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

Cathrine Seidelin	Yvonne Dittrich	Erik Grönvall
IT University of Copenhagen	IT University of Copenhagen	IT University of Copenhagen
Copenhagen, Denmark	Copenhagen, Denmark	Copenhagen, Denmark
Cfre@itu.dk	*Ydi@itu.dk*	*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, administrative 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.
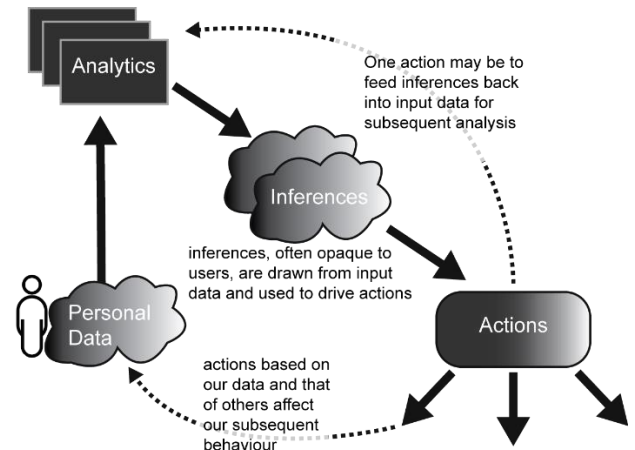

Figure 1. *Human Data Interaction (redrawn from Mortier et al., 2014)*

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.
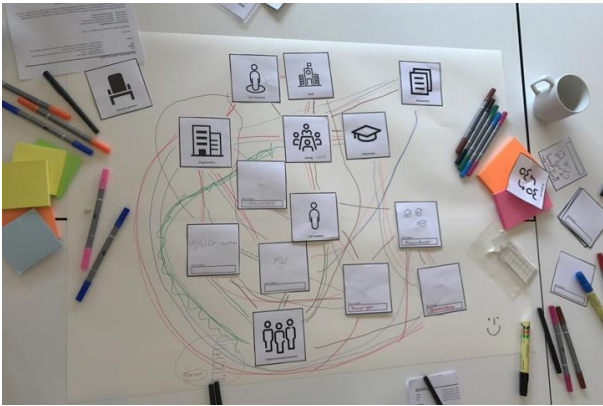

Figure 3. *Final visualisation of how the workshop participants percieve how they relate to each of the data entities in the current LEC database.*

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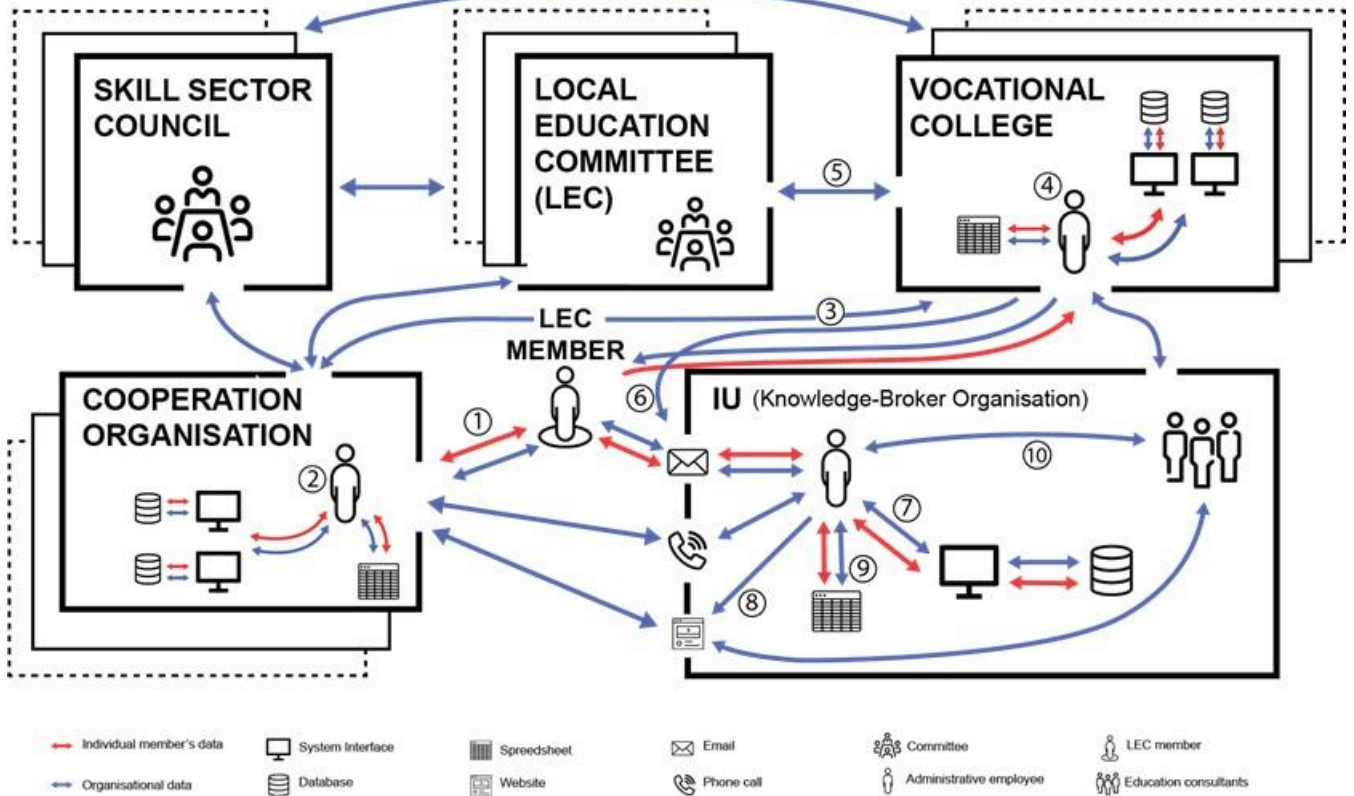
***Figure 4.** Collaboration involving data about LECs and LEC members (Please refer to the text regarding the numbers.)*

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.