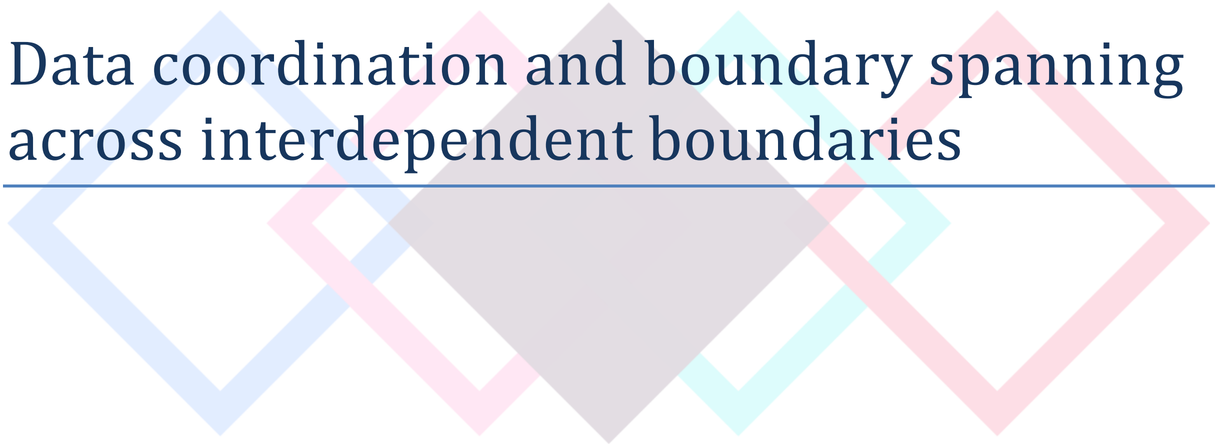
Connected Health Cities – End of Project Report

Workforce Development:

Data coordination and boundary spanning across interdependent boundaries



Contents:

30th March 2020
Victoria Neumann

Abstract:

Is health data the new oil?

The mining and analysis of health data is the subject of large investments in industry and research. The arguments behind this are focussed on improving health care systems. These arguments are normally three-fold:

1) Access to health data will enable and foster innovation in health care.
2) Collecting, archiving and studying data will enable more effective health care. Claims of effectiveness are typically to save costs and public spending for health care, while at the same time improving the treatment regimes and their outcomes.
3) Granting access to health data is envisioned to not only save costs but create monetary returns for public health and, in some cases, even individuals.

The centre-point for these attempts is to create legal ways for data access. This research seeks to understand the political economy of health data in contemporary Britain, within the context of the Connected Health Cities (CHC) programme, which built infrastructure (Data Arks) to house and analyse data.

It specifically concerns emergent regimes of data access. These regimes are operating in political-economic environments that are already characterised by high capitalisation in health care and the promise of new forms of knowledge production through the scrutiny of routinely collected health data.

This research is concerned with consent as data access management in three dimensions:

1) How do CHC Data Arks incorporate consent on a technical level?
2) How are organisational structures contributing to data access?
3) How are existing legal and ethical frameworks impacting design decisions for organisational and technical infrastructures?

30th March 2020
Victoria Neumann

Introduction:

The Connected Health Cities project was exemplary for the developments in the health care sectors. It was an investment into creating data infrastructures that ideally can be utilised by a multiplicity of stakeholders, such as academic researchers, industry and governments. It was part of the Health North pilot project, funded by the UK's Department of Health and delivered by the Northern Health Science Alliance.

The proclaimed aims of the Health North project were to: "(1) establish a social contract with citizens that gives license to use health data for public good; (2) produce timely and actionable information from patient and population data; (3) understand pathways of care across different provider organisations and to target resources to needs in much more agile and specific ways than at present; and (4) to accelerate business growth in the digital health sector in North England."[1]

CHC, as part of Health North, was a pilot initiative in four city regions. The goal was to create a *learning health care system*: "Connected Health Cities accelerates the pace of progress in health and social care services by safely using data and technology to identify health improvements and quickly implementing them into standard practice."[2]

To achieve this goal, the CHC team engaged in various activities such as public engagement, collaboration with existing organisations, infrastructures and systems in industry, academia and health and social care, as well as the building of knowledge infrastructures for specific prioritised diseases.

In the following report, I argue for an examination of data access and consent. The CHC proposed learning health care system is exemplary for the characteristics of its infrastructure as examined by infrastructure studies by various disciplines including in Science and Technology Studies (STS) (Bates et al., 2016; Beer & Burrows, 2013; Bowker et al., 2009; Jackson et al., 2007; Nafus, 2014; Ribes & Polk, 2015), Computer Supported Cooperative Work (CSCW) (Bossen & Markussen, 2010; Dewan, 2001; Grisot & Vassilakopoulou, 2015; Ribes, 2014a, 2014b; Ulriksen et al., 2017), or Information Systems (IS) (Hanseth & Lyytinen, 2010; Pipek & Wulf, 2009; Plantin et al., 2018). Consent in and for information systems has been a

---

[1] See also: https://mrc.ukri.org/documents/pdf/mapping-the-landscape-of-uk-health-data-research-and-innovation-report/

[2] See also: https://www.connectedhealthcities.org/about-us/what-we-do/

long-standing issue in health care and due to new legal requirements such as the General Data Protection Regulation (GDPR), obtaining consent has become more complex. Researching it from the view of infrastructures will allow for alternative insights on how consent is implemented in data access management in practice. By paying attention to how contingencies are resolved, I will draw design recommendations for future systems and data infrastructure.

30th March 2020
Victoria Neumann

Methods:

The notion of information infrastructures has been used to investigate the large-scale information systems and how categories and standards are produced and deployed (Bowker & Star, 2000). Traditionally, studies of information infrastructures were concerned with standards. Data sharing and health care are recurring themes in infrastructure studies as ideas of Electronic Health Records (EHR), or research environments for secondary data use proliferated as digital technologies became cheaper and more broadly available (Grisot & Vassilakopoulou, 2013).

The first characteristics of information infrastructures were developed by scholars around the work of Star (Bowker & Star, 1999; Star & Ruhleder, 1996). They formulate nine distinct features: (1) embeddedness, (2) transparency, (3) reach or scope, (4) learned as part of membership, (5) links with conventions of practice, (6) embodiment of standards, (7) built on an installed base, (8) becomes visible upon breakdown, and (9) is fixed in modular increments, not all at once or globally.

Karasti and Bloomberg (2018) further look into how infrastructures can be studied. From their own empirical investigations and an extensive literature review, they derive five dimensions of information infrastructures: (1) relational, (2) connected, (3) invisible (4) emerging and accreting, and (5) intervention and intentionality. From these dimensions, they give recommendations on how to study information infrastructures ethnographically.

In fact, most infrastructure studies use an ethnographic approach. Further methodological developments, such as infrastructural inversion or adapted versions of anthropological approaches, such as ‘follow the connections’ (from follow the thing) are widespread in order to grasp the complexity and render it researchable.

In the following report, I want to emphasise how infrastructure studies are a fruitful approach to study issues of consent, data sharing and secondary data use. A common denominator of these studies is the acknowledgement that infrastructures are hard to single out as they are integrated with other (often older) systems and non-technical entities such as funding regimes, as well as human actions and practices. The entanglement of various elements (technical and non-technical), sites, and their general messiness make infrastructures challenging to study. The practical implications for my research are, first and foremost, the selection process of what sites and what participants to study in the CHC. One reason why I choose infrastructure studies as a lens for my research is, particularly, the connection between consent and a multiplicity of technical systems within the project. In the following, I present three field sites:

1. Technical Infrastructure:

The CHC is developing a number of technical support systems for its idea of a learning health care system. One of them is the so-called *Data Arks*, of which each region is supposed to build one. These are conceptualised, not simply being databases or archives, but also function as a 'Trusted Research Environment' (TRE), which the CHC defines as a “[...] place for researchers to acquire, process, share and analyse health and care data with all relevant governance and safeguards.”

2. Information Governance:

In the CHC, much of the personnel labour circulates around information governance. The work is characterised by two major work tasks:

I. Making sure everything is legal and compliant with current regulations.
II. The second work task is data access management. While data access is informed by the first work task, it contains more layers as it also tries to connect to all other endeavours within the initiative, including public and industry engagement.

I will conduct research using qualitative methods to understand researchers within the CHC and their experiences of getting data access. Researchers were envisioned to greatly benefit from the setting up TREs and Data Arks with the promise of relatively easy access to data specific to their research area. I wanted to know how these systems and organisational structures work for them in practice and report back their views, needs and values for health data infrastructures.

Methods of Data Collection and Analysis

Typically, the data collection approach in qualitative research is to use multiple information sources, in this case, mainly interviews and document analysis. Each of the selected methods have their own methodological considerations:

1. Document Analysis:

For the document analysis, I use the approach as suggested by Proir (2004) who highlighted the functions of documents in contemporary life and how they stand in relation to human interaction. This method is mainly concerned with identifying and interrelating three characteristics of documents, namely (1) production, (2) use/function, and (3) content analysis.

2. Interviews:

Interviews follow an 'active interview' approach (Charmaz, 2006): here I understand the interviewees are not mere informants but partners in a reflexive process of shared meaning-making. Interview guides were developed and cover topic areas outlined in the sub-questions of the project. For each interviewee, an adapted interview guideline was developed.

The analysis is largely based on recent developments in situational analysis in order to make sense of the complexity and the mess of research.

Situational analysis was developed by Adele Clarke as a supplement to grounded theory (Clarke, 2007; Clarke et al., 2015). Through the creation of situational maps, it engages in the complexity of theoretical and methodological concerns, including materiality, discourses, narratives, and so forth.

To facilitate this approach, I used computer-supported content analysis tools to support my analysis, namely MaxQDA. Such tools offer valuable help when dealing with larger data sets as they help to order, categorize and relate codes, nodes and memos.

Results:

The results formulated in the following sections are preliminary as the research and analytic process is still ongoing.

Data User Attitudes and Perceptions

Researchers reported challenges in getting data and/or data access for their research projects. In health care research, ethical requirements and information governance are not only stringent, but often come from multiple organisations. Each of these organisations has its own ethic process that researchers or other data users have to undergo in order to do their research and/or gain access to data.

The interviewed researchers viewed these processes as ambiguous. On the one hand, they reported that the focus on detail and scrutiny in the application forms and processes helped them to think through what their research is about and what they want to achieve from an early stage. Similarly, participants reported that the TRE was helping them as they lifted the administrative burden of having to manage data themselves.

On the other hand, ethics and information governance processes, while seen as necessary, were described as an obstacle to research, mainly for the administrative overhead they were creating. Research is often limited to a specific time, in particular, funding structures have created specific timelines for research that resemble neoliberal project cycles.

In particular, researchers found it frustrating that they were already part of a project, which promised to enable almost automatic access to relevant data. However, even with close communication with data providers this promise was often not fulfilled.

Similarly, technical issues with the TRE and other data arks caused constraints and delays for researchers. Time pressures and problems with data access shaped decision-making about how to carry out research. What data to pursue was also noticed by IG personnel as a major obstacle.

While the role of data users was mainly formulated in the challenges they faced, there were advantages to working with data repositories. Researchers recognised the usefulness of a TRE or Data Ark in terms of distributing the burden of managing data security and comply with data protection. This had several advantages, from freeing them of the workload of collecting their data or individual consent themselves to the additional measurements to keep data secure on the researcher or university's devices. However, mitigating the risks for data being unintentionally exposed, became one of the key values in the interviews.

Challenges for IG

IG personnel found it hard to act as an interface between multiple stakeholders and balancing act between competing interests.

Firstly, in crude oversimplification, IG personnel viewed the communication practices of academic researchers as problematic. Researchers seemed to like to keep certain details of their data and data collection secret, a participant recalled. This was excused by some participants with concerns over intellectual property rights and increased pressure to publish results fast in order to be the first (also commonly called 'publish or perish').

On the one hand, the value of transparency was being upheld to ensure traceability as to which data was used by whom. This was not only part of the data protection compliance, but rather seen as preparing and ensuring data quality and completeness. On the other hand, this openness was rejected by some researchers due to fears that some might replicate their research or even expose vulnerabilities in methods or increase the risk of re-identification of individuals.

Moreover, depending on the field and discipline, there was a spectrum of preparedness and openness for exploration. While some projects were writing detailed data catalogues and refined them together with ethics and IG workers, other researchers wanted to enter the field "[…] without knowing what to measure."

Formulising procedures that allow for both extremes of the spectrum was generally being pin-pointed as the toughest challenge. However, there was a dividedness around the topics of standards amongst interviewees. While some were calling for more standardised procedures, especially ones that focus specifically on health data, others were more sceptical that that's the solution to current IG problems. This necessary generalisation was deemed to be unsuitable for innovative fields, such as data science, which is still developing methodologically.

Another topic, which was discussed a lot, was the challenges around GDPR. With regards to consent in the form of data permission management, there are still uncertainties around whether individual consent is appropriate and necessary. Furthermore, the greatest uncertainty was how to deal with pseudonymised data and its increased risk of re-identification for individuals. While almost all interviewees argued that the GDPR hasn’t changed any fundamental requirements in health care, it was viewed as an improvement as it brought data permission management into the spotlight.

Conclusion/Discussion:

For data users, issues of consent in the data permission management for data users don't only manifest at specific moments (such as the formal ethic or IG procedures), but are ongoing and therefore shape decision-making in research – ie. which approach to take, what to study, and how to study it. This has effects on the researchers and their profile as their results have a direct impact on how they establish themselves within their field, how and where they publish or otherwise disseminate results.

Furthermore, this impacts and depends on what kind of networks researchers are part of or trying to build relationships with. How this impacts research in the long-term is currently understudied. However, researchers see value in both information governance and technical advantages that TREs or Data Arks provide.

Expressed from many different sites were calls for a change in current practice and culture around data sharing in health-related research, as well as a change in academic research culture to integrate and allow for reasonable IG processes. Most participants found that putting the bureaucracy in place is creating a large administrative overhead that focuses mainly on legal compliance. Instead of focusing on audit in data access, many wished to see a change of focus on swift authorisation.

Participants advocated for better training and education that would not only teach data protection, such as how to better navigate ethics and IG processes, but also how to curate data for secondary research. There is a future-orientation build into this ideal as this would encourage thinking of data collection and caring for data early on. Proposed could be iterative processes in exchange with technical operators and IG personnel to distribute the costs and burdens for making data (including metadata) more interoperable and re-usable in the future.

The research community needs to come together to participate in the creation of open standards and protocols for data sharing in their respective fields. In this case, there could be fruitful discussions on how to mitigate and re-distribute risks for all involved stakeholders. Moreover, these could also crystallise where standards are not useful and more dynamic and responsive structures are needed in order to undertake innovative research that can respond to changing values and scientific methodological developments (Edwards et al., 2009).

Building functional infrastructures is an iterative and cost-intensive process. In current funding structures, financial support typically stops when the project ends. In a time span of three to five years, this is when a system and the organisational structures around them mature. A better resource exhaustion could be realised when long-term funding would be dedicated to maintain and care for those data infrastructures. Cost-sharing and education for data handling for secondary research could be one step to enable better quality and completeness of data for future research.

30th March 2020
Victoria Neumann

References:

Bates, J., Lin, Y.-W., & Goodale, P. (2016). Data Journeys: Capturing the Socio-Material Constitution of Data Objects and Flows. *Big Data & Society*, *3*(2), 1–12. https://doi.org/10.1177/2053951716654502

Beer, D., & Burrows, R. (2013). Popular Culture, Digital Archives and the New Social Life of Data. *Theory, Culture & Society*, *30*(4), 47–71. https://doi.org/10.1177/0263276413476542

Bossen, C., & Markussen, R. (2010). Infrastructuring and Ordering Devices in Health Care: Medication Plans and Practices on a Hospital Ward. *Computer Supported Cooperative Work (CSCW)*, *19*(6), 615–637. https://doi.org/10.1007/s10606-010-9131-x

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. MIT Press.

Bowker, G. C., & Star, S. L. (2000). Invisible Mediators of Action: Classification and the Ubiquity of Standards. *Mind, Culture, and Activity*, *7*(1–2), 147–163. https://doi.org/10.1080/10749039.2000.9677652

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2009). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In *International Handbook of Internet Research* (pp. 97–117). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-9789-8_5

Charmaz, K. (2006). *Constructing Grounded Theory*. Sage Publications.

Clarke, A. E. (2007). Grounded Theory: Critiques, Debates, and Situational Analysis. In W. Outhwaite & S. Turner, *The SAGE Handbook of Social Science Methodology* (pp. 423–442). SAGE Publications Ltd. https://doi.org/10.4135/9781848607958.n23

Clarke, A. E., Friese, C., & Washburn, R. (2015). *Situational Analysis in Practice: Mapping Research with Grounded Theory*. Left Coast Press.

Dewan, P. (2001). An Integrated Approach to Designing and Evaluating Collaborative Applications and Infrastructures. *Computer Supported Cooperative Work (CSCW)*, *10*(1), 75–111. https://doi.org/10.1023/A:1011266229161

Grisot, M., & Vassilakopoulou, P. (2013). Infrastructures in Healthcare: The Interplay Between Generativity and Standardization. *International Journal of Medical Informatics*, *82*(5), 170–179. https://doi.org/10.1016/j.ijmedinf.2012.08.010

Grisot, M., & Vassilakopoulou, P. (2015). The Work of Infrastructuring: A Study of a National eHealth Project. In N. Boulus-Rødje, G. Ellingsen, T. Bratteteig, M. Aanestad, & P. Bjørn (Eds.), *ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work,* 205–221. Springer International Publishing. https://doi.org/10.1007/978-3-319-20499-4_11

Hanseth, O., & Lyytinen, K. (2010). Design Theory for Dynamic Complexity in Information Infrastructures: The Case of Building Internet. *Journal of Information Technology*, *25*(1), 1–19. https://doi.org/10.1057/jit.2009.19

Jackson, S. J., Edwards, P. N., Bowker, G. C., & Knobel, C. P. (2007). Understanding Infrastructure: History, Heuristics and Cyberinfrastructure Policy. *First Monday*, *12*(6). https://doi.org/10.5210/fm.v12i6.1904

Karasti, H., & Blomberg, J. (2018). Studying Infrastructuring Ethnographically. *Computer Supported Cooperative Work (CSCW)*, *27*(2), 233–265. https://doi.org/10.1007/s10606-017-9296-7

Nafus, D. (2014). Stuck Data, Dead Data, and Disloyal Data: The Stops and Starts in Making Numbers into Social Practices. *Distinktion: Journal of Social Theory*, *15*(2), 208–222. https://doi.org/10.1080/1600910X.2014.920266

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*(5). https://doi.org/10.17705/1jais.00195

Plantin, J.-C., Lagoze, C., & Edwards, P. N. (2018). Re-Integrating Scholarly Infrastructure: The Ambiguous Role of Data Sharing Platforms. *Big Data & Society*, *5*(1), 1–14. https://doi.org/10.1177/2053951718756683

Prior, L. (2004). Doing things with documents. In Silverman, D. (Ed.), *Qualitative analysis: Issues of Theory and Method* (pp. 76–94). Sage Publications.

Ribes, D. (2014a). Ethnography of scaling, or, how to fit a national research infrastructure in the room. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '14*, 158–170. https://doi.org/10.1145/2531602.2531624

Ribes, D. (2014b). The kernel of a research infrastructure. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '14*, 574–587. https://doi.org/10.1145/2531602.2531700

Ribes, D., & Polk, J. B. (2015). Organizing for ontological change: The kernel of an AIDS research infrastructure. *Social Studies of Science*, *45*(2), 214–241. https://doi.org/10.1177/0306312714558136

Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, *7*(1), 111–134.

Tempini, N., & Leonelli, S. (2018). Concealment And Discovery: The Role of Information Security in Biomedical Data Re-Use. *Social Studies of Science*, *48*(5), 663–690. https://doi.org/10.1177/0306312718804875

Ulriksen, G.-H., Pedersen, R., & Ellingsen, G. (2017). Infrastructuring in Healthcare through the OpenEHR Architecture. *Computer Supported Cooperative Work (CSCW)*, *26*(1), 33–69. https://doi.org/10.1007/s10606-017-9269-x

Author/Main Contact:

Victoria Neumann,

PhD candidate, School of Computing and Communications,

Connected Health Cities, North West Coast

v.neumann@lancaster.ac.uk

30th March 2020
Victoria Neumann