Dec 12th, 12:00 AM

What Flows Through Data Infrastructures: Blockages, Bends, and Bottlenecks in Sharing Gender Data Between Institutions

Marta Stelmaszak
Portland State University, stmarta@pdx.edu

Erica Wagner
Portland State University, elwagner@pdx.edu

What Flows Through Data Infrastructures: Blockages, Bends, and Bottlenecks in Sharing Gender Identity Data Between Institutions

Short Paper

Marta Stelmaszak
Portland State University
Portland, Oregon, USA
stmarta@pdx.edu

Erica Wagner
Portland State University
Portland, Oregon, USA
elwagner@pdx.edu

Abstract

There is an increased interest in data infrastructures that accompany digitalization in the public sector where such infrastructures support serving the next generation of citizens better. Literature on information infrastructures provides a robust foundation, but so far theorization of what differentiates data infrastructures has been limited. We conducted a case study of a data infrastructure to share gender identity data in U.S. higher education. By tracing how a university navigated around the cultural, structural, content and material layers of the data infrastructure to share student gender identity data with the state and federal government, we uncover how data were flowing through the infrastructure. Because of differences in the layers of the data infrastructure between institutions, the flow was subject to blockages, bends and bottlenecks. Our findings demonstrate that the nature of data brings challenges in developing data infrastructures across four levels with implications for theory and practice.

Keywords: Data infrastructure, public sector, information infrastructure, data artifacts, digital infrastructure, gender, higher education

Introduction

Data infrastructures that accompany digitalization support data collection and decision-making to better serve the next generation of citizens. For example, in 2021, the U.S. government issued the first-ever National Strategy on Gender Equity and Equality, where equity was defined as the consistent and systematic fair, just, and impartial treatment of all individuals, including, among others "LGBTQI+ persons". The strategic vision emphasized striving for a world where equal opportunity is afforded to all citizens (“National Strategy on Gender Equity and Equality” 2021). The strategy recognized various elements contributing to the injustice suffered based on gender identity, that is an individual’s sense of having a particular gender, including economic disadvantage and challenges in the educational system. The government emphasized the need for "adequate collection of gender data" to realize the strategy. It embarked on "a government-wide effort to strengthen data collection and analysis", supporting "collection and analysis of sex-disaggregated data for all programs" and encouraging "more analysis and dissemination of the gender data we do collect, and we will support efforts to close gender data gaps" (“National Strategy on Gender Equity and Equality” 2021, p. 40). Recognizing that the next generation of citizens presses for recognition of gender identity beyond the male and female binary, the U.S. government embarked on an initiative to expand and update the collection of gender identity data in the education system. This is in line with a revised interpretation of the definition of discrimination "on the basis of sex" in the Title IX of the Education Amendments of 1972, which as of 2021, encompasses discrimination based on sexual orientation and gender identity. Information systems (IS) research on data infrastructures can contribute to such

efforts, bringing to the forefront the issues in the development, implementation, and use of data infrastructures in government and the public sector.

Data infrastructures are commonly defined as "the institutional, physical and digital means for storing, sharing and consuming data across networked technologies" (Kitchin 2014, p. 32; see also Parmiggiani and Grisot 2019). While related to information infrastructures, data infrastructures are recognized as a distinct phenomenon (Kitchin 2014; Parmiggiani and Grisot 2019; Rothe et al. 2019; Vassilakopoulou et al. 2019), and researchers point towards the need to account for "data as a central and evolving concern" (Parmiggiani and Grisot 2019, p. 2) in such infrastructures especially in the public sector and higher education (Gulson and Sellar 2019; Hartong and Förschler 2019). Research on data highlights that they are not only digital artifacts (Faulkner and Runde 2019; Kallinikos et al. 2013) with an editable, reprogrammable, malleable nature, but also semantic artifacts that are embedded with specific meanings and structures they are given (Aaltonen et al. 2021; Aaltonen and Penttinen 2021; Alaimo and Kallinikos 2022; Jones 2019). However, research on data infrastructures has not yet reflected these advances in how data are understood. Further, little is known about how data cross the boundaries of organizations (Jarvenpaa and Markus 2020), as their flow has been assumed mainly as unproblematic. To further our understanding of data infrastructures, we thus ask *how do data flow through data infrastructures between organizations*?

To answer this question, we report on an in-depth study of the data infrastructure in the U.S. higher education sector focusing specifically on the collection and sharing of student gender identity data. We investigate how gender identity data flow between university, state, and federal systems in the data infrastructure. We find that data infrastructures can be conceptualized into semantic, structural, content, and material layers. Differences between these layers lead to disruptions to data flows in the form of blockages, bends, and bottlenecks. We present background literature first, then report on our preliminary findings, and finally, sketch out a preliminary discussion and indicate further work.

Background Literature

Public sector infrastructures receive increasing attention both in research and in practice as a result of progressing public sector digitalisation (Aanestad and Jensen 2011; Grisot et al. 2014; Hanseth and Bygstad 2017; Vassilakopoulou et al. 2017). Information infrastructures, especially nation-wide and public sector, rely on collective visions and goals, and it remains a core issue to involve key actors in reaching these goals (Aanestad and Jensen 2011). Stakeholders, their contexts of work, technologies and routines, as well as their skills and beliefs, are often seen as an installed base for these infrastructures that requires cultivation (Aanestad and Jensen 2011; Hanseth and Bygstad 2017). Installed base cultivation strategies have been identified as one of the success factors for infrastructure innovations (Grisot et al. 2014). In other words, successful infrastructures hinge on the mobilization, coordination and cultivation of various stakeholders and their needs. While involving multiple stakeholders may be desirable for the success of an infrastructure, it also brings challenges associated with multiple actors contributing to the same project (Vassilakopoulou et al. 2017). Infrastructures involving numerous stakeholders may create tensions and contestations, as their functioning requires striking a balance between multiple, sometimes divergent interests (Hanseth and Bygstad 2018; Hepsø et al. 2009; Vassilakopoulou et al. 2019). Involving stakeholders and accommodating their needs is "both an ethical concern for public service and a requirement for platform growth and sustainability" (Vassilakopoulou et al. 2017, p. 2), and public sector is characterised by a large number of stakeholders with varying perceptions that may lead to tensions (Hanseth and Bygstad 2018).

Data infrastructures in this context emerge as a related but distinct phenomenon, one that is an object of a growing number of studies. Providing a differentiating view, Parmiggiani and Grisot's (2019) define data infrastructures as enabling the implementation, delivery and use of digital services and consisting of "the institutional, physical, and digital means for storing, sharing and consuming data across networked technologies" (Kitchin 2014, p. 32) where data are an essential resource. In data infrastructures, data are "a central and evolving concern" (Parmiggiani and Grisot 2019, p. 2), and are shaped over time through their construction and evolution in infrastructures against the background of stakeholders' differing concerns (Rothe et al. 2019; Vassilakopoulou et al. 2019). To this end, Parmiggiani and Grisot (2019) put forward an agenda for IS research encouraging more in-depth investigations into the nature of (data) infrastructuring work in the design and development stage carried out by multiple stakeholders, focusing on who does this work and how, and the role of infrastructure designers. Specifically, the authors call for more attention to 1) actors – who is involved in data infrastructuring work at the point of design, who does

what in development, and whose data practices are incorporated in emergent data infrastructures; 2) episodes – where actors come together and where their data practices are investigated; and 3) actions – how data practices are identified as part of data infrastructuring work (Parmiggiani and Grisot 2019).

Literature so far has focused more on infrastructuring and less on data in investigating data infrastructures. We argue that the study of data infrastructures should account more fully for the nature of data that flow though such infrastructures, especially as they enable inter-organizational flow of data (Jarvenpaa and Markus 2020). Data infrastructures that allow the flow of data between organizations need to cope with the complexities that arise from data instability and indeterminacy, for example where data originally generated for different contexts and purposes, with differing collection and processing practices in varying organizations, need to be combined (Jarvenpaa and Markus 2020). Data flowing through data infrastructures travel through different production and use contexts (Winter and Davidson 2019). This is even more evident when considering the nature of data themselves. Recent research in this area highlights that data are not only digital artifacts that are editable, interactive, reprogrammable (Kallinikos et al. 2013), but also that they are supported by material and nonmaterial media (Faulkner and Runde 2019) that prescribe how and what kind of data can be created along pre-programmed lines (Aaltonen et al. 2021; Alaimo and Kallinikos 2017), that is they are involved in shaping the content of data. At the same time, data are governed by logical structures that both enable and constrain them into specific formats: "data do not 'have' a structure but are made by a structure that confers data their capacity to represent contextual facts" (Aaltonen and Penttinen 2021, p. 5922). Finally, "data have a material embodiment or support, but also a specific format that makes them a unit of expression (syntactic unit) within a broader cultural (semantic) system. (…) Although basically material entities, data objects remain at the same time semantic artifacts, cognitive or cultural constructs, recurring arrangements of data ordered according to certain logics, criteria or schemata that serve cognition and knowledge aims" (Alaimo and Kallinikos 2022, pp. 22–26). Taken together, these advances suggest that material, content, structural, and semantic layers coalesce in bringing data into effective existence.

These findings are important for the study of data infrastructures, especially in the light of prior literature that discusses the tensions that arise between stakeholders who need to be engaged in developing information infrastructures. If infrastructures require managing tensions between numerous stakeholders' interests, perspectives, and capacities, then data infrastructures need to balance such tensions across material, content, structural, and semantic layers for data to flow between organizations. Thus, we investigate how data across these layers flow within a data infrastructure in a public sector.

Research Design

We report on an in-depth, revelatory case study (Yin 2018) that follows the changes in collecting and sharing of gender data in data infrastructures underpinning U.S. higher education. The case study design is appropriate as we were seeking an understanding of a complex phenomenon, the boundaries of which are not clearly defined and separable, we had no control over the course of events, and the phenomenon required an intensive description (Eisenhardt 1989; Yin 2018). The single case study design enables us to investigate in detail the workings of a data infrastructure which is essential for the analysis and theorizing from the data we adopted, but it raises some limitations we discuss at the end. The focus of the case study is the flow of gender identity data in student population reporting at a time when the definition of gender identity was changing, spurring efforts to expand gender data collection. We follow the flow of data from a public university (StateU, pseudonym) known for progressive social values, to a U.S. state (State, pseudonym) in which it is located, and the U.S. federal government (Government). In 2016, StateU embarked on the Expanded Bio-Demographic Data (EBD) project to change its internal infrastructure to accommodate the collection of expanded gender identity data alongside the State, which was one of the first in the U.S. to officially recognize non-binary gender identity. StateU is known for and ranks highly in U.S. university rankings as a queer-friendly institution due to its values and initiatives aligned with justice, equity, diversity, and inclusion.

Our data collection included interviews and document analysis. We conducted 17 interviews with employees and students at StateU and one interview with a representative of the off-the-shelf product used at StateU. The interviews were conducted between June 2021 and January 2022 and were partially retrospective, aiming to elicit interviewee recollections of past events from 2016 and present reflections. We asked interviewees about gender identity data within their roles and responsibilities, and specifically about the

EBD project. We obtained 11 documents, such as project charter and internal reports, to complement the data collection regarding StateU. We built our understanding of the state and federal levels drawing from interviews with employees of StateU responsible for reporting to state and federal authorities. We complemented these accounts with 35 documents, including federal and state reporting requirements, public policy statements, gender data practices followed by other universities, and user manual documentation from the software vendor. The data summary included in Table 1 was maintained in a case study database (Yin 2018) by the first author in Atlas.ti as a data management tool.

Type of data	**Quantity of data**			**Period of data collection**
Narrative interviews	19 interviews in total with project team members, IT professionals, administrators, data stewards, beneficiaries, vendor. In total, approximately 18 hours of audio data, 210 pages of transcripts.			June 2021-January 2022
	Interviewee Role(s)	**Unique Individuals**	**Number of Interviews**	
	Beneficiary (I_017, I_018, I_019)	3	3	
	Beneficiary/Admin (I_014)	1	1	
	IT/Steward (group interview) (I_012)	4	1	
	IT/Vendor (I_004)	1	1	
	Project Member/IT (I_001, I_007)	2	2	
	Project Member/IT/Steward (I_003, I_015, I_016)	3	3	
	Project Member/Steward/Admin (I_006, I_008, I_009, I_011)	2	4	
	Project Member/Steward/Beneficiary/Admin (I_005)	1	1	
	Steward (I_002, I_013)	1	2	
	Steward/Beneficiary/Admin (I_010)	1	1	
	Totals	**19**	**19**	
Internal documentation	11 documents in total including the project charter, meeting notes, aggregated system use reports, data reports, approximately 50 pages. (D_001 – D_046)			June 2021-January 2022
External documentation	35 documents in total such as governmental bills, regulations, user manual, policy recommendations, best practice guidelines from another university. In total, approximately 150 pages. (D_001 – D_046)			June 2021-January 2022

Table 1. Summary of data collected

We conducted an inductive qualitative analysis of interview transcripts and documentation following the three-step process recommended by Miles and Huberman (1994). First, we conducted data reduction to highlight sections of text within and across documents linked to the research question. We moved from our preliminary list of codes to the data display phase to make connections between the data and reduce the initial set of codes to a smaller number of first-other themes. We began to identify the semantic, structural, content, and material layers as pertinent to the flow of data in the infrastructure we studied. We started to perceive the blockages, bends, and bottlenecks that impacted the flow of data. During the conclusion, drawing, and verification phase, we settled on the layers and mechanisms that emerged, and we report on them in the preliminary findings below.

Preliminary Findings

Collecting Gender Identity Data at StateU

As part of the EBD project, StateU modified and expanded its internal infrastructure to enable the collection of gender identity data to reflect its stated values. **Semantically**, StateU saw gender as fluid, multiple, plural, and non-binary and was regularly ranked as one of the most queer-friendly universities in the United States. This cultural backdrop informed several initiatives, for example, establishing the Queer Center (pseudonym) and continuing work carried out by the Commission on Gender Equity (pseudonym). For

example, the Commission advocated for the inclusion of gender identity in StateU's affirmative action statement. It was one of the first universities in the U.S. to offer transgender healthcare to students. On its website, StateU explained that it aimed to collect legal sex, gender identity, and sexual orientation data from students noting that "*legal sex designation (…) may be different than your gender*" (D_015). These changes to the collection of gender data were widely perceived as important for the university, as "*we wanted students to feel seen and validated, to see them for their full diversity, we wanted to validate and affirm their identities*" (I_005). University employees widely supported expanding gender identity data collection and expressed "*anything that we can do to make people feel important, respected, included, we should do*" (I_015).

In the **structural** layer, an expanded collection of gender identity data was implemented by creating a web application allowing students and staff to enter answers in three separate fields: legal sex, gender identity, and sexual orientation. The answer in the legal sex field could only be provided once, and subsequent changes had to be requested and implemented through relevant administrative bodies. In cases where these data were not available, administration staff had to reach out to students to complete their data for further reporting purposes, a time-consuming task. The gender identity field allowed a choice from 11 answers, all stored as two-character variables in the data table. An additional open text box allowed to specify identity or identities not listed in up to 30 characters. Changes to the answers, which users could introduce themselves at any time, were not tracked; a history table did not exist. The sexual orientation field was structurally set up identical to the gender identity field.

In terms of **content**, the legal sex field accepted three values: Male, Female, No Response. Users who did not make a selection from the three options were recorded as null values, which resulted in uncertainty around what some of the responses represented, and how to interpret null versus No Response: whether any of them represented non-binary users, especially in the light of the much-discussed X gender designation that the state was planning to introduce as a third legal sex category. The gender identity field allowed the following answers: Agender, Genderqueer, Man, Non-Binary Including Gender Fluid and Gender Non-conforming, Questioning or Unsure, Trans Man, Trans Woman, Transgender, Woman, Identity or Identities Not Listed Please Specify [with text box for write-in], Prefer Not to Answer, while the sexual orientation field similarly included 11 possible answers. These fields were optional and allowed selecting only one response, effectively requiring users identifying with more than one gender identity term or sexual orientation category to use the open text box.

Materially, StateU benefited from having an in-house team capable of implementing the changes to collecting gender identity data stipulated by the EBD project requirements. This included collating an inventory of all university forms and reports which collect gender data, developing a technical solution for storing expanded gender data, updating the required database schemas, and creating a user-facing app to enable users to enter the data. The project team included information technology, business intelligence, and system analyst staff who were able to work on implementing the changes. The project team worked closely with the provider of the off-the-shelf solution that StateU used and the changes required as part of the project. The provider was aware of the requirements and admitted that similar requests were coming in from other universities in the U.S., pushing the company to start incorporating some of the changes into the baseline version of the project: "*we already knew that they had to accommodate pronouns and gender identity*" (I_004). Therefore, the provider was open and receptive to modifications in this area. The expansion of gender data collection impacted the administrative system and other connected systems and applications across StateU.

Gender Identity Data in State Systems

The State requires all universities in the State system to report data about students and courses in a centralized reporting file that supports reporting, analysis, planning, and the development of the educational provision. Each institution in the State system prepares a file that undergoes technical validation and, after approvals, is loaded to the State data warehouse. Generally, universities must submit files twice a term, then once a year with a summary of degrees, financial aid, and dual credit grades. Among many other fields, the file contains data about each student's sex. Before introducing the X gender marker, the State created legal provisions for non-binary sex identification confirmed by judges in court cases. In the **semantic** layer, the State introduced X as an official gender identity marker to be included on all state identification documents, such as driver's licenses. It was aware that collecting data about legal sex is not

sufficient to represent its population appropriately. A senate bill from 2015 required public universities in the State to allow all students, faculty, and staff to identify their sexual orientation and gender identity in university systems. As a result, the State's Higher Education Commission convened a working group consisting of all community colleges to determine the way to collect the data, implement policies, and identify barriers to implementing the senate's bill.

Structurally, the file that each university, including StateU, is required to submit concerns individual-level student data. Regarding gender identity, the file requests data about each student's sex as one character with a binary choice and unknown response. The working group established a common structure for web form fields across all universities in the state. Gender identity was included as a field from which 11 possible answers (including an open text field) were available. Similarly, sexual orientation was added with 11 possible answers (including an open text field) available for selection. It was noted that these were the minimum options that had to be included. The working group indicated that "*the answers to a legal sex designation question shall align with the current federal reporting requirements*" (D_010). On the **content** level, the State systems allowed the reporting of M, F, and U as single characters standing in for Male, Female, Unknown concerning legal sex, while gender identity and sexual orientation were collected and stored using two-character variables. Notably, at the time of the case study, the State did not require reporting gender identity and sexual orientation data. However, efforts were made to collect data with the same content "*in an already agreed-upon format for all public institutions should that request arise in the future*" (D_010).

The **material** layer related to the changes in reporting gender data required collaboration across universities, including StateU, and State agencies. The State system required each university to submit a flat text file delimited by the pipe ("|") character and was accompanied by detailed instructions regarding the extraction, transformation, and loading process for each submitted file. The expansion of gender identity data collection entailed changes to internal infrastructures, for example, as described at StateU. However, not every university in the system could cope with the workload at the same pace as StateU. Some universities estimated several hundreds of hours of work required and costs over 50,000 dollars to modify their systems. An additional technical complication was posed because universities used different off-the-shelf products. As a result, the recommendations of the state working group were to encourage universities using the same off-the-shelf software to collaborate on developing expanded gender data collection systems. In the meantime, "*an anonymous survey could be conducted among all potential and current students and employees in order to collect the data and comply with the requirements of the bill. This would be relatively inexpensive and simple to administer and could provide the requested data quickly while allowing time for institutions and vendors to implement changes to the software*" (D_010).

Gender Identity Data in Federal Systems

At the time of the changes to gender identity data collection, the federal Government changed and then reverted its interpretation of gender with consequences for the Department of Education and Title IX. Public universities receiving federal funds are bound by Title IX, which prohibits sex discrimination in educational programs and activities of recipients of federal financial assistance. In 2016, the U.S. Departments of Justice and Education issued formal correspondence to educators signaling a change in the interpretation of sex under Title IX. The text clarified that sex "*encompasses discrimination based on a student's gender identity, including discrimination based on a student's transgender status*" (D_002), and explained terminology around gender identity, sex assigned at birth, gender transition, and transgender identities. With the change of the Government in 2017, the **semantic** layer surrounding gender identity data shifted. In 2017, previous communication was withdrawn after a federal district court in Texas held that the term sex unambiguously referred to biological sex, while the U.S. Court of Appeals ruled that the new interpretation of the term in the letter provided much-needed disambiguation of sex. Further research was requested on how the term should be interpreted, while the previous definition prevailed in the context of higher education under Title IX.

This cultural climate was reflected in the required reporting to the Integrated Postsecondary Education Data System (IPEDS), a large-scale survey collecting institution-level data from post-secondary institutions in the U.S. IPEDS is a web-based collection system through which all universities have to report, including many other fields, the gender of their students to the Government. The usage of the term gender in the system was seen as a source of confusion: "*Much would be solved if we had clearer definitions around*

exactly what we're supposed to report in those fields. Are we supposed to report legal sex, are we supposed to report gender, if we're supposed to report sex is this sex [that] was assigned at birth or sex on a state ID?" (I_002). In general, it was felt that the IPEDS system perpetuated a lack of clarity around terminology by requiring universities to submit what they called gender statistics but collected legal sex data.

On the **structural** level, the IPEDS system required universities to report on whole populations rather than individual students and collected only legal sex (erroneously termed gender) as a binary response. Thus, the only allowed **content** was Male or Female. Reports to IPEDS were required several times throughout each year and were prepared by relevant staff members. Because IPEDS only accepted Male or Female as possible responses to the legal sex questions, universities, including StateU and universities in other states, developed ways to report unknown or null legal sex data. It was common practice to reassign students whose answers were other than Male or Female to one of the two categories: "*the federal government recommends dealing with students who you do not know the legal sex of by report[ing] them all as one or the other, or to split the numbers between male and female based on the know ratio of males to females*" (I_002). At StateU, such students were reported as Male. On the **material** layer, IPEDS is a web-based survey where the representatives of the responding universities ender data, and the system automatically calculates totals, averages, and percentages and compares the responses to previous submissions for consistency. During the case study, there were no signs of any planned changes to IPEDS to facilitate the collection and reporting of expanded gender identity data. The Government relied on its recommendation to reassign answers according to university procedures.

Discussion and Further Work

The main interest of this project is to understand how data flow through data infrastructures between organizations. We have outlined above a sketch of a data infrastructure that underpins the sharing of gender identity data in higher education in the United States. StateU, a university that provided the focal point for our study, evolved its internal organizational infrastructures (Hanseth and Bygstad 2018) to support the collection and circulation of gender identity data between systems and stakeholders. As part of a broader infrastructure, StateU was required to report on gender identity data to a progressive State, aware of and advocating for a more expansive conceptualization of gender identity. At the same time, StateU was obligated to report gender data to the Government.

We analyzed how the four layers of data infrastructures that we conceptualize as the semantic, structural, content, and material layers differed between StateU, the State, and the Government. As a result of these differences tensions emerged, and the flow of gender identity data was subject to blockages, bends, and bottlenecks, as shown in Figure 1. We define as blockages all instances where gender identity data – again defined as data about a student's sense of their gender identity – cannot flow through the data infrastructure. The clearest example of this is that the Government only accepts data about students' binary legal sex rather than their gender identity or even non-binary or unknown legal sex. Bends correspond to the instances where gender identity data need to change shape to flow through the data infrastructure, which in the case study conducted is exemplified by the reassignment of unknown and null responses, at StateU corresponding to non-binary legal sex, to Male or Female. Bottlenecks capture instances where the flow of gender identity data is slowed down in the data infrastructure. In the case that we observed, this includes, for example, the need by StateU staff to contact students whose legal sex is unknown before the data can be reported to the state. From a student's perspective, they report their gender identity to StateU by recording it in the system as Identity or Identities Not Listed Please Specify, and they specify their gender using the write-in box. When StateU shares these data with the State, they are bent because the contents of the write-in box are not shared, only two letters that designate Identity or Identities Not Listed are communicated to the State. StateU has to report data to the Government, but in this case gender identity data undergoes a blockage as the Government does not accept gender identity data. Consequently, a bottleneck is created as StateU needs to contact the student to obtain these data.

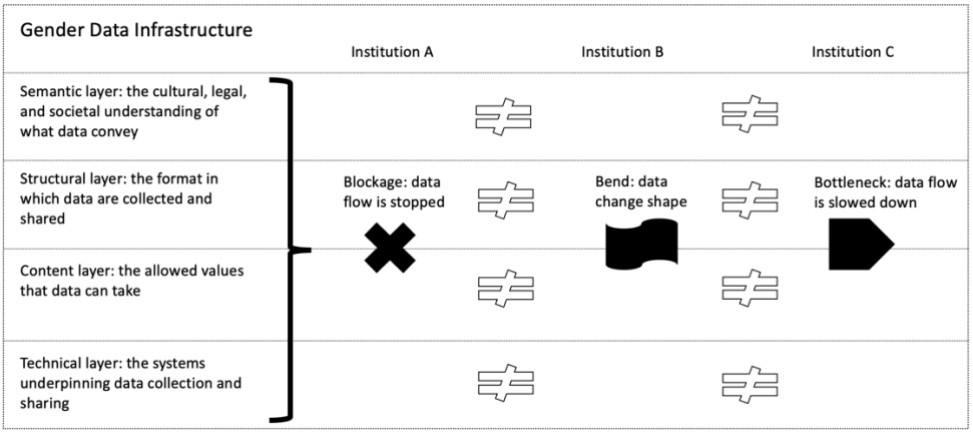

Figure 1. Blockages, bends, and bottlenecks in the gender data infrastructure

Our preliminary findings and theorization contribute to the nascent literature on data infrastructures. By taking into account the multi-faceted nature of data (Aaltonen et al. 2021; Aaltonen and Penttinen 2021; Alaimo and Kallinikos 2022; Jones 2019), we have shown how, in the case of data infrastructures, the semantic, structural, content, and material layers all come to bear on the flow of data within the infrastructure. We posit that investigations of data infrastructures have to go beyond just "the institutional, physical and digital means for storing, sharing and consuming data" (Kitchin 2014, p. 32) and engage closely with the meanings and structures that data travel in. Our preliminary findings indicate that the material and content layers may not be the most challenging for the evolving data infrastructures. However, the structural and, most significantly, semantic layers weigh in more regarding data flows. We also begin to see tensions emerging where the layers differ between organizations and stakeholders, as well as tensions between the layers within organizations: for example, the semantic layer, that is, the understanding of the meaning of data at an organization, may not be congruent with the structure and content of the data that the organization collects and shares. These findings also contribute to growing scholarship on data sharing between organizations. They highlight other three aspects, beyond just the technical feasibility of data sharing, that come into play when exchanging data. These findings are significant for the growing data infrastructures in the public sector (Jarvenpaa and Markus 2020; Parmiggiani and Grisot 2019, 2020; Rothe et al. 2019) that need to respond to the needs of the next generation of citizens.

While we report on preliminary work in this short paper, we are conscious of some limitations. For instance, the single case study design we adopted gives us access to data mainly from one organization and therefore we indirectly draw inferences about the flow of data between various organizations. We do triangulate the data with external documentation, but concede that the empirics are more limited outside of the focal organization. We aim for further data collection to address this limitation. Second, the case concerns gender identity data which may be a specific enough type of data with particular data sharing and reporting requirements. Therefore, our findings may not extend to other types of data. As we develop the paper further, we aim to ground our theorization in a suitable conceptual framework and connect it more closely with existing literature, while remaining focused on gender identity data.

References

Aaltonen, A., Alaimo, C., and Kallinikos, J. 2021. "The Making of Data Commodities: Data Analytics as an Embedded Process," *Journal of Management Information Systems* (38:2), pp. 401–429.

Aaltonen, A., and Penttinen, E. 2021. "What Makes Data Possible? A Sociotechnical View on Structured Data Innovations," *Proceedings of the 54th Hawaii International Conference on System Sciences*.

Aanestad, M., and Jensen, T. B. 2011. "Building Nation-Wide Information Infrastructures in Healthcare through Modular Implementation Strategies," *Journal of Strategic Information Systems* (20:2), pp. 161–176.

Alaimo, C., and Kallinikos, J. 2017. "Computing the Everyday: Social Media as Data Platforms," *The*

Information Society (33:4), pp. 175–191.
Alaimo, C., and Kallinikos, J. 2022. "Organizations Decentered: Data Objects, Technology and Knowledge," *Organization Science* (33:1), pp. 19–37
Eisenhardt, K. M. 1989. "Building Theories from Case Study Research," *The Academy of Management Review* (14:4), pp. 532–550.
Faulkner, P., and Runde, J. 2019. "Theorizing the Digital Object," *MIS Quarterly* (43:4), pp. 1278-1302.
Grisot, M., Hanseth, O., and Thorseng, A. A. 2014. "Innovation of, in, on Infrastructures: Articulating the Role of Architecture in Information Infrastructure Evolution," *Journal of the Association for Information Systems* (15:4), pp. 197–219.
Gulson, K. N., and Sellar, S. 2019. "Emerging Data Infrastructures and the New Topologies of Education Policy," *Environment and Planning D: Society and Space* (37:2), pp. 350–366.
Hanseth, O., and Bygstad, B. 2017. "The EPrescription Initiative and Information Infrastructure in Norway," in *Information Infrastructures within European Health Care*, M. Aanestad, M. Grisot, O. Hanseth, and P. Vassilakopoulou (eds.), pp. 73–87.
Hanseth, O., and Bygstad, B. 2018. "Platformization, Infrastructuring and Platform-Oriented Infrastructures. A Norwegian e-Health Case," in *Working Papers in Information Systems*, P. Nielsen (ed.), University of Oslo.
Hartong, S., and Förschler, A. 2019. "Opening the Black Box of Data-Based School Monitoring: Data Infrastructures, Flows and Practices in State Education Agencies," *Big Data and Society* (6:1), pp. 1–12.
Hepsø, V., Monteiro, E., and Rolland, K. H. 2009. "Ecologies of E-Infrastructures," *Journal of the Association for Information Systems* (10:5), pp. 430–446.
Jarvenpaa, S. L., and Markus, M. L. 2020. "Data Sourcing and Data Partnerships: Opportunities for IS Sourcing Research," in *Information Systems Outsourcing: The Era of Digital Transformation, Progress in IS*, R. Kirschheim, A. Heinzl, and J. Dibbern (eds.), pp. 61–79.
Jones, M. 2019. "What We Talk about When We Talk about (Big) Data," *Journal of Strategic Information Systems* (28:1), pp. 3–16.
Kallinikos, J., Aaltonen, A., and Marton, A. 2013. "The Ambivalent Ontology of Digital Artifacts," *MIS Quarterly* (37:2), pp. 357–370.
Kitchin, R. 2014. *The Data Revolution. Big Data, Open Data, Data Infrastructures and Their Consequences*, London: SAGE Publications.
Miles, M. B., and Huberman, M. A. 1994. *Qualitative Data Analysis: An Expanded Sourcebook*, Sage.
"National Strategy on Gender Equity and Equality." 2021. *The White House*.
Parmiggiani, E., and Grisot, M. 2019. "Data Infrastructures in the Public Sector: A Critical Research Agenda Rooted in Scandinavian Research," in *10th Scandinavian Conference on Information Systems (SCIS), Nokia, Finland*, pp. 1–16.
Parmiggiani, E., and Grisot, M. 2020. "Data Curation as Governance Practice," *Scandinavian Journal of Information Systems* (32:1), pp. 3–38.
Rothe, H., Jarvenpaa, S. L., and Penninger, A. A. 2019. "How Do Entrepreneurial Firms Appropriate Value in Bio Data Infrastructures: An Exploratory Qualitative Study," in *Proceedings of the 27th European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden*, pp. 1–17.
Vassilakopoulou, P., Grisot, M., Jensen, T., Sellberg, N., Eltes, J., Thorseng, A., and Aanestad, M. 2017. "Building National EHealth Platforms: The Challenge of Inclusiveness," in *ICIS 2017 Proceedings*.
Vassilakopoulou, P., Skorve, E., and Aanestad, M. 2016. "A Commons Perspective on Genetic Data Governance: The Case of BRCA Data," in *24th European Conference on Information Systems (ECIS 2016) İstanbul, Turkey, June 12-15, 2016*, pp. 1–17.
Vassilakopoulou, P., Skorve, E., and Aanestad, M. 2019. "Enabling Openness of Valuable Information Resources: Curbing Data Subtractability and Exclusion," *Information Systems Journal* (29:4), pp. 768–786.
Winter, J. S., and Davidson, E. 2019. "Big Data Governance of Personal Health Information and Challenges to Contextual Integrity," *Information Society* (35:1), pp. 36–51.
Yin, R. K. 2018. *Case Study Research and Applications. Design and Methods*, Thousand Oaks, CA: SAGE Publications.
Yoo, Y. 2015. "It Is Not about Size: A Further Thought on Big Data," *Journal of Information Technology* (30:1), pp. 63–65.