how to sustain this approach replacing it with previous practices in non-design intensive SMEs.

When designing learning activities as a way to build SD capabilities in an organisational context it is essential to recognise that while the majority of development work emerge from collaborative practices, the ZPD is different for each member of the organisation. This means it should not be expected that "generic" introduction to SD tools and methods will result in a growing design culture from within the organisation. This finding reflects that of

Holmlid and Malmberg (2018) who also found that knowledge about SD "is not enough to drive the aspired transformation and integration" (ibid. p. 46). Moreover, Blomberg and Darrah stress "no matter how well we understand the practices of a community, it is dangerous to assume that the objects of our designing can simply be inserted in those practices" (2015, p. 52). This emphasises the importance of understanding how SD tools and methods need to be appropriated for a specific organisational context in order for them to be embedded in everyday practice and thereby drive organisational change and prompt service innovation. In our case, it was initially the authors that proposed suggestions for the appropriation of the tools and methods. However, this changed during the 2nd and 3rd wave of the SD initiatives, as the involved employees began to act more as capable peers and, in this way, disseminated knowledge to additional members of the organisation. This transformation occurred due to the organisational members that participated in more than one SD initiative (figure 2). On this basis, we suggest designing for repetition (e.g. to plan for employees' repeated involvement in SD initiatives) as a way to scaffold the organisational members. At the same time, they learn to adapt and apply SD tools and methods in their everyday work practices.

Another way to support the integration of SD is by developing a joint, contextualised knowledge base that supports the temporality of the ZPD. Despite the individual nature of the ZPD, there are times where a group needs to coincide. These moments can be promoted when learners are provided with explicit material about a relevant topic. During the initial phase of this study, the management team needed to develop a mutual understanding of the value of SD in order to decide on whether to establish a formalised SD group or not. To support this decision-making process, the authors created a compendium on "SD at X" that provided the managers with explicit and carefully selected resources. After proper appropriation, the compendium became a central object, which guided the group to discover their interpretations and expressions of the tools and methods concerning their organisation. This shows it is highly relevant to question what and how much is necessary for organisational members to know in order to embed the knowledge in their everyday practice. In this way, the "need to know" object became a structure, a guideline for how to make sense of SD. However, such a structure should only be considered temporary. Once an individual learner or group has grasped the new knowledge, it is necessary to update or even remove the structure in order to create a new scaffold at the given time and space. An example of how this temporality manifested itself, in this case, is the evolution of the short document to a shared book (see micro case #2, figure 1). This enabled the department to expand their ZPD on their initiative. This further exemplifies how the process of adapting SD tools and methods at a local level implies a reflective process among the organisational members, which can lead to the development of a local learning process. In our case, the various learning activities, which were initiated during the first and second wave of SD micro cases, supported the development of such local learning processes, fostering the emergence of a sustainable design culture at IU.

Before presenting the conclusions, it is interesting to come back to our failed attempt to integrate SD at IU through a formalised and centralised structure. The need to change from a formalised to a decentralised approach suggests that it might be necessary for SMEs to adopts SD differently compared to large organisations (Kurtmollaiev et al., 2018). It was not until we addressed the integration of SD as an intrinsic part of everyday practice that we observed organisational transformation and the emergence of a design culture at IU. Thus, it is crucial to acknowledge that service designing includes participating in a social context and therefore, it is necessary to appropriate tools and methods to this context. This helps to embed local conventions in the emerging SD practices that, in our case, fostered a sustainable SD culture.

Conclusion

This paper shows that to integrate SD in organisations as an approach to advance service innovation is not merely a question of providing a SD toolbox. On the contrary, it is crucial to adopt an understanding of SD as ambiguous, diffuse and as an intrinsic part of everyday practice. This allows SMEs to divert from the need to establish end-to-end SD projects or specialised SD teams, which may be too resource-demanding for a smaller organisation. Instead, taking a decentralised and informal approach to the integration of SD enables the members of the organisation to apply relevant tools and methods as a part of their work practices step by step. Our study shows that this approach develops design capabilities, and over time fosters a sustainable SD culture within the organisation. We propose three lessons learned for practice that can help non-design intensive SMEs to integrate SD as an approach to service innovation successfully. First, it is essential to actively involve organizational members in the appropriation of service design tools and methods as this helps to embed local conventions in the emerging service design practices. Second, to design for repetition, meaning that members of the organization are involved in several service design initiatives, can function as a way to scaffold the organisational members. At the same time, they learn to adapt and apply service design tools and methods. Finally, we propose to develop a common, contextualized knowledge to support the temporality of the employees' zone of proximal development. Overall, this study contributes to our understanding of how SMEs can appropriate SD tools and methods to their cultural practices in order to build sustainable SD culture.

References

Antons, D., & Breidbach, C. F. (2018). Big Data, Big Insights? Advancing Service Innovation and Design With Machine Learning. *Journal of Service Research*, *21*(1), 17–39. https://doi.org/10.1177/1094670517738373

Bailey, J., & Lloyd, P. (2016). *A view from the other side: Perspectives on an emergent design culture in Whitehall*. 13.

Bailey, S. G. (2012). *Embedding service design: The long and the short of it*. 11.

Bitner, M. J., Ostrom, A. L., & Morgan, F. N. (2008). Service Blueprinting: A Practical Technique for Service Innovation. *California Management Review*, *50*(3), 66–94.

Blomberg, J., & Darrah, C. (2015). *An Anthropology of Services: Toward a Practice Approach to Designing Services* (Vol. 8). Morgan & Claypool Publishers.

Blomkvist, J., Holmlid, S., & Segelström, F. (2010). *Service Design Research: Yesterday, Today and Tomorrow*.

Brown, T. (2019). *Change by design: How design thinking transforms organizations and inspires innovation* (Revised and updated edition; B. Katz, Ed.). New York: HarperBusiness, an imprint of HarperCollinsPublishers.

Cole, M. (1985). The zone of proximal development: Where culture and cognition create each other. In J. V. Wertsch (Ed.), *Culture communication, and cognition: Vygotskian perspectives*. Cambridge [Cambridgeshire] ; New York: Cambridge University Press.

Curedale, R. (2019). *Design thinking: Process et methods*.

Holmlid, S., & Malmberg, L. (2018). Learning to design in public sector organisations: A critique towards effectiveness of design integration. *ServDes2018. Service Design Proof of Concept, Proceedings of the ServDes.2018 Conference*, 12. Milano, Italy: Linköping University Electronic Press.

IDEO (Ed.). (2015). *The field guide to human-centred design: Design kit* (1st. ed). San Francisco, Calif: IDEO.

Julier, G. (2006). From Visual Culture to Design Culture. *Design Issues*, *22*(1), 64–76. https://doi.org/10.1162/074793606775247817

Kalbach, J. (2016). *Mapping experiences: A guide to creating value through journeys, blueprints, and diagrams*. Beijing ; Boston: O'Reilly.

Kurtmollaiev, S., Fjuk, A., Pedersen, P. E., Clatworthy, S., & Kvale, K. (2018). Organizational Transformation Through Service Design: The Institutional Logics Perspective. *Journal of Service Research*, *21*(1), 59–74. https://doi.org/10.1177/1094670517738371

Lima, F., & Sangiorgi, D. (2018). *Fostering a sustained design capability in non-design-intensive organizations: A knowledge transfer perspective*. 13.

Madden, H. (2017). *Designing Change in a Higher Education Institution* (Thesis, Cardiff Metropolitan University). Retrieved from https://repository.cardiffmet.ac.uk/handle/10369/9930

Malmberg, L. (2017). *Building Design Capability in the Public Sector: Expanding the Horizons of Development*. Linköping University Electronic Press.

Meroni, A., & Sangiorgi, D. (2011). *Design for Services*. https://doi.org/10.4324/9781315576657

Miettinen, S. (2016). *An Introduction to Industrial Service Design*. https://doi.org/10.4324/9781315566863

Osterwalder, A., Pigneur, Y., Bernarda, G., & Smith, A. (2014). *Value proposition design: How to create products and services customers want*. Hoboken: John Wiley & Sons.

Ostrom, A. L., Parasuraman, A., Bowen, D. E., Patrício, L., & Voss, C. A. (2015). Service Research Priorities in a Rapidly Changing Context. *Journal of Service Research*, *18*(2), 127–159. https://doi.org/10.1177/1094670515576315

Patrício, L., Gustafsson, A., & Fisk, R. (2018). Upframing Service Design and Innovation for Research Impact. *Journal of Service Research*, *21*(1), 3–16. https://doi.org/10.1177/1094670517746780

Perecman, E., & Curran, S. R. (Eds.). (2006). *A handbook for social science field research: Essays & bibliographic sources on research design and methods*. Thousand Oaks, Calif: Sage Publications.

Polaine, A., Lovlie, L., & Reason, B. (2013). *Service design: From insight to implementation*. Brooklyn, NY: Rosenfeld Media.

Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers* (2nd ed). Oxford, UK ; Madden, Mass: Blackwell Publishers.

Sangiorgi, D., & Prendiville, A. (Eds.). (2017). *Designing for service: Key issues and new directions*. London ; New York, NY: Bloomsbury Academic, an imprint of Bloomsbury Publishing Plc.

Snyder, H., Witell, L., Gustafsson, A., Fombelle, P., & Kristensson, P. (2016). Identifying categories of service innovation: A review and synthesis of the literature. *Journal of Business Research*, *69*(7), 2401–2408. https://doi.org/10.1016/j.jbusres.2016.01.009

Stickdorn, M., & Schneider, J. (Eds.). (2011). *This is service design thinking: Basics - tools - cases* (2. pr). Amsterdam: BIS Publ.

Stringer, E. T. (2014). *Action research* (Fourth edition.). Los Angeles, California: SAGE.

Thompson, N., & Pascal, J. (2012). Developing critically reflective practice. *Reflective Practice*, *13*(2), 311–325. https://doi.org/10.1080/14623943.2012.657795

UK Design Council. (2019, September 3). The Double Diamond: 15 years on. Retrieved 8 October 2019, from Design Council website: https://www.designcouncil.org.uk/news-opinion/double-diamond-15-years

Wetter-Edman, K., & Malmberg, L. (2016). *Experience and expertise: Key issues for developing innovation capabilities through service design.* 6.

Witell, L., Snyder, H., Gustafsson, A., Fombelle, P., & Kristensson, P. (2016). Defining service innovation: A review and synthesis. *Journal of Business Research*, *69*(8), 2863–2872. https://doi.org/10.1016/j.jbusres.2015.12.055

Publication 5

Seidelin, C., Dittrich, Y., Grönvall, E. – **Co-designing Data Experiments: Domain Experts' Exploration and Experimentation with self-selected Data Sources**. [Submitted to the NordiCHI Conference 2020]

Co-designing Data Experiments

Domain Experts' Exploration and Experimentation with self-selected Data Sources

Cathrine Seidelin†, Yvonne Dittrich, and Erik Grönvall

IT University of Copenhagen, Copenhagen, Denmark, cfre@itu.dk

ABSTRACT

Today, organisations have to deal with multiple heterogeneous data sources from different systems and platforms to maintain and develop their services. However, there is a need for tools to support organisations to determine what data can advance and innovate their services. As part of a larger action research project, we addressed this need by creating two design tools – the Data Sphere and the Data Experiment Template – aim to support domain experts' exploration and experimentation with self-selected data sources. We describe how we tested and evaluated the tools with employees in a Danish organisation. We find (1) the tools' comprehensive and tangible guidance support domain experts to work creatively with data and (2) that data experimentation reveals the benefit of co-design to the domain experts. We use the results to reflect on our process and propose directions for future investigations on tools that can support domain experts to co-design with data.

CCS CONCEPTS

• Insert your first CCS term here • Insert your second CCS term here • Insert your third CCS term here

KEYWORDS

Co-creation, Data Work, Design Things, Data Experiments, Data Sphere, Organisation

INTRODUCTION

Several scholars have pointed out that we are living in an increasingly connected world as we continue to interact with digital technologies and data [27,33,40]. The vision for organisations to become more "data-driven" has created a need for tools that can help domain experts, who are not IT-professionals, to engage in the exploration of and experimentation with data and data sources in their organisations, for instance, to improve their own (data) work practices or support innovation of digital services [37]. Currently, it often requires highly specialised skills to work creatively to make further sense of what data sources are useful and how they might support the innovation of processes, products, and services [4,5,43]. This prevents non-ICT domain experts from engaging with data in innovative ways [6,26].

In line with the tendency to consider data as increasingly important in society, at work and everyday life, there is an emerging body of work that explores "data interactions" within the ICT design research community (including conferences like DIS, CHI and CSCW). Notable examples are Feinberg [16] who has proposed a design perspective on data as a lens that emphasises, e.g. data collection as a design activity, Kun et al. [28] who examines how designers incorporate data work in the design process or Dove [12] who examines the use of domain data in the context of co-design workshops. Other HCI scholars have also begun to examine how data can be (re-)presented in ways that enable domain experts to better make sense of data [13,51], how data influences participatory processes [8,45], or explore ways to enhance people's awareness about wirelessly transmitted data [20]. A commonality for these studies is that the data sources used to enhance data literacy, prompt data work in design activities or improve data visualisation, are predefined by the designers. Thus, little is known about how to support domain experts', who are not IT professionals, identification, exploration, and experimentation of (new) data sources that can prompt data-driven innovation in an organisational context. This paper addresses this research gap by exploring how to encourage domain experts to help themselves on the road to identify and work creatively with data sources that could be useful to their work and their organisational context.

In this paper, we report on a project which was part of a larger 3-year action research project [21,22,41,47]. The action research project was situated at Industriens Uddannelser (English: The Education Secretariat for Industry, henceforth: IU), which is an organisation, based in Copenhagen, that works to maintain and develop vocational education and continuing educations in the Industrial sector in Denmark. The project we report on in this paper constituted the third and final action research intervention, which had two overall objectives: (1) it aimed to develop a design process that could enable the domain experts in the organisation, who are not IT-professionals, to identify, explore, and experiment with self-selected data sources, as a way (2) to advance innovation of data-based services in the organisation. The intervention was designed as a process which included six workshops and three so-called "engagement elements". The project was executed and evaluated in close collaboration with a project group that constituted five employees from IU.

Throughout the project and in this paper, we consider data as "design things" [15] to acknowledge that data have 'agency' which to different extents modify the design process, its outcome(s), and its subsequent uses. We take inspiration from works that discuss data as a "design material" to inform the design of our design tools, concepts and activities [14,16,24]. In this paper, the focus of our analysis is on two design tools we created as part of this design process to prompt experimentation with data. (1) *The Data Sphere* aimed to involve the whole organisation in collectively designing ideas for potential data sources that could be further explored and experimented. Based on the co-designed "data source ideas" from the Data Sphere, (2) *The Data Experiment Template* aimed to support domain experts' ability to concretise and implement experimentation with data alongside their other tasks and projects. Thus, the principal contributions of this paper are the two tools and the description of (1) how they were developed and (2) how they can be used.

In the remainder of the paper, we detail our research activities and present our findings from the field. This paper contributes to the growing body of HCI research that considers data from a design perspective, by discussing three key insights from our exploratory design process and suggests directions for future work.

RELATED WORK

To consider data from a design perspective is still a nascent research area in HCI with a rather small catalogue of examples (including [12,16,29]). Most research focuses on the design of (digital) interfaces and artefacts that can represent data and make it manipulable for domain experts and end-users [52]. However, it does not question the data source itself. Because of this, common data visualisation and data exploration design exercises do not specifically account for the opportunities and challenges of tools that support domain experts' exploration and experimentation of self-selected data sources. We address this gap by proposing two design tools for this purpose. To situate this work, we provide an overview of work that touch upon challenges for co-creation in data interdependent settings, consider data from a design perspective, and discuss design activities for data exploration.

Co-design and Data Work

The growing amount of data production, collection and usage have generated an increasing level of "data interdependence" between organisations, which is challenging to comprehend [45]. Previous research has emphasised that the increased level of connectivity creates challenges for how to establish and co-design for such complex settings [9,33]. Degnegaard [9] argues that these complex settings, where 'value' constitutes a "dynamic, liquid, ever-changing potential across stakeholders and between stakeholders", which imply that a single individual or organisational entity can no longer be pinpointed as the centre of concern. To account for this high level of connectivity, we have applied a co-design approach in this work of designing tools that support domain experts to work creatively with data sources. In this work, we refer to co-design in a broad sense to conceptualise when "*people come together to conceptually develop and create things/Things that respond to certain matters of concern and create a (better) future reality."* [53:12].

Furthermore, we make use of the concept of Data Work [2,17,18,35] as a lens to help us think about the complexity that is included when considering identification and experimentation of data sources in an organisational context. The notion of data work has been coined to address the significant increase in the amount of work that is related to data in some sense, in recent years [2]. Data work has been defined as "any human activity related to creating, collecting, managing, curating, analysing, interpreting, and communicating data" [2:466]. By emphasising these many aspects of data related work, it becomes clear that it requires various encounters between people, technologies, and data, to make data 'work' (e.g. to enable data collection or application of useful data). As Bossen et al. [2] emphasise, these encounters are situated in particular places at a given time. When organisations wish to be able to work innovatively with data for a given purpose, it implies that the organisation is well aware of the many encounters and processes that go into its current data practices. Thus, the notion of data work is relevant in our case as it addresses the need for local knowledge about current data practices, which is important in order to develop tools that can support an organisation's (and the people within it) exploration and experimentation with data. In the next section, we elaborate on our understanding of data.

Data as 'design things'.

The prevalence of data as a mind of information has led to enormous growth in collecting data that is used to influence decisions in various aspects of society and social life [19]. This development has meant that data play a critical role in people and organisations' empowerment [23]. As one response to this tendency, the area of "Human-Data Interaction" has emerged to emphasise research that examines how people interact with data [23]. The emerging body of work seems to assume that in order to get insights from data, people need to interact with data rather than passively consuming them [7,34,48]. This suggests that this form of interaction goes beyond data analysis and includes exploration of data [23].

Although the emergence of Human-data Interaction research points to a need for a distinct research area, HCI researchers have also begun to address the growing necessity to consider data as a fundamental component that shape how people (can) interact with technologies. For instance, Feinberg [16] propose a design perspective on data and show how data collection can be considered a design activity. Muller et al. [36] propose to develop a human-centred study for data science practices. Others have explored the role open data play for local policymaking processes [8,25]. Together, these perspectives open up for discussion on whether or how can data be considered as "design material" that can act as subject to co-design like other physical or functional dimensions of an IT solution in the design process. Ehn [15] questions what is meant when we consider objects and things in design. Drawing on work by Latour [30], he emphasises that "design things" are essential when we deal with 'agency' of both human and non-human actants. To build on the perspective of data as design material, we consider data as 'design things' in this paper. We do so to acknowledge that data have agency that to different extents modify the design process itself, the design outcome of the design process, and its subsequent use. In the next section, we take a closer look at work that, in different ways, have attempted to give agency to data through representation in the design process.

Design activities for data exploration

Previous HCI research show examples of proposed different design concepts that aim to support people in making sense of data to enable them further to work exploratively with data. Data Literacy is a concept, which describes the competencies around the use of data in order to reason, e.g. for problem-solving. As such, Data Literacy is increasingly considered to be a vital skill to gain and maintain in order to be able to make sense of data, data analysis, and data representations[10,52]. Many works on Data Literacy have emerged in literature on informatics, education, and information literacy research, which has generated a focus on how people interpret and evaluate the effectiveness of digital data [11,32,39]. However, some Data Literacy studies have also examined ways to prompt more explorative aspects of making sense of data [50]. One example includes Wolff et al. [50], who designed a board game to support people's understanding of "the relationship between data, the environment from which it derives, and the questions it can be used to answer".

Another branch of research that attempts to foreground that data has agency and thus influence the design of digital interactions, is Information Visualization. HCI researchers have examined how Information Visualisations can be applied in design projects as tools that can increase accessibility, and thus support people's engagement with and understanding of data [12,13,49]. Others have attempted to prompt data exploration by appropriating a data science workflow to the early stages of the design process [28,29].

The abovementioned work gives preference to visual materialisation. However, Lupton (2017) emphasises a countertrend in HCI that explores how other senses can contribute to making sense of data. One example is "FeltRadio", a program that gives haptic sensations consisting of electrical impulses every time an app or a website transmit information to third parties. Thus, the project invites an individual to explore their data usage and (unintentional) data production by making use of haptic sensation [20]. More broadly, the attention to rendering data as 3D artefacts is known as Data Physicalization [24,54]. The underlying assumption of this work is that multisensory experiences are better understood than those where only the visual dimension is used.

As an addition to this prior work, our proposed design tools expand the space for design activities for data exploration by supporting domain experts' exploration and experimentation with self-elected data sources.

A DESIGN PROCESS FOR EXPLORATION AND EXPERIMENTATION WITH SELF-SELECTED DATA SOURCES

One of the main objectives of this project was to design a process that would enable domain experts at IU to identify and further explore and experiment with data sources. Based on earlier insights from the action research project, we designed a process which revolved around six workshops and included three additional "engagement elements". Figure 1 below illustrates the overall process and primary research activities. The process was designed to support the project group during their venture into the unknown work of questioning what constitutes data in their everyday work practices (WS1, Figure 1) to develop recommendations on how to engage creatively with data sources as part of innovating services in the organisation (WS6, Figure 1).

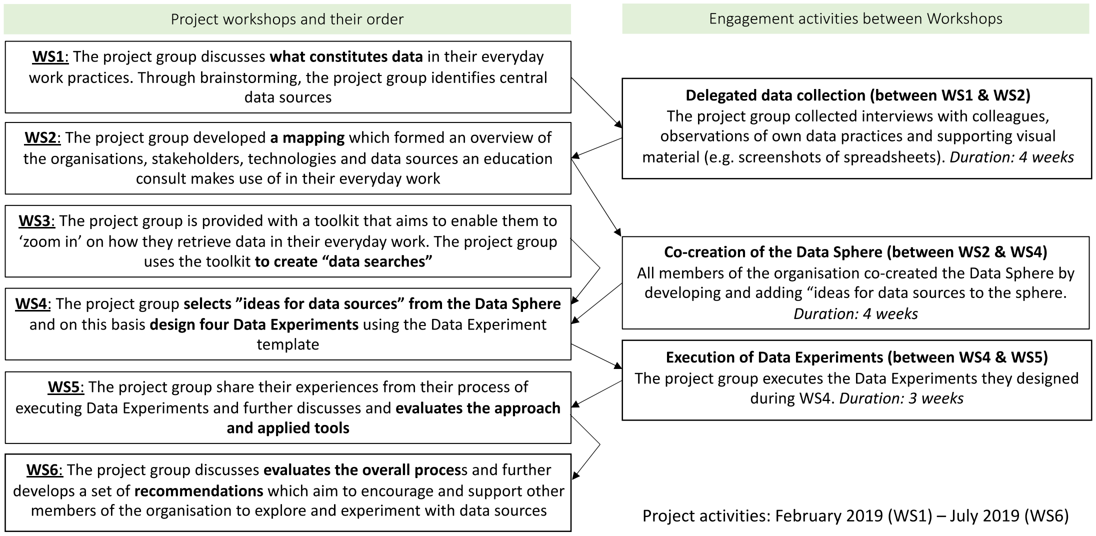


Figure 1: Overview of the design process and activities. The project started with Workshop 1 (WS1) and ended with Workshop 6. Three additional activities took place between the workshops.

To support progress in each workshop constituted the basis of the following workshop. The workshops, therefore, had different objectives and included changing exercises. The engagement elements were included to create awareness of the project in the organisation and to involve knowledge of additional domain experts beyond the project group. However, the members of the project group were the ones that were most involved throughout the process. The project group comprised of five employees (referred to as E1-5), who worked as education consultants at IU. Education consultants comprise more than half of the workforce at IU. Their respective manager had appointed these employees because they had shown particular interest in improving data work or new data-intensive technologies. Given our overall action research approach, we aimed to involve the members of the organisation as a way to develop and design the process based on co-created knowledge throughout the process [21,22]. Thus, the members of the project group played an active role in the design of the process, concepts and tools through their actions and ongoing feedback. However, the project was managed, and the design tools were developed by the first author and implemented with the assistance of IU's internal service designer.

The project ran from February to July 2019. All six workshops lasted between 2-3,5 hours and were video and audio recorded as a way to document our research activities. During the process, the first author also conducted ad hoc interviews and observations to support the project group and to understand how the process and design tools could be adapted and improved. Interviews were audio-recorded and transcribed, and observations were documented as fieldnotes [21]. Overall, this constitutes a rich body of empirical material. This paper focuses on the later stage of the process and builds on empirical data from the co-creation of the Data Sphere to WS6. We do so because these parts, in particular, emphasise the experimental aspects of our design process. Before we elaborate on the research activities, we explain our two design tools - the Data Sphere and the Data Experiment Template - in the following sections.

The first design tool: The Data Sphere

The Data Sphere aims to prompt domain experts in an organisational context to generate ideas for new (use of) data sources that might improve work practices and means for service innovation. The notion of sphere refers to a space over or within which someone or something exists or has influence [38]. Organisations are existing and navigating in an increasingly connected world as a result of the growing use and implementation of digital technologies and data [27,33,40]. Therefore, it could be argued that data to a varying degree influence an organisation's sphere. Thus, the first author and IU's internal service designer developed the Data Sphere as a way to explore how an organisation and its members can engage with this increasingly influential space of data.

Figure 2 shows how the Data Sphere is made up of a wall poster (3,5*3 meter) with a mapping at its centre. The mapping was, in this case, developed by the project group in WS 2 and visualised "the world of an education consultant at IU". Thus, the mapping depicts human actors (colleagues, stakeholders, organisations, businesses, etc.) and non-human actors (technologies and data sources) that an education consultant interacts within their everyday work life. The purpose of placing the mapping at the centre of was to situate and spark creativity for the development of data source ideas. The Data Sphere in itself is represented as the space surrounding the mapping.

As a means to populate the Data Sphere, we designed a form for "data source ideas", which included the following factors: *Name of the data source, Where does the data come from?, What kind of data is it?, Why is it an inspirational data source?*. The design of the form was intended to guide the members of the organisation. Based on insights from earlier interventions in the action research project, we had learned that it was essential to design tangible and somewhat structured tools in the context of promoting the domain experts to work creatively with data [45,46]. Additionally, the management required that the forms should include space for the employees' name because they wanted to get an idea of who participated in the co-design of the Data Sphere.

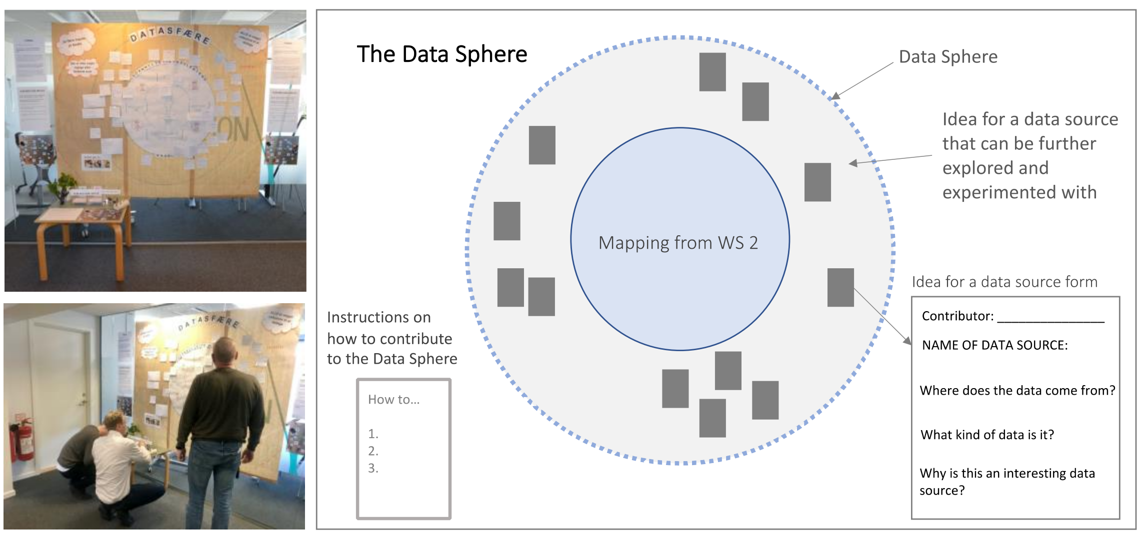


Figure 2: From left: The pictures show the original Data Sphere and how members of the organization engaged with the initiative and contributed with ideas for data sources. The right side shows a generic representation of the Data Sphere and its main components: the mapping at the centre, the sphere and data source forms.

The second design tool: The Data Experiment Template

The Data Experiment Template aims to support domain experts to explore and experiment with data sources they have identified as potentially useful or valuable. The action research project had made visible that the organisation had minimal experience with prototyping and testing ideas and concepts in "designerly" ways [46]. To accommodate this condition, the first author and IU's internal service designer developed a template that reminded of a "recipe" to support the project group step by step in order for them to progress from identifying a data source idea to proposing a design for how this idea might be explored. The template included the following sections: Title, Aim, How is this experiment challenging current or pointing to new ways of working with data?, Assumptions, Lessons learned, Practicalities (see also Figure 3).

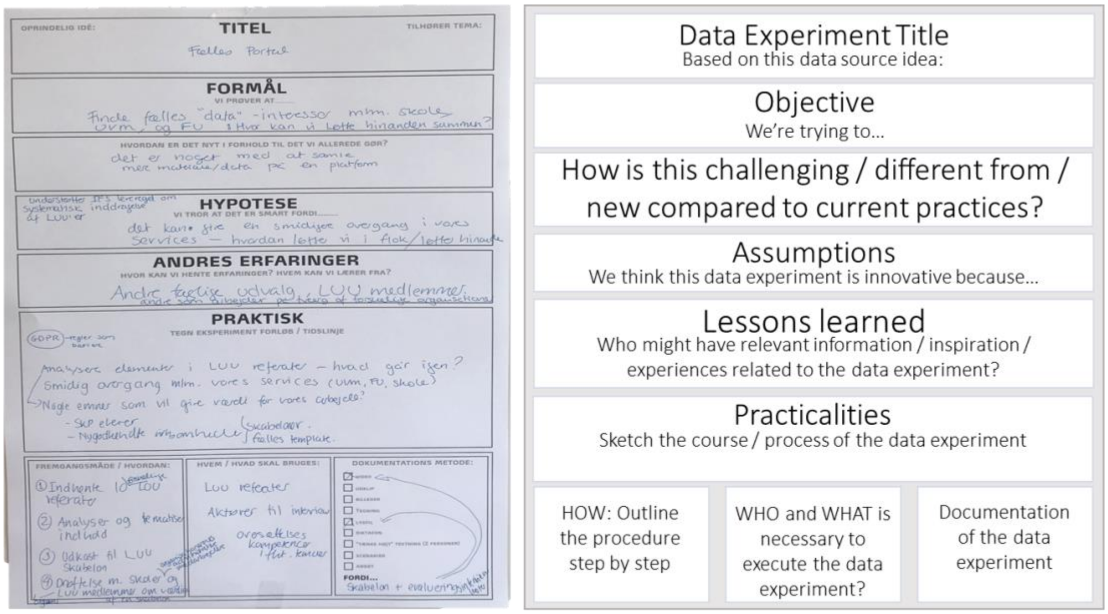


Figure 3: The Data Experiment Template. On the left: an example from the design process. On the right: a translated remake of the template.

From Data Sphere to Data Experiments

The two proposed design tools were implemented during the process in the following way: The Data Sphere was placed the central hallway at IU and introduced to the members of the organisation the monthly information meeting. It was announced that the Data Sphere would constitute the foundation for the project group's next workshop (WS 4) one month later. During this time, the members of the organisation created and contributed with 40 different data source ideas using the forms.

Then, during the initial stage of the fourth workshop, the project group was asked to process these ideas in two rounds: In the first round, the group was asked to examine each of the 40 idea forms and physically move it from the Data Sphere to a whiteboard. The second round focused on idea selection based on weighted parameters [42]. Thus, the members of the group were asked to discuss five parameters to identify which of the ideas were useful for further exploration and experimentation. The aim was for the group to prioritise and select four ideas that should constitute the point of departure for the design of Data Experiments. The group decided on the following: *Quality*, *Value*, *Resources*, *Competences*, and *News Value*. Then, they collectively graded each idea on each parameter from 1-5. The ideas with the highest score were discussed amongst the group, who selected four ideas to use for the design of Data Experiments.

The project group collectively used and populated Data Experiment Templates to design the four Data Experiments based on the selected ideas (see Table 1). Then, the Data Experiments were divided between the members of the group: each member was involved in the execution of two Data Experiments. After three weeks, the project group gathered again to share their experiences and evaluate the exploration and experimentation phase of the process (WS 5). During this time, they had managed to carry out two of the four Data Experiment in full. Finally, the group evaluated the process of co-designing Data Experiments, and on this basis developed recommendations to support other members in the organisation to set out to identify, explore and experiment with self-selected data sources (WS 6).

Table 1: Overview of the four Data Experiment designed by the project group.

Experiment	1	2	3	4
Title	Data about Elective Specialization Courses	Colleagues as a data source	Phone interviews for trend spotting	Open Source System for all
Data source	Structured, quantitative data	Unstructured qualitative data	Structured qualitative data	semi-structured quantitative and qualitative data
Objective	To explore how 1) identify relevant users of this data source and 2) to develop a suitable way to share the data	To explore how colleagues can be considered a valuable data source	To explore how phone interviews can be used to spot trends in Industry in a structured way	To explore and identify common "data interests" with selected external stakeholders to improve cross-organisational data work
Newness	Existing data source, exploring new ways of organising the data work	Existing data source, exploring whether/how colleagues can be considered a data source	Existing data source, exploring new ways of organising the data work	New data source

ANALYSIS OF THE DESIGN ACTIVITIES

Through conducting the project, we explored two design tools and related design activities for creating pathways may support domain experts to explore and experiment with data sources to promote data-driven service innovation in the organisation. The process uncovered specific strengths and limitations of the Data Sphere and the Data Experiment Template. In this section, we summarize our observations and present key challenges and benefits.

The Data Sphere: Understanding the organisation's data usage

A great benefit from the implementation of the Data Sphere was its ability to create a design space in the organisation that prompted discussions related to data usage between managers and employees across divisions and teams. Informal conversations with and observations of people in the hallway showed that they were happy to be included in the process. Moreover, the involvement increased their curiosity about the project. The display of the project group's mapping of "the world of an education consultant at IU" also enabled feedback and questioning of this representation. However, the Data Sphere also challenged the organisation in the sense that it invoked discussions and reflections about what constitutes as data in the context of various work practices and cooperation both with other members of IU and external stakeholders. Thus, the inclusion of the entire organisation ensured both generation of ideas from various members of IU, increased awareness about the project, and prompted reflection concerning the data work undertaken in the organisation.

During the processing of the Data Sphere, the project group categorised the 40 data source forms by dividing them into three overall categories: "Already applied data sources" (20), "New data sources" (17), and "Data Attitude Statements" (3). Based on a joint discussion, the group members decided to exclude the "Data Attitude Statements" because they were expressions of opinions about data and did not include or refer to any particular data source. The group also discussed whether to include "Already applied data sources" because, they argued, the Data Sphere was meant to expand how they include new data sources in their work and the organisation's service innovation. However, the group chose to include the category of ideas in the further process because "*what is a well-known [data source] to me might be new and unfamiliar to my colleague"* (E4, WS 4).

Working with the ideas from the Data Sphere also made visible that it is challenging to distinguish between one's Data Sphere and the organisation's Data Sphere. When discussing how to categorise the ideas, a group member elaborated on their understanding of the difference between one or the other form of Data Sphere: "*We are talking about what constitutes the employee's data sphere, and what constitutes the organisation's data sphere. There is much qualitative information that is part of your data sphere; for example, something that we do not have time to document or informal conversations. And then there is the organisation's data* *sphere, which includes what we report, and which forms we populate*" (E2, WS 4). The need to distinguish between one personal data sphere and the organisation's data sphere is an interesting finding because it illustrates how data work happens at different "levels" in an organisation, and thus how it is necessary to be aware of what "level" you are designing for.

Collectively designed parameters are a useful way to select data source ideas

Our design process differs from previous work because it aimed to support organisational members to identify data sources for subsequent exploration and experimentation themselves. As a way to help members of the project group to assess the Data Sphere ideas, they were asked to decide collectively on five parameters. This approach supported the group to state their reasons for and reflected on the 40 data sources. Interestingly, we observed how this approach created common ground within the group by prompting them to argue for their verdict of each of the parameters for each of the selected ideas. The excerpt below illustrates how the group members aligned the meaning, which was attributed to each of the parameters (E2 and E3, WS4):

E2: *"I immediately think of data quality."*

E3: *"Yes, that is a good start. Write that [on the board]. I also think 'Validity' is important."*

E4: *"But I think that [validity] is a subcategory of data quality. Reliability and Validity belong under Data Quality, right?"*

E2: *"Yes, but that also depends on how many parameters we have, oh right five, then yes, I agree."*

Through these ongoing negotiations among the group members, the activity supported the group's ability to select ideas that suited their situatedness. The discussions on and use of the selected parameters made it easier for the individual group members to argue why, e.g. "Data about Elective Specialisation Courses" (Experiment 1, see table 1) should score 5 for "Value". This indicates that the activity is a useful way for a group of domain experts to analyse the co-designed Data Sphere.

Another interesting observation was how this activity made visible to the group that they, as an organisation, are dependent on other stakeholders in the broader network: If IU changes their data work, it will most likely influence other actors' data practices [45]. When evaluating ideas for the subsequent design of Data Experiments, the project group discussed, for instance, the competencies that it requires to retrieve data about "Elective Specialisation Courses". The excerpt below illustrates part of this discussion (E2 and E3, WS 4):

E3: *"Yes, that is a 1 [very easy to retrieve]! I have been in contact with [employee at the governmental agency for IT and learning], who generates this data. And when I get the data, I can easily make a pivot table [in a spreadsheet] – it is just a matter of a few clicks."*

E2: *"But can we retrieve the data on our own?"*

E3: *"No, we cannot extract the data ourselves because it is from EASY-A [governmental IT system, which is inaccessible for the employees at IU]"*

E2: *"Okay, so we depend on them [contacts at the governmental agency for IT and learning], but that makes things much more complicated."*

This excerpt exemplifies how the project group, again and again, were made aware of the organisation's data interdependences with other stakeholders in the network.

Overall, the project group found the exercise very helpful as a way to structure their discussions about the organisation's co-designed Data Sphere. When evaluating the design activity, one participant stated that "*I would very much like to have a picture of this board. It is fascinating to articulate these different dimensions of data. Maybe I can use it [the parameters] for other tasks"* (E2, WS 4). Our observations from this data exploration exercise indicate that collectively selected parameters help to align the workshop participants understanding of whether a data source should be explored further and how to prioritise amongst many different proposals.

The Data Experiment Template reveals the need for tangibility and challenge of data sources' level of abstraction

To work "an experimental mindset" is uncharted in this organisational context, where the employees and managers often need tasks and solutions to be approved by external stakeholders [45]. The organisation's limited resources and need to provide a high level of accountability to multiple stakeholders constitute a barrier for experimenting with different possible solutions [46]. This means, in this case, the members of the project group had to both comprehend how to explore data sources and to learn what it can mean to experiment. This challenge became visible during the design of the Data Experiments, where the project group were making use of the Data Experiment Template. Here, we noticed that they preferred and addressed the template's specific questions while skipping the more open-ended part that allowed the participants to sketch the data experiment (see Figure 3). This observation was confirmed when the group evaluated the Data Experiment Template. They emphasised that they enjoyed the tangible format and the guiding structure, however, they "*did not know how to tackle the drawing exercise*" (E5, WS 4).

Furthermore, we observed that the specification of things to consider regarding their Data Experiments prompted discussions about the context in which the experiments were to be implemented. For instance, the section on 'Lesson Learned' and thus whom they could learn from opening up for discussions about other stakeholders in the network that would also be affected if IU changed the way a particular data source was handled. A similar situation occurred when the project group discussed the questions related to 'Practicalities'. This aspect of the template made the group members reflect on how to implement the Data Experiments in their everyday work practice. These observations suggest that domain experts benefit from rather comprehensive and specific instructions in order to grasp how they can work exploratively and experiment with data.

One concern regarding the exercise of designing Data Experiments was the data source's level of abstraction. We noticed that the data source's level of abstraction influenced the project group's ability to experiment as well as the possibilities for implementing the Data Experiment within the time frame of the project. For example, Experiment 2 revolved around how the organisation could consider colleagues as a data source to promote best practices and insights about data work across teams and departments in the organisation. One of the group members explained how the data source's "fluffiness" had created challenges for their implementation of the experiment: "*This experiment did not revolve around a new data source per se but focused on developing a way to better structure an existing data source. This data source idea was abstract, and it made it difficult to make a concrete experiment with it*" (E4, WS 5). In contrast, Experiment 1, which focused on data about "Elective Specialisation Courses" and revolved around structured, quantitative data had made it easier for the participants to carry it out the Data Experiment. This observation suggests that it is relevant to consider a data source's level of abstraction in relation to the domain experts' knowledge of and experience with creating and conducting experiments

Unfolding two Data Experiments

To illustrate the project group's design and experimentation with data, we elaborate on two Data Experiments (see experiment 1 and 4 in Table 1), which were carried out within the scope of the project.

These two Data Experiments demonstrate different ways the project group explored and experimented with data.

Data Experiment 1: Quantitative Data about Elective Specialisation Courses

Data Experiment 1 was based on the idea to explore how members of the organisation could work differently with a data source that was only being applied to a limited extent at the time of the workshop. The data source constituted structured data about "Elective Specialisation Courses". These courses are a mandatory part of all vocational educations in Denmark. They are formally developed by Sector Skills Councils, who are responsible for making sure that the vocational educations are developed according to the needs of the labour market. The education consultants at IU work to support and facilitate meeting structures for 12 Sector Skills Councils. The dataset containing information about the Elective Specialisation Courses had been applied by employees at IU to provide information about which courses are being offered at the vocational colleges and to generate insights, e.g. about which courses are in high demand.

The project group decided to explore how this data source could be made more available both for more members in the organization, but also to consider for this data could be shared with organizations similar to IU that could also benefit from making use of the data. They wanted to explore and reflect on the process, in detail, that the consultant (E3) goes through from requesting the data to making use of the data source at a Sector Skills Councils meeting. They also wanted to prototype and test a "data guide" that could enable people in other organizations to explore the dataset.

The project group thought it was easy to carry out Data Experiment 1, but they thought it was difficult to document and reflect on their process because "*it was so straight forward"* (Interview with E3, June 2019). The group members, who were in charge of the experiment, chose to create an explicit guide which was a part of the dataset/spreadsheet to support other education consultants' and external stakeholders' ability to make use of this data source. They received positive feedback on the added guidance from the external stakeholders, who agreed to test their proposal. Another interesting outcome from this Data Experiment was the project group's reflection on how this more explorative and experimental way of working had helped them to establish better the need for this data source amongst stakeholders in the network. This particular data source is fragile because IU is dependent on other stakeholders in order to be able to get access, but also because the stakeholder whom they dependent on (in this case a governmental agency) has decided to close the IT system that enables this data source. Thus, by being able to develop and establish a joint need among many stakeholders in the larger network, IU might be able to ensure this data in the future.

Data Experiment 4: An Open Source System for All

Data Experiment 4 aimed to explore and identify common "data interests" with selected stakeholders to examine whether there is potential for creating a cross-organisational open-source system. Compared with Experiment 1 above, this experiment had a much more abstract point of departure. The idea did not include a specific, structured and quantitative data source. Rather, it constituted a grand vision for cross-organisational data management of multiple and heterogeneous data sources. To develop such an open-source system was far beyond the scope of this project. Therefore, the project group needed to design a Data Experiment that addressed the issue; however, in a way that would be manageable within the scope of the project and concurrent with their other tasks. With guidance from the first author, the group decided to focus on meeting minutes from Local Education Committees, as these are an important source of information for many stakeholders in the network and thus could exemplify a common data interest. However, these minutes are not very well structured, and there are no standards for what should be included in the minutes. The project group, therefore, designed an experiment where they would analyse ten meeting minutes to identify themes that could provide a framework for a more generic minutes template that could ideally create a pathway for joint data interests – and data collection. The project group also expressed the need to get feedback from external stakeholders as a part of exploring this idea. They agreed to get inputs on their prototype minutes template from three committee members from different Local Education Committees and a principal at local vocational college to learn more about their needs and use for the meeting minutes.

Findings from their analysis of the meeting minutes showed great inconsistency amongst the ten most recently submitted minutes. As they were not able to create a prototype template based on their sample, they instead chose to develop a list of data categories, which they assumed could generate valuable insights for the education consultants at IU. The categories ranged from broad topics related to "development activities" to more specific aspects such as "discussions on Local Education Development Plans". The group members, who were driving this experiment, asked for feedback from three members of different Local Education Committees. They conducted the feedback by sending an email (which included a short explanation about the experiment, five questions about the list) and followed up with a phone interview.

The feedback from the selected stakeholders varied significantly. Two of the committee members replied that it seemed like an interesting idea but did not think they had a say in a potential development process. The third committee member was in line and stated, "*this is interesting; it's something we can use as a structure"* (Email from Local Education Committee member 1. June, 2019). However, this member also chose to share the list and interview questions with his affiliated local vocational college to make sure he did not overstep his role. This resulted in a surprising email to the members of the project group. The email was from a principal at the local college, who was very frustrated about "*having been left out of the decision-making process of this new initiative*" (Email from the principal. June 2019). The members of the project group at IU attempted to solve the situation by emphasising that the suggestions for data categories in the minutes *only* constituted an experiment. However, the situation escalated, and the project group was contacted by the chairperson of the largest trade union in Denmark, who requested clarification of IU's meeting minutes initiative. This development of a 'simple' Data Experiment is a significant finding that clearly shows some of the challenges of co-creating in such complex settings [45]. As one of the members from the project group expressed: "*This tells something about the network we are navigating in and how politically sensitive it is, because they perceived it as criticism… I <u>did</u> write that this was just an experiment, I wrote it was just some ideas, but she [the principal at the local vocational college] interpreted it as a criticism of their minutes*" (E2, WS 5). Although the project group emphasised their experimental approach and objective, it challenged existing power structures amongst the "data interdependent stakeholders" in the wider network. This underlines the relevance of exploration and experimentation when innovating in more 'data-driven ways' in a cross-organisational context.

In sum, our findings presented in this section suggests benefits and limitations for how the two proposed design tools support domain experts' identification, exploration and experimentation with self-selected data sources. The next section moves on to discuss our three key insights from these findings, which may inform future investigations.

DISCUSSION AND FUTURE DIRECTIONS

In this section, we discuss three key insights which emerged from our analysis of co-designing Data Experiments with domain experts in an organisational context. Finally, we point to directions for future work.

Data exploration makes data interdependence visible for domain experts

Our work suggests that data exploration supported our domain experts' understanding of the interdependence between different stakeholders that manifests itself in the data and its usage. The Data

Sphere provided a stepping stone for the organisation to consider the many "low-hanging fruits" consisting of potential data sources that might be interesting to probe. As a tool, the Data Sphere prompted the organisations' collective awareness of data as something that can be explored. The parameters that the project group collectively chose to guide their exploration of different data sources helped them to understand and articulate the complexity of the suggestions. It was, in particular, the parameters "resources" and "competences" that promoted the group's discussions about how a given data source would also imply considerations for specific stakeholders and their data work. Moreover, the actual implementation of the Data Experiments made this data interdependence visible to the project group. For example, Experiment 1, made the members of the project group aware of the fragility of the data source, because it required other stakeholders (and their IT systems) to get access to the data. This Experiment also showed the project group how they could make use of the broader network in order to secure their data needs better. Likewise, Experiment 4 made visible that the proposed joint creation of data is recognised as a change in the relationships between stakeholders, warranting a more comprehensive deliberation process.

Both experiments point to the need to consider the interaction between different stakeholders, both when designing data and making use of data in a specific context.

Experimenting with data promote the value of co-creation

The experiments presented above indicate that in order to develop new ways of data usage and to reap the benefits of specific data and data analysis, the wider context even beyond the organisation needs to be considered. This is in line with Degnegaard's [9] argument for creating settings that support co-design amongst stakeholders, or Bean and Rosner, who state that often benefit or value of design *"is relational, and it needs to be continuously created and re-created. This is the work of design*" [1:18]. In the above reported experimentation with data, the benefit of co-design became visible not only to us researchers but also to the project group. Especially, Experiment 4, which created unforeseen ripple effects that forced the project group to respond to the concerns and needs of other external stakeholders. As one group member explained when evaluating the implementation of the Data Experiments: "*I've begun to look into the notion of co-design – I mean what is it really? Now I understand that when you collaborate, you create something for the target group, but when you co-design, you develop together with the target group… we need to be open to the possibility that they [other stakeholders] might have a different agenda… I think it is crucial that we cocreate in the future: we need to understand when to throw in the towel and say "we cannot control this, we need to co-design these [data] solutions with others that want to control just as much as we do…*" (E3, WS 5). Other members of the project group echoed this realisation by emphasising the need to expand the involvement of external stakeholders in development work of IT and data solutions. It could be argued that the participants benefitted from the overall co-design approach of the longitudinal action research project, and thus was motivated to learn more about this topic. However, we interpret the empirical evidence to suggest that the Data Experiment Template as a tool and the process of implementing Data Experiments supported the domain experts' perception on value as something that is continuously co-created amongst multiple the stakeholders in the network. Recognising the need for co-design in order to reap the benefits of data and analytics points to the need to not only acknowledge that data needs to be designed [16] but to develop a *co*-design perspective on data [44], which is also further discussed in the next section.

Cooperative exploration of and experimentation with data sources expose data's ambiguity as "design things"

Ehn's interpretation of "design things" is rather ambiguous. On the one hand he draws on the Scandinavian tradition, meaning that 'thing' refers to meeting spaces where concerns and political decisions are addressed. On the other hand, he also makes use of a more object-oriented understanding of the concept that refers to "*the object of concern in design, the design object and its many 'representatives', the design* *of things as matters of concerns and possibilities of experiences"* [15:92]. This ambiguity resonates with our observations.

Data, on the one hand, became an object of design during our process. In the Data Sphere both existing and new data sources were collected and placed. We observed how, in the workshops, the many dimensions of data were discussed in order to decide which data sources to choose. Through the use of the Data Experiment Template, data also became a malleable 'material' that allowed the project group to consider new and improved ways to make use of data. On the other hand, the tools and methods to promote the design with data brought about a space to discuss and explore data as a common issue. For example, the Data Sphere created a physical space that allowed all members of the organisation to gather and to question the status quo by discussing data-related possibilities and constraints. Likewise, the design workshop created meeting places where the members of the project group could consider the interconnectedness with other stakeholders through data and discuss ways to address the political sensitivity in this cross-organisational context.

Furthermore, through the reaction of the environment it became very visible that if Experiment 4 were to be implemented, it would require a thorough deliberation and co-design process – with other words: another design thing – to agree on the use of minutes as data. In another part of the project, we explored notations that facilitate the co-design of concrete structures of data in relation to the needs supported through it [44]. Also in that case the careful choice of notations allowed the joint exploration and discussion of data needs and data. As Bowker [3] points out "*any "thing" that we create (object, way of looking at the world) irreducibly embodies theory and data"*. Thus, the further exploration of representation, tools and methods that let data design things emerge, can contribute to address the political dimension of data and data analytics [27,31].

Conclusion

The aim of our work was to gain a better understanding of how we can design tools that can support domain experts in organisations to explore and experiment with self-selected data sources. The paper propose and examine two in situ two tools for non-IT experts to design and work with data, which we term the Data Sphere and the Data Experiment Template. We identify both benefits and limitations for how our proposed tools affected domain experts' ability to work creatively and design with data. Our design process generated the design and implementation of two distinct Data Experiments. Both experiments indicate the need to consider the interaction between different stakeholder when making use of data in a specific context. Moreover, through the experimentation with data sources, the benefit of a co-design approach became visible to the domain experts. Recognising the need for co-design in order to realise the benefits of data and analytics reveal the need to not only acknowledge that data needs to be designed [16] but to develop a *co*-design perspective on data [44]. Finally, our corporative exploration of and experimentation with data render its ambiguity as "design things" visible. Thus, this work represents an effort to stimulate future investigations of representation, tools, and methods that can help to enable the emergence of data design things. With the current focus on data, for individuals, organisations and societies alike, and how we design (for) data (use), we encourage the design community to join this conversation on how to expand the human centred design approach to enable, facilitate or craft design work that articulate and incorporate data to a greater extend.

References

[1] Jonathan Bean and Daniela Rosner. 2014. Big data, diminished design? *interactions* 21, 3 (May 2014), 18–19. DOI:https://doi.org/10.1145/2592267

[2] Claus Bossen, Kathleen H Pine, Federico Cabitza, Gunnar Ellingsen, and Enrico Maria Piras. 2019. Data work in healthcare: An Introduction. *Health Informatics J* 25, 3 (September 2019), 465–474. DOI:https://doi.org/10.1177/1460458219864730

[3] Geoffrey C Bowker. 2014. The Theory/Data Thing Commentary. *International Journal of Communication* 8, 5 (2014), 5.
[4] Danah Boyd and Kate Crawford. 2012. Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15, 5 (June 2012), 662–679. DOI:https://doi.org/10.1080/1369118X.2012.678878
[5] Elizabeth F. Churchill. 2012. From data divination to data-aware design. *Interactions* 19, 5 (September 2012), 10. DOI:https://doi.org/10.1145/2334184.2334188
[6] Shirley Coleman, Rainer Göb, Giuseppe Manco, Antonio Pievatolo, Xavier Tort-Martorell, and Marco Seabra Reis. 2016. How Can SMEs Benefit from Big Data? Challenges and a Path Forward. *Quality and Reliability Engineering International* 32, 6 (October 2016), 2151–2164. DOI:https://doi.org/10.1002/qre.2008
[7] Andy Crabtree and Richard Mortier. 2015. Human Data Interaction: Historical Lessons from Social Studies and CSCW. In *ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work, 19-23 September 2015, Oslo, Norway*, Nina Boulus-Rødje, Gunnar Ellingsen, Tone Bratteteig, Margunn Aanestad and Pernille Bjørn (eds.). Springer International Publishing, Cham, 3–21. DOI:https://doi.org/10.1007/978-3-319-20499-4_1
[8] Joshua Ddamba and Yvonne Dittrich. 2015. Data for Participation and Participation as Data: Supporting Incremental Participatory Decision-Making in Urban Planning. *Selected Papers of the IRIS, Issue Nr 6 (2015)* (January 2015). Retrieved from https://aisel.aisnet.org/iris2015/5
[9] Rex Degnegaard. 2014. Co-creation, prevailing streams and a future design trajectory. *CoDesign* 10, 2 (April 2014), 96–111. DOI:https://doi.org/10.1080/15710882.2014.903282
[10] Catherine D'Ignazio. 2017. Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal* 23, 1 (July 2017), 6–18. DOI:https://doi.org/10.1075/idj.23.1.03dig
[11] Catherine D'Ignazio and Rahul Bhargava. 2015. Approaches to Building Big Data Literacy. (2015), 6.
[12] G. Dove. 2015. CoDesign with data. Retrieved August 13, 2019 from https://core.ac.uk/reader/29017176
[13] G. Dove and S. Jones. 2014. Using Information Visualization to Support Creativity in Service Design Workshops. (2014). Retrieved August 13, 2019 from https://core.ac.uk/reader/76980229
[14] G. Dove and S. Jones. 2014. Using Data to Stimulate Creative Thinking in the Design of New Products and Services. (2014). Retrieved August 13, 2019 from https://core.ac.uk/reader/29017095
[15] Pelle Ehn. 2008. Participation in design things. (2008), 10.
[16] Melanie Feinberg. 2017. A Design Perspective on Data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, ACM Press, Denver, Colorado, USA, 2952–2963. DOI:https://doi.org/10.1145/3025453.3025837
[17] Joel E. Fischer, Andy Crabtree, James A. Colley, Tom Rodden, and Enrico Costanza. 2017. Data Work: How Energy Advisors and Clients Make IoT Data Accountable. *Computer Supported Cooperative Work (CSCW)* 26, 4–6 (December 2017), 597–626. DOI:https://doi.org/10.1007/s10606-017-9293-x
[18] Joel E. Fischer, Andy Crabtree, Tom Rodden, James A. Colley, Enrico Costanza, Michael O. Jewell, and Sarvapali D. Ramchurn. 2016. "Just whack it on until it gets hot": Working with IoT Data in the Home. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), Association for Computing Machinery, San Jose, California, USA, 5933–5944. DOI:https://doi.org/10.1145/2858036.2858518
[19] Jonathan Gray, Carolin Gerlitz, and Liliana Bounegru. 2018. Data infrastructure literacy. *Big Data & Society* 5, 2 (July 2018), 205395171878631. DOI:https://doi.org/10.1177/2053951718786316
[20] Erik Grönvall, Jonas Fritsch, and Anna Vallgårda. 2016. FeltRadio: Sensing and Making Sense of Wireless Traffic. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16*, ACM Press, Brisbane, QLD, Australia, 829–840. DOI:https://doi.org/10.1145/2901790.2901818
[21] Gillian R. Hayes. 2011. The relationship of action research to human-computer interaction. *ACM Trans. Comput.-Hum. Interact.* 18, 3 (July 2011), 1–20. DOI:https://doi.org/10.1145/1993060.1993065
[22] Gillian R. Hayes. 2018. Design, Action, and Practice. In *Socio-Informatics: a practice-based perspective on the design and use of IT artifacts*. 302–318. Retrieved April 11, 2020 from

https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1724959
[23] Heiko Hornung, Roberto Pereira, M. Cecilia C. Baranauskas, and Kecheng Liu. 2015. Challenges for Human-Data Interaction – A Semiotic Perspective. In *Human-Computer Interaction: Design and Evaluation*, Masaaki Kurosu (ed.). Springer International Publishing, Cham, 37–48. DOI:https://doi.org/10.1007/978-3-319-20901-2_4
[24] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper Hornbæk. 2015. Opportunities and Challenges for Data Physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, ACM Press, Seoul, Republic of Korea, 3227–3236. DOI:https://doi.org/10.1145/2702123.2702180
[25] Ian G. Johnson, Aare Puussaar, Jennifer Manuel, and Peter Wright. 2018. Neighbourhood Data: Exploring the Role of Open Data in Locally Devolved Policymaking Processes. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (November 2018), 1–20. DOI:https://doi.org/10.1145/3274352
[26] Rochelle King, Elizabeth F. Churchill, and Caitlin Tan. 2017. *Designing with data: improving the user experience with A/B testing* (First edition ed.). O'Reilly Media, Inc, Sebastopol, CA.
[27] Rob Kitchin. 2014. *The data revolution: big data, open data, data infrastructures & their consequences*. SAGE Publications, Los Angeles, California.
[28] Peter Kun, Ingrid Mulder, Amalia de Götzen, and Gerd Kortuem. 2019. Creative Data Work in the Design Process. In *Proceedings of the 2019 on Creativity and Cognition - C&C '19*, ACM Press, San Diego, CA, USA, 346–358. DOI:https://doi.org/10.1145/3325480.3325500
[29] Peter Kun, Ingrid Mulder, and Gerd Kortuem. 2018. Design Enquiry Through Data: Appropriating a Data Science Workflow for the Design Process. BCS Learning and Development Ltd., 1–12. DOI:https://doi.org/10.14236/ewic/HCI2018.32
[30] Bruno Latour and Peter Weibel (Eds.). 2005. *Making things public: atmospheres of democracy*. MIT Press ; ZKM/Center for Art and Media in Karlsruhe, Cambridge, Mass. : [Karlsruhe, Germany].
[31] Yanni A. Loukissas. 2019. *All data are local: thinking critically in a data-driven society*. The MIT Press, Cambridge, Massachusetts.
[32] Deborah Lupton. 2017. Feeling your data: Touch and making sense of personal digital data. *New Media & Society* 19, 10 (October 2017), 1599–1614. DOI:https://doi.org/10.1177/1461444817717515
[33] Ezio Manzini. 2015. *Design, when everybody designs: an introduction to design for social innovation*. The MIT Press, Cambridge, Massachusetts.
[34] Afra Mashhadi, Fahim Kawsar, and Utku Günay Acer. 2014. Human Data Interaction in IoT: The ownership aspect. In *2014 IEEE World Forum on Internet of Things (WF-IoT)*, 159–162. DOI:https://doi.org/10.1109/WF-IoT.2014.6803139
[35] Donald McMillan, Arvid Engström, Airi Lampinen, and Barry Brown. 2016. Data and the City. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, ACM Press, Santa Clara, California, USA, 2933–2944. DOI:https://doi.org/10.1145/2858036.2858434
[36] Michael Muller, Melanie Feinberg, Timothy George, Steven J. Jackson, Bonnie E. John, Mary Beth Kery, and Samir Passi. 2019. Human-Centered Study of Data Science Work Practices. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, ACM Press, Glasgow, Scotland Uk, 1–8. DOI:https://doi.org/10.1145/3290607.3299018
[37] OECD. 2019. *The Path to Becoming a Data-Driven Public Sector*. OECD. DOI:https://doi.org/10.1787/059814a7-en
[38] Oxford Dictionary. sphere noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com. Retrieved April 24, 2020 from https://www.oxfordlearnersdictionaries.com/definition/english/sphere?q=sphere
[39] Alex Young Pedersen and Francesco Caviglia. 2019. Data Literacy as a Compound Competence. In *Digital Science*, Tatiana Antipova and Alvaro Rocha (eds.). Springer International Publishing, Cham, 166–173. DOI:https://doi.org/10.1007/978-3-030-02351-5_21
[40] Johan Redström and Heather Wiltse. 2019. Changing Things: Innovation through Design Philosophy. (2019), 11.
[41] Colin Robson. 2002. *Real world research: a resource for social scientists and practitioner-researchers* (2nd ed ed.). Blackwell Publishers, Oxford, UK ; Madden, Mass.

[42] N. F. M. Roozenburg and J. Eekels. 1995. *Product design: fundamentals and methods*. Wiley, Chichester ; New York.

[43] Daniel M. Russell, Gregorio Convertino, Aniket Kittur, Peter Pirolli, and Elizabeth Anne Watkins. 2018. Sensemaking in a Senseless World: 2018 Workshop Abstract. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, ACM Press, Montreal QC, Canada, 1–7. DOI:https://doi.org/10.1145/3170427.3170636

[44] Cathrine Seidelin, Yvonne Dittrich, and Erik Grönvall. 2020. Foregrounding Data: An Exploration of how Data can become an Object of Design. (2020).

[45] Cathrine Seidelin, Erik Grönvall, and Yvonne Dittrich. 2018. Data Work in a Knowledge-Broker Organisation: How Cross-Organisational Data Maintenance shapes Human Data Interactions. In *HCI '18 Proceedings of the 32nd International BCS Human Computer Interaction Conference*, Belfast, United Kingdom, 1–12. DOI:https://doi.org/10.14236/ewic/HCI2018.14

[46] Cathrine Seidelin, Stine M. Sivertsen, and Yvonne Dittrich. 2020. Designing an organisation's design culture: How appropriation of service design tools and methods cultivates sustainable design capabilities in SMEs. In *ServDes.2020 – Tensions, Paradoxes, Plurality*, RMIT University, Melbourne Australia.

[47] Ernest T. Stringer. 2013. *Action research* (Fourth edition. ed.). SAGE, Los Angeles, California.

[48] Eliane Zambon Victorelli, Julio Cesar Dos Reis, Heiko Hornung, and Alysson Bolognesi Prado. 2020. Understanding human-data interaction: Literature review and recommendations for design. *International Journal of Human-Computer Studies* 134, (February 2020), 13–32. DOI:https://doi.org/10.1016/j.ijhcs.2019.09.004

[49] Francesco Vitale, William Odom, and Joanna McGrenere. 2019. Keeping and Discarding Personal Data: Exploring a Design Space. In *Proceedings of the 2019 on Designing Interactive Systems Conference - DIS '19*, ACM Press, San Diego, CA, USA, 1463–1477. DOI:https://doi.org/10.1145/3322276.3322300

[50] Annika Wolff, Matthew Barker, and Marian Petre. 2017. Creating a Datascape: a game to support communities in using open data. In *Proceedings of the 8th International Conference on Communities and Technologies* (C&T '17), Association for Computing Machinery, Troyes, France, 135–138. DOI:https://doi.org/10.1145/3083671.3083686

[51] Annika Wolff, Kortuem Gerd, Daniel Gooch, Elias Giaccardi, and Chris Speed. 2016. Designing with Data: A Designerly Approach to Data and Data Analytics. In *Proceedings of the 2016 ACM Conference Companion Publication on Designing Interactive Systems - DIS '16 Companion*, ACM Press, Brisbane, QLD, Australia, 53–56. DOI:https://doi.org/10.1145/2908805.2913017

[52] Annika Wolff, Daniel Gooch, Jose J Cavero Montaner, Umar Rashid, and Gerd Kortuem. 2016. Creating an Understanding of Data Literacy for a Data-driven Society | The Journal of Community Informatics. (2016). Retrieved October 14, 2019 from http://ci-journal.net/index.php/ciej/article/view/1286

[53] Theodore Zamenopoulos and Katerina Alexiou. 2018. *Co-design as collaborative research*.

[54] Jack Zhao and Andrew Vande Moere. 2008. Embodiment in data sculpture: a model of the physical visualization of information. In *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts - DIMEA '08*, ACM Press, Athens, Greece, 343. DOI:https://doi.org/10.1145/1413634.1413696

Publication 6

Seidelin, C., Lee, CP., Dittrich, Y. – ***Understanding data and cooperation in a public sector arena***. [Resubmitted to the 18th European Conference on Computer Supported Cooperative Work 2020]

Author(s) (2020): ECSCW 2020 Exploratory Papers Instructions. In: Proceedings of the 18th European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centred Computing on the Design of Cooperation Technologies - Exploratory Papers, Reports of the European Society for Socially Embedded Technologies (ISSN XXX-XXXX), DOI: 10.18420/ecscw2019-to-be-added

Understanding data and cooperation in a public sector arena

Cathrine Seidelin[1], Charlotte P. Lee[2], Yvonne Dittrich[1]
[1]IT University of Copenhagen, Denmark. [2]University of Washington, WA, USA.
cfre@itu.dk, cplee@uw.edu, ydi@itu.dk

Abstract. This note explores how data work takes place in a public sector arena. We report on findings from a 3-year research project with a Danish organisation, which, amongst other things, aimed to improve current data practices in the organisation. We make use of the notion of 'social arenas' as a lens to understand the complex setting the organisation is situated in. We find that data work in this context takes place among multiple stakeholders and requires cooperation across organisational boundaries. Moreover, changes in data practices in one site changes cooperation among multiple stakeholders in the arena. Additionally, we develop a diagram of this complex setting, which constitutes an analytical tool that supports our understanding of the site (or sites) of intervention where data work is examined. Our study contributes to the field of CSCW by proposing and showing how the notion of sub-arena helps to comprehend the cooperation and interaction within the surprisingly complex public sector and locate the (sub-)arenas and stakeholders affected by a change in how data is provided and used.

Introduction

The growing development and use of digital technologies and data are transforming societies with great implications for how daily operations are (and can be) run in the public sector. This development has generated an increasing number of organisations, who are trying to improve practices and implement tools to transform data into ‘insights’ or ‘innovation’ (Bright et al., 2019; OECD, 2019; Ostrom et al., 2015). However, while data is becoming increasingly important in society, at work, and in everyday life, little is known about how the increased focus on data, and thus the increased work with or related to data affect cooperation in the public sector. Therefore, we explore how data practices influence cooperation and impact the organisation of stakeholders in the public sector. Moreover, we question the role data play in this (re-)organisation.

In this paper, we draw on a perspective of data as defined through the ways data are embedded and enacted in everyday practices. As Bossen et al. (2019, p. 465) points out ‘data do not sit in ready repository, fully formed, and easily harvestable. Data must be created through various forms of situated work’.
Furthermore, we argue, to research data and data-based services provided by and integrating whole sectors, research as well as design of such services has to develop ways to conceptualise practices and work beyond individual organisations and across societal sectors. We make use of the concept of ‘data work’(Bossen et al., 2019; McMillan et al., 2016) as a lens to consider what such conceptualisation of cross-organisational data practices may look like in the public sector.

Our study is situated in a public sector arena that deals with vocational education and continuing education. This arena involves many different stakeholders, including ministries, governmental agencies, trade unions, employer associations, and education secretariats. As our point of departure, we focus on an organisation, Industriens Uddannelser (English: The Education Secretariat for Industry, hereafter the acronym IU is used), which assists the collaboration between these diverse stakeholders to develop, among other things, educational programs for vocational education and continuing education in the industrial sector in Denmark. In this paper, the notion of “stakeholder” is used to indicate that any specific person does not only contribute with his/her expertise, but also represents the interest of e.g. a labour market organisation, a vocational college, or the student body of a specific program.

During our longitudinal study with the goal to develop methods and tools that enable the employees of IU to design data based services, we came to understand that most of IU’s activity as well as the respective data needs includes other organisations and stakeholders like vocational colleges, labour market organisations, and other governmental agencies. We recognised that the concepts around data and data work did not provide us with a way to conceptualise these cooperation structures and the interaction between organisations and people.

Therefore, we draw on the concept of ‘social arena’ (Strauss, 1985) as a way to frame the stakeholders that work and collaborate in this particular part of the public sector in Denmark around vocational education and continuing. We make use of this lens to better understand the types of multiple-stakeholder environments that are common in the public sector in order to further to understand data work and data practices in this context. As any such sector in society, the sector of vocational education and training is further structured to allow for cooperation around more specific concerns. IU facilitates particular meeting structures that enable representatives from different organisations in the arena to work and collaborate in order to address certain shared concerns. We propose to use the concept of ‘sub-arena’ in order to describe the interaction between stakeholders around specific tasks, e.g. specific educational programs and their implementation at specific vocational colleges, and the interaction of between these sub-arenas and the overall arena, where these sub-arenas are decided on and their mandate is framed.

The note’s core contribution is our demonstration of how and that these concepts can help to comprehend the cooperation and interaction within the surprisingly complex public sector and locate the (sub-) arenas and stakeholders affected by a change in how data is provided and used. We propose the set of concepts adopted from sociology as a tool to make sense of and design for cross organisational data work. The remainder of the note is structured as follows: First, we relate our study to previous work in CSCW that has considered the role data play in and for collaboration in different context. Moreover, we elaborate on the concept of social arenas. Then, we present our field site and method before turning to our findings which shed light on the data work in this particular arena on the Danish public sector. Finally, we discuss our proposal to use the concept of sub-arenas and how our diagram may constitute a tool for scoping the site (or sites) of intervention in multi-stakeholder environments.

Related Work

In this section, we elaborate on the notion of data work and present very brief accounts of studies that examine data practices and the role of data CSCW research. Then we explain on the notion of social arena and how we make use of it as our conceptual frame.

Data consists of symbols that are stored to support specific activities, e.g. by representing relevant aspects of a specific domain (Kitchin, 2014). In this paper, what constitutes data reflects the people working in this arena’s understanding of data. Thus, we look at data with a broad lens, including a diverse set of data types that encounter both qualitative and quantitative, unstructured and structured forms of data. Moreover, we refer to “data work” as complex and distributed human activities related to data practices (Bossen et al., 2019; Fischer et al., 2017). Specifically, the notion of data work has been conceptualized to address “any human activity related to creating, collecting, managing, curating, analysing, interpreting, and communicating data" (Bossen et al., 2019, p. 466). This form of work is complex, distributed, and often interdependent of other stakeholders (Bossen et al., 2019; Fischer et al., 2017). The literature on data work and digital data practices cover various contexts. Examples includes studies examining data practices in the context of civic engagement, which emphasise that although data are often 'broken' (Pink et al., 2018), they are essential to the work of activists because it supports actions around social issues (Alvarado Garcia et al., 2017). Thus, data and data work strongly influences how non-profit organisations can work and coordinate future initiatives (Erete et al., 2016). In the context of distributed collective practice and scientific data collections, scholars addresses the opportunities and challenges that data sharing and collaboration hold for the design of data directories and more broadly scientific communities (Birnholtz & Bietz, 2003; Paine et al., 2015). Moreover, examples in the literature include investigations into the growing current work practices related to data science (Muller et al., 2019; Passi & Jackson, 2018; Tanweer, 2018). These studies examine amongst other things what constitutes current data science practices and they develop in different organisational contexts.

These different perspectives on data work emphasise practices related to work and cooperation around data as recognised activity and show data as an acknowledged entity that to various degrees shape how work (can) take place. Our study contributes to this discourse by demonstrating how data work takes place in a multiple-stakeholder environment in the public sector.

The notion of distributed organizations is well-known in CSCW. The concept is often used to shed light on the various social and technical aspects of work and coordination that is needed in order to support work across distance (e.g. Becker, 2001; Hinds & Kiesler, 2002; Ribes et al., 2013). Previous research has examined data sharing and collaboration in dispersed contexts (Paine et al., 2015). In our case, data work also takes place across organisations. We therefore considered if we could conceptualise our case as a distributed organisation. However, we were not able to identify one organisation or governance body, but a set of independent and cooperating heterogeneous stakeholders.

In our attempts to make sense of and describe this highly connected field site, we made use of the notion of 'social arena' (Strauss, 1985). The concept of social arena has been defined as 'a place in which different communities of actors meet to discuss shared or overlapping projects or concerns' (Balka et al., 2008, p. 517), and thus constitutes a field that is contained by dominant processual and structural conditions (Strauss, 1985). The place is here meant in a metaphorical sense as a forum for discussion and negotiation. Gärtner and Wagner (1996) apply the notion of social arena as a lens to consider different forms of participation in industrial research and design projects. They propose a framework, which describes three arenas for participatory design in this context. The arenas are characterized as follows: 'the political and policy-making context (Arena A); the institutional/organizational context for action (Arena B); and the context of design – support of work practice, public spaces for community involvement, and so on (Arena C)' (Wagner, 2018). The authors argue that the social arenas, where systems and workplace design take place, have to be thought of as local interpretations and understandings of processes that cut across the arenas and are adapted and embedded within them (Gärtner & Wagner, 1996). They propose to use the concepts to make sense of the the highly situational context of a project. In this note, we will not apply their framework per se; however, we will draw on their idea that the notion of an arena emphasises the political and organisational context of social action in a large network of distinct organisations.

Method

This note builds on data from a 3-year action research project, which focused on how organisational members of IU could improve their data practices as a means to deliberately promote the organisation's design and innovation of data-based services. Hayes states "action research offers a systematic collaborative approach to conducting research in HCI that satisfies both the need for scientific rigour and promotion of sustainable change" (2011, p. 2). We draw on this perspective and understand Action Research as a methodology that implies that the research aims to induce change and improvement of certain aspects of the targeted research domain (Hayes, 2011; Reason & Bradbury, 2013; Robson, 2002). In this case, the primary research domain constitutes IU. To engage with the research domain, the first author was working in the organisation approximately three days a week from September 2016 to July 2019. During this period, the author used different methods to understand the field site, in particular, the stakeholders involved, and the data practices used by different stakeholders to collaborate, negotiate, and make decisions. Overall the fieldwork consisted of more than 250 units of observation, including (1) design, facilitation, and documentation of 22 workshops, (2) participation and observation of 51 meetings, (3) 12 in-depth interviews, (4) approximately 70 documents (emails, reports, presentations), (5) images, and (6) ongoing field notes to document informal conversations, observations and reflections throughout the project period. The result of the action research is discussed in other articles. This note addresses a challenge, we as researchers and designers were confronted with: How to understand and relate to the complex network of stakeholders that the organization collaborated with in order to solve its core tasks. We observed that this organization fell outside the category of a 'normal' organization that mainly use data (at least in part) for internal tasks. As argued above, the concept of distributed organizations did not fit either. On the contrary, IU is an organization that is put into being – in a specific location – to support public governance of a specific domain, and this organizational constellation influences how data are used. For this reason, we chose to make use of our body of material to analyze the complex collaborations between different stakeholders and how data are used in these collaborations within particular area of the public sector domain. We developed our analysis in two main ways, which happened in parallel and influenced each other.

One way we developed our analysis was by identifying specific examples that could help us to develop our thinking about what constitutes collaboration in this arena, and whether/how data are used. We categorized the examples, and on this basis four themes emerged: (1) Data work underpins much of the cooperation in this public sector arena, (2) data interdependence shapes data work, (3) data are used to support negotiation and decision-making, and (4) enables new forms of data work to emerge which further prompts new forms of cooperation to emerge in this context. We drew on the whole dataset to develop our categorization and especially looked out for examples that would not fit. We elaborate on the themes in the Findings section.

The other way we developed our analysis was by trying to depict the arena. The fieldwork generated rich empirical material that led to an in-depth understanding of the complex network of actors that constitutes the arena. The complexity of this arena is depicted in the description below, and, especially in the diagram (Figure 1). Initially, the diagram emerged from discussions about how to characterise IU as an organisation. As the diagram developed through 10 iterations, it became an analytical tool for relating the data work at IU with the cooperation of different stakeholders in the arena. As a way to prevent researcher bias in this flexible design, the first author checked the understanding the diagram represents by discussing with organisational members at IU (Robson, 2002). This occurred in two rounds; the first round included the CEO and a manager, and the second round involved the three employees in the IT-department (a senior IT developer, a senior IT consultant, and a junior IT-consultant). In both instances, the organisational members related instantly to the model, which they thought reflected a good understanding of "their world". The CEO and manager asked if the trade associations could be named so they could print the diagram and display it at IU. The members of the IT-department questioned the "level" of the diagram, and also suggested adding more details, for instance, "the individual student who contacts IU outside of their vocational college or industry employer. However, due to the focus of the paper we decided to maintain the diagram at an organisational level. As such, figure 1 constitutes a significant finding in that it has provided an overview of the arena and its (data) interconnectedness.

Field site

Based on the perspective of IU, this research deals the public sector arena that works to maintain and develop vocational educations and continuing educations in Denmark. In order to make sense of this arena, we briefly introduce the Danish labour market model, which constitute a central governing frame for the stakeholders in this arena. This is followed by an elaboration of IU, as a way to describe this complex space in more depth.

Danish labour market model constitutes a dominant condition for how organisations in Denmark operate and collaborate, and thereby becomes an important aspect for understanding the wider context of our field site. The model is a term for the overall organisation of the Danish labour market, which constitutes a division of labour between the state and the social partners (being employers' organisations and trade unions) (Danish Business Authority, 2019). In our case, it is, in particular, the model's inherent requirement for Tripartite Cooperation that governs the ways in which vocational educations and adult vocational educations are negotiated, regulated, and developed in Denmark. Tripartite Cooperation refers to the embedded obligation for the social partners to be accountable for agreements being made, e.g in relation to negotiations regarding topics such as 'work environment' or 'education'. The public sector arena which we focus on this paper can be considered an outcome of the Danish Labour Market Model because the social partners of the labour market are required to develop the educations in accordance with the Tripartite Cooperation.

In order to bundle interests and expertise, the governance of vocational education and training is organised according to four main fields: 1) Food, agriculture, and experiences, 2) Office, trade, and business, 3) care, health, and pedagogy, and 4) technology, construction, and transportation. This study specifically focuses on the organisation of the 4th field, which includes Industry-related educations. The central stakeholders in this arena include the government, in particular the Ministry of Education, the governmental agency for Learning and IT, employer associations, trade unions, industry companies, vocational colleges (and students), and education secretariats, such as IU. The many different stakeholders represent varying and different interests in the arena. They all cooperate on an ongoing basis to solve their shared or overlapping projects and concerns related to vocational educations and continuing education courses. Much of this cooperation takes place in committees like Sector Skills Councils, Local Education Committees, and Development Committees. In the following, we elaborate on IU, which constitutes a particular organisation that exists to support and facilitate much of the cross-organisational collaboration in this arena.

IU is an education secretariat based in Copenhagen, Denmark. IU was founded as a self-governing institution in 2000 by three major employer and employee associations. As such, these core stakeholders gave IU a mandate to facilitate and support the corporation that is necessary in order for them to meet the requirements of the Danish labour market model. The aim of the organisation is to improve the utilization of resources in order to enhance efficiency and improve the quality of processes related to the maintenance and development of vocational education programs and continuing education courses.

IU has six overall tasks that emphasise the work the organisation performs in this public sector arena. These overall tasks include: 1) Education development, 2) Operations of educations, like e.g. approval of companies to train apprentices, 3) Events to promote vocational industrial educations, 4) Communication with the same purpose, 5) Policy-support, and 6) Administration. IU provides and facilitates particular meetings structures that enable representatives from different organisations in the arena to work and collaborate in order to address certain shared concerns. We term these cross-organisational fora as sub-arenas to make this specific collaborative character of the arena visible.

Findings

This section presents the main findings from our exploration of data work in a public sector arena and the role data play in this context. First, we make the complex setting in which IU is situated visible by presenting a diagram that depicts the public sector arena. On this basis, we show how data work underpins much of the cooperation in this large network of stakeholders. Furthermore, we show how data interdependence shapes data work and how data support cooperation amongst the many different stakeholders in this setting.

Data work underpins cooperation among stakeholders in the complex world of vocational educations

To maintain and develop vocational education and continuing education requires involvement of multiple stakeholders for IU to solve its core tasks. We have attempted to visualise the complexity of the arena in Figure 1, which illustrates how IU interacts with the many different stakeholders in order to maintain and develop the organisation's service provisioning. Every circle is an actor in the arena. Every line indicates collaboration and participation. The triangles represent sub-arenas, formally established as well as temporary committees of cross-organisational collaboration. Considering the model in this way emphasises the complexity of the arena in which IU exists and navigates.

For example, the way in which IU maintains and develops the education programs is through highly organised committee work. IU handles and facilitates 12 Sector Skills Councils (see triangles in figure 1), which constitute authorities that are responsible for making sure that the vocational education programs and continuing education courses are developed according to the needs of the labour market. A sector skills council consists of representatives from employer associations and unions, and an education consultant from IU who coordinates and support the council and its members. Altogether, IU handles 39 vocational educations and more than 1000 continuing education courses. Our examination of data work in this public sector arena is primarily based on the perspective of IU. Thus, in the process of understanding what constitutes data work in this particular arena, the diagram enabled us to consider which stakeholders might be involved and/or affected by the data work we examined.

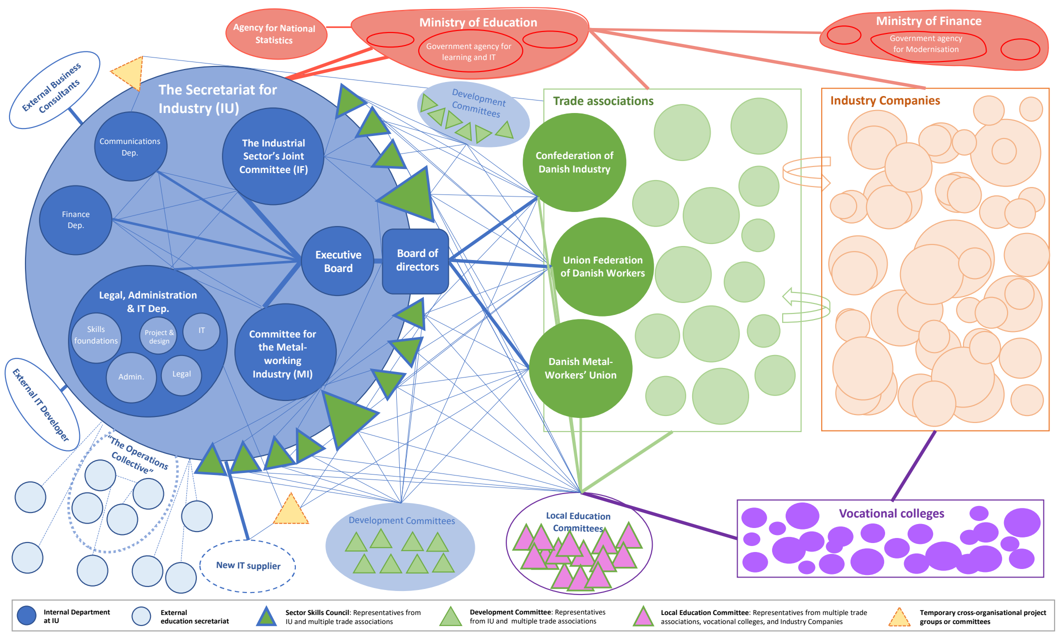


Figure 1. Diagram of the public sector arena for vocational education and training in Denmark.[1]

To illustrate what constitutes data work this complex setting, we elaborate on an example were data practices in and across multiple organisations support cooperation in the arena. The example revolves around Local Educations Committees (LECs, visualised as pink triangles in figure 1) that exist to strengthen the relations between the local industry and vocational colleges to ensure agreement between the labour market's needs and the vocational educations. LECs work locally to implement the legal frameworks provided by the Sector Skills Councils and the Ministry of Education (Danish Ministry of Education, 2019). LECs are made up of 4-8 committee members that represent both employer associations and unions, and additionally, two representatives from the local vocational college. The representatives from the employer associations and unions are often local people who have been appointed by the association or union they are affiliated with. There are 165 LECs alone in the industrial sector in Denmark (IU, 2019). As shown in previous work (Seidelin et al., 2018), it requires careful organisation and cross-organisational data work to audit the members of the LECs and to make sure that each LEC is equally staffed with representatives from employer associations and unions, as required by law. IU acts as a "neutral" part between the stakeholders, and has been trusted with the task to collect, store, and maintain all relevant data about the LECs in the so-called LEC database. In order for IU to be able to maintain the data, it is necessary to coordinate with other stakeholders in the arena. When a LEC member retires, or a new member is appointed, an administrative worker at IU initiates an array of data practices that involves multiple stakeholders, leading

[1] The size of the figures in the diagram does not indicate the actual size of the organisations. Due to the situatedness of the research project, the diagram highlights the perspective of IU. This means that the figures might have been depicted differently in the diagram if another stakeholder in the arena had been the focal point of the project.

to the formal assignment of a new member and update of related data in the LEC database (Seidelin et al., 2018). Consequently, the LEC data and the related maintenance work constitute a system that assists ongoing cooperation in the arena. This example demonstrates how cross-organisational data work supports the collaboration around the shared task to maintain the LECs.

Data work shapes negotiation and decision-making in the arena

Data practices related to certain tasks play a key role in how negotiations (can) develop and how decisions are made in this public sector arena. To substantiate this observation, we highlight an example that show how data work inform negotiations and support decision-making.

The example revolves around data work which was undertaken to investigate the state of automatization in the Danish Industry. Industry 4.0 is a concept that has been used to describe the automatization of the industrial sector (Schwab, 2018). Industry 4.0 is expected to have a major influence in terms of which skills will be needed and in order to support an increased level of atomisation in Industry (Tænketanken Mandag Morgen & Teknisk Landsforbund, 2018). This development has also attracted attention amongst stakeholders in the arena. The trade associations (depicted as green circles in figure 1), in particular, were very concerned about how Industry 4.0 will affect for instance the need to upskill workers in industry. IU was therefore commissioned by the board, and thus multiple trade associations, to develop an analysis of the current level of digital competencies in various industry companies.

An education consultant at IU explained how the data work they undertook both shaped and supported the following negotiation process among the stakeholders:

> "We were talking a lot about Industry 4.0, and therefore it was decided that we should do a "digital check-up", which consisted of us [education consultants] interviewing a number of industry companies about their understanding of Industry 4.0. The purpose was to develop an analysis and a report that described the current state in various Danish companies. Based on the interviews, we concluded that "Industry 4.0" is primarily a concept that is used in big cities and in academia. For me, it was a realization of how we play a central role in the conceptualization of this concept… Most companies did not have an organizational narrative about "we are 4.0", but we needed "company profiles" to provide the "digital check-up". So, by questioning them [industry companies], we are also shaping the need to be 4.0… When we question this system [the arena], we disturb the system so that it begins to reflect on why, for instance, our machine operators are not learning about Big Data. This changes things". (Education Consultant. Workshop video recording. May 2019)

The citation illustrates how organisational members of IU created data and insights through their data work. The outcome of these practices was eventually included in negotiation and decision-making processes related to how the many involved stakeholders should address the requirements of Industry 4.0. The data created and interpreted by the education consultants at IU influenced multiple stakeholders in the network through their data practices. Specifically, this array of data practices resulted in, amongst other things, new continuous education courses for plastic processing technicians about, e.g. data-driven production and maintenance (3Dprinting). Consequently, industry companies are now upskilling their employees in technologies and techniques that prepare them for Industry 4.0.

During our research, we observed similar situations, for instance, when IU consultants were discussing the development of educations and new courses with external stakeholders in sub-arenas; when management was developing a new strategy; or when vocational students would make a complaint about their apprenticeship. Thus, the example emphasised here demonstrates that when certain data practices are undertaken in the arena, it is likely to influence what future steps are (and can be) taken in negotiation and decision-making processes.

Changes to data practices changes cooperation in the arena and sub-arenas

Data and cooperation are tightly intertwined; changes to data practices changes cooperation in the area and sub-arenas. To illustrate this finding, we elaborate on an example where a specific dataset was included to support routine cooperation, initially, in one sub-arena. The example deals with Elective Specialization Courses (ESCs), which constitute a mandatory part of all vocational education programs in Denmark. ESCs are developed by the sub-arenas, who are responsible for making sure that the vocational education programs are developed according to the needs of the labour market. The ESC arrangement is therefore designed to be dynamic to make sure the education programmes meet current needs and future industry demands. The demand for a new ESC can emerge from different stakeholders in the arena. However, the vocational education act states that there can only be a certain number of ESCs per vocational education program. This means in order for a council to develop new courses, they need to close down others. It used to be very difficult for the sub-arenas to decide whether to maintain, develop, or close down an ESC. Education consultants at IU used to share a spreadsheet with relevant vocational colleges and ask which ECSs they offered. The vocational colleges often replied that they offered all courses, and this prevented any action. To improve this work practice, an education consultant at IU reached out to an acquaintance at the governmental agency for IT and learning. This person developed an SQL query that provided a dataset that contained the number of gradings for each course. This data was used as an indicator for whether and to which degree an ESC is actually taken. The underlying assumption was that '*if you get a grade, then you have most likely attended the course*' (Education Consultant at IU. June 2019). The availability of this dataset has allowed the sub-arenas to get new insights about the ESCs in order to update the education programs continuously. Today, this dataset is used regularly both to close down courses in order to develop new ones, and likewise, to identify popular ESCs that might become a mandatory course due to the documented increased demand. Thus, the example demonstrates how the changed data work changed the cooperation amongst involved stakeholders in the area and sub-arenas.

Discussion

Based on our empirical findings, we discuss three key points that contribute to a better understanding of the role data play and how data work takes place in a public sector arena. First, we discuss how the organisation of this particular arena involves sub-arenas and how it requires IU to use data both on a routine basis and in emergent ways. This is followed by how data constitutes a form of participation in the arena. Finally, we discuss stable and emergent data needs in the arena and point to future work.

Data interdependence and Sub-arenas

The stakeholders in this public sector arena work together – though in different ways – to maintain and develop vocational education that addresses the needs of the labour market in the industrial sector in Denmark. Figure 1 emphasises the complexity the actors of the arena navigate in. The diagram reveals how many different sites of collaboration exists and are needed in order to maintain and develop the tasks determining the arena. In this way, we shed light on how data work takes place and the role data play in the creation and maintenance of the interdependence among stakeholders in this particular public sector arena. The diagram also reveals the importance of IU's role to facilitate and support different meeting structures in order to ensure the cross-organisational collaboration that enables representatives from different organisations in the arena to cooperate around shared concerns.

We have proposed the concept of sub-arenas to describe the regular interaction between stakeholders around specific tasks. Furthermore, our empirical findings show that there are two types of sub-arenas in this context. We categorise these as 'fixed sub-arenas' and 'temporary sub-arenas' (Figure 1, green and yellow triangles). The Sector Skills Councils and LECs constitute fixed sub-arenas in that these entities are well-established and formally organised. This form of sub-arena primarily involves routine-based data needs that support continuous committee work. However, sometimes this form of sub-arena addresses emergent data needs, for example, when IU was commissioned to develop the analysis of the current level of digital competencies companies. With 'temporary sub-arenas' we refer to forms of organisation, where different stakeholders collaborate within a provisional time frame to define and/or solve a specific problem. The temporal aspect of this form of sub-arena creates situations were discussion about what data should be included for a specific project are explored and defined "on the go".

Our study reveals that most of the data usages were concerned with making specific aspects of the domain of industrial vocational education and training accessible to the stakeholders of the arena. Thus, rather than informing and supporting one organisation, data was in most cases collected, used and acted upon across different organisations.

A tool for scoping the site of intervention in multi-stakeholder environments

This section discusses how the diagram (Figure 1) that emerged through our explorations of data work in the public sector might constitute a way to support researchers and designers when scoping the site (or sites) of interventions in multi-stakeholder environments. In this study, the diagram has constituted an analytical tool that has allowed us to model (sub-)arenas and stakeholders and in this way grasp the complexity of a particular public sector domain. Stakeholder mapping and analysis are part of many project management and (service) design methods. The concept of social arenas enables one to more easily recognize the shared interests and objectives that constitute social arenas when identifying and involving stakeholders, instead of relying on simple checklists.

When first studying the data practices around one specific set of data in this context, we 'followed the data' to identify relevant domain experts as a way to make sense of the data work related to the LEC database (Seidelin et al., 2018). Initially, we perceived this databased and its related services as a relatively simple. However, this intervention unfolded into a complex interorganisational cooperation, which also influenced stakeholders who were not directly involved in the data work round LECs. Over time, we learned that this high level of interdependence and complexity was the norm, rather than the exception, when it comes to data practices at IU. In this context, any data-based service design will involve a heterogeneous network of actors who are either directly involved in the data practices or effected by the change. We would argue that a tool, such as Figure 1, from the very beginning of the research process could have helped us to
identify both stakeholders and individuals directly involved in the data practices as well as stakeholders who are affected by the project and thus would have to be involved. For example, in our research, vocational colleges did not figure as directly involved in the data practices in the beginning. Including them in the redesign would have allowed stakeholders to address collaboration through the LEC data in a more comprehensive manner early on. In sum, the figure that emerged from our explorations of data work in the public sector and the concepts of arenas and sub-arenas point to a useful way to shed light on the fact that there are many different ways to scope the site of intervention. This could help designers and researchers to not only acknowledge the complexity, but also to better understand and furthermore to be able to be more precise about our scoping of the site of intervention.

Conclusion

The aim of this note was to develop a better understanding of the role data play and how data work takes place in a public sector arena. By examining some of the overall tasks of a central stakeholder in such an arena, our findings highlight how data work in this context takes place among multiple stakeholders and require cooperation across organisational boundaries. We propose to use the notion of sub-arena to describe the interaction between stakeholders around specific tasks, as a way to comprehend the cooperation and interaction in a multi-stakeholder environment such as the public sector. Moreover, we provide a complex figure of the public sector arena, which we argue constitutes an analytical tool for understanding the site of intervention. Thus, we offer these concepts as a way to make sense of and design for cross-organisational data work.

References

Alvarado Garcia, A., Young, A. L., & Dombrowski, L. (2017). On Making Data Actionable: How Activists Use Imperfect Data to Foster Social Change for Human Rights Violations in Mexico. *Proceedings of the ACM on Human-Computer Interaction*, *1*(CSCW), 1–19. https://doi.org/10.1145/3134654

Balka, E., Bjorn, P., & Wagner, I. (2008). Steps toward a typology for health informatics. *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work - CSCW '08*, 515. https://doi.org/10.1145/1460563.1460645

Becker, M. C. (2001). Managing Dispersed Knowledge: Organizational Problems, Managerial Strategies, and Their Effectiveness. *Journal of Management Studies*, *38*(7), 1037–1051. https://doi.org/10.1111/1467-6486.00271

Birnholtz, J. P., & Bietz, M. J. (2003). *Data at work: Supporting sharing in science and engineering*. 339–348. https://doi.org/10.1145/958160.958215

Bossen, C., Pine, K. H., Cabitza, F., Ellingsen, G., & Piras, E. M. (2019). Data work in healthcare: An Introduction. *Health Informatics Journal*, *25*(3), 465–474. https://doi.org/10.1177/1460458219864730

Bright, J., Ganesh, B., Seidelin, C., & Vogl, T. M. (2019). *Data Science for Local Government* (SSRN Scholarly Paper ID 3370217). Social Science Research Network. https://papers.ssrn.com/abstract=3370217

Danish Business Authority. (2019, September 28). *The Danish labour market model* [Text]. Danish Business Authority. https://danishbusinessauthority.dk/danish-labour-market-model

Danish Ministry of Education. (2019). *Faglige udvalg og lokale uddannelsesudvalg*. https://www.uvm.dk:443/erhvervsuddannelser/ansvar-og-aktoerer/raad-og-udvalg/faglige-udvalg-og-lokale-uddannelsesudvalg

Erete, S. L., Ryou, E., Smith, G., Fassett, K. M., & Duda, S. (2016). Storytelling with Data: Examining the Use of Data by Non-Profit Organizations. *Proceedings* *of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, 1271–1281. https://doi.org/10.1145/2818048.2820068

Fischer, J. E., Crabtree, A., Colley, J. A., Rodden, T., & Costanza, E. (2017). Data Work: How Energy Advisors and Clients Make IoT Data Accountable. *Computer Supported Cooperative Work (CSCW)*, *26*(4–6), 597–626. https://doi.org/10.1007/s10606-017-9293-x

Gärtner, J., & Wagner, I. (1996). Mapping Actors and Agendas: Political Frameworks of Systems Design and Participation. *Human-Computer Interaction*, *11*(3), 187. https://doi.org/10.1207/s15327051hci1103_1

Hayes, G. R. (2011). The relationship of action research to human-computer interaction. *ACM Transactions on Computer-Human Interaction*, *18*(3), 1–20. https://doi.org/10.1145/1993060.1993065

Hinds, P., & Kiesler, S. (Eds.). (2002). *Distributed work*. MIT Press.
I
U, I. U. (2019). *Overview of Local Education Committees (In Danish: 'Overblik over Lokale Uddannelsesudvalg')*. https://iu.dk/udvalg/lokale-uddannelsesudvalg/oversigt-over-luu/

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures & their consequences*. SAGE Publications.

McMillan, D., Engström, A., Lampinen, A., & Brown, B. (2016). Data and the City. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 2933–2944. https://doi.org/10.1145/2858036.2858434

Muller, M., Lange, I., Wang, D., Piorkowski, D., Tsay, J., Liao, Q. V., Dugan, C., & Erickson, T. (2019). How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–15. https://doi.org/10.1145/3290605.3300356

OECD. (2019). *The Path to Becoming a Data-Driven Public Sector*. OECD. https://doi.org/10.1787/059814a7-en

Ostrom, A. L., Parasuraman, A., Bowen, D. E., Patrício, L., & Voss, C. A. (2015). Service Research Priorities in a Rapidly Changing Context. *Journal of Service Research*, *18*(2), 127–159. https://doi.org/10.1177/1094670515576315

Paine, D., Sy, E., Piell, R., & Lee, C. P. (2015). *Examining Data Processing Work as Part of the Scientific Data Lifecycle: Comparing Practices Across Four Scientific Research Groups*. https://www.ideals.illinois.edu/handle/2142/73644

Passi, S., & Jackson, S. J. (2018). Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–28. https://doi.org/10.1145/3274405

Pink, S., Ruckenstein, M., Willim, R., & Duque, M. (2018). Broken data: Conceptualising data in an emerging world. *Big Data & Society*, *5*(1), 205395171775322. https://doi.org/10.1177/2053951717753228

Reason, P., & Bradbury, H. (2013). *The SAGE Handbook of Action Research: Participative Inquiry and Practice* (H. Bradbury-Huang, Ed.; Second edition). SAGE Publications Ltd.

Ribes, D., Jackson, S., Geiger, S., Burton, M., & Finholt, T. (2013). Artifacts that organize: Delegation in the distributed organization. *Information and Organization*, *23*(1), 1–14. https://doi.org/10.1016/j.infoandorg.2012.08.001

Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers* (2nd ed). Blackwell Publishers.

Schwab, K. (2018). *Shaping the fourth industrial revolution*. World Economic Forum ; Crown Business.

Seidelin, C., Grönvall, E., & Dittrich, Y. (2018). Data Work in a Knowledge-Broker Organisation: How Cross-Organisational Data Maintenance shapes Human Data Interactions. *HCI '18 Proceedings of the 32nd International BCS Human Computer Interaction Conference*, 1–12. https://doi.org/10.14236/ewic/HCI2018.14

Strauss, A. L. (Ed.). (1985). *Social organization of medical work*. Transaction Publishers.

Tænketanken Mandag Morgen, & Teknisk Landsforbund. (2018). *Industri 4.0 i praksis: Hvordan griber vi mulighederne?* Tænketanken Mandag Morgen : Teknisk Landsforbund.
Tanweer, A. (2018). *Data Science of the Social: How the practice is responding to ethical crisis and spreaading across sectors* [University of Washington]. https://digital.lib.washington.edu/researchworks/handle/1773/43343?show=full

Wagner, I. (2018). *Critical Reflections on Participation in Design* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780198733249.003.0008

Publication 7

Bright, J., Ganesh, B., Seidelin, C., Vogl, T (2019) ***Data Science for Local Government***. Oxford Internet Institute. Oxford University. Available at: https://smartcities.oii.ox.ac.uk/wp-content/uploads/sites/64/2019/04/Data-Science-for-Local-Government.pdf

Data Science for Local Government

Jonathan Bright
Bharath Ganesh
Cathrine Seidelin
Thomas Vogl

Oxford Internet Institute
University of Oxford
March 2019

About this report

The Data Science for Local Government project was about understanding how the growth of 'data science' is changing the way that local government works in the UK. We define data science as a dual shift which involves both bringing in new decision making and analytical techniques to local government work (e.g. machine learning and predictive analytics, artificial intelligence and A/B testing) and also expanding the types of data local government makes use of (for example, by repurposing administrative data, harvesting social media data, or working with mobile phone companies). The emergence of data science is facilitated by the growing availability of free, open-source tools for both collecting data and performing analysis.

Based on extensive documentary review, a nationwide survey of local authorities, and in-depth interviews with over 30 practitioners, we have sought to produce a comprehensive guide to the different types of data science being undertaken in the UK, the types of opportunities and benefits created, and also some of the challenges and difficulties being encountered.

Our aim was to provide a basis for people working in local government to start on their own data science projects, both by providing a library of dozens of ideas which have been tried elsewhere and also by providing hints and tips for overcoming key problems and challenges.

This report was supported by funding from Google.

Key findings

- Data Science is still in a nascent stage in UK local government work. For example, few authorities are exploiting the potential of machine learning to enhance service delivery, or exploring the use of artificial intelligence to enable different forms of interaction with customers and citizens. Hence there is enormous potential for the use of these techniques to be expanded, and thus to deliver better services to citizens.

- The key reason for this is that doing 'data science' in local government faces a number of crucial barriers. People we spoke to consistently highlighted the difficulty of finding time (and support from senior management) to produce innovative data science projects. Whilst in theory the context of austerity provides stimulus for innovation, in practice the dramatic reductions in budgets have meant that back-office analysts who have retained their positions are almost exclusively focussed on statutory reporting, with hardly any possibility of engaging in new work (especially with any risk of failure).

- Despite all these barriers, local government is also a site of considerable innovation, with a huge number of pilot projects in progress in areas such as machine learning, artificial intelligence, data merging and A/B testing. There is often talk of a skills gap in local government, with people unable to hire the staff they need. But we found lots of examples of skilled analysts and business intelligence specialists working on remarkable projects with shoestring budgets. Hence, we would encourage local governments to invest more in the people they currently have by providing them with training and space to innovate, whilst looking less to third party contractors and consultants.

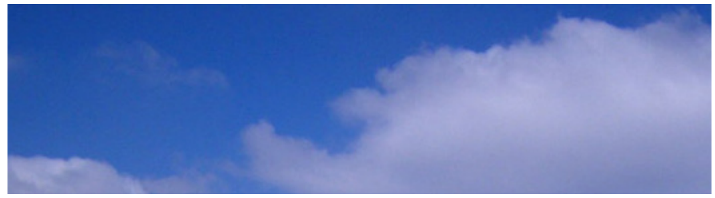
- It is also important to be clear about the potential outcomes of data science projects. The case for many such projects is often built around the idea that they will save money. In the current climate of intense financial difficulty this is understandable. But we also believe this is fundamentally the wrong way to conceive data science in a government context: many useful projects will not, in the short term at least, save money. For example, data science projects which identify areas for early interventions still need to be supported by funds to actually carry out those interventions; whilst data science projects that identify needs more efficiently may also identify needs which were previously unknown. In short, data science should be conceived of as something that improves services for citizens, and allows people working in local government to optimise their time, rather than something which will save money.

- Data science projects are inevitably people focussed: they might be about supporting a frontline social worker in their day to day activity, providing insight and intelligence to senior management, or making decisions about intervention pathways for particular citizens. So, it's critical that these people are involved in the projects! The best examples we found in our work involved close collaboration with agencies and citizens, with data science conceived of as a service rather than something that tells people what to do. Interestingly, when people who are generating the data can see how it is being used, then the quality of the data (and acceptance of systems) gets a whole lot better.

- There are strong concerns about privacy, ethics and accountability in the introduction of new data science technologies. The practitioners we spoke to were acutely conscious of issues such as potential bias when (for example) deploying new decision making technology. However, there was uncertainty about the best way to avoid these problems Clear and open standards and guidance about how to use data science techniques in a way compliant with existing legal and ethical frameworks would be a really important enabler for the sector.

- Finally, though many people have highlighted concerns about both the quality and quantity of data in local government, we found that while 'big data' might be desirable small data is often enough. It is true that many advanced analytical techniques are being developed in an industry context where having hundreds of millions of data points would be the norm. But we found encouraging examples of machine learning projects leveraging datasets of a much smaller scale. Hence, even though pooling data (and getting access to more) is tricky, people working in the area should be encouraged to start small and work with what they have, to develop quick proofs of concept, and to not be put off by potentially limited access to data.

Introduction



It is an exciting time to be working in local government. The last ten years have brought wholesale digitisation, first of back office systems and then of front office service interactions, with more and more citizens 'channel shifting' onto digital ways of connecting with their local municipality. These shifts have brought with them a wealth of data on citizen preferences and behaviours which is more open and tractable than ever; and added to this, new sources of data such as social media are emerging.[1]

At the same time, advances in analytical techniques have opened up new ways of understanding this data and putting it to use (for example, the rise of predictive analytics, artificial intelligence and A/B testing), raising the possibility of a host of new ways of doing local government work. These advances have been accompanied by significant developments in the availability of tools: for example, it is now possible to install sophisticated, open source software (such as R and Python) which enables advanced machine learning at very little cost. These three shifts: greatly enhanced data availability, new analytical techniques, and the availability of tools to put them together are components of what people are increasingly referring to as 'data science', something which stands positioned to revolutionise the way government interacts with citizens.

It is also an incredibly challenging time to work in local government. By 2020, central government funding will have decreased by almost 80% compared to its 2010 level according to some figures,[2] meaning that local authorities face enormous financial pressures. And the problems local authorities are required to deal with have largely been on the rise. To take just a few examples from the hundreds of services local authorities deliver,[3] increases in longevity have meant that demand for adult social care is projected to increase by 67% in the period 2015-2040;[4] contacts to children's services have increased by 78% in the last 10 years;[5] and rough sleeping has almost tripled since 2010.[6] Many local councils are facing huge difficulties to balance budgets under these conditions, and reductions in services and staff members have been widespread.[7] Although in a sense these challenging conditions have stimulated innovation, they have also meant that there is little time or appetite for real risk taking in local government work (and innovation often becomes a synonym for projects which might save money).

The aim of this report is to help promote the expansion of data science in local government, whilst being conscious of the background and pressures people face. On the basis of desk research, a practitioner survey, and interviews, we have sought to map out how data science is currently being used, and capture common problems and challenges in its implementation. In particular, we are aiming to support and enable people working in local government who would like to get a 'data science' project off the ground but have been unable to find the time and space to make it work, or aren't quite sure what the best avenue to pursue is.

There are a huge number of these people out there (and we were lucky enough to talk to some of them during the course of this work) who have good ideas and often the data and skills to execute them: but they lack the time and the support from senior management to innovate and be creative. This report is designed to support their work: to provide ideas for projects to execute, tips for solving common problems, and above all to showcase the many fascinating things being done with data science around the UK (and beyond), to help others get similar projects off the ground.

The report has two main sections. In the first part, we look at different types of technique which fall under the broad heading of 'data science'. In the second, we consider cross-cutting challenges (and responses to those challenges) for the sector.

1 Giest, S. 2017. Big data for policymaking: fad or fasttrack? Policy Sciences 50(3), 367-382; Daas, P., Puts, M., Buelens, B. and P. van den Hurk. 2015. Big Data as a Source for Official Statistics. Journal of Official Statistics 32 (2), 249–262; Malomo, F. and Sena, V. 2017. Data Intelligence for Local Government? Assessing the Benefits and Barriers to Use of Big Data in the Public Sector. Policy & Internet, 9, 7-27; Lavertu, S. 2016. We All Need Help: 'Big Data' and the Mismeasure of Public Administration. Public Administration Review, 76, 864-872.
2 English councils brace for biggest government cuts since 2010 despite "unprecedented" budget pressures. The Independent, 1 October 2017.
3 Local Government Services List. The Local Government Association
4 Adult social care at a glance. The National Audit Office. p 22.
5 Child protection services near crisis as demand rises. BBC News, 6 November 2018.
6 Rough sleeping – explore the data. Homeless Link.
7 Councils under financial strain. BBC News, 7 September 2018.

Types of Data Science

In this section, we review the different types of 'data science' technique which are currently being used in local government in the UK that we unearthed through our desk research, survey instrument and interviews. For each one, we look first at the problem area it addresses, its general definition, and then provide some typical use cases of the technology, before addressing common implementation problems and challenges.

1. Predictive Analytics and Decision Support Technologies

A considerable proportion of local government work involves deciding when and where to apply services and interventions (and who to apply them to). Much of this work happens in a reactive fashion, following some kind of referral or request. For example, when child services receive a safeguarding report, they must decide whether to follow up with a social care assessment. Adult social care workers must decide when conducting needs assessments when individuals can be assigned support services. Police officers may decide after an arrest whether to proceed to a charge or assign an individual to some other pathway of intervention. Some of this work also happens proactively: for example, housing officers may decide which properties to inspect in search of 'Houses in Multiple Occupation' (HMO) violations, whilst food standards agency inspectors might have to choose which restaurants to investigate.

These decisions occur in a wide variety of contexts and situations, yet they all typically share a number of common features. First, the decision about how to allocate services isn't straightforward, such that considerable expertise is required to conduct it correctly and considerable time is required from one or more experts. Adult social care referrals, for example, may take in information from healthcare professionals, social workers and family members, as well as independent advocates.

Second, the overall volume of cases is typically high, meaning that the decision making process itself is a significant drain on resources and there is pressure to take complex decisions quickly. For example, a fifth of children in England are referred to children's services before the age of five,[8] meaning that over half a million referrals are made around the country each year.[9] Third, the consequences of making the 'wrong' decision are significant. If people are incorrectly given an intervention they didn't need, this costs the service money, and may well be upsetting or inconvenient for the person involved. However, if an intervention isn't assigned where it could have been useful, then an opportunity may be missed to help someone in need or to prevent an act of wrongdoing.

One way that data science can start to help in this area is through the introduction of decision support technologies.[10] These technologies are computerised systems which seek to guide people making service intervention decisions. While these systems can take many forms, currently there is growth in the use of machine learning techniques to produce predictions or risk scores for individual areas or different cases: 20 of our survey respondents (16%) mentioned that their local authority is experimenting with some kind of predictive analytics.[11]

Phil Canham, a data scientist working at Barking & Dagenham's corporate insight hub, explained some of the aspirations behind predictive analytics:

> *"Ultimately it's about ensuring residents in need get the right service at the right time. Where the data protection laws allow us to, the idea would be that certain front line staff would have access to the data so they can make the most appropriate decisions. But we'd need to do this carefully, and make sure there was appropriate training around how to interpret results."*

Machine learning, in this context, is a family of methods that involves making use of past data and experience to derive algorithms for the prediction of future outcomes. These algorithms can be derived from data in multiple different ways, but the essential principle is that 'features' of past cases are compared with past outcomes to explore how

8 Bilson, A., Featherstone, B. and Martin, K. 2017 How child protection's 'investigative turn' impacts on poor and deprived communities. Family Law Journal, 47 (4), 416-419.
9 [Rise in child protection cases ups pressure on services](). CommunityCare.
10 Rogge, N., Agasisti, T., & Witte, K. D. (2017). Big data and the measurement of public organizations' performance and efficiency: The state-of-the-art. Public Policy and Administration, 32(4), 263–281. Wise Council: Insights from the cutting edge of data-driven local government. NESTA.
11 The Benefits of Predictive Analytics in Councils. Catalyst Project, University of Essex.


characteristics of particular cases (either individually or combined) correspond to results. This process produces an algorithm which can then produce a prediction of the outcome of a new case, based on its characteristics. Hence, rather than being explicitly programmed, the algorithm (or at least certain parameters of the algorithm) are ‘learnt’. These predictions can then be used as a decision making aid.

Of course, local government has always had a need for forecasting and prediction. However, historically forecasting has largely taken place at a policy or strategic level, and has involved forecasting demand for a given service which needs to be provisioned in advance (for example, demand for special educational needs schooling).[12]

The novelty here is that predictive analytics can also be applied to an operational level, providing a tool which frontline managers can use to allocate resources (e.g. by directing inspections) and perhaps even one which frontline workers themselves can use to aid decisions (for example, deciding when to allocate a particular citizen to a given pathway), by providing more context and background information or even offering up a ‘risk score’ which could supplement existing judgment or provide a summary of existing data.[13]

For example, in the case of social work, Anna Crispe (Suffolk) said that: “as an individual social worker … you work with individual children and families and you document the work you have done … but there might be something else, a more strategic view that the data can offer, which would support your decision-making.” This is what decision-support tools seek to achieve.

One interviewee working in the area, who preferred not to be named, highlighted the particular importance of this type of ‘personalised’ prediction:

> *“We have been doing some work on risk of homelessness … the problem is not knowing how many homeless people will there be in general, its which people will it be, or what pathways will have led them to the stage? That is a more important question … and this is where machine learning approaches become really useful.”*

Indeed, sometimes the separation between strategic policy functions and decision support is also not always clear. As Jon Gleek (Doncaster) put it: “There is a bit of blurring going on in research and intelligence, between what’s performance information and what’s business intelligence - who is the customer of data science? The manager or frontline workers?”

The potential benefits of predictive analytics in a government context are threefold. First, the deployment of scarce resources can potentially be optimised, such that frontline staff time is spent more where it actually matters and less on interventions that make little difference. Second, citizens themselves will hopefully have a better experience, in the sense that services delivered will more quickly match their needs.

Finally, there is the potential for interventions to occur before problems develop, thus potentially both improving outcomes and saving scarce resources. Fran Bennett (Mastodon C) provided an example of the use of this type of technology in the area of strategic forecasting.

> *“We found through our work with various local authorities that one of the areas that they struggle with is special educational needs … The authority has a big task in trying to figure out what needs are going to arise, in what age children will go to school, where in the area the children will be living, and therefore where they need provision. We built a machine learning model to simulate future demand for places and how that varies if the local authority changes their policy on something, or if other external factors change such as housing … we help them think through this problem which is just impossible using something like Excel.”*

12 Reddick, C. 2004. Assessing Local Government Revenue Forecasting Techniques, International Journal of Public Administration, 27, 597-613.
13 Pratchett, L. 1999. New Technologies and the Modernization of Local Government: an Analysis of Biases and Constraints. Public Administration. 77, 731-751.

Use case 1: Children's social services

The area where predictive analytics is currently being most frequently applied (albeit only in a trial form) is in children's social services, particularly at the 'front door' of the service where social workers must decide whether to refer cases for further action or not (indeed, welfare and social care areas were the biggest application domain for data science reported in our survey: 44 of our respondents, or 35%, said that welfare and social care was making use of data science; **Figure 1**).[14]

However, much of this work is exploratory, and there are few examples of technologies genuinely changing frontline practice. As Jon Gleek (Doncaster) said: "I'm not sure anyone has really strong uses of machine learning in local government right now." Here, decision support technologies could provide a useful supplement to this complex decision making area, potentially enabling social workers to concentrate their effort on higher risk cases whilst sparing low risk families the intrusion of being screened.

One example of such a trial is provided by the Behavioural Insights Team, who have developed a structured topic model which is applied to the case notes of social workers.[15]

They are currently developing the model into a risk assessment tool which will inform decision making in the area. A similar project was undertaken by PricewaterhouseCoopers in West Sussex, where they reviewed past patterns of contact to identify risk and inform early intervention in children's social care using machine learning and natural language processing to analyze both structured and unstructured administrative data at the individual level.

Meanwhile, Hammersmith & Fulham have developed a predictive model which is used to assess the risk that children will become "looked after" by the state.[16] Outside of the UK, a similar effort has been made in the county of Allegheny in the United States.[17,18]

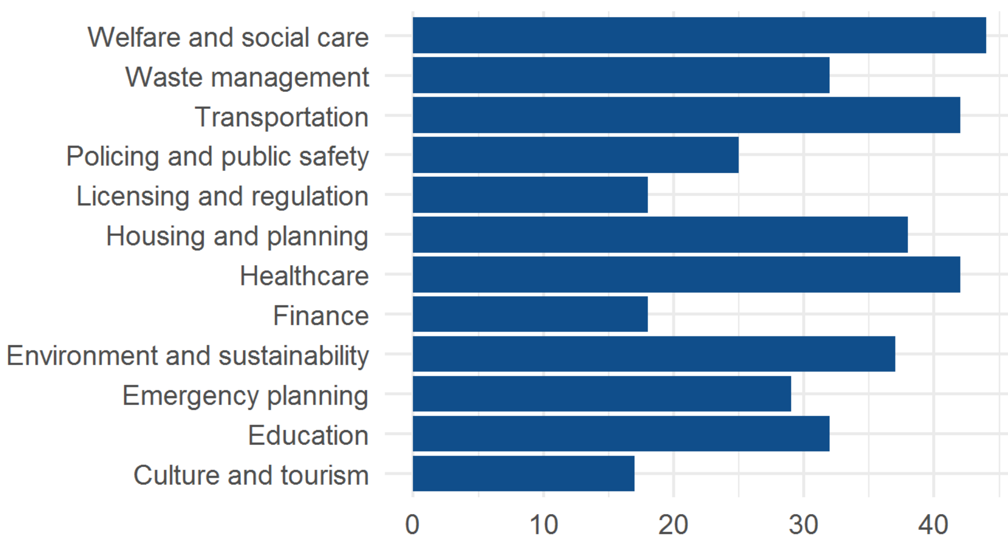


Figure 1: Application Domains for Data Science Projects

14 London uses data to predict which children will be abused. apolitical.

15 Using Data Science in Policy. The Behavioural Insights Team. pp. 16-20.

16 Business Intelligence - transformational services.

17 Can an algorithm tell when kids are in danger? New York Times, 2 January 2018.

18 Chouldechova, A., Benavides-Prado, D., Fialko, O. and Vaithianathan, R. 2018. A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81:134-148.

Use case 2: Emergency Services

A second potential use case is in the area of emergency services. In a criminal justice context, predictive algorithms are already widely used in the United States to inform bail hearings, sentencing and parole decisions.[19,20,21,22] In the UK, applications are starting to appear, albeit in a much more experimental fashion. One example is provided by the HART tool in Durham, which provides a risk score to custody officers when they process individuals who have been arrested.[23,24] Making use of data on past offending as well as demographic characteristics, it divides arrestees into low, moderate and high risk categories, with moderate risk individuals eligible for an out-of-court rehabilitation programme. Marion Oswald (University of Winchester), who has studied the tool, said:

> *"one of the motivations [of the HART tool] is to try and bring together information that, say, a new custody sergeant may find very difficult to analyze because they don't have that long-term knowledge of doing the job. So, it's to try and bring together some consistency in decision-making."*

Another example is provided by the Braunstone Blues programme in Leicester.[25] This project unified data from Fire, Police and Ambulance services to understand which individuals, households and streets were placing the most demand on emergency services. Lynn Wyeth (Leicester) commented on the motivations behind the project:

> *"What we wanted to do was to target those people using the resources the most. We wanted to reduce the number of people that would ring in … because it's a strain on resources, so it was definitely to be more efficient, but also it was to give them the right service. Because often it wasn't the police they needed, it was social services."*

The areas identified are then targeted with preventative home visits to help assess and understand their situation and potentially stop problems before they develop (for example, by fitting window, shed and smoke alarms). In the third year of the project, the area showed a 1% decrease in calls to both the Police and Fire & Rescue (whilst calls to a comparator area had increased). Similar projects in terms of fire safety prevention have been trialled at Suffolk.[26]

Use case 3: Targeted Inspections

A third potential use case is the area of targeted inspections. The need to enforce local rules falls on a variety of different branches of local government, for example the need to make sure council tax is paid correctly or the need to find Houses of Multiple Occupation (HMOs). Inspections are one potential way of enforcing these rules, and one potential use of predictive analytics is to improve the efficiency of these inspection operations.

One example of this is a project in Belfast, which made use of a company called Analytics Engines to develop a tool for identifying properties potentially paying incorrect amounts of business rates.[27] The software improved the efficiency of inspection teams by more than 200% and found almost £400,000 of unclaimed rates in just the first weeks of operation.

In London, similar work has been done in the context of HMO inspections.[28] Software has been developed in conjunction with NESTA which aims to help find hidden HMOs, which are a major source of both unclaimed rates and potential health and safety risks. The software provides a probability for each property, and allows inspectors to potentially guide decisions with respect to which properties to inspect. Newham has also done work in this area enabling them to

19 Sent to Prison by a Software Program's Secret Algorithms. New York Times, 1 May 2017.
20 Kehl, D., Guo, P. and Kessler, S. 2017. Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School.
21 Machine Bias. ProPublica.
22 Berk, R., Sorenson, S. and Barnes, G. 2016. Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions. Journal of Empirical Legal Studies, 13(1), 94-115.
23 Oswald, M., Grace, J., Urwin, S. and Barnes, G. .2018. 'Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality' Information & Communications Technology Law.
24 Durham police criticised over 'crude' profiling. BBC News, 9 April 2018.
25 Public service: state of transformation. Public Service Transformation Academy. p 45.

26 Interview with Anna Crispe (Suffolk).
27 COBALT in its first two weeks identified £390k of unclaimed non-domestic business rates. Analytics Engines.
28 London Office of Data Analytics pilot - now for the hard part. NESTA.

find rogue landlords.[29] Other examples abound. In the UK, predictive analytics are being used to help assign police to specific patrol routes and investigations[30,31] and to target Ofsted inspections.[32] In the US, a wide variety of similar 'targeted inspection' projects have been trialled, in the areas of identifying potentially problematic law enforcement officers,[33] targeting food inspections,[34] identifying lead pipes for removal,[35] and finding areas at a high risk of fire.[36,37] In Canada, a similar project has been used in building inspection works.[38]

Issues in the deployment of predictive analytics

The use cases above bring together four common themes which are worth considering in the deployment of predictive analytics technologies. An obvious first one of these is the quantity of data which is available.[39] Many machine learning technologies have been developed in academic and business contexts where access to datasets with millions of records (or more) would not be unusual. In the context of a local council service, by contrast (such as child or adult social care), it would be more common to have a few thousand cases per year. Hence possibilities for extensive model testing and development may be more limited.

However, even in these limited data contexts, our interviewees highlighted that modelling is not impossible. For example, James Lawrence (Behavioural Insights Team) said: "even with a few thousand records per year, it still seems like it is possible to develop models." Rhema Vaithianathan (Auckland University of Technology), who has worked closely with Allegheny County in the US on the implementation of these technologies, agreed, saying that "what we feel now is that, we tended to start where data is rich, but...now we are working in areas with far fewer 'features' [variables upon which predictions can be built], and you can still achieve strong predictive power."

A second and closely related issue concerns the quality of data. As Matthew Cain (Hackney) put it: "I get the impression we are trying to fly before we have learnt to walk with predictive analytics…the quality of data in local government is often not yet high enough to support this type of technology—garbage in, garbage out." Anna Crispe (Suffolk) agreed, saying that "predictive analytics might just be a little blip, if we can't sort out all the data underneath it." For example, if data about results on outcomes from adult social care is not highly trustworthy, then predictive models built on that data will be similarly flawed.

Furthermore, many interviewees highlighted the need to combine quantitative data with subject expertise. For example, on the topic of predicting rough sleeping, Si Chun Lam (Coventry) explained that "it has got to be a balance between using what the data shows us and combining that with professional expertise of front-line staff as well as the lived experience of rough sleepers to understand why those social services are not working for them and what can we do differently."

A third critical issue is how models will be used by frontline staff. All interviewees who addressed the subject of predictive analytics were careful to highlight that these tools should supplement rather than replace existing skilled insight, and hence act as a kind of secondary check on decisions already made. James Lawrence (Behavioural Insights Team) said:

> *"a machine alone cannot make a decision that has legal consequence for an individual … even the legalities of it aside, I think it's absolutely correct that the human makes the final decision because … there may be some pieces of a particular case that are very unique to that case which are not reflected by the model … so we very much view this as a decision aid."*

Anna Crispe (Suffolk) also supported the idea that predictive analytics should act only as a decision-support tool, saying that: "It's a safety netting approach, but it's not perfect and

---

29 The London Borough of Newham Efficiency Plan
30 Palantir has secretly been using New Orleans to test its predictive policing technology. The Verge, 27 February 2018.
31 PredPol software which targets crime down to small zones has slashed north Kent crime by 6%. KentOnline, 14 Aug 2013.
32 Ofsted to use artificial-intelligence algorithm to predict which schools are 'less than good'. Tes. 29 March 2018.
33 Benchmark Analytics and the University of Chicago to Create National Research Consortium on Police Early Intervention and Outcomes. Benchmark Analytics.
34 Food Inspection Forecasting. City of Chicago.
35 How a Feel-Good AI Story Went Wrong in Flint. The Atlantic, 3 January 2019.
36 Predicting Fire Risk and Prioritizing Fire Inspections. Firebird.
37 Can Algorithms predict House Fires? Data Smart City Solutions.
38 Non-Profit Safety Regulator Uses Machine Learning To Improve Public Safety. Finance Digest, 9 February 2018.
39 Rogge, N., Agasisti, T., & Witte, K. D. (2017). Big data and the measurement of public organizations' performance and efficiency: The state-of-the-art. Public Policy and Administration, 32(4), 263–281.

the practitioner's judgement would hold sway at all times; it's just trying to give practitioners another piece of information to help them make better decisions." Rhema Vaithianathan (Auckland University of Technology) noted that in practice this seems to be how the technology is used: "the most common response about the impact of the decision support tool is that it made case workers stop and think in certain cases where previously they might have gone faster, rather than replacing their judgment". However, Vaithianathan also highlighted that "how our algorithms combine with human judgement and decision-making to get us closer to the 'ideal world' is an open question at the moment."

Marion Oswald (University of Winchester), who has been studying Durham's HART model, also highlighted that "Durham are clear that they do not regard this as a decision-making tool. They're clear with their custody sergeants that it's one factor that they should consider when thinking about whether a person is appropriate for the 'Checkpoint' intervention…As this type of technology comes more into practice, the decision making processes of frontline workers themselves may change." Oswald also highlighted that it is important for them to retain a role in decision making, saying that:

> *"the role of the human has got to be thinking, 'well, does that output actually fit the circumstances in which I am operating and what other factors aren't datafied but are relevant to the decision I'm making?' I think that's an important continuing role for the human, in these really difficult public sector decisions where you've got lots of discretion and lots of different circumstances that you're likely to be encountering."*

Related to this, there is also the question of how people generating the data underlying the tool will respond to its introduction. James Lawrence (Behavioural Insights Team) explained that:

> *"it's very important that any kind of tool or decision aid that comes about as a result of this work is not used as a performance management tool, or anything to beat social workers about the head with because the moment you do that, then it starts to open the possibility that they will begin to game the predictions...so the tool itself will not be making effective recommendations because it's being fed information that's designed to trick it."*

Equally, the expectations of those using the tools also need to be managed. Phil Canham (Barking & Dagenham) gave the example of an externally run pilot project which looked at predicting the likelihood a property was a 'House in Multiple Occupation' [HMO]. He explained:

> *"the problem is, if you set this up as a service, people expect it to be very accurate. Now maybe by using predictive analytics the accuracy has improved from 1/200 to 1/7—but still it isn't the case that every property it comes up with was an HMO."*

Canham explained how, in one of the pilots of the projects, inspection officers were unimpressed because the system was recommending things which were (to the officers) obviously not HMOs. "Through no fault of their own, the company who developed this particular model simply didn't have the detailed knowledge of the borough," he said. "But this knowledge is crucial."

One issue also worth considering in this context is the explainability of results. Some machine learning techniques are more or less 'black boxes', with the precise reasons for decisions very hard to discern. Others are much more transparent: for example, the Behavioural Insights Team prototype tool, which uses structural topic models, highlights specific passages which were of relevance in case notes when making its decision. This explainability can be very important in getting people to trust results.

A final area of relevance is the issue of bias. Applying algorithms to intensely personal and sensitive decision making areas such as child protection and criminal justice raises complex ethical issues of fairness.[40,41,42] Another interviewee, who asked not to be named, said: "I used to say that we don't make predictions about individuals. This is increasingly untenable as a position because of the potential benefits. The moral obligation is to do it but be really careful."

One major issue is the extent to which individual belonging to social groups becomes determining in decisions made. For example, whether ethnic or racial characteristics have a pre-determining impact on the decision of the algorithm, or whether the area where they live might exhibit a strong

40 Eubanks, Virginia. 2017. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York, NY: St. Martin's Press.
41 Machine Bias. ProPublica.
42 Voigt, C. and Bright, J. 2016. The Lightweight Smart City and Biases in Repurposed Big Data. Proceedings of HUSO, The Second International Conference on Human and Social Analytics.

influence.[43] Another issue is whether the algorithm exhibits certain types of bias against specific individuals, perhaps because of deficiencies in the data or the way the model is designed. For example, Marion Oswald (University of Winchester) said: "data can be biased because it's often not got everything in it that's relevant for the public sector's decision."

However, it is worth bearing in mind that the technology may also offer the potential to correct existing (human) biases in systems and perhaps spare people from unnecessary investigations. For example, Rhema Vaithianathan (Auckland University of Technology) highlighted the potential importance of choosing not to perform an investigation in the context of child welfare. "Our child welfare system is incredibly prevalent and one of the challenges is that it's not random. They're hugely present in families of colour and poorer communities. There's a huge presence of child welfare and the child welfare system is not consistent in its decisions. It's like we're dragnetting kids into a system. So, I have real concerns about that front door needing to be much more systematic and consistent than it is. That would be one part of what better decisions look like."

What is clear is that any introduction of such systems needs to be treated cautiously, and that measurement of potential bias needs to be integrated into the way the systems are rolled out.

43 UK police are using AI to inform custodial decisions – but it could be discriminating against the poor. Wired, 1 March 2018.

2. Artificial intelligence

Interaction with citizens is at the heart of local government work. These contacts can be quite generic and fleeting, for example many services will operate call centres which field queries on routine matters such as parking permits, council tax payments, and school places, amongst a huge list of other matters. They can also be highly specific and personalised, for example home care visits in the context of an adult social care programme which may help put people to bed or prompt them to take medication. However, in both cases they can be an enormously costly area of government work. Generic call centres in large councils routinely field hundreds of thousands of calls per year,[44] whilst in the context of adult social care many councils have been forced to commission visits which last just 15 minutes as a means of saving money.[45] In many cases, citizens can struggle to communicate adequately with government on their own terms, and hence may miss out on the possibility of being connected to useful services.

Artificial intelligence is a potential technique which may help alleviate some of the above problems, or at least provide a supplement to existing services. Although artificial intelligence is a term that has taken on many meanings, in this case we refer to it as a technique that is used to create 'autonomous agents' which are capable of having interactions with humans in written or spoken language. The interactions may be used to complete tasks or solve problems, or to connect the human to an appropriate service or piece of information.

The technology behind autonomous agents has advanced considerably over the last few years, with machine learning techniques being used to help improve both the capacity of the agents to understand language and their ability to identify the correct response (for example, Google recently released a demonstration of their Google Assistant booking a hairdresser appointment in human language over the phone).[46] And the technology is increasingly starting to be used in government work.[47]

44 Customer Insight Report 2016-2017. Brighton & Hove City Council.
45 Home care visits should last at least 30 minutes, says official guidance. CommunityCare, 23 September 2015.
46 Google's Latest AI Booked a Hair Appointment Over The Phone, And People Are Freaked Out. ScienceAlert, 9 May 2018.
47 Androutsopoulou, A. et al. 2018. Transforming the communication between citizens and government through

Use case 4: Chatbots in customer call centres

One clear use case for these technologies is the creation of 'chatbots'—autonomous agents which typically interact through a website and make use largely of text-based communication.[48,49] The aim of chatbots is to take pressure off of face-to-face and telephone services by allowing people to conduct transactions online, and also potentially increase engagement and accessibility to services amongst demographics who might not use other digital channels, explained Rocco Labellarte (Oxford City Council), who has worked closely with these technologies.

They are hence in many ways similar to online forms and other digital 'channel shift' strategies, and in some senses simply provide an alternative interactive way to fill in a form. However, they may present advantages over digital forms: some people may prefer a more interactive experience, and it may be that they are able to simplify more complex tasks by presenting questions in a staggered fashion. They also present the possibility of making it simple for interactions to be conducted in any language, something which is of increasing relevance for many councils.

One example of a chatbot is provided by Enfield, which developed a bot to facilitate the process of applying for planning permission for loft development.[50] Another example of this was the 'housing helper' in Hackney, which facilitated the reporting around social housing (for example, raising repair orders),[51] and Transport for London's travel bot which operates over Facebook.[52] A further example is provided by the NHS, which is planning to launch a chatbot type app to help with diagnosis.[53]

Ritchie Somerville (University of Edinburgh) also highlighted how this type of chatbot could be potentially used to simplify extract, transform and load tasks in a variety of local government application areas such as statutory reporting.

AI-guided chatbots. Government Information Quarterly.
48 Ibid.
49 USCIS Launches a Virtual Assistant and her name is EMMA. Immigration View.
50 Enfield joins Microsoft in CitizenBot project. UK Authority, 21 June 2017.
51 What we learnt from prototyping. HackIT.
52 Facebook Travelbot. Transport for London.
53 When will the NHS medical advice smartphone app launch, what services will it offer and what other NHS apps are there? The Sun, 11 September 2017.

Use case 5: Adult social care

Another application domain of these technologies is in the area of adult social care. In Hampshire, trials are underway with the deployment of Amazon Echo smart home devices in homes of adults receiving some kind of care.[54] Mark Allen from Hampshire explains: "what this technology does is to provide a safe guard that is there 24/7 and that actually provides, in some cases, that reassurance that if something happens somebody will be informed, and therefore somebody can do something about it." In addition to this safeguarding function, these technologies have also been enormously enabling for individuals with limited mobility: at voice command, they can change a television channel, or put the radio on, or even read a book.

They thus fill in a gap between visits from professional carers (though no-one suggests they will actually replace them). Steve Carefull (PA Consulting) who also worked on the trial, gave another example:

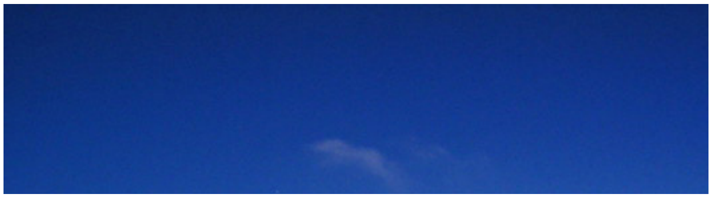
> *"For one gentleman who needs to be lifted into and out of bed every day, the last thing a carer would do at night would be to put the tumble dryer on. His dryer has an anti-crease cycle that turns over every 15 minutes all night and it keeps him awake. With this technology he can now turn it off with his voice."*

They may even alleviate social isolation, for example making it easier to place a phone or skype call to a family member. Mark Allen elaborates on Hampshire's results and highlights how they "found that people—both the people receiving care and the carers—really began to feel in control of this stuff [the Amazon Echo]. This wasn't about Social Services coming and going… this was something they could use and control directly."

They can also act as a point of contact between various care professionals who may have overlapping responsibility for an individual—allowing them to leave messages and notes for each other. Finally, and importantly, they are also much less costly than bespoke technology enabled care devices, and are, as Allen puts it, "something you would actually want to have in your home." Hence, in future roll-outs it may even be the individual themselves who purchases the device. Carefull confirms this, saying that:

> *"technology in social care often isn't especially appealing or attractive. It tends to look old fashioned*

54 See: https://www.youtube.com/watch?v=YL-nQGPxc68



and institutional; beige boxes with red buttons on, etc. A device that people want to have in their house rather than something they have to have in their house makes a difference. A device like Echo with Alexa is also multi-functional. So it might be something that users actually want to buy and use to support their needs, and this could make a difference to social care, which is under huge workforce and financial pressure."

Issues in the deployment of artificial intelligence

As artificial intelligence technologies start to develop, a number of issues recur which may affect their eventual deployment. One obvious area is the extent to which the technology requires human intervention and supervision. Rocco Labellarte (Oxford City Council) cautions: "a digitally non-savvy procurement exercise might not recognise the amount of implementation work which is required." In Enfield, the chatbot required almost a year of development to deal with one application area. Although the technology has certainly developed since then, it is clear that chatbots may require significant upfront training and investment before being launched. In many industry applications, chatbots are being built alongside existing customer service centres which also use web chat: the transcripts of past interactions thus provide training data for future automated agents. However, this is not the case in all local government contexts.

Related to this are the demographics and issue areas that chatbots and autonomous agents are expected to target, which are often much wider ranging than those found in private industry. As Rocco Labellarte (Oxford City Council) puts it: "a chatbot for mortgages focusses on a specific demographic…a chatbot for a local council has a huge and wide ranging demographic."

This diversity in the potential user base creates diversity in the types of cases seen by the chatbot and also increases the type and volume of potential answers coming back and different processes which might be initiated as a result. And when there is more variety in potential questions and answers, the chatbot itself needs to become more sophisticated. One interesting point in this respect was the fact that, when introducing chatbots, business processes are often simplified to make them fit into the technology (rather than making the technology more complicated to fit into the business process).

A third area to consider is how citizens may react to interacting with a chatbot rather than a real person. Citizens may feel that their concern is not being taken seriously if presented with a chatbot. Indeed, there are anecdotal reports of some chatbots being specifically trained to try and address this issue, for example by building in some waiting time before a response to give the impression the bot is thinking about the issue. However, Matthew Cain (Hackney) also highlighted that "there are some areas where a citizen may prefer interacting with a chatbot" for example in reporting financial difficulties or medical conditions. One thing which citizens seem to appreciate from chatbots (as opposed to telephone or face to face interactions) is their ability to provide an audit trail, which demonstrates that an interaction took place.

Related to this is the question of whether chatbots should make it clear that they are 'automated agents' rather than real people. Most people we spoke to on the subject felt that making it clear that you were conversing with a bot was an important part of building trust in the process. For example, Rocco Labellarte (Oxford City Council) said: "even if you have an agent which could blend into the conversation, I still think it would be important to know...it wouldn't be a positive feeling to find out later you hadn't known you were talking to a chatbot." As Matthew Cain (Hackney) said:

"We found it was important for people to know that the bot is a bot...it was also really important that the bot left an audit trail so people could prove that the transaction had happened."

Another area concerns the extent to which artificial intelligence can replace human intervention. In the adult social care example, Steve Carefull (PA Consulting) highlighted that "the cohorts of people this works well for are those with physical disabilities or sight impairment. Many may still need hands-on support from human carers; these consumer devices clearly cannot replace that". So, people making use of the technology need to be conscious that while it might improve outcomes it is unlikely to save money. There is also the question of how developed the technology is. Carefull said:

"The smart home ecosystem is still quite immature, and this type of use in social care is a small area of the market. The technology doesn't do everything we might want—for example, we can't yet manage an 'estate' of Alexa devices outside of an experimental setting, to enable us to 'push' care-related messages such as health or severe weather alerts to all of them at once."

So it will be important to see the directions the technology develops in before making large investments in it.


3. Data merging and centralisation

One of the characteristics of local government work is the volume of different services which are provided for citizens (almost 1000),[55] and the variety of different operators which are involved in their provision. In individual domains such as adult social care, dozens of providers may be involved in offering home visits, operating care homes or providing transport. During their life course, citizens will make use of multiple different services, for example making use of education, hospitals, waste management services, etc.

The complexity of the local government ecosystem was enhanced (some would say exacerbated) by the wave of reforms under New Public Management,[56,57] which have been strongly criticised both for making services often more difficult for citizens to understand and navigate on their own and for not having realized the benefits of mutual support offered by complimentary services.[58]

The fragmented nature of local government work creates a number of critical data issues. Key data can be held in multiple different locations, owned by different individuals, and stored in different formats. At the managerial level, it can be challenging to obtain a complete picture of what is happening in an individual service domain (for example, exactly where money is being spent or where challenges or critical issues are likely to occur). In terms of individual citizens, it can be difficult for service providers to act in a joined up way or recognise problems which may only be evident when perspectives from multiple different service providers are joined up.

In response to this, a variety of governments are working on master data management technologies which will allow them to join up data, either at the level of an individual service or across multiple services. As Andrew Ramsay (Bradford) puts it: "when you think about the services that a council is responsible for, the big move at the minute is to go to individual records, so it becomes like an Amazon account so that data about someone is all in one place… it's not held in the same place, but it can be viewed in the same place." Phil Canham (Barking & Dagenham) echoed this, saying that:

> *"The council has recently undergone a huge structural change, where lots of siloed services have been brought together to become more resident-centric. This didn't happen overnight and it potentially enables us to get a clearer picture of things like individual households, and to build service models based on need. The idea would be to support people before they fall into crisis, for example debt problems, or homelessness, and potentially do early interventions in a more cost effective but impactful manner".*

This is particularly important because it allows the council to work in a much more joined up fashion. Canham continued:

> *"In the past a lot of things would have been treated as separate incidents. A family might be in crisis from the point of view of one service, while another arm of the council is completely unaware of this."*

Sometimes this can involve creating dashboards with services such as PowerBI or Tableau which unify multiple different data sources into a single area: 58 of our survey respondents (46%) reported using dashboards in their local authority. Areas such as Oxfordshire, Surrey, Solihull, Derbyshire, Suffolk, Kent, Sunderland, and Tarragona in Spain, have combined datasets at the client level to improve analysis related to initiatives such as the Troubled Families Programme and the Affordable Warmth programme.[59] Local authorities such as Surrey and Sunderland have also integrated services data using digital tools—such as Tableau and Orbis applications supported by OpenCalais, graph dB, 5* open data,[60] and noSQL—to allow service providers to better understand their clients' contexts. Such efforts have a variety of potential use cases.

55 Local Government Services List. The Local Government Association
56 Hood, C.. 1995. The 'New Public Management' in the 1980s: Variations on a Theme. Accounting, Organizations and Society 20, 93.
57 Elgin, D. and Bushnell, R. 1977. The Limits to Complexity: Are Bureaucracies Becoming Unmanageable? The Futurist, December 1977.
58 CQC, Care Quality Commission. 2017. Review of Children and Young People's Mental Health Services. Phase One Report. Newcastle upon Tyne: Care Quality Commission.
59 How information sharing is improving help for troubled families. Centre of Excellent for Information Sharing. Middlesbrough Affordable Warmth Partnership. NICE.
60 Lee, S., Bright, J., Margetts, H., Wang, N. and Hale, S. 2018. Explaining download patterns in open government data: Citizen participation or private enterprise? International Journal of Electronic Governance .

Use case 6: Single views of the customer

An obvious use case of data merging is to create 'single views' of customers of the local authority. One example of this is provided by the 360 tool in Sunderland. Sharon Lowes, Senior Intelligence Lead at Sunderland City Council, explained:

> *"One of the biggest challenges we always have in adult social care is front-line staff are having to make decisions about individuals, often in the backdrop of huge time pressures and system pressures. So, one of the things we've done is we've brought together a range of datasets from across a range of services...and we've produced a tool called the 360, which is a web-based tool that allows a practitioner to get a 360 degree view of an individual, and their family, and their services, and their interaction with all services, some of which are commissioned and some of which are delivered in-house."*

This tool allows practitioners to identify where a service is not working, for example where a client has repeatedly gone through a procedurally mandated programme with no benefit (a pattern which may have been previously hidden in unconnected records), and make data-driven adjustments to what would have traditionally been done. "The impact has been the practitioners feel much more confident in their decision-making and more confident to challenge what they traditionally would do." This change was also found to have improved data quality, as front-line service providers saw that the data was useful, allowing for better decision-making and learning.

> *"I also know that the data quality has improved. The minute that our social workers and our occupational therapists saw the information displayed, it suddenly had a different purpose to it, and not just a purpose in terms of the use of data, but actually a purpose in their own head around why they write something or how they write something. So, we certainly saw a shift in data quality in the early days."*

Practitioners can also have greater confidence in safely discharging clients because they can see that the other supports are in place, where this information may have previously been distributed among providers in the system and thus unavailable.

> *"The other thing about the tool is that we developed it with the practitioners, so it wasn't a tool that we built and then submitted to them, we got them in from day one. That was one of the reasons why I think we got so much buy-in. That's not how they traditionally worked with IT in the past."*

Another example is provided by North Lanarkshire. They have taken the approach of centralising only core customer information (that is, name and address details), which allows the council to operate a 'tell us once' service for things like a change of address. Peter Tolland (North Lanarkshire) explained the importance of this approach:

> *"We decided on having an index rather than a data warehouse, and we did that for a practical reason: we had about 60 to 70 databases worth of customer information which weren't being kept up to date. What we didn't want to do was to create yet another one, which we would have done with the data warehouse...so the selling point was to create an index where individual departments would still have full control over their backend database systems, but we would then create a way where we would keep all the personal information current and up to date."*

By centralising records of customer-citizen interaction, the service also creates the potential for citizens to have much more satisfying engagements with local services, as they feel that government is acting in a joined up way. Avoiding the creation of data warehouses also allows some records about people to be different if there is a good reason for them to be. Tolland explains: "there may be times when certain services ought to have different information about individuals, for example citizens who are escaping from a domestic abuse situation whose future address needs to be hidden."

Finally, in addition to transforming day-to-day routines, single views also seem to offer enormous research potential. Si Chun Lam (Coventry City Council) said:

> *"Potentially, with a number of sources, we could get our data to a point where we can start identifying, these are the people who might likely come into contact with social care and / or might benefit from early intervention services. There's big data that could outline that if we are able to track a cohort of people through, let's say, five years, and compare the outcomes, are we able to demonstrate some sort of impact of working in the long-term and more preventative way using services more suited to them, does that have better outcomes and lower cost for the public purse as well?"*

Sam Buckley (Enfield Council) agreed, saying that: "previously, we've just kind of done silo analysis in a sense, you know, housing uses data just for housing, children's for children's, and what we want to try and do is bring our data sets together really. So, we've got an all-encompassing view of our customers."

Use case 7: Dashboards for managing private service providers

Another area where data centralisation can produce enormous benefits is in terms of managing spending, particularly in terms of the complex web of private entities who are engaged in providing various aspects of local government work such as house visits for adult social care, or social housing. Keeping on top of these providers (and which ones are more or less efficient) can be a real challenge. Dashboards which are automatically updated and which centralise all the relevant information can hence provide enormous benefit by providing the ability to both anticipate problems early and to see areas where things are being done inefficiently.

An example of this was given by James Rolfe (formerly Executive Director of Resources at Enfield Council), who describes the use of a PowerBI dashboard to manage a privatised social housing company wholly owned by Enfield Council. The data provided by the dashboard highlighted where they were paying over the odds for temporary accommodation and allowed them to manage a scheme which overall was successful in saving more than £4 million.

Rolfe's colleague Sam Buckley says the benefit is that "it just really illustrates the outliers for the service, so it's really staring you in the face rather than being hidden in lines of data, it's actually a visual representation of the issues."

Warwickshire, meanwhile, have been pioneering the use of dashboards for managing adult social care quality assurance. Spencer Payne (Warwickshire) explains: "they provide the capacity to, for example, understand quickly if a provider is getting into financial difficulty, and take appropriate action." This enabled them to behave in a much more proactive fashion: "previously we would be much more reactive, and not necessarily notice problems before they occur."

One of the key benefits of dashboards is that they make data instantly available, something which facilitates productive management and decision making. Even though this data might have been previously accessible, making it immediately available makes certain types of conversation feasible. For example, dashboards have recently been introduced into children's social care in Rutland. Previously, managers would ask a front line worker why some key performance indicators weren't being met. However, without the data readily available, it would be difficult to find exactly which cases were raising the average.

Now, managers can find out exactly what happened, and decisions can be made much quicker. Jon Adamson (Rutland) explained:

> *"the design and use of Tableau dashboards for children's social care has changed the way that managers work, and they use them on a regular basis. Previously, most of the conversation around performance information ends up focusing on whether the figures are right … We've moved beyond that and said, 'no, no, the figure is right. Why is the data that way? Let's understand it a bit more. Let's understand what the impact of that means'."*

The dashboards have also contributed to improving the quality of the data, as front line workers (who often input the data themselves) can see it being used by management in meetings and appreciate the importance of getting it right, much more than just being told by a data analyst that data quality is important. Adamson added:

> *"it [the introduction of a new case management system Liquidlogic] forces a specific workflow (the order in which tasks have to be completed by a social worker) and that was the biggest change and that was the hardest thing for people to get used to, but it's also the thing that improves data quality, makes the system work, and gives transparency and oversight."*

https://www.flickr.com/photos/dafyddbach/9223596446/

Use case 8: Different views of local service needs

A final data merging use case concerns the ability to understand more about which areas and regions are placing the most demand on services, and perhaps react and plan accordingly. We have already referred to the Braunstone Blues programme in Leicester which is an example here.[61] Another example is provided by the county of Suffolk, which has established Suffolk Office of Data & Analytics (SODA).[62] SODA was set up to provide data insight services to Suffolk County Council, seven district councils, Suffolk Constabulary and the Suffolk Clinical Commissioning Groups. Liz Barnard (West Suffolk) explained: “the idea is to do something distinct from ‘single views of the customer’ work—it is about gaining new insights to support policies that transform people’s lives.”

Michaela Breilmann (Suffolk) explained how the Suffolk Office of Data & Analytics (SODA) started its work. “One of the first projects we worked on was called ‘Data on a Place’, and essentially the aim was to see if we could bring together case-level data from all public service organisations to understand all the interactions we have, for a given ward. This was a huge task and we were unable to do this as we did not have the right information governance in place. We also identified gaps in both capacity and capability to extract the data at that level. This is why SODA now formalises our approach to information governance and IT architecture to enable the sharing and combining of data.”

Issues in the deployment of data merging

There are a variety of common challenges and issues which recur in data merging projects. The most obvious of these concerns establishing access to data. Many organisations struggle to have a bird’s eye view of the types of data that are actually held across the council, or who is responsible for owning and managing them (NESTA’s data maturity model has established a useful list of benchmarks in this regard).[63] Even once this has been established, securing access is by no means straightforward, as those who have responsibility for the data may be hesitant about sharing it. For example, Jon Gleek (Doncaster) described a project making use of both Health and Social Care data:

> *“this project...looks at flows of people across health systems into care systems, and helps us see, for example, what happens to people six months after they’ve been discharged (of course we don’t see individual names—it’s more about demographics, their cohort etc.). The project has been on the ground for a couple of years because it takes so long to get the data in the right shape and all the information governance sorted. But now that we are getting data out, it is becoming really interesting as a strategic intelligence tool.”*

In this respect, the business process established by Sunderland is interesting. Sharon Lowes (Sunderland) explains: “when we started out, we kept things small and focussed on quick wins: this enabled us to set a precedent and get known around the organisation. Now, people and products come to us.”

Part of enabling access is also about maintaining confidence around the privacy and security of the data. In this respect, the North Lanarkshire model, which establishes a common set of core data which enables merging in specific instances (rather than actually merging the data in a data warehouse), is again worth highlighting. One key question in terms of privacy is however how much councils should seek to intervene, even if this is to the direct benefit of citizens. For example, Si Chun Lam (Coventry) comments: “it is conceivable we could bring together data which would allow us to identify people who could benefit from free school meals—but should we actually do it? It’s not clear that people would actually want us to do that.”

A second crucial challenge is of course actually connecting and merging data which may be held in many different formats. It is interesting to note that many councils are investing in so called ‘ETL’ software (extract, transform and load) such as Talend.[64] These bits of software can act as a middle layer between lots of different datasets, potentially automatising complex data connection operations. However, it is also the case that increasingly working with high quality structured data formats is useful. The example of Hackney provides a case in point: one key area they have been working on is exposing simple but fundamental bits of council information (such as bin collection times) as APIs. This allows other services to be built on top and reduces individual software dependencies (for example, the

---

61 Public service: state of transformation. Public Service Transformation Academy. p 45.
62 https://www.healthysuffolk.org.uk/soda
63 The model can be found at: https://datamaturity.esd.org.uk/

64 https://www.talend.com

database software behind the API can change without other knock on services being affected).[65]

Another interesting angle here is the potential use of automatic text processing technologies to simplify the ETL stage of the process. Robert Steele (Reigate & Banstead) explained that: "Text detection was a great tool to illustrate to people that it was possible to sniff out the salient parts of large unstructured text in seconds. It seems to have great potential in terms of client care notes in areas of social care." Our survey research supported this idea: 26 respondents (20%) said their local authority was making use of some kind of automatic text or content analysis. Data integration tools are particularly important in terms of collaboration. As Anne Kearsley (Oxfordshire) said: "When individual teams collect data for their own use, a spreadsheet is perfectly fine. If you then try to share that with multiple teams...it rarely scales."

A third challenge is related to getting people to use the new services. While James Rolfe (formerly of Enfield Council) spoke to the successful use of dashboards, he did also indicate that more work was needed:

> *"I think where we haven't yet fully delivered is getting managers across the council to actively use this data. They still need to be presented with it in a more traditional and slightly linear way. And if we're to become truly data driven then managers should be entirely comfortable digging through reports, having a look around them, asking questions, understanding trends, and all of that sort of stuff. And therefore, learning new tricks and new skills is vitally important."*

James' colleague Sam Buckley (Enfield Council) was positive about the future: "Most of the people in this field are used to having the information that they historically had and then when you start to show them the other possibilities, it opens their eyes up to things that they might not historically have asked for. So, it's very much a gradual building process." Spencer Payne (Warwickshire) also commented "we invested a lot of effort in helping colleagues shift their mindset. There is still work to do here and it takes time to completely change the way people work."

65 Interview with Matthew Cain (Hackney)

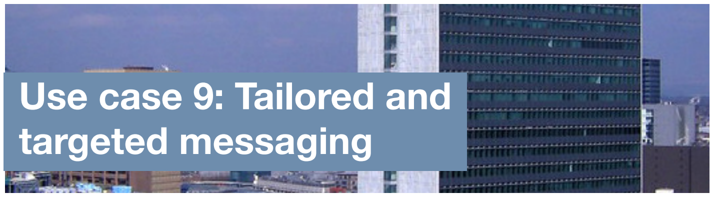
4. Experimentation and personalisation

One characteristic of local government services is that they often have a one size fits all nature. When encountering adult social care services, individuals will be assessed for particular packages of care (e.g. help preparing food) which last for fixed amounts of time (e.g. six weeks). When distributing messaging, the same communication may be sent to all customers regardless of need or circumstances. Or when applying for benefits or means tested services, forms are standardised rather than reflecting complex different situations of individuals. This inflexible nature of delivery creates potential waste, as services are applied for longer than they might be needed. It also potentially makes services less effective, by not tailoring them to the individual circumstances of customers, or indeed the wider area in which they live.

In response to this, some councils are starting to look at the potential for service personalisation. This involves, simply, attempting to tailor a service to an individual or group situation. This could occur in a variety of different ways. Survey research, or indeed local knowledge, might inform where and when to apply different types of personalisation. For example, service managers may simply know that cash payments may be more appropriate for certain areas and groups. However, there is also the growing use of experimental techniques as ways of informing and optimising service delivery. Recently, A/B testing has become a growing means of experimenting with different approaches to a service or message and seeing which one works best, although it has yet to reach widespread diffusion: only 13 of our survey respondents mentioned making use of it (10%; **Figure 2**).

Much of this testing was popularised by the Behavioural Insights Team, which has pioneered the technology at both central and local government level.[66] Broadly speaking, these techniques involve separating out different messages or services into different groups, and randomly assigning individuals to the groups, to see which one works best (with success often described in terms of take-up of the service).

66 https://www.bi.team/bi-ventures/testbuild/



Use case 9: Tailored and targeted messaging

An obvious application of experimentation has been in terms of tailored messaging, particularly around promoting ‘channel shift’, a process which describes moving citizens from one means of achieving a service to another (typically moving from an offline or telephone service to a digital one).

One of the earliest examples produced by the Behavioural Insights Team [BIT] concerned the use of messages encouraging people to pay their council tax by direct debit, a service which produces considerable financial savings for councils. The BIT showed how different mailshots could be experimented with and the result (in terms of increased take-up of direct debit) could be measured.[67]

More recently this type of technique has been applied to things like payment demands and green waste subscriptions. As James Rolfe (Enfield) commented: “there is a need to understand the different demographics and citizens that a council deals with. Some may be happy paying by direct debit, whilst others might prefer paying in cash, for example.” Creating this type of understanding may help increase the successfulness of payments and subscriptions.

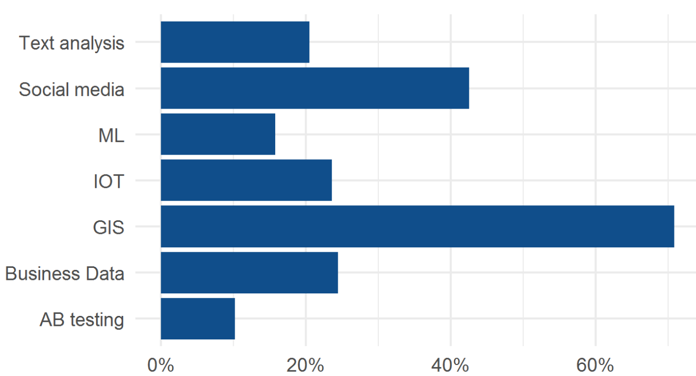


Figure 2: Types of Data Science

67 John, P. and Blume, T. 2017. Nudges That Promote Channel Shift: A Randomized Evaluation of Messages to Encourage Citizens to Renew Benefits Online. Policy & Internet 9, 168-183.

Use case 10: Personalised care and support

Another potential application domain is in the area of care personalisation. As Steve Carefull (PA Consulting) explains, areas of care (such as adult social care) are extremely generic:

> *“in adult social care, you can be assessed once, then assigned a pattern of dom-care visits of X times a week for a year. This doesn’t allow for variation in what the service user may need. There’s also a misconception that a person’s needs are inevitably only going to increase. But we know from our own lives this isn’t always the case.”*

In this respect, an experimental project known as the ‘study supporter’ programme (recently piloted by the Behavioural Insights Team) is interesting. Andy Hollingsworth, who works at the BIT, explained:

> *“adult learners who are enrolled in a course are asked to nominate two friends or family to help provide them support. These people then receive text messages reminding them that this person is taking the course and encouraging them to stick with it.”*

The project ended up producing a considerable improvement in attendance. Although the particular focus was on learning, Hollingsworth described how this “light touch” approach could be rolled out to all sorts of other health and lifestyle areas where people are required to persist with some kind of programme. And having tested the intervention in an experimental paradigm was vital, because it provided rigorous evidence that it actually worked.

Issues in the deployment of experimentation and personalisation

There are a few key issues here. One is the extent to which citizens will actually appreciate the ‘personalisation’ of a given service. This could raise questions of fairness and justice, as some people might appear to be getting more than others out of the state. It could also relate, simply, to perceptions of privacy. Si Chun Lam in Coventry notes that identifying people for targeted mail-outs about direct debit council tax payments or using online services can be perceived negatively by the public: “it’s okay if people are okay with that, but in other cases people can get quite upset and say, ‘well, how did you know that?, and get quite concerned about what government is doing.”

Another issue is the way results are presented. As Andy Hollingsworth (Behavioural Insights Team) explains,

randomised control trials can often appear to have smaller effects than people were hoping for, and these more rigorous methods can hence sometimes lead to disappointment amongst those designing service interventions. This isn't a reason not to do them, of course, but thought needs to be put into how 'null results' are reported. Hollingsworth commented:

> *"randomised control trials can tell you whether an intervention works or not, but not necessarily why. So over time we've learnt to line up the experimentation with process evaluation, and also to mix our reporting in with stories from real people in the trials, which people inevitably respond to more and which help put results into context."*

5. New Forms of Data

One of the core difficulties surrounding several areas of local government work is a lack of accurate data and information about policy contexts. For example, 60% of all journeys happen by road, and yet there is little available information on the source and destination of these trips or the extent to which they proceeded smoothly.[68] In the environmental context, local governments have important responsibilities in terms of monitoring local air and noise pollution, yet again data can be complicated and difficult to collect. Without these types of information, all sorts of policy problems become less tractable.

In response to these problems, a considerable number of local governments are investing in new forms of data collection.[69] One important area of growth here is in the use of "Internet of Things" (IoT) enabled smart sensors which can enable detection of all sorts of novel metrics which might previously have been very hard to capture. 30 of our survey respondents (24%) mentioned the use of this kind of technology in areas as diverse as transport and parking management, monitoring of vehicle fleets, waste management services, building monitoring and environmental monitoring.

Saqib Yasin (Southampton) explains that the use of sensors "offers opportunities in terms of manual and labour intensive processes." He went on to say that

> *"the sensors enable organisations … so they don't need to perform checks and if something goes wrong or needs attention, we'll be alerted to it, which allows for more efficient monitoring."*

This can save an authority on cost and time. These sensors are typically placed on physical infrastructure which is owned or operated by the council itself, though in some instances citizen volunteers may also be co-opted into networks.[70]

Another growth area is in the form of repurposed data from sources such as social media and mobile phone companies.[71,72,73,74] These 'soft' datasets are potentially even more significant than IoT data because they are potentially much cheaper to obtain and process.[75] Fully 54 of our survey respondents (43%) mentioned making use of social media data, often in the context of public relations, whilst 31 (24%) mentioned using third party business datasets such as mobile phone data.

68 Transport Statistics for Great Britain. Department for Transport.
69 Bright, J. and Margetts, H. 2016. Big Data and Public Policy: Can It Succeed Where E-Participation Has Failed? Policy & Internet, 8, 218-224

70 For example, the Oxford Flood Network is dedicated to monitoring flood water levels around Oxford, and makes use of sensors placed both in public spaces and in individual properties which overhang the river owned by volunteers.
71 Poel, M., Meyer, E. T. and Schroeder, R. 2018. Big Data for Policymaking: Great Expectations, but with Limited Progress? Policy & Internet, 10, 347-367.
72 Bright, J., Hale, S., Margetts, H. and Yasseri, T. 2014. The use of social media for research and analysis: a feasibility study. DWP Ad-hoc Research Report 13.
73 Nash, V. , Bright, J. , Margetts, H. and Lehdonvirta, V. 2017. Public Policy in the Platform Society. Policy & Internet, 9, 368-373
74 Agostino, D. and Arnaboldi, M. 2017. Social media data used in the measurement of public service effectiveness: Empirical evidence from Twitter in higher education institutions. Public Policy and Administration.
75 TVoigt, C. and Bright, J. 2016. The Lightweight Smart City and Biases in Repurposed Big Data. Proceedings of HUSO, The Second International Conference on Human and Social Analytics.

Use case 11: Smart Street Bins

One interesting use case is in the area of smart street bins. These have been trialled by the City of Edinburgh who, working with a Swedish company called Inovo, have installed sensors on the underside of the ceiling of 300 street bins around the city.[76] The sensors measure the amount of material in a street bin and report when the bin is empty or full. They provide alerts when the bin is nearing capacity, and were initially networked with sim cards and relied on existing telecommunications services to send data. The sensors are now being considered in other areas such as Glasgow[77] and Perth and Kinross.[78]

One of the first things these sensors enabled was more efficient waste management services. Ritchie Somerville (University of Edinburgh, formerly of the City of Edinburgh Council), who was involved in the pilot explained: “there was a clear proposition in the waste service team: they knew they were undertaking journeys that added no value”. The sensors quickly provided the data which enabled them to understand these journey patterns, and eventually resulted in the fleet of waste management trucks being cut from four to three. However, interestingly, the bin data also enabled all sorts of other analysis to be investigated. The location of bins themselves could start to be optimised, allowing them to be placed on critical routes (for example, in between the train station and the city centre). Furthermore, the bin data also started to provide a picture of more general population movement around the city, and how it varies with things such as special events. Hence, the data provided a broad picture of human behaviour in the city.

76 [Sensors alert Council when bins need emptied](). Edinburgh.gov.uk

77 [‘Smart’ sensors fitted to litter bins in Glasgow](). BBC News, 6 March 2017.

78 [Council will trial smart bin sensors and fly tipping app](). PKC.gov.uk.

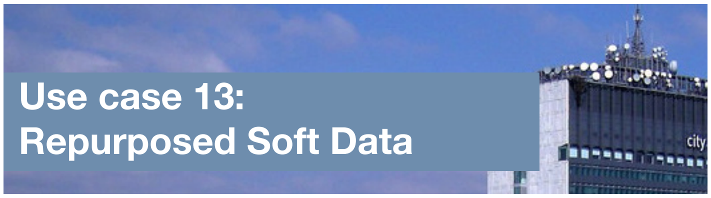
Use case 12: WiFi and environment monitoring sensors in Southampton

Working with a private company named Barter for Things,[79] Southampton City Council sought to increase the number of WiFi-enabled sensors to monitor environment and equipment in council-owned buildings in the city. The company makes small devices can be used by private homeowners that take a small electric current and can be connected to extend the range of a WiFi network. Homeowners are compensated based on the traffic passing through their antenna, providing a small benefit to users who set up these devices in their homes. The council, after a series of conversations with Barter for Things, has also appropriated these devices to manage equipment. In an ongoing pilot project with the company, these devices are being rolled out as internet-enabled sensors in residential units. Saqib Yasin of Southampton City Council explained:

> *“We have facilities people who go out to check heating systems, water supply, and its temperature...rather than someone physically going out and checking the temperature of water, the device will send out an alert if the temperature falls outside a [given] tolerance.”*

The devices, which can increase WiFi accessibility in the city, are also useful for the council to manage equipment and monitor environmental conditions that might require inspection or other action. This allows for cost savings for the council by preventing the need for regular in-person inspections. Alternatively, in public buildings the devices allow measurement of the use of doors and mechanical failures and are equipped to send alerts when repairs are needed. The sensors are intended to send information to a cloud server which then handles notifications and allows for data analysis. The project was conceived as an exploratory pilot and it is too early to see the effects. However, it presents a new data source that can help optimise services and prevent overuse of energy in buildings.

79 https://www.barterforthings.co.uk/



Use case 13: Repurposed Soft Data

In addition to IoT devices, a considerable amount of experimentation is taking place in the area of what might be called ‘soft’ data[80] from mobile phone platforms and social media companies. Mobile phone companies have some history already in terms of selling tranches of their cell phone tower data for enhancing understanding about population movement.

Recently these business models have started to transition to data from GPS enabled devices and apps. For example, Google’s Better Cities programme has been partnering with cities such as Amsterdam to show how anonymised, aggregate data from its Android mobile phone platform can be used to understand mobility patterns on its road network (with data validated by comparing it to traffic cameras).[81] Mobility companies such as Waze[82] (a journey management app) and Strava (a fitness social media network) have launched similar initiatives.[83]

In each case, data shared by users of the platform contribute to building up a picture of different types of movement around a city. Recently, research has started to look at whether similar insights could be obtained from social media platforms.[84,85,86]

In addition to their use in transport, these types of soft data have also been used in the area of understanding citizen opinion. In Coventry, use is starting to be made of social media monitoring software such as Hootsuite to understand social reactions to the city. As Si Chun Lam (Coventry) explains:

> *“one use of this technology is to understand reaction to specific initiatives; for example, Coventry was named UK city of culture for 2021—we can immediately track the reaction to that in terms of media benefit.”*

Another potential use case was provided by Lucy Knight (Devon), who has looked at collecting feedback from Twitter around various different issues such as local libraries. She explained:

> *“on two or three separate occasions colleagues have mentioned to me that they are concerned about traditional ways of capturing feedback: we only hear from the people who have time to sit down and fill out forms. Social media provides an opportunity to do something different in this area.”*

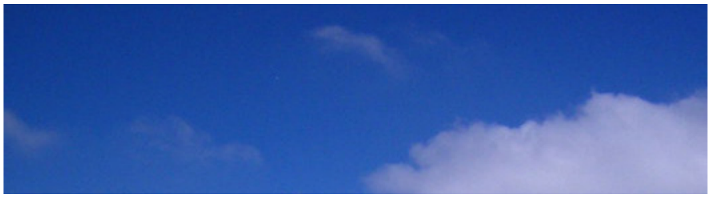
The town of Jun in Spain has gone even further, encouraging all public officials and citizens to sign up to a Twitter account, enabling rapid and transparent communication between citizens and government.[87]

Issues in the deployment of new forms of data projects

A variety of common issues recur in the deployment of novel types of data. Privacy is an obvious issue: with new means of collecting data come potentially novel intrusions on the privacy of individuals. For example, in the case of WiFi and environmental sensors in Southampton, the implementation seemed to raise relatively few concerns because the types of data collected were centred specifically on the monitoring of particular metrics, such as temperature. The sensors only collect data that the council was previously measuring using much more expensive in-person monitoring. Further, the sensors are designed to use very little bandwidth, and are limited in the amount of data they can send as they rely on transmitting very few packets of data. Thus, no personal data could be collected and the council is only alerted when what is being monitored falls outside a certain threshold that requires the council’s attention. In other cases, such as social media, whilst data may be open and accessible, individuals may not have realised that this means it can be harvested and made use of either to understand citizen perceptions or population movement. This was something many of our interviewees were acutely conscious of.

80 Severo, M. , Feredj, A. and Romele, A. 2016. Soft Data and Public Policy: Can Social Media Offer Alternatives to Official Statistics in Urban Policymaking? Policy & Internet, 8, 354-372.
81 Tackling Urban Mobility with Technology. Google.
82 https://www.waze.com/en-GB/ccp/casestudies
83 https://metro.strava.com/
84 McNeill, G., Bright, J. and Hale, S. 2017. Estimating local commuting patterns from geolocated Twitter data. EPJ Data Science, 6.
85 Bright, J., De Sabbata, S., Lee, S., Ganesh, B. and Humphreys, D. 2018. OpenStreetMap data for alcohol research: Reliability assessment and quality indicators. Health & Place, 50, 130-136.
86 Bright, J., Camargo, C., Hale, S., McNeill, G. and Raman, S. 2018. Estimating traffic disruption patterns with volunteer geographic information. 2nd International Conference on Advanced Reserach Methods and Analytics (CARMA 2018).

87 The Incredible Jun: A Town that Runs on Social Media. Huffington Post, 6 December 2017.



A second issue, particularly relevant in the use of soft data, is in the validation of novel forms of measurement. One of the motivations for the use of this type of data is that existing 'gold standard' datasets do not exist. For example, mobile phone data is interesting in a road traffic context because we lack existing road traffic data. Social media data is interesting in a customer relations context because we expect our existing feedback is biased.

However, this then makes it complex to know exactly how to interpret the new measures developed, or how much to trust them as a basis of decision making. A key related issue here is the extent to which there might be bias in these datasets: for example, the majority of citizens do not make use of Twitter, and those that do are likely to be drawn from younger, urban demographics.[88]

6. Spatial Analysis

Local government is inherently place-based in nature and as such most local decisions have an important geographic component. Services must be provided at a certain distance from given population centres. Individuals have to be connected with these services in a reasonable amount of time (for example, children need to access school facilities). And of course personnel themselves must be allocated and routed around cities in the most efficient way.

Spatial analysis aided through the use of GIS (Geographic Information Systems) is the most common data analysis process underway in local governments and enables answers to many of these types of questions. For example, Danny McAllion (Renfrewshire) commented:

> *"one of the most useful tools that we have got is GIS. GIS teams are embedded in data analytics, they already pull information from several different services...what we want to do in the future is develop the analytical side of things, so we can start doing more sophisticated work on the actual data. For example, we want to do more with customer services information...if we could get a better view of where calls are coming in from and the different choke points this would be really useful."*

Much of the data available to local governments in the UK is spatially tagged, and the data captured by sensors and surveys often feed into spatial analysis. GIS predates data science, though many of the statistical techniques used in data science are prevalent in the everyday work of local governments, particularly visualization and geographic regression.

There are numerous uses for GIS that we encountered through the survey: 90 of our survey respondents (71%) mentioned its use in their organisation, which made it the most used technique. The most common approaches have been the combination of local datasets with national datasets that can be mapped onto spatial boundaries. This allows for the profiling of neighbourhoods through various indices to analyse the availability of local government services as well as certain risks that neighbourhoods face, such as deprivation. Many national indices, such as the Indices of Multiple Deprivation[89] are widely used to profile districts of a municipality to better plan and optimize service provision.

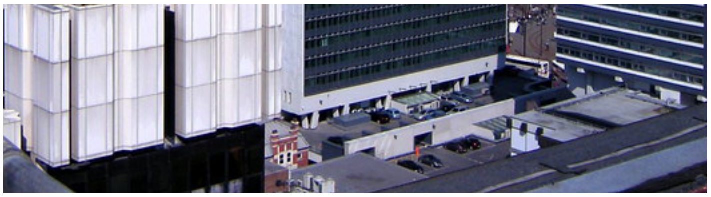
The borough of Doncaster, for example, uses GIS alongside a wide range of indicators to produce a 'State of the Borough' report which combines data from the census, local surveys, health records, and internal benchmarks to report on progress on a range of topics (such as housing, social care, education, and labour). Such uses of spatial information allow local government to produce benchmarks and set attainable and measurable goals for planning and development. Based on the free text responses to our survey, we find that GIS is primarily used for thematic mapping to understand variation between places in a municipality, though this data is often used to inform predictive analytics, experiments, and client-facing tools such as dashboards (which are covered in other sections of this report).

One interview participant, Saqib Yasin (Southampton) says that "It makes for more interesting reading when you can overlap maps and show things visually, and plus, we can begin to corroborate anecdotal reports or assumptions with the data." Dan Carpenter (Oxfordshire) agreed, commenting that:

> *"I have increasingly stopped using printed, static maps and switched to more interactive tools that people*

88 Blank, G. 2017. The Digital Divide Among Twitter Users and Its Implications for Social Research. Social Science Computer Review, 35, 679-697.

89 https://socialvalueportal.com/indices-of-multiple-deprivation-in-the-uk/



Use case 14: Customer segmentation

can use to explore for themselves. People find this interactivity much more engaging, and it helps them understand what is going on much faster. With a web map, people can interact with the data in a new way and it helps to communicate about the data that we hold.”

https://www.flickr.com/photos/motti82/3408044354/

One example of a use case in GIS is customer segmentation: building granular geographic maps of how different types of residents are spread throughout an area. Sarah Tonks (Hull City Council) explained

> *“We’ve recently produced our own local segmentation model, as we found national segmentation models didn’t represent Hull accurately. So, we used output area level socio-demographic census data and thousands of records from our customer relationship management software and overlaid that with our own transactional and attitudinal research data. This has been really helpful in identifying particular groups, for example people who were ‘non-participants’ in the city’s cultural life, who can then be targeted with specific events which were visible to the community.”*

Oxford is another example of work done in this area. Analysts have developed heat maps of the impact of benefit caps around Oxford, which have helped councillors understand which areas will be most affected. They have also developed maps of HMO locations, which have enabled ward officers to understand their wards better. As Tiffany Ko (Oxford) explained, “often there is a need to communicate this information to council members who are busy and need data to be presented in a way that they can easily understand: maps can be a very useful visual tool for doing this.” Indeed, a number of local authorities, including Wealden, Swindon, Rutland, Suffolk, Shropshire, Redbridge, Gloucester, Doncaster, and Oxfordshire have integrated systems and developed tools to geospatially plot service users and their proximity to service provision to help make decisions about where to locate services and how to respond to changing needs.

A further example is Reigate & Banstead’s ‘boundary review tool’,[90] which puts real-time, high-resolution statistical information in the hands of senior leaders within a gamification-style environment, allowing them to prepare boundary proposals themselves. Robert Steele (Reigate & Banstead) explained: “we reaped officer time savings, our councillors (as primary users) were empowered to lead on this work on behalf of their communities and everyone benefited from greater transparency and wider engagement.”

90 Linking people and places. geoplace.

Use case 15: Vulnerable People Emergency Response Programme

The Vulnerable People Emergency Response Programme involves the automated processing of vulnerable person data to prepare priority lists for emergency planning.[91] It was enabled through a partnership between Surrey County Council, Kainos (for their Datactics program), and Skyskape, who built a data-sharing hub which allows relief agencies to share information during emergencies to ensure vulnerable people are protected. The programme established links between health, emergency services, and local authority services, linking data between 30 providers under a data governance policy, also linking data with addresses to map clients. If there is a risk of fire or flooding in any part of the county, the fire and rescue providers could quickly obtain a list of vulnerable people in the area, including their name, address, and needs, so that the rescue team can bring the right equipment, for example if the person needs a sterile environment, uses a wheelchair, or has mental health needs. The system uses the NHS number as the main identifier and includes a postcode address file, so that there can be GPS mapping and a dashboard with an interactive map. Robert Steele (Reigate & Banstead Borough Council, formerly Digital Platform Manager at Surrey County Council) explained the usefulness of the project:

> *"Previously obtaining lists of vulnerable people was a hugely time-intensive process; emergency planners would have to sift through lots of different data sources and make record-by-record judgements about whether the information related to the same person. This could result in the risk of false positives and negatives, consuming time that would have been better allocated to incident planning and response. This project completely changed the situation, delivering prompt, reliable information and freeing up a huge amount of time for emergency planners."*

91 Vulnerable People Emergency Response Programme. Datactics.

Issues in the use of spatial analysis

Spatial analysis has widely been used in local government in recent decades. Consequently, the technique is more mature than the new forms of data described in the preceding section. However, there remain a number of issues that present challenges that analysts using spatial data and GIS need to overcome. These issues are not in themselves unique to spatial analysis and reflect challenges raised elsewhere, including issues related to privacy and security of spatial data, the impacts that findings from spatial analysis might have on individuals and communities, and challenges in centralising and merging datasets.

As much of the data used in spatial analysis involves addresses and postcodes, the privacy and security of spatial data is of significant concern. At times, this can be related to sensitive information about individuals and places, such as poverty, health, and sociodemographic characteristics such as ethnicity, religion, and nationality. As local governments are increasingly interested in the development of interactive maps for public use, the possibility that privacy and security could be compromised depending on how data is presented could increase.

While interactive mapping technologies are exciting prospects for developing interactive tools to encourage the participation of residents, careful attention ought to be paid to the scale at which data is presented. Too high of a spatial resolution could help in the identification of particular people or reveal sensitive information about them (e.g., national origin, religion, income, etc.). A balance is required, and different scales of data presentation should be considered in particular applications.

For example, Danny McAllion (Renfrewshire) commented on how GIS software and interactive web tools can help residents report problems and reduce the number of duplicate records that the council has to sort through:

> *"For example, five or six people [living on] a street might request the same repair. GIS-based software allows people to report something, and others can see that it has been reported, instead of reporting the same thing over and over."*

In such a situation, identifying the exact location of the repair is necessary to avoid duplicate reports. In others, such as a spatial model based on GIS and predictive analytics, too high a resolution could lead to unintended identification of individuals in public-facing tools.

The findings from spatial data analysis can be sensitive in nature and could have unintended consequences on neighbourhoods and communities. For example, in the use case above, spatial analysis might identify particular neighbourhoods, communities, and even individual addresses as vulnerable. This might have impacts that problematise certain spaces in ways that might lead to disproportionate responses.

When issues such as health, poverty, deprivation, and other factors are taken into account, it is important that analysts anticipate how the spatial data they present, and the findings that can be drawn from it, may affect how a community is viewed and what policy impacts labeling a portion of a city or town as "vulnerable" may have on its residents.

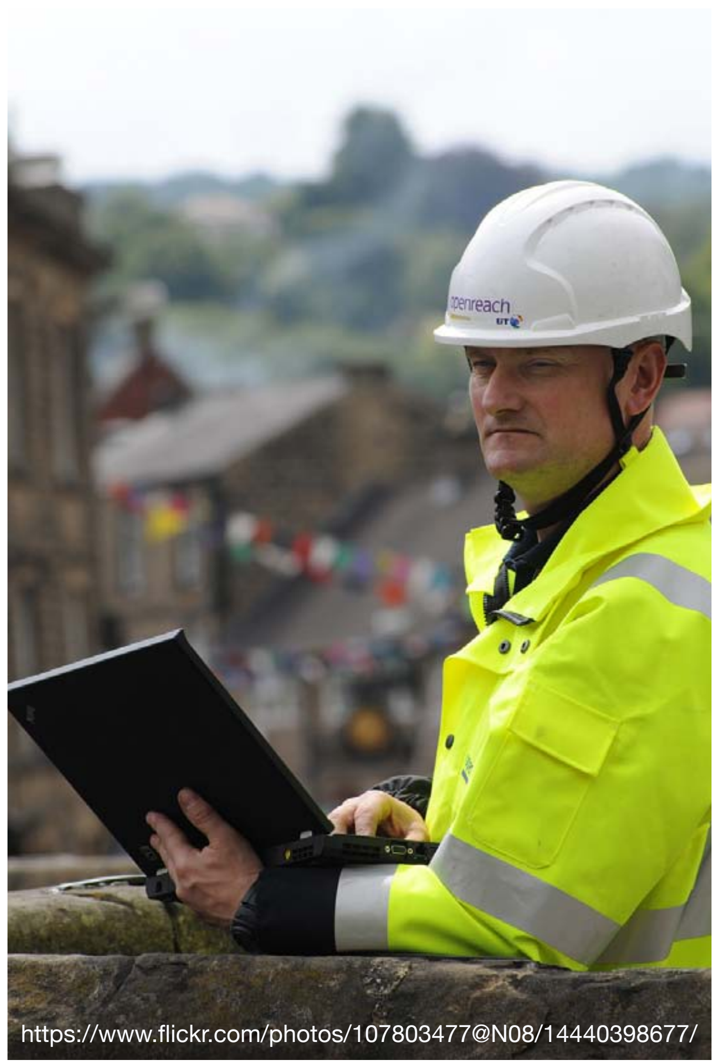
While it is necessary to understand where services need to be delivered across different communities, there are issues that can arise from spatial analysis that might identify one area as more likely to face crime or health issues than another. While this is a necessary part of identifying challenges that a local government faces, when presenting data it is important that analysts are mindful of how particular models might represent a place and consider how findings can be presented in a nuanced manner that might challenge the labelling of places as problematic by focusing on specific issues rather than identification of neighbourhoods, postcodes, or communities as 'problem' areas.

Finally, the primary technical problem facing spatial analysis in local government remains the centralisation and merging of datasets held by a local government. Interoperability of datasets is a persistent problem that we identify across the data science techniques covered in this report. Effective data warehouses and centralised systems can significantly speed up planning processes and spatial analysis.

At times, the use of very different types of data can lead to important cost savings. For example, Dan Carpenter (Oxfordshire) reflects:

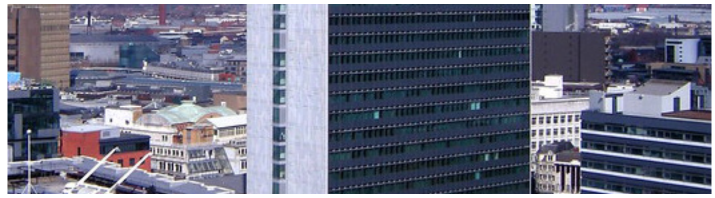
> *"I think planners are suddenly becoming aware of what the data that they already have can do...and what we're increasingly doing is helping them to use that data to answer different questions. For example, we have been modelling the distribution of species like bats, and this is one area where we have been able to say, 'we know where bats are, but can we think about where they might be so that we reduce the number of applications we then send back that require a bat survey'. So, this helps to lessen the burden of work that planners have to do every day."*

Building on existing datasets, centralising, and merging them into large datastores are necessary for all types of data science in local government. The collation of various types of data can have significant, and unexpected cost savings, such as in this example above that involves cross-referencing the modelled distribution of bat species with planning applications.



https://www.flickr.com/photos/107803477@N08/14440398677/

Doing Data Science



In the second part of the report, we have a look at some general issues facing those who want to do data science in local government, together with some tips on how to overcome them.

1. Making the Case

Probably the most fundamental challenge in the area of data science in local government is the need to 'make the case' to senior management to get them to buy in and allow staff time and resources to be dedicated to a new project. This is not to say that senior management are implicitly hostile to data science projects. Rather, in our survey research more than 40% of respondents referred to a lack of commitment to developing a culture of innovation that fosters the use of data analytics in creative ways (**Figure 3**). The most frequently highlighted reason for this is the budget pressures that local governments face, which often makes departments reluctant to fund innovation and risky projects.[92]

One interview participant, Saqib Yasin (Southampton) noted that it can be difficult to justify data science projects relating to non-statutory services in the context of financial pressures. Another interviewee, James Rolfe (formerly from Enfield), told us that "what we are finding is that it's sometimes just too expensive to do the full data analytics piece. Diminishing returns undoubtedly come into play." Indeed, austerity pressures may mean that business insight and analysis functions are the first to be targeted for cuts. One interviewee, who preferred not to be named, said: "It is difficult politically to support analytical and intelligence functions if they come at the expense of front line staff". Anna Crispe (Suffolk) agreed when commenting on her local authority's IT strategy, saying that

> *"We have an IT strategy, which is the right direction of travel …all about putting things in the data warehouse in a structured way and…producing dashboards and much better analytics. But given how tight local government funding is at the moment, we have struggled to find the resources to implement that strategy."*

Many survey and interview respondents argued that the soft skills required to make the business case for a data science project were a core part of the work. Numerous respondents indicated that management were positive about using data science, but were reluctant to dedicate resources to it because they did not clearly understand its potential.[93] Sometimes this can require an outside push. For example, Rhema Vaithianathan (Auckland University of Technology) said that

> *"part of my job is to bridge the gap between the technology and the leadership because it is all too easy for technology to end up in the corner, gathering dust, and never getting used. Getting leadership buy in is a huge challenge with this type of work."*

While there is broad agreement that the corporate culture in local government is shifting to one that embraces more use of data, such projects are often not prioritised. To do so, presentation and argumentation skills are extremely important for analysts in local government. Tiffany Ko (Oxford) highlighted an important issue here, saying that:

> *"whether people you are communicating the data to also have the skills or background to understand it is really crucial. For example, there will sometimes be a lack of understanding around statistical uncertainty, so the onus is on analysts to communicate the distinction between 'definitive' data from data that has some degree of uncertainty."*

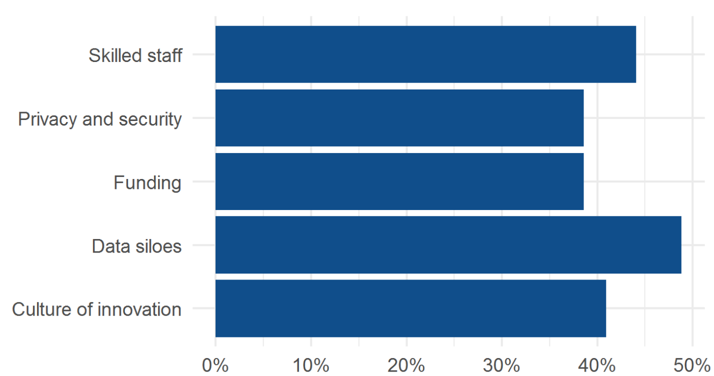


Figure 3: Barriers to data science

92 Interview with Robert Steele (Reigate & Banstead), who said that 'financial constraints' are a number one concern in data science in local government.

93 Berman et al. 2018. Realizing the Potential of Data Science. Communications of the ACM, 61, 67-72.

From the surveys and interviews, we find that the most effective way to make the case for such projects combines case studies and initial analysis which can demonstrate some tangible benefits. As Ritchie Somerville (University of Edinburgh, formerly of Edinburgh Council), put it, "We always start with something small. I was running an innovation team and our job was to do small things that were meaningful to someone … there were things you could do—without blowing the budget—that'll allow you to demonstrate the opportunity and allow people to experience what is possible."

Sam Buckley (Enfield) suggested a similar approach: "If you can start to demonstrate where a particular service, or a particular manager has used this data and use them as a case study, I think other people then buy into that a little bit more quickly." This is a challenge for many analysts in local government as they note that often they are so overwhelmed with performance indicators and day-to-day analysis that they do not have time to develop new projects.

Jon Adamson (Rutland) agrees, saying that: "Although I'm saying this will work, from a business intelligence perspective, other people have to be convinced about that and see it happening." And of course, having already delivered a successful project helps:

> *"We can say with confidence, 'we'll resolve that, we'll get over that, it will work, it's worked before', and I think the challenges in that sense will be less because we'll know that we can resolve them. Whereas, last time around, it was harder to have that same confidence in what we were doing."*

Another core issue is that, because there is little time or space to do experimental work, projects may only take off if they can demonstrate that they resolve an immediate, urgent need. Reflecting on her experience in developing a tool which performed text mining on social media, Lucy Knight (Devon) commented that:

> *"we developed a quick prototype … but in this kind of case, there has to be some severe consequence of not having listened to or been in touch with the mood expressed on Social Media. They [decision makers] would have to have been stung basically."*

Hence space for proof of concept projects (which might realise larger gains further down the line) is often limited by their ability to also display an immediate impact. Knight felt that part of the problem was that her prototype was a solution looking for a problem rather than one that addressed an immediate need. The story tells us that exploratory work—even when it provides useful tools—do not always get taken up. Rather, as she notes, what is important in driving the use of these types of tools is developing the right partnerships with relevant teams and having a clear explanation of what the tool can do.

The approach taken here, which focuses on supporting decision-makers rather than telling them what to do, makes the development of these partnerships more effective. For example, when describing the installation of smart bin sensors, Ritchie Somerville says:

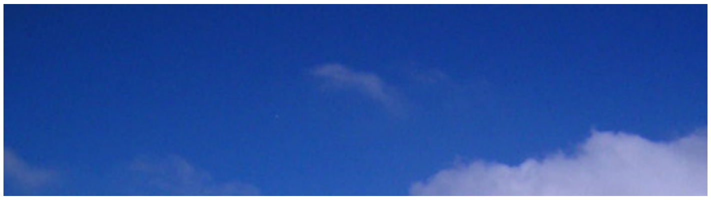
> *"it started from a very humble beginning, of wanting to see what is possible … what was fascinating about it was that it was done through a service, so the service was the lead agent of this; the technology guys were there supporting, but it wasn't a 'Smart City' project, it was a 'we want to make our bins more efficient' project. They were asking all the critical questions of how the service operated. I think they were much more inquisitive about how they wanted to change than about the technology … It was actually all focused on 'How do I make this service more effective?'"*

Sam Buckley also supports showing the benefits, saying that "If you show real benefit and value to people of using those insights, you're much more likely to get that kind of buy-in that you need to make that step to becoming a data-driven organization."

However, while the need to make a compelling case is clear, it is also important to manage expectations. Because of the difficulty of getting management buy-in, there may be a tendency to over promise with respect to the outcomes of a project.[94] Data scientists need to be acutely aware that management might sign off on a data science project expecting major results without an understanding of the time it might take to put it together, the lengthy process of getting data sharing agreements approved, and the time required to clean data and train algorithms. As Anna Crispe (Suffolk) put it when talking about starting her own projects,

> *"I've tried to be really clear from the outset that this is exploratory, this is new stuff. It is difficult, given how*

94 Gil-Garcia, J., Chengalur-Smith, I. and Duchessi, P. 2007. Collaborative E-Government: Impediments and Benefits of Information-Sharing Projects in the Public Sector. European Journal of Information Systems. 16, 121-133.



hard-pressed local government is at the moment, to be spending time and resource and thought-power on something that might not deliver, but keeps being dangled in front of our noses as the keys to the promised land. We have a strong partnership with a University, and it is only through that partnership that we have been able to make progress with this. It's a challenge managing expectations, but still making the case that it's a useful thing to do."

In addition, even with successful projects, there is a need to make sure the results are actually used. Danny McAllion (Renfrewshire) commented on a project which extracts vehicle telemetry data from the council fleet:

"On the technical side, we now have the capacity to analyse the data and collect it … but we're not fully exploiting it to the extent we perhaps could. It's about more putting in place the operational systems that would allow managers to use the data … so we've found it's not so much the data analytics part of it but it is getting that embedded in operational practice."

Sharon Lowes (Sunderland) describes how this has come about in her context: "It was very much for me around an organizational culture change programme, rather than an IT data programme. The significant changes that I've seen are with people who now really want to make decisions, whether they are very small operational decisions, or strategic decision making, using data and using evidence, but requiring support to enable them to do that. The fact that people are coming, knocking on our doors, it's becoming embedded in our way of working, which is fantastic, but which was no easy feat." In order to achieve this, Lowes said,

"we went really back to basics and did a lot of work around the value of data, the value of evidence-based decision-making, the value of making their jobs easier, whether that was their data practices, or their data collection, or their front-line work, whatever it was, we were able to tally the value for the individual, kind of, the 'what's in it for me?'"

One really crucial issue here is the fact that the results of data science projects might not save money that was initially hoped and may even increase demand on public services. As Lynn Wyeth (Leicester) said: "Data science projects don't always save money … sometimes they just open a can of worms and then you're a victim of your own success, because you identify all the needs. But addressing those needs costs money." One interviewee, who preferred not to be named, gives the case of a mobile phone app which allowed reporting of city maintenance issues such as potholes. While this streamlined back office processes considerably, a lot more issues were reported, meaning that in the end the result was more about increasing citizen satisfaction than saving money.

2. Procurement

Procurement of appropriate tools is a critical challenge in the local government data science context. There is no 'one size fits all' data science solution, which means that local government bodies need to adapt and/or develop more or less customized tools that enable them to apply data in a meaningful and beneficial manner.[95] Meanwhile, the process of finding suitable software solutions, agreeing to terms and securing services from external suppliers is highly complex and thus creates a barrier for local government to accelerate new data science initiatives.[96]

Our interview research shows that local governments are approaching the complex process of procuring data science tools in two main ways: either by purchasing off-the-shelf analytics software (e.g. Microsoft PowerBI or Tableau) or by installing open source software packages (such as R and Python). Both of these approaches imply advantages and disadvantages. In our survey, we found more use of off-the-shelf solutions (especially from Microsoft), but also a small but significant group of people making use of open source packages such as R & Python (**Figure 4**).

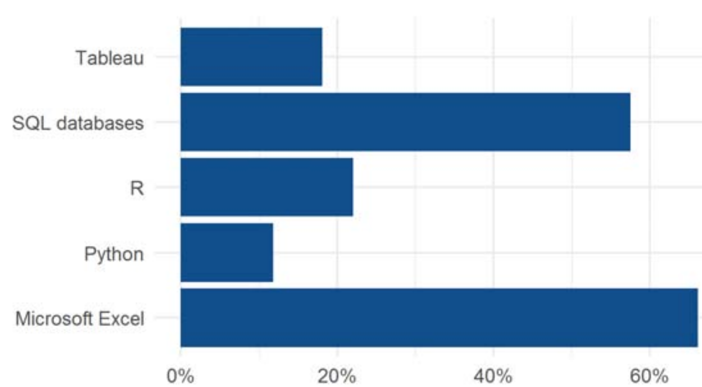


Figure 4: Data Science Tools

95 Kitchen, R. (2014) The Data Revolution. London: Sage.
96 Malomo, F. and Sena, V. 2016. Data Intelligence for Local Government? Assessing the Benefits and Barriers to Use of Big Data in the Public Sector. Policy & Internet 9, 7–27.

Off-the-shelf solutions have enormous potential to kick start data science projects because of their accessibility. They are typically much easier to install than open source software packages, often because they seem to more easily satisfy the requirements of IT departments. For example, Dan Carpenter (Oxfordshire) said:

> *"We've only recently got R on our work laptops. It was a bit of a battle to get it on there in the first place, which is often the case … local councils are spending a lot of money just to check if it is OK to install this software. A secure central download site would be a big help!"*

Off-the-shelf solutions do not have this problem. They are something which thus enables fast implementation, most likely built on proven technology and often include access to support and potential upgrades. They may well also be more user friendly and make it easier to quickly generate high quality, impressive outputs (as one interviewee commented, a 'wow' factor is necessary in the early days of a project to generate momentum and interest). However, they come with drawbacks as well: due to the solution's generalisability there is a 'natural' lack of focus on specific requirements, which might result in missing features.

Another potential disadvantage of off-the-shelf software is that the software company / developer retains the ownership of the software. This creates risks if for example the developer decides not to support or develop the product at any point. Finally, there is a price implication.

Open source solutions present a different picture. They are typically license free, which may mean that they are lower cost than off-the-shelf solutions (though, like any piece of software, there will be costs in terms of staff time for installation and maintenance). They may well be more flexible and easier to tailor to particular solutions or instances. Moreover, they offer an increased level of ownership and control of the software product and accompanying data. However, there can also be concerns about security and privacy. For example, Jon Adamson (Rutland) said: "The issues around confidentiality are significant for us."

The majority of our interviewees had chosen to engage with off-the-shelf software solutions such as Power BI, GIS solutions, Tableau and case management systems. Several of the interviewees expressed how these types of software have been valuable tools to support a growing trust in the data, which has further induced organisational changes. However, some also touched upon the limitations that these solutions imply. For example, one interviewee pointed out that gaining access to a third-party system does not solve the issue of getting access to skills to collect, process and apply the data.

There are also challenges in the procurement process when government may lack key skills and knowledge, though the situation may be improving. As Fran Bennett (Mastodon C) commented: "I hope we are moving towards a moment when government agencies are intelligent buyers and users of this technology." Finally, an interesting point was made by Rocco Labellarte (Oxford), who said that "a lot of government technology comes from firms who specialise in consumer electronics. But applying the logic of the consumer domain to government isn't always straightforward." For example, government technologies will have to work across wide demographics and often cater for a wider variety of use cases.

3. Skills and Training

Skills and training were consistently mentioned as a key barrier to doing data science in local government: 56 of our survey respondents (44%) highlighted that this was an important challenge. Several also pointed to cutbacks which have been made in recent years in local government administrations. Many of these cutbacks have fallen on non-frontline staff, which can often mean people with analytical skills. Cuts in back office services have meant that, as Anne Kearsley (Oxfordshire) puts it

> *"... while some of the materials are recorded, the actual knowledge capital of interpretation and context could be lost, and you have to start from scratch, even when the data is there. So, the question is how do you build on these pieces of work."*

While a desire to preserve frontline staff is understandable, these cuts may have been counterproductive in the long term. As Andrew Ramsay (Bradford) said: "Where you are not in control of your … insights and have to make budget savings, you end up making budget savings in the wrong places and in the wrong way."

Furthermore, a challenge for contemporary analysts can be to keep up with developments in new technology. Sam Buckley (Enfield) said:

> *"It's challenging just keeping up to speed with technology. It's very different even after the last three or four years. Where historically it was very much like, 'do you know your way around a spreadsheet' and you were kind of okay. Where now suddenly you've got an influx of specialist reporting software, statistical packages. I think the thing*

that's really exciting is you can really easily demonstrate value with these now, where before they were kind of seen as nice-to-haves, where now actually when you can start using them to real effect you can really demonstrate that you can invest in those types of products."

Lynn Wyeth (Leicester) stresses in particular that part of the challenge is freeing up time so that training in new skills can occur. One survey respondent agreed, reporting that analysts were overwhelmed with "routine performance or management information" indicators that took time away from developing more creative projects. But Buckley indicated that if time can be carved out, results can be good:

"We've often found that actually just showing them a tool and talking them through it, is basically the best approach to do. I think this breaks down to two stages for me. So, my experience so far has been showing people actually dashboards and visualizations, gets people quite excited quite quickly on and they can see the immense benefit of it. I think the biggest challenge and the next step is actually in making them feel comfortable in doing it themselves. And that's the bit that I think requires more support, more training."

And James Rolfe comments that many skills are transferable from other areas of work: "this isn't learning a new skill that's completely alien to people, it's about applying the things they probably learn in other parts of their lives to day-to-day work." Previous reports and studies have emphasized the major skills gap that is prevalent in society as well as the need to address this vital issue.[97,98,99] As a response to this growing need, data science training courses targeting local authorities have started to emerge. For instance, the UK government has initiated a Data Science Accelerator Program, which specifically aims to upskill and teach people in local government about data science based on the issues that participants have identified in their local setting.[100]

Furthermore, it is also important to remember the advantages of working in a local government context for this kind of technology development. Part of this comes from potential flexibility. As James Lawrence (BIT) said "in local government I'd say that the story is that they don't usually have the capacity to do advanced machine learning themselves, but that they have more leeway in how they actually implement it." And part of it also comes from the fact that government has interesting, hard problems to solve. Brian HIlls (The Data Lab) commented:

"Our skills development programmes … help put public sector organisations like the NHS on a level playing field with the big technical firms. Data science students are interested in solving hard problems … and the NHS have a lot of hard data problems that have a direct impact on people. So, we have found that students are keen to work with a lot of public sector bodies, there are a lot of challenges they can lend their talent to rather than just the big technical firms."

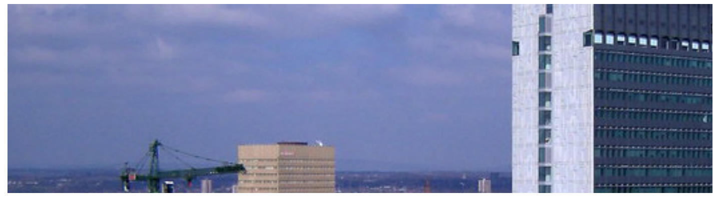
Another aspect to bear in mind is the potential use of outside partners as a way of tackling projects. Some rely on procurement and contracting out to consultants, often from multinational corporations (MNCs), while others work directly with local small-and-medium enterprises (SMEs) and invest directly in staff. However, while some respondents had positive experiences with consultants, some were also sceptical about this type of relationship. Sarah Tonks (Hull) commented:

"Preserving key analytical skills in local authorities is a new challenge due, in part, to public sector cuts, which in turn degrades organisational memory: for example the value system that grades jobs doesn't fully appreciate these skills, and is indicative of a lack of understanding with regards to the possibilities that the innovative use of data provides. It's also often difficult to get this type of professional into the public sector."

However, she said: "Making use of private sector analysts and consultants prevents the organisation itself from learning, which is quite important." Liz Barnard (West Suffolk) agreed, saying that: "I think it is fair to say we have found there to be limited benefits from external consultants doing this sort of work, compared to those with existing local knowledge in-house."

Working with the university sector can also be of significant benefit. Si Chun Lam in Coventry says that: "I think that universities working with local government to train people to do more and potentially for partnerships to be built as well could be really productive." Many local governments reported bringing in postgraduate students from local universities with an interest in the area to work on projects.

97 Government Transformation Strategy: better use of data. Cabinet Office.
98 Skills of the Datavores: Talent and the data revolution. NESTA.
99 Mind the data skills gap: UK businesses warn of a shortage of talent able to transform big data into big value. NESTA.
100 Introduction to the Data Science Accelerator programme. Government Digital Service.



Andrew Ramsay (Bradford) also gave the example of collaborations with academia: he developed a partnership with a business school where students on relevant courses were given data science projects. Of course, there were concerns over data sharing and privacy, but Ramsay said that these "weren't unsolvable—we got past that."

An internship approach can be beneficial for councils because it provides specialised resources with technical training at a relatively low cost for short periods of time. Highly specified projects have been reported as very important, though in some cases local governments felt that the students participating in these programmes did not have the requisite skills and some competitive selection process would be useful. Developing ongoing partnerships with university departments with a focus in policy and data (which are proliferating across the UK) can help to facilitate useful knowledge exchange. In other cases, working directly with academics has been shown to also be useful in developing tools and conducting exploratory work.

However, despite the potential of working with outside partners, most people we spoke to also highlighted the importance of having in-house expertise. For example, Si Chun Lam (Coventry) said: "In Coventry's case, we're lucky that IT is not a barrier for us, because it's our own in-house IT team. We talk to them and we say, 'we need this', and they understand it and work with us to make it happen."

4. Ethics, privacy and data protection

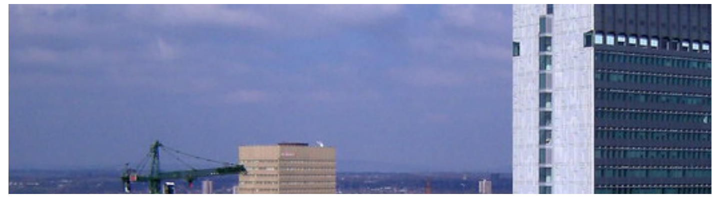
A further data science challenge concerns the need to respect requirements in terms of ethics, privacy and data protection. Privacy and data protection concerns include inadequate security and privacy safeguards that undermine public confidence, absence of clear data protection guidelines that create uncertainty around rules, and direct statutory barriers to potentially useful information sharing, processing, and use.

New concern has arisen surrounding the implementation of the General Data Protection Regulation (GDPR). Some local authority staff members have suggested that poor understanding of the law within local authorities has made people reluctant to share data and that in preparation for GDPR implementation certain historic datasets have been protected or deleted, reducing the amount of information that could otherwise have been available, for example, as training data for machine learning processes (indeed, 39% of survey respondents mentioned privacy and security concerns as a barrier to data science). For example, Tiffany Ko (Oxford) said that "GDPR has made people more cautious about sharing data. It doesn't mean you can't do it, but perhaps we still lack the confidence that collaborative work will be able to proceed past the initial planning stages."

Some organisations argue that these concerns are overblown and that the new regulation clarifies and improves rules around information processing, suggesting that organisations should not see the GDPR as a threat, but as a way to improve information sharing by reviewing existing processes and engaging with staff and clients.[101] Further, some argue that concerns about privacy and information protection are just foils for more fundamental organisational challenges surrounding cooperation, coordination, and information sharing between distinct organisational units.[102] Of course, since as far back as the 1970s, there has been concern that laws about privacy and information protection could place a freeze on information sharing and legitimate research in health and human services.[103] As public sector organisations have pursued digitisation efforts, concerns around privacy and information protection law have persisted, held up as reasons why data, system, and service integration have not been successfully completed.[104]

That being said, there are genuine concerns related to privacy in the context of data science, for example, how to share child safety data with machine learning specialists. In 1995, the Information and Privacy Commissioner of Ontario, Canada in partnership with the Netherlands Data Protection Authority coined the term 'Privacy-Enhancing Technology'

101 The General Data Protection Regulation - an Opportunity for Change. Centre of Excellence for Information Sharing.
102 Horsley, T. 2014. Troubled Families Update Part 1 – the Background. Housing Quality Network.
103 Doll, R. 1974. Public Benefit and Personal Privacy: The Problems of Medical Investigation in the Community. Proceedings of the Royal Society of Medicine, Symposium on Constraints on the Advance of Medicine, 67, 1281–85.
104 Landsbergen D., and Wolken, G. 2001. Realizing the Promise: Government Information Systems and the Fourth Generation of Information Technology. Public Administration Review 61, 206–20.
Eynon, R. and Dutton, W. 2007. Barriers to Networked Governments: Evidence from Europe. Prometheus 25, 225–42. Yang, T. and Maxwell, T. 2011. Information-Sharing in Public Organizations: A Literature Review of Interpersonal, Intra-Organizational and Inter-Organizational Success Factors. Government Information Quarterly 28, 164–75.



in their paper ‘Privacy-Enhancing Technologies: The Path to Anonymity’. Since then new technologies have emerged to supplement the suite of options to address privacy concerns. Agencies can, for example: encrypt data, attach meta-data tags to their data to indicate whether it is personal information or sensitive data, and build in or take advantage of privacy rules that can be or have been programmed into applications.[105] There is also the Our Data, Our Way project, which was adopted by Rhema Vaithianathan in her work on predictive analytics. This project makes four recommendations for data use and sharing: 1) avoid a deficit approach, 2) people’s needs come before data, 3) provide evidence of sound practice, and 4) build evidence of sound practice with communities who want to make data work for them.[106]

Si Chun Lam in Coventry speaks to practical steps that can be taken to include privacy in the design of systems, make them transparent and accountable, and give the public confidence:

> *“our ICT colleagues have restructured our data warehouse so that at every stage, a privacy impact assessment is conducted before datasets are combined and permissions sought at an appropriate level so we can determine, ‘if we were to combine these two sets of data together, what’s going to happen?’ So, there’s some accountability that helps us to determine how data is combined together. So, we’ve got a process to get that sorted out so that where there’s a challenge or a deficiency, we can actually illustrate how it was reached so that it can give people some degree of trust in what we’ve done.”*

Rhema Vaithianathan (Auckland University of Technology) also highlighted the importance of consultation and engagement in addressing this type of concern: “I’m a firm believer that you must, first and foremost, go and talk to and listen to the community members who are most likely to be subject to your algorithms and listen to their concerns.”

Finally, there is the need to remember that projects such as data merging should not just be pursued for the sake of it. Peter Tolland (North Lanarkshire) said:

> *“We’re trying not to bring datasets together, just because you can. Because really you can bring any datasets together. So, what we’re saying is that we wouldn’t do that. What we say is that we have a defined problem that we’re trying to solve and then we will bring that information together. We’ve talked to customers, and they are OK with that.”*

5. Sharing data

The ability to share data (between different branches of an agency, between different agencies within a local authority and even between local authorities) is a fundamental enabler of some of the data science techniques we have mentioned above; it’s also one of the most difficult challenges; indeed, difficulties with ‘data siloes’ were the most frequently mentioned barrier to data science projects in our survey (49% of our respondents mentioned having problems here). For example, as one interviewee, who preferred not to be named, commented in relation to a project about building a centralised data warehouse:

> *“there is a lack of integration between different parts. Every single office and administration only thinks about their part of the job. For example, if I am in the office that is overseeing all the private regulation works, I’m not interested in sharing my information with other officers.”*

Sometimes this sharing can also be complicated by the fact that work takes place in markets with multiple suppliers, who often have systems that are not compatible between each other, which can create problems. As Andrew Ramsay (Bradford) put it: “Across the region we have six or seven different providers and that causes issues as the systems begin to determine social work practice rather than the social work practice determining the systems”.

In the UK, these challenges were arguably made worse by the adoption of “New Public Management” in the 1980s, which resulted in government being run more like a business with characteristics of disaggregation, competition, and incentivization.[107] Disaggregation reduced inter-agency collaboration by separating functions and siloing services. It

105 An Introduction to Privacy Enhancing Technologies. iapp.
106 Data Futures report: Our Data, Our Way - What New Zealand people expect from guidelines for data use and sharing. Data Futures Partnership.

107 Hood, C. 1995. The ‘New Public Management’ in the 1980s: Variations on a Theme. Accounting, Organizations and Society 20, 93; Dunleavy, P., Margetts, H., Bastow, S. and Tinkler, J. 2006. New Public Management Is Dead - Long Live Digital-Era

also attempted to put providers and their administrations in competition, all things which have made data sharing more difficult.[108]

Furthermore, disaggregation has resulted in lots of different systems which can complicate sharing even when there is agreement. Anna Crispe (Suffolk) highlighted the problem, saying that "just within Children's services, we use eleven different systems, none of which talk to each other, and most of which don't use the same unique identifier. We were getting increasingly frustrated that we thought there was value in the data, but we couldn't make people understand that value, because they can't physically see the information, because it's all hidden in these systems".

The same applies to multiple private systems. Anne Kearsley (Oxfordshire) gave the example of adult social care: "it's a complex market with about 50 home care providers, each delivering care to around four or five of the 14 localities. As a result, providers' care delivery overlaps a lot and there is lack of capacity due to inefficiency in the amount of time being travelled. Solving this is a really complex data problem … and so we need to share understanding of the inefficiencies with our providers so that they trust us to reallocate contracts geographically."

One impact of the complexity of sharing data is that it dramatically slows down development cycles. Brian Hills (The Data Lab) gives the example of the 'delayed discharge' project, which identified ways in which the discharge process in NHS Scotland could be sped up, thus freeing up crucial resources.[109] Hills said:

> *"The project was not too much about data science but more about organisational navigation and data sign off, etc. We took one year's health board data and we tested a model against that data, and it was 98% accurate…the NHS tested this model against other health boards and it held up in the other test cases. The challenge is how do you make that innovation cycle go faster? It took a year to two years since the project kicked off."*

Governance. Journal of Public Administration Research and Theory 16, 467–94.
108 Malomo, F. and Sena, V. 2017. Data Intelligence for Local Government? Assessing the Benefits and Barriers to Use of Big Data in the Public Sector. Policy & Internet, 9, 7-27.
109 Collaboration between Scottish Government, NHS/NSS and The Data Lab Wins Award. The Data Lab.

There are sometimes no easy answers to these problems. Some of the bits of advice we have given above also hold true: starting small and making a case to the individual agencies about benefits.

It is worth highlighting that many analytics projects may not need extensive data joining. For example, Rhema Vaithianathan (Auckland University of Technology) is optimistic about potential uses of data science even in areas that do not have large, pre-existing data warehouses. "What we feel now is that, we tended to start where data is rich, but … now we are working in areas with far fewer 'features' [variables upon which predictions can be built], and you can still achieve strong predictive power."

Finally, having senior management support is also crucial for getting data sharing agreements in place. For example, the Humber sub region has a successful data sharing agreement between local authorities, police and health agencies. Sarah Tonks (Hull) commented that: "We still have barriers in individual cases…but we have CEOs wanting to work in principle and that helps when you are butted up against a wall."

https://www.flickr.com/photos/kenjonbro/3218106225/

Future trends

This report has been about considering the current state of data science in UK local government. In the conclusion, we want to look to the future a little bit and consider what directions data science might take next.

One key area concerns joined up work across local authorities. As Andy Hollingsworth (Behavioural Insights Team) said: “One of the most promising avenues...is to work across authorities. This provides the scope for larger trials and potentially enables you to solve problems that single authorities couldn’t manage on their own.” However, he also cautioned that working across authorities is by no means straightforward: “things such as different IT systems, differences in the way data is collected and collated and small differences in the way services themselves are delivered all make this type of collaboration a real challenge.” Brian Hills (The Data Lab) agreed, commenting:

> *“there are 32 local authorities in Scotland, and they all have the similar problems…many have the same suppliers and they don’t have internal budget to exploit their data or understand it. We’ve been working with a couple of them directly, but what we found is that we need to try and scale that to have greater impact. For example…so many local authorities will pay to bus or taxi schoolchildren, and they want to look at ways to optimise it: should we replace a taxi with a bus, for example? That is a generic problem for all 32 local authorities which organisations such as the National Improvement Service for local government in Scotland have a remit to tackle.”*[110]

What is interesting in this regard is whether solutions developed in one borough can be ported to another, perhaps as a service. For example, the London Borough of Hammersmith & Fulham have a branch which seeks to offer data science services to other councils.[111] However, it is also worth highlighting the difficulties some machine learning projects in the US have experienced when being ‘ported’ from one context to another.[112] Often, training algorithms on local data is key.

A second key area concerns the appropriate structure for data science services to take within local government. We found a lot of variety in the people we talked to: some people favoured data scientists embedded in particular teams and departments, able to take advantage of substantive knowledge and local expertise. Others favoured centralised data science ‘services’, offering expertise to all departments and building economies of scale. Still others looked at even wider ‘offices’ which would cut across multiple local authorities, which comes with several advantages.[113,114]

As Michaela Breilmann (Suffolk) explained: “having even a virtual office for data analytics formalises everything and gives everybody that funds it a much higher stake in it.” A formalised office also has the ability to apply what Liz Barnard (West Suffolk) called a “quality stamp” on individual pieces of research, because they have developed a good reputation more generally. All of these models have strengths and weaknesses, and the best way to incorporate data science into government remains a source of lively debate.[115]

A final and perhaps most important area is the extent to which local government can develop into a place which actively fosters innovation (something which 41% of our survey respondents mentioned as a challenge). Part of this is about allocating budget for analysts to do more than just statutory reporting: to engage in training courses, to experiment with new projects, to come up with their own ideas (39% of survey respondents also mentioned funding issues as a key barrier).

But part of it also involves shifting the way projects are thought about. As Lucy Knight (Devon) put it: “Leadership is beginning to get this. But we have to think quite hard about the language we use when trying to explain the possibilities. This [area of work] is not a tech piece it is a culture piece.” Or, as Ritchie Somerville (University of Edinburgh) said: “How do you engender a greater sense of curiosity within the public sector? All the examples have involved someone who has been prepared to be curious.” Bringing in this type of culture is not easy: it requires a potential acceptance of failure, and a willingness to spend time creatively on projects without a guarantee of success. Ultimately however it is this type of culture that will truly enable data science in local government.

110 http://www.improvementservice.org.uk/
111 Business Intelligence - transformational services.
112 Machine Bias. ProPublica.

113 See e.g. the West Midlands Office for Data Analytics.
114 The Worcester Office for Data Analytics.
115 What’s the ideal model for an Office of Data Analytics? NESTA

Research method

The report is based on work that has been ongoing since 2017. Following extensive desk research, we created a survey instrument asking some fundamental questions about the types of data science techniques being used and common barriers and challenges to using them. We sent a personal email invite to at least one person in all of the (almost 450) local authorities in the UK, asking them to complete the survey. The individuals were chosen because they worked in areas in and around 'data science' such as business intelligence, analytics, open data and government digitisation.

We received over 120 responses, of which 82 were complete. At least 64 different local authorities were represented (**Figure 5**). In the second half of 2018, we conducted almost 40 in depth interviews with professionals working in the area. Some of these professionals were selected from the survey, and others were contacted because they had been identified in our desk research as prominent speakers or thinkers in the area of government data science. All responses were given in a personal capacity.

The work we did aimed to provide a snapshot of the types of data science going on in the UK. However, we want to note that, while the empirical work we undertook is quite extensive, it should not be interpreted as 'representative' or generalisable to the UK as a whole. We could not control who responded to the survey, and it is likely that we got over representation from people who already have an interest in data science. In addition, the individuals involved responded in a personal capacity, rather than on behalf of their organisation: they may well have been unaware of (for example) the exact extent of use of a particular analytical technique in other parts of their organisation. Therefore, the results of the survey should be considered indicative rather than conclusive. Nevertheless, considering that there are around 450 local authorities in the UK as a whole, 120 survey responses and 40 interviews is a significant volume.

About the Authors

Jonathan Bright is a Senior Research Fellow at the Oxford Internet Institute. He conducts research in the areas of digital government and digital politics.

Bharath Ganesh is a political geographer focusing on data science and local government and the ethics and politics of researching violent online extremism.

Thomas Vogl is a doctoral student at the Oxford Internet Institute investigating public sector organizational memory in the digital age. Prior to his studies, he worked in provincial public service in Canada.

Cathrine Seidelin is an industrial PhD student at the IT-University of Copenhagen and the Education Secretariat for Industry (Denmark). Her research aims to develop co-design methods to support SMEs in the process of designing data-driven services.

https://www.flickr.com/photos/didbygraham/4360031385/

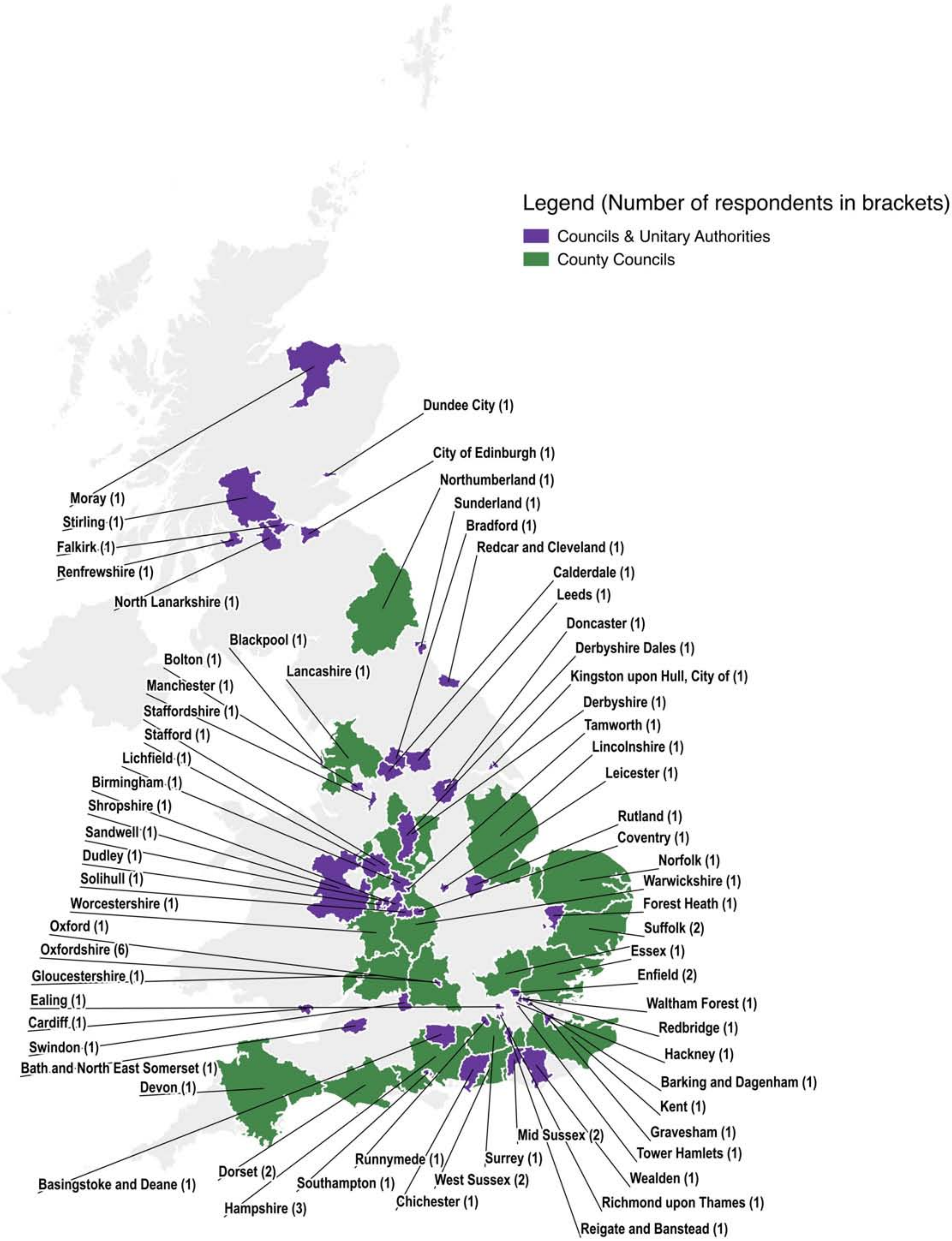
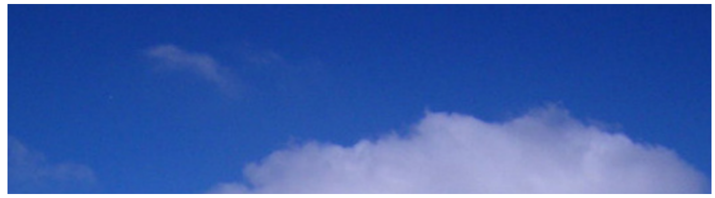


Figure 5. Locations of Survey and Interview Respondents

Interviewees



Jon Adamson, Business Intelligence Manager, Rutland County Council

Mark Allen, Head of Strategic Commissioning, Hampshire County Council

Liz Barnard, Service Manager (Corporate Policy), West Suffolk Council

Fran Bennett, CEO & Co-founder, Mastodon C

Michaela Breilmann, Data and Insight Manager, Suffolk Office of Data & Analytics

Sam Buckley, Head of Data and Management Information, Enfield Council

Matthew Cain, Head of Digital and Data, London Borough of Hackney

Phil Canham, Data Scientist, Corporate Insight Hub, Barking and Dagenham

Steve Carefull, Director, PA Consulting Group

Dan Carpenter, TVERC Projects Manager, Oxfordshire County Council

Anna Crispe, Head of Knowledge & Intelligence, Directorate of Health, Wellbeing and Children's Services, Suffolk County Council

John Gleek, Head of Research and Intelligence, Doncaster Council

Brian Hills, Head of Data, The Data Lab

Andy Hollingsworth, Senior Advisor, Behavioural Insights Team: North

Anne Kearsley, Digital GIS Solutions Manager, Oxfordshire County Council

Lucy Knight, Data Lead, Devon County Council

Tiffany Ko, Data Analyst and Policy & Partnership Officer, Oxford City Council

Rocco Labellarte, Chief Technology and Information Officer, Oxford City Council

Si Chun Lam, Insight Development Manager (Place and Public Sector Transformation), Coventry City Council

James Lawrence, Head of Data Science, Behavioural Insights Team

Sharon Lowes, Senior Intelligence Lead, ICT and Intelligence Service, Sunderland City Council

Danny McAllion, Data Analytics & Research Manager, Policy & Commissioning, Renfrewshire Council

Marion Oswald, Senior Fellow in Law, Director of the Centre for Information Rights, Department of Law, University of Winchester

Spencer Payne, Insight Service Manager, Insight Service, Warwickshire County Council

Andrew Ramsay, Corporate Programme Lead, Bradford Council

James Rolfe, former Executive Director of Resources, Enfield Council (now Chief Operating Officer at Anglia Ruskin University)

Ritchie Somerville, Data Innovation Director, University of Edinburgh (formerly Edinburgh Council)

Robert Steele, Geographic Information Manager, Reigate & Banstead Borough Council

Peter Tolland, Chief Information Officer, North Lanarkshire Council

Sarah Tonks, Customer Insight and Engagement Advisor, Hull City Council

Stefano Tripi, Planning and Control Unit, City Manager's Office, Municipality of Modena

Prof. Rhema Vaithianathan, Co-Director of the Centre for Social Data Analytics, Auckland University of Technology, NZ

Lynn Wyeth, Head of information Governance & Risk, Leicester City Council

Saqib Yasin, Service Lead, Data Integration and Performance, Southampton City Council

This project has received funding from Google.
Oxford Internet Institute | University of Oxford | March 2019
Download: https://smartcities.oii.ox.ac.uk/data-science-for-local-government-report/

Publication 8

Vogl, T., Seidelin, C., Ganesh, B., Bright, J. – **Smart Technology and the Emergence of Algorithmic Bureaucracy: Artificial Intelligence in UK Local Authorities**.
[Resubmitted to Public Administration Review - Special Issue on Transformation in Government]

Smart Technology and the Emergence of Algorithmic Bureaucracy: Artificial Intelligence in UK Local Authorities

Thomas Vogl[†], Cathrine Seidelin[*], Bharath Ganesh[†], and Jonathan Bright[†]

[†] Oxford Internet Institute, Oxford University, Oxford, United Kingdom, thomas.vogl@oii.ox.ac.uk, bharath.ganesh@oii.ox.ac.uk, jonathan.bright@oii.ox.ac.uk

[*]Department of Computer Science, IT University of Copenhagen, Copenhagen, Denmark, cfre@itu.dk

Abstract:

In recent years, local authorities in the UK have begun to adopt a variety of 'smart' technological changes in support of service delivery. These changes are producing profound impacts on the structure of public administration. Focusing on the particular case of artificial intelligence, and specifically autonomous agents and predictive analytics, a combination of desk research, survey questionnaire, and interviews were used to better understand the extent and nature of these changes in local government. Findings suggest that, while at a nascent stage, local authorities are beginning to adopt smart technologies and they are having an unanticipated impact on how public administrators and technology become imbricated in the delivery of public services. This imbrication is described as algorithmic bureaucracy, recognizing the many interactions between offices held by public administrators and computational algorithms that are increasingly part of the everyday working environment. This construct provides a framework to explore how these technologies are transforming the socio-technical relationship between workers and their tools, as well as the way that work is organized in the public sector.

Evidence for Practice

- A new form of bureaucratic organization enabled by computational algorithms is beginning to emerge in local authorities
- Autonomous agents can assist citizens with their service needs, but they can also be used to help public administrators to carry out their tasks.
- People using smart technologies in local authority service provision are attempting to deal with complexity not by simplifying problems into set procedures, but through adaptive predictive algorithmic models that can learn from new inputs and changes in conditions.
- When introducing new computational algorithms, identify the relevant social groups that are impacted by its implementation, understand the contextual implications from their perspectives, and leverage internal capacity as much as possible in order to address local needs and challenges about which outsiders may not be aware.

In the past decades, local governments have developed digital information technology infrastructures, which create an environment that allows for the development of new applications to support efficient digital service delivery. However, these innovative possibilities create new socio-technical challenges (Rodríguez Bolívar and López-Quiles 2018). This article focuses on the adoption of new technologies that are enabled by computational algorithms in local authorities, in particular looking at autonomous agents and decision assistance tools. It explores how computational algorithmic technologies offer an opportunity to enhance Weberian machine bureaucracy while preserving key public sector values of fairness, impartiality, and standardization (Cordella and Tempini 2015). In this way, these types of tools could produce profound impacts on the structure of public administration in local authorities.

While smart technology could easily be interpreted as a 'neat and stylish term', smart technology, in this article, is understood as computational algorithmic tools that are programmed so as to be capable of some independent action, whereby they are quick at learning and are able to react or respond intelligently to their informational environment, including differing requirements, varying situations, or past events (Oxford English Dictionary 2019). Following this definition, we refer to autonomous agents and predictive analytics decision assistance tools as smart technology. Four key questions guide this research: 1) to what extent are smart technologies being adopted in UK local authorities? 2) what are the characteristics of these technologies? 3) what are the ways in which smart technology integrates into the organization of work in local authority public administration settings? 4) what are the implications of this change for how we conceptualize the study of public administration in the era of smart technologies?

This study suggests that smart technologies are at an early, but foundational, stage of adoption in local authorities and argues that smart technologies add a new element to the socio-technical organization of public administration in local authorities. It is not just a shift from street-level to system-level discretion (Bovens and Zouridis 2002). Instead, where there is a shift to the system level, multiple stakeholders, representing different relevant social groups with different forms of knowledge and perspectives, are involved in design and implementation (Pinch and Bijker 1987). Where a tool in use remains at the street-level, attention is needed to how smart technology mediates informational feedback loops and collective intelligence. Based on these results, this article then offers a socio-technical framework for the continued study of smart technologies in public administration.

Administrative systems have a long history of evolution in response to the demands of modernity. Machine bureaucracy embedded ideals of impartiality, procedural fairness, and efficiency in a hierarchy of rule governed offices supported by files, an enhancement over previous systems, such as patrimonialism (Weber 1968). However, since the middle of the 20th century, commentators have questioned the ability of traditional bureaucracy to deal with the increasing complexity of modernity and have worried about undesirable inertia (Elgin and Bushnell 1977). Scholars began to argue about new approaches to public administration (Pollitt and Bouckaert 2011), some of which focused on an approach that emerged in the 1980s and came to be known as the New Public Management (Lynn 2001), which was characterized by managerialism and the use of market mechanisms, such as outsourcing, as a means to overcome some of the challenges associated with modern complexity and make government more efficient (Hood 1995).

In parallel with New Public Management changes, there were advancements in the development of information technology that were impacting the infrastructure of public administration (Margetts 1999), in particular, the development of the internet as a means to communicate information quickly between computers (Naughton 2001). These changes had begun much earlier with the introduction of computation (Wilkins 1968; Simon 1973) and scholars realized that “[t]o design effective decision-making organizations, we must understand the structure of the decisions to be made; and we must understand the decision-making tools at our disposal, both human and mechanical - men and computers" (Simon 1973, 272). Unfortunately, while some had made early predictions of the valuable role that computation would play in decision support (Hadden 1986; Danziger and Kraemer 1985; Hurley and Wallace 1986); there was a period where it was seen to have been woefully neglected in the study of public administration (Dunleavy et al. 2006; Pollitt 2011), with only a select few scholars suggesting that information technology (IT) was changing the fundamental paradigm of public administration to one with digitalization at its core (Dunleavy et al. 2006).

More recently, there is renewed interest in the impact of new developments in IT on the very structure of public administration (Agarwal 2018; Margetts and Dorobantu 2019). Where written rules and procedures are not fast enough and there are too many for people to remember, algorithms are seen as a way to provide support. Some research has begun to look at how more sophisticated algorithms that rely on a foundation of computation, administrative data collection, and information communication create new ways to use data in public administration (Mergel, Rethemeyer, and Isett 2016; Allard et al. 2018), though not always for the better (Lavertu 2016). Other research suggests that smart technologies could displace work through automation (Bovens and Zouridis 2002). However, algorithms may do more than improve analytics and automation, they may also change the nature of public administration.

With the emergence of smart technology, this article suggests that a new model of bureaucratic administration is combining people, computational algorithms, and machinereadable electronic files and forms to deal with complexity and overcome some of the limitations of traditional bureaucracy, while preserving core public sector values. This change necessitates a new framework within which to structure research of digital public administration. In the following section, we will situate the concept of smart technology broadly within technological change and then within public administration, highlighting lacunae in the current literature on technology in local authorities. We will then set out the approach we took to explore the current state of smart technology in local authorities and its impact on the way public administration is organized, which includes a survey and two illustrative case studies. Following that, we will present the results of the research, discuss their implications, and conclude.

Theory

This study is situated in the context of evolutionary theories of digital government progress, and the associated theories around a shift from street- to system-level bureaucracy and from values of procedural equality to equality of outcomes. These three constructs are elaborated below.

Historically, changes through digitization were seen as the prerogative of central government, with smaller orders of government lacking the skills and capacity to deliver major technological change (Dunleavy et al. 2006). Some scholars suggested that as digital changes progressed, the environment would evolve to include greater digitization in, first, state, or regional, and then local government (Gil-Garcia and Martinez-Moyano 2007). Some began to test this hypothesis in the context of websites in municipalities (Moon 2002). While IT may have been more centralized early on, with those closer to local matters less IT intensive, this has begun to change (Malomo and Sena 2017; Rodríguez Bolívar and López-Quiles 2018). In the UK, with austerity and digital strategies, local authorities are looking for efficiencies using technology and there are diverse approaches across the country (Symons 2016; Dencik et al. 2018; Bright et al. 2019). While some comment on the persistent challenges that local authorities face (Malomo and Sena 2017; Fischer et al. 2019), there appear to be examples where local authorities are charting a new course through the use of smart technologies. Despite this renewed interest, some of the best publicly available data in government technological change is focused on outward facing e-service delivery (UN 2018), rather than on how technology can transform work internally across the organization. As a result, there is a gap relative to our understanding of smart technology adoption in local government.

There has always been a balance between rules and discretion in bureaucratic organizations (Crozier 1964; Lipsky 2010; Zacka 2017). With the ability to embed rules in code (Lessig 2006), some have argued that discretion has moved from front-line workers to system designers for routine operations (Bovens and Zouridis 2002). Others suggest street-level bureaucrats may be cut out of some interactions entirely as autonomous agents over web interfaces support isocratic services for individuals (Dunleavy and Margetts 2015). However, in some cases, the reality of effective smart technology use may continue to include a role for both human and machine agents. While there is a substantial literature on the imbrication of social and material agencies in organizations (Orlikowski and Scott 2008; Leonardi 2012; 2013), this understanding has not broadly translated into the public administration field.

A key characteristic of smart technologies is their ability to learn from continuous real-time data inputs and adjust their responses accordingly. Previous studies have looked at early conceptions of artificial intelligence, such as expert systems (Hurley and Wallace 1986; Hadden 1986) or the functional simplification and closure of procedures in technologies (Cordella and Tempini 2015). Some have begun to explore the potential impact of machine learning technologies on work and decision-making in public administration (Agarwal 2018). Others suggest that this shift towards learning technologies could be accompanied by a change from systems that deliver procedural equality to those that can provide equality of outcomes (Dunleavy and Margetts 2015). However, empirical study of the implications of smart technologies are more limited.

With these theoretical positions, this study aims to explore three theoretical constructs: the extent to which digital transformation has evolved towards increased smart technology use in local government; the social and material implications of smart technologies for the relationship between street-level and system-level bureaucrats; and the replacement of rule-based systems of procedural equality with outcomes focused learning technologies. The findings have implications for the conceptualization of public administration in the era of smart technology.

Methodology

This study is based on research that took place between November 2017 and December 2018. It adopts a similar methodological approach to other research in digital government change (Eynon and Dutton 2007). In particular, it included three techniques: a survey, desk research, and subsequent in-depth interviews conducted with people working in the area of local government data science in the UK. The underlying assumptions for the research are that the introduction of smart technology has socio-technical impacts on the nature of public administration in local authorities and that those impacts can be better understood by eliciting the experiences of people working with these technologies.

Desk research was used in two stages of the research. In the first stage, it was used to find organizational charts, contact lists, and other names associated with smart technology projects in local authorities to create the survey invitation list. Where specific names could not be found, invitations were sent to the generic email address of the local authority. In the second stage, it was used to find publicly reported information about smart technology projects in local authorities to enhance our understanding of the breadth of the projects, to identify people and cases of innovative practice or relevant IT change in the area of artificial intelligence and algorithms within local government in the UK, and to corroborate information from the cases discussed during interviews.

A survey instrument was developed, which included a mixture of closed and open text responses (the full survey is available in appendix A1). The survey was originally designed to provide a broad overview of the spread of data science[1] technologies, reasons for their uptake, barriers to their implementation, and the impact of these technologies. In this paper, the focus is on a subset of questions about smart technologies. Personal email invitations to complete the survey were sent to a list of at least one person in 285 of the 408 local authorities in the UK[2]. In total, 402 invitations were sent. Individuals who were invited to participate were selected because they worked in areas related to algorithms and artificial intelligence such as information technology, business intelligence, analytics, open data, and government digitization. Local authority organizational charts and contact lists were used wherever possible. The survey was at least partially completed (29% or more of the survey was completed) by 93 respondents, for a response rate of 23%. Of those respondents who provided their position (55), 32 were from intelligence, data, and research positions, 10 were from digital and IT, 7 were from policy and strategy, and 6 were from service or projects. 72 different local authorities with small, medium,

[1] Data science is defined as bringing in new decision making and analytical techniques to local government work (e.g. machine learning and predictive analytics) and also expanding the types of data local government makes use of.
[2] 343 local authorities in England (36 Metropolitan districts, 32 London boroughs plus the City of London, 55 Unitary authorities plus the Isles of Scilly, 26 County Councils, and 192 District councils), 32 Unitary authorities in Scotland, 22 Unitary authorities in Wales, and 11 Unitary authorities in Northern Ireland.

and large populations and geographical areas were represented. The breakdown by type of local authority is presented in table 1 below.

Table 1: Breakdown of number of participants and local authorities by local authority type[3]

Local authority type	Number of participants	Number of local authorities of this type in the sample
County council	30	18
District council	17	13
Metropolitan districts	13	11
Unitary authorities	26	24
London boroughs	7	6

Survey results suggested which topic areas were most common for case study follow-up, and provided an overview of trends in UK local authorities.

34 audio recorded semi-structured interviews of between 30 and 60 minutes in length were conducted with individuals who were selected based on their survey responses or their online profile found during the earlier desk research phase. These individuals were either working in UK local authorities, central government, or in enterprises providing algorithmic services to these authorities. Conversations were about the characteristics of projects and included details about concrete examples (semi-structured interview questions can be found in appendix A2). Of those contacted, 34 were interviewed by phone or over Skype. Participants by

[3] For more detail about the structure of local authorities in the UK, see (Ministry of Housing, Communities & Local Government 2019; Minister for Local Government, Housing and Planning 2019; Law Wales 2015; nidirect 2015)

position type included 20 in intelligence, data, and research, 4 in digital and IT, 6 in policy and strategy, and 4 in service or projects. Interview recordings were collectively reviewed in order to identify key themes and quotes, which were noted or transcribed. These three steps - desk research, survey results, and interview responses - were used to triangulate common themes and form more nuanced understandings of the data. Drafts of analysis were shared and discussed as a group to identify gaps, confirm relevance, and compare interpretations (Ospina, Esteve, and Lee 2018).

Two representative cases were selected based on survey results, interview responses, and publicly available documentary evidence to illustrate the breadth of phenomena that are occurring with the implementation of smart technology projects in local authorities. These two cases are autonomous agents and predictive analytics. Specifically, chatbots as autonomous agents and predictive analytics related to Houses in Multiple Occupation (HMO) and Children's Social Care risk. The case on predictive analytics looks at two instances of use because these use cases are the responsibility of different tiers of local authority, while chatbots are occurring across all types. 10 out of the 34 interview respondents from intelligence, data, and research, digital and IT, and service or projects positions were able to comment on these specific cases. Their commentary is supplemented by documentary evidence. While a multi-method qualitative approach offers opportunities to corroborate information and enhance credibility (Lincoln and Guba 1985; Seale 2002; Klein and Myers 1999), the limitations are that survey respondents were self-selected and that the individuals involved responded in a personal capacity. There could be selection bias such that those who responded might be from the most innovative or the only local authorities in the UK pursuing such changes.

Results

There is evidence of smart technology adoption in local authorities, particularly in the categories of autonomous agents and predictive analytics for decision assistance. 25 survey respondents (27%) mentioned that their local authority is experimenting with some kind of automatic text or content analysis. 15 survey respondents (17%) mentioned that their local authority is experimenting with some kind of predictive analytics. For overall data science approaches, welfare and social care was the biggest application domain reported in the survey (see table 2 below for a breakdown of the top application domains).

Table 2: Number and percentage of respondents reporting on data science use by application domain

Application Domain	**Number of Survey Respondents**	**Percentage of Survey Respondents**
Welfare and social care	43	46%
Transportation	38	41%
Healthcare	38	41%
Housing and planning	33	35%

Our case study examples are drawn from these service areas. The following two sub-sections look at cases of smart technology in local authorities and their impact on work and organization. The first section will specifically look at chatbots as a form of autonomous agent. The second will look at predictive analytics for HMOs and children's social care.

Autonomous Agents and Do-it-yourself Service Delivery. Chatbots are autonomous agents which typically interact through a website and make use largely of text-based communication to facilitate citizen-government interactions (Androutsopoulou et al. 2018). A Chief Technology and Information Officer who has worked closely with these technologies explained that the aim of chatbots is to take pressure off of face-to-face and telephone services by allowing people to conduct transactions online, and also potentially increase engagement and accessibility to services amongst demographics who might not use other digital channels (telephone interview, 3rd quarter, 2018). Examples of areas adopting chatbots include public transportation (Transport for London n.d.), planning permission applications for loft development (UKAuthority 2017), health diagnosis (Burnip 2017), and social housing issues reporting (Swainston 2017). But in this case, the focus is on (a) a resident facing chatbot for planning permissions, (b) the use of virtual assistants for adult social care, and (c) the use of middleware bots to assist public administrators. Results indicate that chatbots may replace some human work, but in practice significant human work will be needed behind the scenes in order to keep the technology useful and usable. Results also indicate that there may be new dynamics between street-level workers and their clients and a refocussing of street-level work, as chatbots take over some routine tasks.

(a) Resident facing chatbots for planning permissions: The London borough council of Enfield decided to implement IPsoft's Amelia AI assistant to deal with some planning permission applications (Everett 2017). The Assistant Director of IT at Enfield Council indicated that in the context of a growing population and continued austerity, some people still rely on in person or phone services because they struggle with digital, but "if they could do all that by talking to a device at home, their personal assistant at home with no keyboard or screen that connects to a digital ecosystem at the back-end and handles their request, it would do so much to remove the digital divide” (Everett 2017). A user friendly chatbot such as this could enable more isocratic service delivery, even among those who need additional assistance with digital.

The Assistant Director also indicated that rather than merely making workers redundant, “AI has the potential to take out repetitive admin processes that are too complex and nuanced for regular automation, freeing people up to do more sophisticated and gratifying work” (Everett 2017). Rather than replacing workers outright, automation with AI chatbots can deal with more complex service than procedural tools, such as web-forms and expert systems, and add another tool that could be used, in conjunction with existing staff, to provide more efficient responses to public service queries of a range of different levels of complexity. This suggests that work may not shift entirely to the system-level, but that responses will be made up of an imbrication of material and human agents.

The Assistant Director also explained how the process of development and implementation required the involvement of more than just technical staff: “You need three things, so people who understand the system, that is IPSoft and us in terms of the dictionary of terms. You need the business planning team, who map out processes, work out the kinds of questions people will ask in what order and what kind of words they use. And then you have to work with residents and people doing the testing to feed back into the project” (Everett 2017). Even a system-level bureaucracy requires input from multiple parties suggesting that a front-end that may look like an isocratic tool, may in fact require a significant socio-technical administrative infrastructure in order to function. The Assistant Director explained: “When you have more than 600 processes touching on multiple applications, it’s easy to underestimate the time it will take to get it right. But to get it to work, you have to be able to plug everything into the system and build on that. So AI may become the face of the Council, but behind the scenes is a whole body of things that have to work together" (Everett 2017). Work at the system level is complex in local authorities because of the diverse range of processes that smart technologies need to address and the stakeholder groups that need to be involved.

While the council did experience challenges with Amelia, it was able to find a solution (UKAuthority 2017). An individual working on the project explained that there was "very much a massive divide between what was advertised as being this blonde Scandinavian avatar that could speak to anybody and the reality of actually a text driven system that required a mass of expert coding at the back end to actually deliver it" (telephone interview, 3rd quarter, 2018). The vendor over-promised and under-delivered and the local council had to take over. In the end, the natural language processing only applied to the structure of the question, while behind the scenes the chatbot "was still fed by a logical workflow that was coded into the system" (telephone interview, 3rd quarter, 2018). Even at the system-level, there is a need for continuous human supervision and input into a smart tool. This may demonstrate that the current status of chatbots does not meet the threshold for pure learning technologies that can deliver equality of outcomes through appropriate responses to citizen requests. In this example, there is a blending of the ability to learn from natural language with a reliance on procedures embedded in code.

(b) The use of virtual assistants for adult social care: In adult social care, while chatbots and automated assistants, such as Amazon's Alexa, can provide some support and allow for some isocratic service delivery in clients' homes (Taylor 2018), there are still many instances where a human needs to be involved, for example where a chatbot cannot answer a question, where an automated assistant cannot carry out a physical or social task, or where the workers who fulfill these tasks may generate valuable data in their documentation. An IT consultant highlighted that "[m]any may still need hands-on support from human carers; these consumer devices clearly cannot replace that" (telephone interview, 3rd quarter, 2018). In these cases, a chatbot can supplement a service by enabling communication with caregivers and family, as well as providing control over connected smart home technologies, but it cannot replace the components of that service that require human presence. In addition to giving clients more independence, these technologies can free caregivers from routine tasks, allowing them to instead "do more of the human touch" (Taylor 2018).

While virtual assistants support some simple coordination tasks, for example by helping caregivers to leave messages for one another, there are also certain limitations. The smart home ecosystem is immature, the devices have difficulty understanding social care terminology and requests, and proprietary technologies do not allow councils to push messages out to their clients (telephone interview, 3rd quarter, 2018). The aspiration is that as these smart technologies develop, intelligence could be shared more effectively between caregivers, Internet of Things data could provide early warning signs that enable improved preventative services, and this human and material data could potentially enable a shift to equality of outcomes, but this is not currently the reality. An IT consultant predicted that "inside that data, that technology will give us the capability to say, 'I am going to vary this depending on need, I am going to intervene earlier on, before a crisis, and I am going to make this personalized rather than standardized'" (telephone interview, 3rd quarter, 2018). The interesting point it that in both the current and the aspirational settings, the integration of workers and smart technology is key.

(c) The use of middleware bots to assist public administrators: Some automated services are not directed towards citizens, but instead support public servants or professionals. A local government official highlighted how this type of chatbot could potentially be used to simplify and automate 'extract, transform and load' tasks in a variety of local government application areas, such as statutory reporting, acting as a kind of automated assistant for public servants, so that they could focus on more complex analytical tasks rather than more mundane routine administrative ones (telephone interview, 3rd quarter, 2018). An innovation team lead explained, "I've started to see the use of software bots, effectively one bit of software driving another, to try and optimize certain processes where, to be honest, human activity isn't, perhaps, the best use of resources" (telephone interview, 2nd quarter, 2018). Statutory returns create a significant draw on analyst time. If the processing of these returns could be automated, workers could have more time to work on more complex, innovative, and meaningful data analysis projects (telephone interview, 3rd quarter, 2018).

Such automated tools can also help to highlight where practices differ across teams, such as where a workaround may have been adopted in day-to-day work to deal with a problem. The innovation team lead explained that "a bigger opportunity or bigger challenge, certainly in the use of AI or robotics, is that the initial implementation will expose that there have not been standard practices at play" (telephone interview, 2nd quarter, 2018). The interaction between the technology and the people working with it can lead to greater efficiency, not only by automating routine tasks, but by highlighting biases and inefficiencies in human information practices that could be improved upon (Mittelstadt et al. 2016). By revealing instances where workarounds were developed to deal with various data deficiencies, local governments can begin to tackle the root causes.

Autonomous agents represent an ongoing evolution in the adoption of technologies within local governments; however, the aspirations for these technologies are not being met as expected. While there is some transition from street-level to system-level administration, the transformation is actually much more complex. The involvement of multiple stakeholders is needed for chatbots to work, virtual assistants complement street-level work, and middleware bots can also provide support to public servants and other professionals. Further, current autonomous agents combine new learning techniques that focus on outcomes atop a foundation of procedural equality. Emergent findings unrelated to the theoretical framework suggest that autonomous agents can free workers from routine tasks, provide feedback on the consistency of human practice, such as when things are not being done as expected, and reveal that vendors may over-promise and under-deliver, while local authorities may have internal capacity that can develop smart technologies that are sensitive to local context. In the next section, we will look at tools that are specifically designed to support workers by enhancing their decision-making.

Predictive Analytics and Learning in Complex Systems. In addition to providing professionals with assistance when carrying out routine tasks, smart technologies are also helping with decision making. A considerable proportion of local government work involves making complex decisions about when, where, and to whom to deliver services and interventions. These decisions are complex because they involve a wide variety of contexts and situations. A combination of rich data, large caseloads, and significant consequences taxes the information processing capacity of professional staff, which could lead them to make decisions based on heuristics or the limited information that they have been able to retrieve and analyze within the available timeframes (Cuccaro-Alamin et al. 2017; Sanders et al. 2017). A Senior Intelligence Lead explained, “front-line staff are having to make decisions about individuals, often in the backdrop of huge time pressures and system pressures” (telephone interview, 2nd quarter, 2018).

Smart technologies are beginning to help in the context of complexity and time pressure by means of predictive decision assistance (Rogge, Agasisti, and De Witte 2017). These technologies are computerized systems which seek to guide people making service intervention decisions. Often, they rely on machine learning techniques that make use of algorithms and past data to make predictions about future outcomes. Crucially, rather than being explicitly programmed, the algorithm learns from training data and responds to new data inputs. While these systems can take many forms, currently there is growth in the use of machine learning techniques to produce predictions or risk scores. Two examples of service areas where local authorities are applying predictive models are HMOs and decisions in children's social care. In these cases, the focus is on (d) worker feedback to designers, (e) collective knowledge, and (f) worker-smart-technology feedback loops. Results indicate that predictive analytics depend on the contextual knowledge of street-level workers during design. Results also indicate that predictive analytics enable collective intelligence and generate positive feedback loops around data collection, processing and presentation for use.

Targeted inspections of HMOs are one area of predictive analytics adoption. A variety of different branches of local government need to enforce local rules. Inspections are one potential way of enforcing these rules, and one potential use of predictive analytics is to improve the efficiency of these inspection operations. In the case of HMOs, there were examples showing how a reliance on data and smart technologies alone were not sufficient to deliver the results that local authorities were looking for. Software was developed in conjunction with Nesta which aimed to help find hidden HMOs, which are a major source of both unclaimed rates and potential health and safety risks (Dragicevic et al. 2018; Copeland 2017). The software provided a probability for each property and allowed inspectors to potentially guide decisions with respect to which properties to inspect. Nesta had hired a company that the local authorities would work with on a predictive model.

(d) Worker feedback to designers: A data scientist in a local authority explained their experience of the HMO inspection project saying that “we provided the data that they requested and they came up with a predictive model that was tested and it basically fell flat on its face” (telephone interview, 3rd quarter, 2018). The data scientist went on to explain that “there are a couple of big authorities that had a similar experience, they were not impressed with the results, and others had problems providing the data in the first place” (telephone interview, 3rd quarter, 2018). While overall prediction accuracy might improve, one data scientist explained that if the system makes predictions that are clearly absurd to seasoned workers, trust in the system as a whole could be undermined (telephone interview, 3rd quarter, 2018). The data scientist gave an example of how, in one of the pilots, inspection officers were unimpressed because the system was recommending things which were (to the officers) obviously not HMOs: “Through no fault of their own, the company who developed this particular model simply didn’t have the detailed knowledge of the borough … But this knowledge is crucial” (telephone interview, 3rd quarter, 2018). Vendors lacked the relevant contextual knowledge to design predictive tools in which workers had confidence.

In order to resolve this issue, the data scientist explained how “alongside the Nesta model, we’re developing our own using a couple of random forests and logistic regression, but the key thing was that we were very careful to work with service about what variables to include, which properties to include in the test and the training set, and so on and so forth and we actually came up with much better results” (telephone interview, 3rd quarter, 2018). Making a useful predictive model involves not only the data and smart technology, but an understanding of the context and what street-level factors should be included in the model, which only front-line staff possess. Ultimately, “this was seen as an aid, as a tool, rather than the answer. So, together with that and the local knowledge of the officers and phone calls from members of the public or councillors, the staff have a much better idea of which properties are worth inspecting" (telephone interview, 3rd quarter, 2018).

Predictions can also be made at the individual level and these results can be used for operational purposes, providing a tool which frontline managers and workers alike can use to aid decisions, either by helping to retrieve, analyze, and present more context and background information, or by generating an assessment to supplement existing judgment (Pratchett 1999). A specific example in children's social care is related to decision support technologies that could provide a useful supplement to workers who screen cases to identify if further action is needed, potentially enabling social workers to concentrate their effort on higher risk cases whilst sparing low risk families the intrusion of being screened. One example of such a trial is provided by the Behavioural Insights Team, which has developed a structural topic model that is applied to the case notes of social workers (Sanders et al. 2017). Similar to the HMO example, the Behavioural Insights Team sought feedback from social workers and team managers during the development of the tool to determine if the topics identified by the natural language processing algorithm made sense to the workers (Sanders et al. 2017). They are currently developing the model into a risk assessment tool that can inform decision making in the area. Beyond the feedback from front-line workers, a data scientist cautioned that the algorithms do also need to be retrained from time to time, highlighting the continued role of people in the supervision of smart technology (telephone interview, 3rd quarter, 2018). People continue to play an important role in the development and supervision of machine learning models.

(e) Collective knowledge: An advantage of smart technologies is that they can learn and adapt, improving pattern identification and prediction, in response to inputs documented by different workers. For example, in the case of social work in children's social care, a Head of Knowledge and Intelligence said that: "as an individual social worker ... you work with individual children and families and you document the work you have done ... but there might be something else, a more strategic view that the data can offer, which would support your decision-making" (telephone interview, 3rd quarter, 2018). A consultant working with a local authority shared this sentiment, saying that:

> it's not just about the service user, it's about their circle of support and it's about sharing intelligence. It's exactly the kind of thing that has caused the disasters in public services where the police knew something was wrong, the school knew something was wrong, the social worker knew something was wrong, and actually if they had all spoken to each other they would all know that something was catastrophically wrong, but none of them spoke to each other. So, you end up with a crisis situation. Those little snippets of intelligence and those little insights from other people that mean you can make a more complete judgement about what somebody's needs are and how they need to be supported are really important (telephone interview, 3rd quarter, 2018).

Decision support is needed where the ability of a human alone to parse the vast quantities of data are insufficient to meet the needs of the task at hand or are overwhelming any individual human being's ability to make a decision on that basis.

All interviewees who addressed the subject of predictive analytics were careful to highlight that these tools should supplement rather than replace existing skilled insight, and hence act as a kind of secondary check on decisions already made. A Head of Quantitative Research said "a machine alone cannot make a decision that has legal consequence for an individual ... even the legalities of it aside, I think it's absolutely correct that the human makes the final decision because ... there may be some pieces of a particular case that are very unique to that case which are not reflected by the model ... so we very much view this as a decision aid" (telephone interview, 3rd quarter, 2018). Other interviewees also supported the idea that predictive analytics should act only as a decision-support tool. A Head of Knowledge and Intelligence emphasized that predictive tools are "just trying to give practitioners another piece of information to help them make better decisions" (telephone interview, 3rd quarter, 2018). However, the practitioners' judgement should always overrule a computational decision. A scholar noted that in practice this seems to be how the technology is used: "the most common response about the impact of the decision support tool is that it made case workers stop and think in certain cases where previously they might have gone faster, rather than replacing their judgment" (Skype interview, 3rd quarter, 2018). This highlights the continued importance of front-line workers and their lived experience as a critical factor in service-level decision-making, and how this can be supported by decision assistance based on the collective intelligence of all workers recording data.

(f) Worker-smart technology feedback loops: The socio-technical relationship between those who collect the data, the machines that store and process it, the information system designers, and the people who retrieve and use the relevant information is critical to its functioning. The model is only as good as the input data that is collected and its ability to describe the context of the case. Feedback loops can help support data collection. An intelligence lead explained that "[t]he minute that our social workers and our occupational therapists saw the information displayed, it suddenly had a different purpose to it, and not just a purpose in terms of the use of data, but actually a purpose in their own head around why they write something or how they write something." (telephone interview, 2nd quarter, 2018). The interaction between technology and staff helps to support learning. The technology learns from the data and produces useful analysis and workers learn how to collect data to enable the technology. This creates a positive symbiotic relationship between the workers collecting data and the smart technologies that process it. Getting data collection right is important. As one head of knowledge and intelligence worried, “predictive analytics might just be a little blip, if we can’t sort out all the data underneath it” (telephone interview, 3rd quarter, 2018). Predictive analytics depends on an accurate and reliable foundation of integrated data that is accurately collected.

Making predictive decision-assistance technologies work involves technology and working with staff to identify which data sources need to be brought together. An innovation team lead explained, “it is now entirely technically possible for me to look at natural language processing, so free text, across a social care record, a health record, a police record, DWP [UK Department for Work and Pensions], so from all those different data sources, effectively create a data universe around a particular individual” (telephone interview, 2nd quarter, 2018). The innovation team lead went on to say, “I think there’s a really interesting piece of research yet to be fully undertaken with practitioners about how you create that risk universe and what are all the data points you would need and what would it actually do? Would it instruct an intervention, or would it just flag up the possibility of something, or the probability of something being an issue?” (telephone interview, 2nd quarter, 2018). Making smart predictive tools work, involves not just the technology, but the interaction between the technology, those who feed the information into the technology, and those who use the outputs of that technology to help inform their work. The innovation team lead concluded that “I think the human side of this is going to be more critical than the technology side” (telephone interview, 2nd quarter, 2018). A research agenda for smart technologies needs to look at the technological and the human aspects of the information infrastructure that supports smart technology in local authorities.

Predictive analytics also represent an ongoing evolution in the adoption of technologies within local governments, but, again, the aspirations for these technologies are not being met as expected. While there is some transition from street-level to system-level administration, the transformation is actually much more complex. The contextual knowledge of street-level workers is invaluable in system design, predictive analytics complement street-level work by bringing to bear the collective intelligence of all workers, and positive feedback loops can be developed where public servants and other professionals are shown the value of quality data collection and data integration. More than in the case of autonomous agents, predictive analytics promise greater equality of outcomes by learning from data and offering guidance specific to those inputs, rather than focussing on a procedural logic. Emergent findings unrelated to the theoretical framework suggest that predictive analytics reveals how the smart technology depends on more than the simple dichotomy between street- and system-level administrators, and, again, that local authorities may need to take over from vendors in order to develop smart technologies that are sensitive to local context.

Rather than replacing human intervention outright, for example with the introduction of autonomous agents, smart technologies could be used to enhance the decision-making capabilities of professional service providers by bringing together the knowledge of multiple professionals, lightening the burden of some information retrieval and analysis tasks, and freeing up attention needed for direct service. Smart technologies also crucially depend on the situated knowledge, data collection, and interpretations of front-line workers to be most effective. In the next section, we will look at what these findings mean for current theory and propose a new conception of public administration in the era of smart technology that can act as a framework for future research.

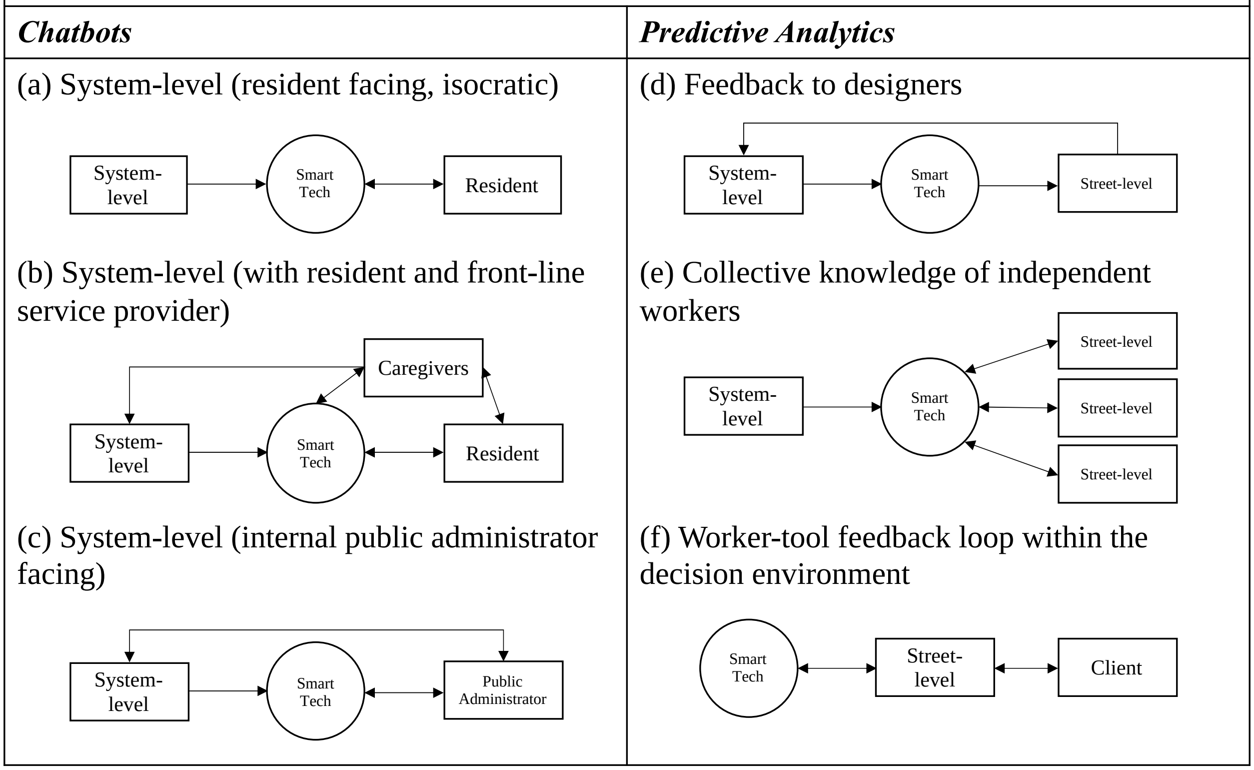
Discussion:

Findings suggest that, while at a nascent stage, local authorities are beginning to adopt smart technologies, though not necessarily in the ways that were envisioned. The findings also suggest that while smart technologies may automate certain types of local government work, there are numerous cases where a new relationship is being created between public administrators and artificial intelligence technologies. This relationship is described here as algorithmic bureaucracy, recognizing the interaction between offices held by public administrators, both street-level and system level, and computational algorithms that are increasingly becoming an everyday part of the working environment. This paper establishes a framework setting out six of the principal ways in which these technologies are not simply replacing people but are transforming the socio-technical relationship between workers and their tools, as well as the way that work is organized in the public sector.

In the context of this research, it is important to consider two key insights, "that bureaucracies are sociotechnical systems; and that the organization of information-processing is key to bureaucratization pushing ahead (for better or worse) the modernization and rationalization of human conduct." (Dunleavy et al. 2006, 40). Public administration in local authorities is experiencing automation and the adoption of predictive tools. Automation is not just about isocratic service delivery (Dunleavy and Margetts 2015), but requires the involvement of multiple stakeholders and supports the work of public servants, both street- and system-level. Predictive tools depend on the contextual knowledge of street-level workers, support decision making by processing and bringing to bear more information than any individual could have, and enable positive feedback loops related to information collection, processing, and presentation, which allows the public sector to handle greater complexity, including in the decision space of front-line workers. Automation and predictive analytics were also found to free workers from routine tasks, provide feedback on the consistency of human practice, demonstrate the need for the involvement of multiple stakeholders, and reveal the latent capacity to deliver smart technologies that exists within local authorities. If algorithms represent a computational procedure or set of rules used in problem-solving, then we may expect them to automate many public service functions. While this may be the case for chatbots that interact with residents, we have also seen from chatbots for public administrators and predictive tools for decision assistance that there is still an important role for people and organization.

The concept of algorithmic bureaucracy provides a framework to understand how computational algorithms are affecting all offices in the public sector, from the street- to the system-level. This framework accounts for: automated processes made possible by algorithms; the system of roles, hierarchy, and files that constitute traditional bureaucracy; and how these two things interact. As Simon stressed decades ago, “we must understand the decision-making tools at our disposal, both human and mechanical - men and computers" (1973, 272). The following figure illustrates the six socio-technical interactions that are suggested by the findings (from (a) to (f)) and the manifold ways in which smart technologies and public administrators can become imbricated into an algorithmic bureaucracy.

Figure 1: Socio-technical relationships involving smart technologies



Under chatbots, there is (a) isocratic system-level relationships, which represents the standard perspective on the adoption of smart technologies in public administration, though even in this case, the picture was more complex, given the number of stakeholder groups involved in design. There are also (b) relationships between front-line caregivers and smart technology that frees those individuals from routine tasks and allows them to more effectively communicate, so that they can focus on the human elements of care work, and (c) internal relationships between system-level designers, middleware bots, and administrators to realign tasks and uncover unproductive practices. Under predictive analytics, there is (d) the critical role of feedback from workers to designers when evaluating the utility of such tools. There are also (e) feedback from many individual workers to the tool, creating collective knowledge for all workers, and (f) feedback loops between workers who collect and use information, and the smart technologies that process and present the client-related data for use within the decision environment.

The theoretical implications of these findings are mixed. They suggest an ongoing evolution in the extent to which digital transformation is occurring in local government. While in some cases they do support the idea that there is a transition from street- to system-level administration, there are many other cases where this picture is not as clear, as multiple stakeholders continue to have some involvement in the design, implementation, and application of smart technologies. Finally, they also show that some applications of smart technologies are made with the intent to focus on equality of outcomes using adaptive learning technologies; however, there are other cases in which smart technologies are built on a foundation of procedural equality. Emergent findings suggest that the introduction of autonomous agents and predictive analytics will free workers from routine tasks, provide feedback on the consistency of human practice, demonstrate the need for the involvement of multiple stakeholders, and reveal the latent internal capacity that exists within local authorities to deliver contextually sound smart technology. Algorithmic bureaucracy suggests that there are multiple ways in which smart technologies and public administrators become imbricated in the delivery of services. This conceptual framework illustrates some of these interactions and provides an example of how to clarify and study the many and diverse implications of smart technologies in public administration settings.

Conclusion:

This study, which included three techniques: a survey, desk research, and subsequent in-depth interviews, indicates that smart technologies are increasingly being adopted and used to automate certain tasks and enhance human work practices and decision making in local authorities in the UK. Smart technologies appear to involve more stakeholders than initially expected, mediate work relationships between professionals and their clients, offer public-administrator-facing in addition to client-facing support, necessitate feedback between street-level and system-level public administrators, enable collective intelligence, and create positive feedback loops with street-level workers. Overall, the more widespread introduction of computational and algorithmic tools across service areas in local authorities is evidence for a change in the nature of public administration towards a form of algorithmic bureaucracy. However, this change is not a wholesale replacement of public administrators and traditional mechanisms of organization in public administration, but a transformation of the socio-technical relationship between workers and their tools, as well as the way that work is organized in the public sector. Thus, an algorithmic bureaucracy is able to handle greater complexity in the decision environment while also enhancing individual and administrator competence when trying to solve problems.

The findings suggest a new way of conceptualizing public administration in the context of smart technologies. The concept of algorithmic bureaucracy calls attention to the need to study the imbrication of computational algorithms with traditional public sector organizing. This paper has presented six different forms of interaction across the two cases of autonomous agents and predictive analytics. These forms of interaction set out the constructs that need to be studied when trying to understand the implications of the introduction of smart technologies in public sector settings. Future research could use these constructs to study similar phenomena in different jurisdictions or service sectors or could expand on these six constructs to further elaborate the concept of algorithmic bureaucracy. Hopefully, this paper provides a robust framework for the continued study of smart technology in socio-technical systems.

References:

Agarwal, P. K. 2018. “Public Administration Challenges in the World of AI and Bots.” *Public Administration Review* 78 (6): 917–21. https://doi.org/10.1111/puar.12979.

Allard, Scott W., Emily R. Wiegand, Colleen Schlecht, A. Rupa Datta, Robert M. Goerge, and Elizabeth Weigensberg. 2018. “State Agencies’ Use of Administrative Data for Improved Practice: Needs, Challenges, and Opportunities.” *Public Administration Review* 78 (2): 240–50. https://doi.org/10.1111/puar.12883.

Androutsopoulou, Aggeliki, Nikos Karacapilidis, Euripidis Loukis, and Yannis Charalabidis. 2018. “Transforming the Communication between Citizens and Government through AI-Guided Chatbots.” *Government Information Quarterly*, October. https://doi.org/10.1016/j.giq.2018.10.001.

Bovens, Mark, and Stavros Zouridis. 2002. “From Street-Level to System-Level Bureaucracies: How Information and Communication Technology Is Transforming Administrative Discretion and Constitutional Control.” *Public Administration Review* 62 (2): 174–84. https://doi.org/10.1111/0033-3352.00168.

Bright, Jonathan, Bharath Ganesh, Cathrine Seidelin, and Thomas Vogl. 2019. “Data Science for Local Government: Challenges and Opportunities.” Oxford, UK: Oxford Internet Institute, University of Oxford. https://smartcities.oii.ox.ac.uk/data-science-for-local-government-report/.

Burnip, Laura. 2017. “When Will the NHS Medical Advice Smartphone App Launch, What Services Will It Offer and What Other NHS Apps Are There?” The Sun. September 11, 2017. https://www.thesun.co.uk/news/2548383/nhs-medical-advice-smartphone-app-111-helpline-latest/.

Copeland, Eddie. 2017. “London Office of Data Analytics Pilot - Now for the Hard Part.” Nesta. January 22, 2017. https://www.nesta.org.uk/blog/london-office-of-data-analytics-pilot-now-for-the-hard-part/.

Cordella, Antonio, and Niccolò Tempini. 2015. “E-Government and Organizational Change: Reappraising the Role of ICT and Bureaucracy in Public Service Delivery.” *Government Information Quarterly* 32 (3): 279–86. https://doi.org/10.1016/j.giq.2015.03.005.

Crozier, Michel. 1964. *The Bureaucratic Phenomenon*. Chicago: University of Chicago Press.

Cuccaro-Alamin, Stephanie, Regan Foust, Rhema Vaithianathan, and Emily Putnam-Hornstein. 2017. “Risk Assessment and Decision Making in Child Protective Services: Predictive Risk Modeling in Context.” *Children and Youth Services Review* 79 (August): 291–98. https://doi.org/10.1016/j.childyouth.2017.06.027.

Danziger, James N., and Kenneth L. Kraemer. 1985. “Computerized Data-Based Systems and Productivity among Professional Workers: The Case of Detectives.” *Public Administration Review* 45 (1): 196–209. https://doi.org/10.2307/3110149.

Dencik, Lina, Arne Hintz, Joanna Redden, and Harry Warne. 2018. “Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services Project Report.” Cardiff, UK: Data Justice Lab, Cardiff University. https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf.

Dragicevic, Nevena, Eddie Copeland, Andrew Collinge, Paul Hodgson, Wil Tonkiss, and Alan Lewis. 2018. “Piloting the London Office of Data Analytics.” London, UK: Greater London Authority in partnership with Nesta. https://media.nesta.org.uk/documents/loda_pilot_report.pdf.

Dunleavy, Patrick, and Helen Margetts. 2015. “Design Principles for Essentially Digital Governance.” In *111th Annual Meeting of the American Political Science Association, 3 - 6 September 2015, American Political Science Association*. San Francisco, USA: (Unpublished).

Dunleavy, Patrick, Helen Margetts, Simon Bastow, and Jane Tinkler. 2006. *Digital Era Governance: IT Corporations, the State, and e-Government*. Oxford: Oxford University Press.

Elgin, Duane, and Robert A. Bushnell. 1977. “The Limits to Complexity: Are Bureaucracies Becoming Unmanageable?” *The Futurist*, December 1977.

Everett, Cath. 2017. “Could AI Chatbots Be the New Face of Local Gov? Enfield Council Thinks So.” Diginomica. February 23, 2017. https://diginomica.com/ai-chat-bots-new-face-local-gov-enfield-council-thinks.

Eynon, Rebecca, and William H. Dutton. 2007. “Barriers to Networked Governments: Evidence from Europe.” *Prometheus* 25 (3): 225–42. https://doi.org/10.1080/08109020701531361.

Fischer, Robert L., Francisca García-Cobián Richter, Elizabeth Anthony, Nina Lalich, and Claudia Coulton. 2019. “Leveraging Administrative Data to Better Serve Children and Families.” *Public Administration Review* 0 (0). https://doi.org/10.1111/puar.13047.

Gil-Garcia, J. Ramon, and Ignacio J. Martinez-Moyano. 2007. “Understanding the Evolution of E-Government: The Influence of Systems of Rules on Public Sector Dynamics.” *Government Information Quarterly* 24 (2): 266–90. https://doi.org/10.1016/j.giq.2006.04.005.

Hadden, Susan G. 1986. “Intelligent Advisory Systems for Managing and Disseminating Information.” *Public Administration Review* 46: 572–78. https://doi.org/10.2307/975579.

Hood, Christopher. 1995. “The ‘New Public Management’ in the 1980s: Variations on a Theme.” *Accounting, Organizations and Society* 20 (2): 93.

Hurley, Michael W., and William A. Wallace. 1986. “Expert Systems as Decision Aids for Public Managers: An Assessment of the Technology and Prototyping as a Design Strategy.” *Public Administration Review* 46: 563–71. https://doi.org/10.2307/975578.

Klein, Heinz K., and Michael D. Myers. 1999. “A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems.” *MIS Quarterly* 23 (1): 67–93. https://doi.org/10.2307/249410.

Lavertu, Stéphane. 2016. “We All Need Help: ‘Big Data’ and the Mismeasure of Public Administration.” *Public Administration Review* 76 (6): 864–72. https://doi.org/10.1111/puar.12436.

Law Wales. 2015. “Local Government Bodies in Wales.” Welsh Government: Law Wales. March 4, 2015. https://law.gov.wales/constitution-government/government-in-wales/local-gov/local-gov-bodies-wales/?lang=en#/constitution-government/government-in-wales/local-gov/local-gov-bodies-wales/?tab=overview&lang=en.

Leonardi, Paul M. 2012. “Materiality, Sociomateriality, and Socio-Technical Systems: What Do These Terms Mean? How Are They Different? Do We Need Them?” In *Materiality and Organizing*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199664054.003.0002.

———. 2013. “Theoretical Foundations for the Study of Sociomateriality.” *Information and Organization* 23 (2): 59–76. https://doi.org/10.1016/j.infoandorg.2013.02.002.

Lessig, Lawrence. 2006. *Code*. 2nd ed. Code Version 2.0. New York: BasicBooks.

Lincoln, Yvonna S, and Egon G Guba. 1985. "Establishing Trustworthiness." In *Naturalistic Inquiry*. Newbury Park: Sage.

Lipsky, Michael. 2010. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. Updated. New York: Russell Sage Foundation.

Lynn, Laurence E. 2001. "The Myth of the Bureaucratic Paradigm: What Traditional Public Administration Really Stood For." *Public Administration Review* 61 (2): 144–60.

Malomo, Fola, and Vania Sena. 2017. "Data Intelligence for Local Government? Assessing the Benefits and Barriers to Use of Big Data in the Public Sector." *Policy & Internet* 9 (1): 7–27. https://doi.org/10.1002/poi3.141.

Margetts, Helen. 1999. *Information Technology in Government: Britain and America*. Routledge Research in Information Technology and Society. London: Routledge.

Margetts, Helen, and Cosmina Dorobantu. 2019. "Rethink Government with AI." *Nature* 568 (7751): 163. https://doi.org/10.1038/d41586-019-01099-5.

Mergel, Ines, R. Karl Rethemeyer, and Kimberley Isett. 2016. "Big Data in Public Affairs." *Public Administration Review* 76 (6): 928–37. https://doi.org/10.1111/puar.12625.

Minister for Local Government, Housing and Planning. 2019. "Local Government - Gov.Scot." Gov.Scot. 2019. https://www.gov.scot/policies/local-government/.

Ministry of Housing, Communities & Local Government. 2019. "Local Government Structure and Elections." GOV.UK. April 3, 2019. https://www.gov.uk/guidance/local-government-structure-and-elections.

Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. 2016. "The Ethics of Algorithms: Mapping the Debate." *Big Data & Society* 3 (2): 2053951716679679. https://doi.org/10.1177/2053951716679679.

Moon, M. Jae. 2002. “The Evolution of E-Government among Municipalities: Rhetoric or Reality?” *Public Administration Review* 62 (4): 424–33.

Naughton, John. 2001. *A Brief History of the Future: From Radio Days to Internet Years in a Lifetime*. Woodstock, NY: Overlook Press.

nidirect. 2015. “Local Councils.” Nidirect. October 27, 2015. https://www.nidirect.gov.uk/articles/local-councils.

Orlikowski, Wanda J., and Susan V. Scott. 2008. “Sociomateriality: Challenging the Separation of Technology, Work and Organization.” *The Academy of Management Annals* 2 (1): 433–74. https://doi.org/10.1080/19416520802211644.

Ospina, Sonia M., Marc Esteve, and Seulki Lee. 2018. “Assessing Qualitative Studies in Public Administration Research.” *Public Administration Review* 78 (4): 593–605. https://doi.org/10.1111/puar.12837.

Oxford English Dictionary. 2019. “‘smart, Adj.’.” 2019. https://www.oed.com/view/Entry/182448?rskey=KAWHOi&result=3&isAdvanced=false .

Pasquale, Frank. 2015. *The Black Box Society*. Harvard University Press.

Pinch, Trevor J., and Wiebe E. Bijker. 1987. “The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other.” In *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, edited by Wiebe E. Bijker, Thomas Parke Hughes, and Trevor J. Pinch. Cambridge, Mass.; London: MIT Press.

Pollitt, Christopher. 2011. “Technological Change: A Central yet Neglected Feature of Public Administration.” *NISPAcee Journal of Public Administration and Policy* 3 (2): 31. https://doi.org/10.2478/v10110-010-0003-z.

Pollitt, Christopher, and Geert Bouckaert. 2011. *Public Management Reform: A Comparative Analysis - New Public Management, Governance, and the Neo-Weberian State*. Vol. 3rd ed. Oxford: OUP Oxford. http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=405923&site=ehost-live.

Pratchett, Lawrence. 1999. “New Technologies and the Modernization of Local Government: An Analysis of Biases and Constraints.” *Public Administration* 77 (4): 731–51. https://doi.org/10.1111/1467-9299.00177.

Rodríguez Bolívar, Manuel Pedro, and José Miguel López-Quiles. 2018. “Smart Technologies for Smart Governments: A Review of Technological Tools in Smart Cities.” In *Smart Technologies for Smart Governments: Transparency, Efficiency and Organizational Issues*. Public Administration and Information Technology 24. Springer International Publishing. DOI 10.1007/978-3-319-58577-2_1.

Rogge, Nicky, Tommaso Agasisti, and Kristof De Witte. 2017. “Big Data and the Measurement of Public Organizations’ Performance and Efficiency: The State-of-the-Art.” *Public Policy and Administration* 32 (4): 263–81. https://doi.org/10.1177/0952076716687355.

Sanders, Michael, James Lawrence, Daniel Gibbons, and Paul Calcraft. 2017. “Using Data Science in Policy.” London: Behavioural Insights Team. https://www.bi.team/wp-content/uploads/2017/12/BIT_DATA-SCIENCE_WEB-READY.pdf.

Seale, Clive. 2002. "Quality Issues in Qualitative Inquiry." *Qualitative Social Work* 1 (1): 97–110. https://doi.org/10.1177/147332500200100107.

Simon, Herbert A. 1973. "Applying Information Technology to Organization Design." *Public Administration Review* 33 (3): 268–78. https://doi.org/10.2307/974804.

Swainston, David. 2017. "What We Learnt from Prototyping." *Housing – HackIT* (blog). March 29, 2017. https://blogs.hackney.gov.uk/hackit/tag/housing.

Symons, Tom. 2016. "Wise Council: Insights from the Cutting Edge of Data-Driven Local Government." London: Nesta.

Taylor, Steven. 2018. *BBC Daily Politics Show 23rd April Alexa Full Version*. https://www.youtube.com/watch?v=YL-nQGPxc68&feature=youtu.be.

Transport for London. n.d. "Facebook TravelBot." Transport for London. Accessed January 27, 2019. https://www.tfl.gov.uk/travel-information/social-media-and-email-updates/facebook-travelbot.

UKAuthority. 2017. "Enfield Joins Microsoft in CitizenBot Project." UKAuthority. June 21, 2017. https://www.ukauthority.com/articles/enfield-joins-microsoft-in-citizenbot-project/.

UN, United Nations. 2018. "E-Government Knowledge DataBase." United Nations, Department for Economic and Social Affairs, Division for Public Administration and Development Management. 2018. https://publicadministration.un.org/egovkb/en-us/#.WowPhoJpGRs.

Wachter, Sandra, Brent Daniel Mittelstadt, and Chris Russell. 2018. "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GPDR." *Harvard Journal of Law & Technology (Harvard JOLT)* 31 (2): 841–88.

Weber, Max. 1968. *Economy and Society: An Outline of Interpretive Sociology*. Edited by Guenther Roth and Claus Wittich. New York: Bedminster Press.

Wilkins, Leslie T. 1968. “Computer Impact on Public Decision Making.” *Public Administration Review* 28 (6): 503–8. https://doi.org/10.2307/973327.

Zacka, Bernardo. 2017. *When the State Meets the Street: Public Service and Moral Agency*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.