



University of
Massachusetts
Amherst

Recognition in Personal Data: Data Warping, Recognition Concessions, and Social Justics

Item Type	Article
Authors	Stelmaszak, Marta;Wagner, Erica;DuPont, Nicolle Nixon
Rights	Attribution 4.0 International
Download date	2025-02-10 07:37:33
Item License	http://creativecommons.org/licenses/by/4.0/
Link to Item	https://hdl.handle.net/20.500.14394/55068

RECOGNITION IN PERSONAL DATA: DATA WARPING, RECOGNITION CONCESSIONS, AND SOCIAL JUSTICE¹

Marta Stelmaszak

Isenberg School of Management, University of Massachusetts Amherst,
Amherst, MA, U.S.A. {mstelmaszak@isenberg.umass.edu}

Erica L. Wagner and Nicolle Nixon DuPont

Portland State University, Portland, OR, U.S.A.
{elwagner@pdx.edu} {nicolle.dupont@pdx.edu}

Data-related harm and injustice are commonly viewed through instrumental, procedural, distributional, or representational theories of social justice. These theories do not account for the social injustice that occurs through the lack of recognition of individuals when data are first conceptualized. We explore the recognition of individuals in data conceptualization by drawing from information systems (IS) literature on data artifacts that acknowledges the fact that data are comprised of semantics and formats. Guided by recognition theory, we studied a project to expand sexual orientation and gender identity (SOGI) data collection at a public, U.S.-based LGBTQ+ welcoming university. We found that while the actors involved worked towards recognizing in SOGI data the fact that identities are layered, non-binary, plural, and fluid, the data themselves still misrecognized individuals due to data warping. We argue that data warping occurs because of recognition concessions between social recognition through data semantics and systems recognition through data formats. Such concessions are both necessary for some recognition but accessory to misrecognition. Our findings have implications for recognition theory, data justice, and information systems research, as well as for personal data in practice.

Keywords: Personal data, data conceptualization, data artifacts, gender, recognition, social justice, SOGI

Introduction

A growing number of organizations have begun to examine their systems, structures, and practices to recognize and support groups that have been historically marginalized in economic, social, political, and cultural terms, such as the LGBTQ+² community (Holvino, 2010; Kelemen & Rumens, 2008; Rumens et al., 2019) with the aim of complying with the principles of social justice, equity, diversity, and inclusion (JEDI). Organizational efforts in this area include prioritizing conversations about JEDI, increasing leadership buy-in and

support for JEDI initiatives, and increasing accountability for JEDI-related progress (Bendl et al., 2009; Bendl & Hofmann, 2015). Many of these rest on the ability to collect, store, and report data about the populations served—clients, employees, investors, community partners, and more (Guyan, 2022; Labuski & Keo-Meier, 2015; Masiero & Das, 2019; Tarafdar et al., 2023). As such, some organizations also aim to promote data justice—that is, “fairness in the way people are made visible, represented and treated as a result of their production of digital data” (Taylor, 2017, p. 1, see also D’Ignazio & Klein, 2020; Dencik et al., 2022).

¹ Min-Seok Pang, Atreyi Kankanhalli, Margunn Aanestad, Sudha Ram, and Likoeb M. Maruping were the accepting senior editors for this paper. Lucas Introna served as the associate editor. The transparency material for this paper is available at <https://doi.org/10.17605/OSF.IO/VN7R8>.

² LGBTQ+ is an acronym that stands for lesbian, gay, bisexual, transgender, and queer, and + is used in recognition of all non-straight, non-cisgender identities. It is used in this paper in accordance with the GLAAD Media Reference Guide available at <https://glaad.org/reference/>.



©2024. The Authors. This work is licensed under the terms of the Creative Commons Attribution CC BY 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

This was certainly the case at the organization we encountered. StateU, a U.S.-based progressive public university, launched a high-profile project to overhaul its data collection and related practices with the aim of achieving more just recording of personal data pertaining to LGBTQ+ individuals in its systems (by launching a web application to collect data about individuals' legal sex, gender identity, and sexual orientation). Despite its best intentions, StateU ended up with an unjust result, where the data were still a source of injustice to individuals (by misrepresenting their gender identities and sexual orientations). In retrospect, stakeholders asked: "Why try to be inclusive when you're going to be more harmful?" (Stakeholder, Interview), and concluded that attempts to create more data justice, if not done right, "can be extremely negative if you give something that people have been asking for, for a long time, in a way that doesn't reflect what they were asking for" (Stakeholder, Meeting). Thus, the organization achieved a socially unjust outcome despite its best efforts to foster more social justice through data.

To understand this puzzling outcome, we turned to research on data justice. The literature primarily draws on four theories of social justice: instrumental, procedural, distributional, and representational; offering means to explain, theorize, and mitigate many instances of data injustice (Dencik et al., 2022; Kitchin, 2022; Kitchin & Lauriault, 2014), including those related to gender (D'Ignazio & Klein, 2020). However, the literature is largely unable to fully explain the empirical phenomenon of injustice that occurs before data are used, resources are distributed, or voices are included in or excluded from data. Further, data justice research does not closely study the information systems involved in the creation of injustice (Bruhn Jensen, 2021; Dencik et al., 2022; Kitchin, 2022). Such detailed study might help to improve understanding of the newly observed form of data injustice, which thus far remains unaccounted for theoretically.

In the spirit of phenomenon-focused problematization (Monteiro et al., 2022), we set out to unpack the empirical phenomenon and formulate a theory on the basis of our findings. We turned to the theory of recognition as social justice (Benhabib, 2018; Fraser, 1997, 2000; Fraser & Honneth, 2003). Recognition, that is, the acknowledgment of the intrinsic value of a person (Honneth, 2003), is fundamental to social justice as "one becomes an individual ... only by virtue of recognizing, and being recognized by, another subject" (Fraser, 2000, p. 109). Furthermore, denying recognition, including of one's gender, may result in harm to self-identification and identity (Fraser, 2000; Taylor, 1994). The need for recognition is often the reason why individuals share and disclose data about themselves, as "driven by an individual's need for recognition of the inner self ..., individuals are able to show who they see themselves as, or who they wish others to see themselves as" (Leidner & Tona,

2021, p. 350). Providing due recognition in personal data ("any information relating to an identified or identifiable natural person," Leidner & Tona, 2021, p. 343) is a matter of human dignity, while misrecognition is an affront to dignity, often leading to microaggression, exhaustion, frustration, anger and anxiety (Spiel, 2021) or an outright risk to health and life (Dhoest & Szulc, 2016; Szulc, 2019).

Given the phenomenon we observed, it became apparent that it is crucial to understand how the process of providing recognition in personal data unfolds. Current information systems (IS) research on data artifacts shows that data are comprised of semantics, that is, the meaning they are supposed to represent (Alaimo & Kallinikos, 2022; Kallinikos, 2006; Kallinikos et al., 2013; Tuomi, 1999), as well as formats, such as data structures, types, and properties (Alaimo & Kallinikos, 2022; Faulkner & Runde, 2019; Hirschheim et al., 1995; Recker et al., 2021). Thus, providing recognition in data is reliant upon these two aspects. Data semantics and formats are worked out when data are first conceptualized (Chua et al., 2022) by the actors involved in the process, i.e., when discovery, discussions, and decisions take place about how to represent phenomena in data. However, we know little about this process in practice; consequently, we have little understanding about how semantics and formats are established in data conceptualization and what the resulting consequences in terms of what the data might signify. Further, data semantics and formats have not been discussed in terms of recognition as an intended outcome for what data represent. Thus, we ask: *How are data semantics and formats involved in providing recognition in personal data?*

To answer this question, we studied a six-month project to expand data collection regarding individuals' legal sex, gender identity, and sexual orientation (SOGI) data at StateU. We analyzed how the actors involved worked towards providing recognition of LGBTQ+ individuals through data semantics and data formats based on the theory of recognition (Fraser, 2000; Fraser & Honneth, 2003; Honneth, 2004; Taylor, 1994). Drawing on in-depth interviews, documentary evidence, and personal communication, we uncovered *data warping*, i.e., a phenomenon that involves data semantics and formats misrecognizing individuals to whom the data pertain, despite the actors' intentions to the contrary, resulting in social injustice through a lack of recognition. Further, we found that data warping results from *recognition concessions*, i.e., trade-offs between competing forms of recognition that are necessary to achieve an overall state of some recognition in data. *Social recognition* of individuals by those who read, analyze, or use the data that are built into data semantics is often at odds with *systems recognition* in data formats that ensure that individuals are legible to the systems that record, process, and disseminate the data. Providing recognition in

data requires both forms of recognition, while concessions need to be managed because the trade-offs are both necessary to achieve some recognition and accessory to some misrecognition in data.

We contribute to the scholarship on social justice and IS by illuminating the crucial nexus of recognition and personal data. We show the complexity of working towards the provision of due recognition in practice amid a variety of actors, and we foreground a systems perspective in a world where recognition is increasingly mediated by data and technologies that heavily rely on them. Our theorization of recognition concessions helps provide a better understanding of how information systems are implicated with social justice in design and demonstrates that achieving social justice in digital technologies may require a careful balancing act. We propose recognition as a lens to study personal data that accounts for their relationship with human dignity and identity. Finally, we contribute to data justice research by advancing the study of social justice as recognition underpinned by IS research. We offer guidelines for IS managers in organizations that want to engage in similar efforts.

Theoretical Background

Data Justice as Recognition

Research on data justice is intensifying as instances of data harm and injustice are becoming more and more common (D'Ignazio & Klein, 2020; Kitchin, 2022). The field is concerned with harm and injustice related to data, and the majority of theorizing draws from the conceptualizations of social justice as instrumental (fair use of data and just outcomes), procedural (fair handling of data), distributional (fair distribution of data and associated resources), or representational (equal voice in data power) (Dencik et al., 2022; Dencik et al., 2019; Kitchin, 2022; Kitchin & Lauriault, 2014). For example, the instrumental justice lens studies how data are used in ways that discriminate, marginalize, or exclude people (Heeks & Renken, 2018), with a focus on outcomes and impacts. Procedural justice focuses on harm rendered through data practices and processes, such as the capture, processing, and sharing of data, or control and consent (Heeks & Renken, 2018; Robinson & Franklin, 2021). Distributional justice relates to the coverage of people in datasets, differential access to data resources, or unequal distribution of data and uneven outcomes (Cinnamon, 2017, 2020). Finally, representational data justice concerns the extent to which data subjects are able to control or participate in data politics and decision-making (Cinnamon, 2017, 2020; Kitchin, 2022).

However, recent political and philosophical thought suggests that a necessary condition for all forms of social justice is recognition. The “recognition-theoretical turn” (Honneth, 2003, p. 125) is based on the idea of justice enacted through acknowledging and being acknowledged by others (Fraser, 2000), and it indicates that justice stems from seeing others as different but equal and from recognizing and valuing diversity and difference (Fraser, 2003). Social justice as recognition means “a difference-friendly world, where assimilation to majority or dominant cultural norms is no longer the price of equal respect” (Fraser, 2003, p. 7). A just society recognizes the individual dignity of all and secures the condition of mutual recognition, whereby personal identity formation and individual self-realization can be accomplished by all (Honneth, 2004). Drawing from Hegelian phenomenology, recognition encompasses a “reciprocal relation between subjects in which each sees the other as its equal and also separate from it” (Fraser, 2003, p. 10). In this sense, recognition is essential in the construction of an individual’s identity (Fraser, 2003). Recognition, then, is a relation of acknowledging and being acknowledged that gives rise to individual identity, integrity, and dignity.

Social injustice arises when recognition is withheld (Honneth, 2004). Injustice can mean cultural domination, where an individual is subjected to patterns of interpretation that are alien or hostile to their own culture; nonrecognition, whereby an individual is rendered invisible within their own culture; or misrecognition in the form of oppression or imprisoning someone in a distorted or reduced mode of being (Fraser, 1995; Taylor, 1994). Such injustice can inflict harm, and due recognition is not just a courtesy but “a vital human need” (Taylor, 1994, p. 25). To be denied recognition or to be misrecognized may result in harm to self-identification and identity (Fraser, 2000). Because experiences of misrecognition violate individual identities, those affected engage in “the struggle for recognition” (Honneth, 1996), that is, the fight for acknowledgment and affirmation of their particular identities. Thus, achieving social justice stems from recognition, and, conversely, social injustice is born from a lack of due recognition.

While recognition is perceived as necessary for social justice, there are two lines of thought as to how it can foster social justice: *recognition and redistribution* and *redistribution as recognition*. The former, proposed by Nancy Fraser (1995, 2000, 2003), stipulates that while recognition is essential for social justice, it has to be followed by the redistribution of economic and political resources for social justice to be achieved. In this dual perspective, recognition and redistribution are both fundamental and irreducible dimensions of justice: “Justice today requires *both* redistribution *and* recognition. Neither alone is sufficient” (Fraser, 2003, p. 9). The latter, monistic perspective advocated

for by Axel Honneth (1996, 2003; 2004), highlights recognition as the primary, fundamental, overarching value, while redistribution is a consequence, a derivative of due recognition. While further discussion of these two perspectives is beyond the scope of this paper, it is important to note that for both, recognition is the first and necessary condition of fostering social justice.

One of the areas where individuals have historically engaged in a struggle for recognition is gender and sexuality (Fraser, 2003; Honneth, 2003). Historically, masculinity has been privileged, and everything associated with femininity has been devalued and denied due recognition. This has resulted in gender injustice that “cries out for redress via recognition” (Fraser, 2003, p. 21). Similarly, “gay and lesbian subcultures no longer struggle mainly for economic equality or material redistribution, but for respect for the characteristics by which they see themselves culturally bound together” (Honneth, 2003, p. 111). Extending the logic of recognition into the current understanding of gender identity and sexual orientation calls for the due recognition of individuals whose identities may differ from the dominant ones. It entails recognizing those who do not fit into clear-cut gender distinctions between men and women and recognizing that their gender identity, legal sex, and sexual orientation may not always align (Wade & Ferree, 2018; West & Zimmerman, 1987); that there may exist gender identities that are non-binary (McCann & Monaghan, 2019; Wade & Ferree, 2018); that people may identify as having more than one gender at any given time (McCann & Monaghan, 2019; Szulc, 2020); and that their gender may change over time (Mackenzie, 2017; McCann & Monaghan, 2019). Fostering social justice for such minoritized and underprivileged populations (often referred to as LGBTQ+), according to the logic of recognition, rests on recognizing such individuals’ gender identities and sexual orientations.

Recognition can be related to personal data (for the calls to use this lens see Braun & Hummel, 2022; Cinnamon, 2020; Dencik et al., 2022). It is the need for recognition of the inner self that drives individuals to decide to share or disclose data about themselves (Leidner & Tona, 2021). By deciding to disclose data, individuals gain the ability to show who they see themselves as and who they wish to be seen as; at the same time, those who see, use, or process these data can see the individuals in question (Leidner & Tona, 2021). Yet, recognition in data is also a locus of struggle, for example, for LGBTQ+ individuals and their SOGI data. On the one hand, collecting and using SOGI data can help affirm the chosen gender and sexual identities of individuals (Johnson, 2015; Mearns et al., 2020), and provide them with more recognition in organizational decision-making (Kelemen & Rumens, 2008; Rumens et al., 2019) while simultaneously educating others about the changing notions surrounding these concepts.

On the other hand, SOGI data often misrepresent individuals (Guyan, 2022) and can contribute to the “erasure of non-binary people” (Spiel, 2021, p. 480). Therefore, attempts at providing recognition in personal data can lead to socially just outcomes, but they can also result in misrecognition.

Thus, viewing social justice as recognition allows us to unpack what it means to recognize an individual in data and take note of the extent to which personal data do or do not allow for due recognition. Whether recognition in data is provided or not is, of course, also a matter of data collection, accumulation, or analytics (which are the focal points of data justice research), but it is first and foremost a matter of how data are designed. In other words, recognition in data depends on actions, choices, and decisions when data are created. Yet, such data conceptualization tends to escape data justice studies, in which information systems and infrastructures are often not considered and the materiality of the systems and data are neglected (Bruhn Jensen, 2021; Dencik et al., 2022). This is where we turn to IS research on data.

Data Semantics and Formats

Research on data conceptualization (e.g., requirements elicitation, conceptual modeling, and representation theory, Chua et al., 2022) provides a starting point to understand the creation of data as a process whereby information requirements are discovered, data generation processes are discussed, and the scope of data is defined (Chua et al., 2022). Studies in this area have traditionally highlighted how data are conceptualized by IS professionals who elicit data requirements from business users and who subsequently model data in databases (Recker et al., 2021). Recent literature has highlighted that more and different people, often without training in IS, are becoming increasingly involved in data conceptualization, and thus “we need to better understand ... the assumptions, beliefs, and design knowledge they bring to bear” (Recker et al., 2021, p. 281, see also Lukyanenko et al., 2019). This may especially be the case with personal data that may require various actors’ involvement in conceptualization. Therefore, different practices, knowledge, assumptions, and beliefs may contribute to how data are conceptualized, even before they are put to any use.

Data conceptualization can be seen as a process of creating data artifacts, i.e., human-made entities that can be acted upon (Aaltonen et al., 2021; Kallinikos et al., 2013) and are structured, i.e., composed of constituent parts (Faulkner & Runde, 2019). Data artifacts have “a specific format that makes them a unit of expression (syntactic unit) within a broader cultural (semantic) system” and “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 & Kallinikos, 2022, p. 22). Conceptualizing data, then, encompasses working out two aspects: semantics and formats.

In terms of semantics, data are meant to stand in for phenomena that have particular cognitive, cultural, social, or physical significance (Alaimo & Kallinikos, 2022). Therefore, data are imbued with meaning (Kallinikos, 2006; Tuomi, 1999): “All computer data ultimately have to be interpreted in terms of their natural language meaning(s). Hence, data can at best convey meaning from someone to someone, but they cannot ‘have’ any objective meaning,” explain Hirschheim and colleagues (1995, p. 62). In this sense, data semantics represent “the deep structure” (Wand & Weber, 1995; Wand & Weber, 1990): the meaning of a phenomenon that data are intended to represent, often corresponding to knowledge, rules, views, and beliefs. The role of data semantics is to contain a relatively stable representation of a phenomenon so that data artifacts can be exchanged, shared, and understood in a similar way by various data users. Data semantics are, then, the meaning that the data intend to convey, which is established by the actors involved in data conceptualization.

In terms of formats, data consist of symbols that adhere to the rules of language (in Faulkner & Runde, 2019; code in Aaltonen & Penttinen, 2021; or grammar in Hirschheim et al., 1995). They are created to function as a form of expression (Alaimo & Kallinikos, 2022; Faulkner & Runde, 2019; Hirschheim et al., 1995). Data “consist of symbols arranged into well-formed expressions, where well-formed means that these expressions adhere to the syntactical and semantic rules of the language in which they are couched” (Faulkner & Runde, 2019, p. 1284). Data format can include several elements: data type (defining the behavior and operations possible, such as integer, character, or Boolean) (Aaltonen & Penttinen, 2021; Tuomi, 1999); data structure (a grouping of various data types in an organized way, such as a list or tree) (Beynon-Davies, 2016; Beynon-Davies & Wang, 2019; Bowen et al., 2006; Tempini, 2020); as well as their attributes, i.e., behaviors and operations on data (e.g., whether mathematical operations can be performed on a datum, or whether it can store characters and can be concatenated) (Aaltonen et al., 2021; Beynon-Davies, 2016; Eriksson & Ågerfalk, 2010).

Data formats are often perceived as derivatives of data semantics (Bera et al., 2019; Chua et al., 2022). Data types, structures, and attributes have been treated in research as largely mundane technical artifacts: “The design, selection, or choice [of format] is typically considered unproblematic” (Eriksson & Ågerfalk, 2010, p. 434; see also Beynon-Davies, 2016). However, some studies show that data structures, for example, can be key actors within institutional order (Beynon-

Davies, 2016; Beynon-Davies & Wang, 2019), dictating processes and procedures. Data types, structures, and attributes carry both ontological (what things are seen as existent) and deontological assumptions (not only what actors are expected and enabled to do, but also what they are prohibited from doing) (Beynon-Davies & Wang, 2019), and they contribute to generating organizational facts. For example, in a study of data identifiers, Eriksson and Ågerfalk (2010) show how the design, choice, assignment, withdrawal, and replacement of identifiers in different formats can have economic and political consequences, calling to study how the format of identifiers relates to physical objects and the rules governing format creation.

Additionally, these two aspects are intrinsically linked, as it is formats that enable data to contain semantics: “any digital inscriptions need to be accompanied by a structure that allows contextualizing and making sense of their semantic content” (Aaltonen & Penttinen, 2021, p. 1; see also Tuomi, 1999). Data are created by putting semantics into formats that underpin the possibility of computationally recording, processing, and analyzing said data. Equally, data formats without semantics are meaningless. Data conceptualization therefore involves working out data semantics (what the data mean) and formats (data types, structures, and attributes).

The current literature indicates that conceptualizing data involves discovery, discussions, and decisions around both data semantics and data formats. By extension, recognition in data concerns both of these aspects. Providing recognition in personal data would then seem to require imbuing both data semantics and data formats with the capacity to recognize individuals, and the lack of recognition in data could result from inadequate data semantics or data formats. Studying the conceptualization of data artifacts as the creation of both data semantics and data formats can help in understanding how recognition in data can be achieved and, conversely, what leads to a lack of recognition.

Research Design

Research Setting

We conducted an in-depth, revelatory, interpretive (Sarker et al., 2018) case study that allowed us to observe and analyze a novel phenomenon (Yin, 2018). The aim was to investigate how personal data (specifically SOGI) are conceptualized in order to provide recognition as part of a project intended to develop a web application to collect such data in an organization’s management information system (MIS). The case study design is appropriate, as we seek to understand a

complex phenomenon whose boundaries are not clearly defined and separable, we had no control over the course of events, and the phenomenon of interest required thorough description (Eisenhardt, 1989; Yin, 2018). We sought a setting in which we would be able to investigate personal data conceptualization against a background where the issues of recognition were at stake. To investigate these issues, we needed access to a setting where recognition in data was at the forefront of concern in an organization that wanted to act upon social injustice, which is a sensitive and often controversial topic. These criteria led us to identify university environments as ideal because of numerous recent projects related to the expansion of collection of personal data pertaining to race, ethnicity, and gender. Specifically, conversations on U.S. campuses related to gender identity and the Title IX educational amendment (protecting people from discrimination based on sex in education programs that receive federal funding) pointed us toward an investigation of SOGI data. Consequently, we obtained research access at a U.S. state university located in a LGBTQ+ welcoming city.

The university, StateU (all names of organizational units and roles have been pseudonymized), is known for progressive social values and regularly tops niche rankings as an LGBTQ+ friendly campus. It is located in one of the first states in the U.S. to officially recognize non-binary legal sex designation and gender identity. StateU is a public university comprising 10 academic units, with approximately 1,700 faculty members and about 24,000 students. There is an active and vocal segment of the campus community that identifies with various legal sex, gender identity, and sexual orientation designations and is supported by the work of the internal Commission on Gender Equality (an internal body fostering gender diversity, equity, and inclusion) and Queer Center (a unit at StateU dedicated specifically to providing support to LGBTQ+ students). We focused our study on the Expanded Bio-Demographic Data (EBD) project that the university embarked on in January 2016 and concluded on August 1, 2016, with the launch of a web application allowing beneficiaries (students, staff, and faculty) to disclose their SOGI data. This project was driven by StateU's mission to improve the sense of inclusion felt by the campus community in terms of their gender identity and sexual orientation, which included recognizing that legal sex, gender identity, and sexual orientation are layered, non-binary, plural, and fluid facets of identity. The objectives of the project were: (1) to collate an inventory of all university forms and systems collecting gender-related data and modify them accordingly; (2) to develop a technical solution for declaring, updating, and storing SOGI data; (3) to update the underlying databases that store gender-related data; and (4) to develop policies for acceptable use of SOGI data.

The project was high on StateU's agenda, with the vice president of Administration and the vice president of Diversity and Inclusion as project sponsors. Because of the project's scope and its impact across the entire campus, a formal IT project was initiated with a dedicated IT project manager, and the project itself followed the official project practices and procedures of the Project Management Office, including strict documentation requirements. The project team consisted of a group of domain experts who had in-depth knowledge and represented the interests of students, staff, and faculty, as well as administrative users of SOGI data. Among domain experts were representatives of the Office of Diversity and Inclusion, Institutional Reporting, the Registrar's Office, Human Resources, and the Queer Center. The five domain experts were accompanied by five information technology (IT) colleagues—(e.g., business intelligence team manager, MIS team manager, developer, analyst)—as technical experts and the project manager. The team met regularly every two weeks for nearly six months to implement the project's objectives.

The launch of the SOGI data web application on August 1, 2016, was generally perceived as a success. Various comments were made on public fora about StateU's commitment and dedication to leading on SOGI data, with StateU being identified as an example for other institutions to follow: "Congrats to [StateU] for certainly doing great work and being a resource for the rest of the institutions as they work toward collecting and reporting on this important data!" (Internal Document). However, the reception of the web application and its perception as a means to recognize minoritized sexual and gender identities in SOGI data were mixed, as we detail in the findings section below. In summary, the EBD project was a well-conceived and properly run IT project that was intended to realize the goal of providing more recognition in SOGI data at a progressive and dedicated organization, yet the result was, to a large degree, a failure to properly recognize individuals' identities in data.

Data Collection

We conducted a field study of the EBD project between June 2021 and January 2022. Our data collection effort was three-fold, including interviews, direct observation, and document analysis. In total, we conducted 20 interviews. Two individuals were interviewed more than once to cover extended discussions and follow-up questions, and there was one group interview involving four interviewees who preferred to be interviewed as a group. The interviewees held various roles at the university and included members of the project team (domain and technical experts), and EBD application beneficiaries, i.e., students, staff, and faculty using a variety of legal sex, gender identity, and sexual orientation designations. We also interviewed the manager

responsible for SOGI data at the vendor of the standard MIS in use at StateU and universities nationwide. We interviewed everyone on the EBD project team who was still employed at StateU at the time of the study (some original team members left and were not available to participate in the study), and we promoted participation among beneficiaries. We deliberately interviewed participants who were closely involved in the project due to our interest in data conceptualization, rather than, for example, later data use or impact. Due to the sensitive nature of SOGI data and the EBD project, data collection was underpinned by “a desire to find ways to gather empirical data about the experiences of politically and socially marginalized people without reproducing such marginalization through practices of research” (Schilt et al., 2018, p. 5, see also Sarker et al., 2018)³. Thus, the majority of the interviews followed the narrative interviewing convention designed to elicit the interviewees’ view of the critical moments from the project without the researcher first defining what is important (Bauer, 1996; Bauer & Gaskell, 2007). This approach aligns with the practices of careful listening and organizing conversations around specific actors, concrete situations, and resonant symbols (Ghaziani, 2019; Love, 2019). Furthermore, uninterrupted narration often highlights controversies and individuals who might not otherwise be considered to participate (e.g., some beneficiaries that we interviewed). Interviews with the project team were conducted between June 2021 and January 2022 and were partially retrospective in nature, aiming to elicit interviewee recollections of past events from 2016 as well as present reflections. Interviews with beneficiaries were conducted between September 2021 and November 2022 and focused on their perceptions of SOGI data.

Throughout data collection, project-related documents obtained from interviewees were collected and analyzed as complementary sources of data. Among others, we obtained the project charter document; detailed technical specifications; detailed meeting agendas and minutes from all 16 project meetings; a master spreadsheet detailing the project schedule, issues tracker, communication plan, and the data cookbook; and the actual SOGI data reports. In total, we obtained 20 internal documents totaling 124 pages. In addition, we investigated external documents referred to by interviewees, including federal and state reporting requirements, public policy statements, SOGI practices followed by other universities, and user manual documentation from the software vendor (11 documents totaling 422 pages). We also directly observed three project

meetings in early 2022 concerning a related personal data collection project in which the EBD project team members were involved. Finally, the first author complemented and triangulated the data by maintaining regular personal communication with two key informants in high-level administrative positions, who had a combined total of 30 years of institutional knowledge and were involved in the EBD project team meetings, participated in decision-making, and attempted to use SOGI data. Crucially, one of these informants participated in all 16 EBD project meetings and provided additional context, explanations, and insights beyond the documented meeting minutes. Communication included informal conversations, email exchanges, and feedback on the draft findings of the study. As per Yin’s suggested approach (2018), the five types of data we collected were kept in a case study database by the first author within ATLAS.ti as a data management tool. All evidence is summarized in Table 1 below.

Data Analysis

We conducted an inductive, interpretive (Sarker et al., 2018) analysis of data collected while following the three-step process recommended by Miles and Huberman (1994). During phase one—data reduction—the first author carefully read and reflected on each piece of data, highlighting sections of text within and across transcripts that related to the research question, drawing on recognition theory as a lens (Sarker et al., 2018). In phase one, it transpired that recognition in SOGI data concerned four elements that were prominently featured in the data: layered identities, non-binary gender, plural genders, and gender fluidity. Therefore, the data were coded according to how these facets of recognition were built into data semantics and into data formats. Specifically, the data were coded based on the way layered, non-binary, plural, and fluid gender identity were reflected in (1) data semantics and (2) data formats. For example, an interview transcript stating “the original implementation committee that included myself and the current director of the Queer Center determined that ... definitions are fluid and evolve over time, and they are very personal” (Domain Expert, Interview) was coded as “domain expert asserting gender fluidity,” and “if someone changes their answer, does this need to be tracked? 1. No” (Internal Document) was coded as “technical expert seeking clarification on recording gender fluidity in data.” Through this process, we traced the four facets of gender.

important element of research design as it engaged us more closely with the subject matter and allowed us to take active steps against social inequality and injustice. We provide details and reflections on this process in Appendix 1.

³ In line with the aim to not only research but also give visibility to the people studied, we developed citation values and aspirational citation metrics to consciously increase the number of sexual and gender minority scholars cited (Belcher, 2019; D’Ignazio & Klein, 2020; Emerson & Lehman, 2022; Gray, 2015; Schucan Bird & Pitman, 2020). This was an

Table 1. Summary of Data Collected				
Type of data	Quantity of data			Period of data collection
Narrative interviews	20 interviews in total with project team members (domain and IT experts), data beneficiaries (students and staff), and the software vendor. Approximately 19 hours of audio data, and 226 pages of transcripts.			June 2021- November 2022
	Interviewee roles (identifiers in findings)	Unique individuals	Number of interviews	
	Student (Student Beneficiary, Interview)	3	3	
	Staff and faculty (Staff Beneficiary, Interview)	2	2	
	Technical staff (Technical Expert, Group Interview)	4	1	
	Software vendor (Vendor, Interview)	1	1	
	Technical project member (Technical Expert, Interview)	5	5	
	Stakeholder project member (Domain Expert, Interview)	5	8	
	Totals	20	20	
Direct observation of meetings	3 project meetings. Approximately 5 hours of audio data and 100 pages of transcripts. (Technical Expert, Meeting or Domain Expert, Meeting)			October 2021- January 2022
Internal documentation	20 documents in total, including project charter, technical specifications, meeting agendas and minutes from all 16 project meetings, a master spreadsheet detailing the project schedule, issues tracker, communication plan, and the data cookbook, SOGI data reports. Approximately 124 pages. (Internal Document)			June 2021- January 2022
External documentation	11 documents in total, including government bills, regulations, user manuals, policy recommendations, best-practice guidelines from another university. Approximately 422 pages in total. (External Document)			June 2021- January 2022
Personal communication	30 years of combined experience of two key informants in high-level administrative positions, accessed as and when needed by the first author through personal communication, including informal conversations and interactions, email exchanges, and drafting the findings of the study.			Throughout field study

Next, we moved from our preliminary list of codes to the data display phase, where connections between the data were made and the initial set of codes was reduced to a smaller number of first-order themes (Miles & Huberman, 1994). For example, all codes referring to domain experts providing (or not) recognition to the aspects of gender (such as the one cited above) were identified with the broader theme of *gender fluidity—data semantics*. Conversely, all codes referring to technical experts providing (or not) recognition to the aspects of gender (including the one cited above) were grouped under *gender fluidity—data formats*. We began to chart data semantics and data formats in order to unpack their impacts on recognition. In the process, we identified situations where recognition in data was not achieved, which is what led us to coin the term *data warping*. We further identified four configurations of these two aspects with

differing impacts on recognition in data, depending on the actors involved, their objectives and intentions, and the scopes of influence (see Table 3 for details of findings).

Finally, during the conclusion-drawing and verification phase (Miles & Huberman, 1994), because of the inductive nature of our study, we oscillated between data analysis and theory construction. This led us to the identification of the two forms of recognition and the theorization of recognition concessions needed to achieve recognition in data. Additional research design details are included in the transparency materials for this study, including the Institutional Review Board materials, interview guides, data structure with codes, additional quotes, and the project timeline (see <https://doi.org/10.17605/OSF.IO/VN7R8>).

Findings

Toward the Recognition of LGBTQ+ Individuals

For over two decades, StateU engaged in a number of initiatives to create a welcoming environment for LGBTQ+ individuals, including creating an internal Commission on Gender Equality, establishing the Queer Center, offering transgender healthcare to students, and developing a number of processes and procedures to improve opportunities and outcomes. For a number of years, the Commission on Gender Equality had advocated for the need to collect SOGI data to support these efforts. This work culminated in a memorandum and proposal submitted to the Registrar's Office in early 2015 regarding changes to data collection that would involve the collection of data on the legal sex and separately on the gender identity of students, staff, and faculty, as well as the additional collection of data on sexual orientation. "It can be very affirming to see yourself represented in data, there is a sense of legitimizing your identity and that does have value" (Staff Beneficiary, Interview), argued beneficiaries campaigning for the collection of SOGI data.

In early 2016, the state where StateU is located introduced a new bill requiring all public universities in the state to allow students, staff, and faculty to disclose sexual orientation and gender identity designations on forms and in systems used to collect demographic data. While the state has been known for its welcoming attitude towards LGBTQ+ individuals, it was long felt by the local population that more could be done. The intentions behind the bill were clear:

The impetus of this bill is not simply to create a mechanism to gather information on LGBTQIA+ students and employees on [the state's] university campuses. But rather, the goal is to create a system that focuses on recognition, respect, and safety for all students. By allowing the voluntary disclosure of sexual orientation [and gender identity], previously invisible students may literally be counted if they choose to be. The data collected will provide a more accurate representation of students and employees in a meaningful way. (External Document)

As legislators perceived the complexity and novelty of the data, they established a work group to settle the details of data collection and reporting. The work group included representatives from every university in the state, including StateU. The work group met several times to discuss, among other things, "the system modifications necessary to

accommodate the questions [about sexual orientation and gender identity] and protect the privacy of the data" (External Document) and to create "recommendations presented in the report for rules to govern the common format and timeframes for collecting and reporting data" (External Document). The main discussion point across the meetings was "the most appropriate phrasing of the questions" (External Document), and the work group arrived at a set of proposed questions (to become data collection fields) and answers (to become response options in the data collection fields). Not all discussions led to consensus. For example, the work group did not agree on the recommended method of reporting, that is, student-record level or aggregate reporting.

The implementation of the bill coincided with the publication of a "Dear Colleague" letter⁴ on transgender students issued by the U.S. Department of Justice and the U.S. Department of Education on May 13, 2016. Focusing on civil rights protections for transgender students, the letter offered significant guidance for educational institutions on a number of issues, including privacy and protection of personally identifiable information held in education records. Obligated to keep students' transgender status confidential, which could include managing data about individuals' legal sex or names that differ from their affirmed gender identity or lived name, institutions across the U.S. scrutinized their systems and procedures to ensure compliance.

StateU faced pressure to collect SOGI data from three directions: its internal bodies, state legislature, and federal guidance. It responded to these pressures by launching its Expanded Bio-Demographic Data (EBD) project in January 2016. The project's high profile on the StateU campus reflected the hopes placed in SOGI data. The recognition of individuals' sexual orientation and gender identity in data was perceived as a necessary first step to develop "the ability to respond to increasing needs for resources" (External Document) and specifically to "help [StateU] provide additional resources and support members of the community" (Internal Document). Above all, however, StateU "wanted students to feel seen and validated, to see them for their full diversity. We wanted to validate and affirm their identities" (Domain Expert, Interview). For beneficiaries, "that in itself is sort of a political statement, kind of allowing people to exist" (Student Beneficiary, Interview). The importance of collecting data in a way that recognizes individuals' identities was "really important for beginning to tell a story, and collecting data, you're potentially collecting power and in that there is power in numbers. And numbers do tell a story" (Domain Expert, Interview).

⁴ Available at: <https://www2.ed.gov/about/offices/list/ocr/letters/colleague-201605-title-ix-transgender.pdf> (accessed on May 10, 2021)

The ambitions for providing recognition guided the EBD project team in its work for nearly six months, up to August 1, 2016, when the web application to collect SOGI data was launched and made available to beneficiaries on campus. However, the new application was not met with universal acclaim. For example, students were disappointed by what they perceived as limited and somewhat arbitrary available options: “I really, really emphasize the importance of being able to identify oneself, however you would like to. And I think that these demographics [currently not included] have a right to be represented” (Student Beneficiary, Interview). The inability to select more than one option was also raised: “I almost feel like checkboxes would make more sense. So then you can select multiple, because you could be a man and transgender” (Student Beneficiary, Interview). The way SOGI data were collected “leaves [out] room for complete autonomy and complete self-determination” (Student Beneficiary, Interview) because of prescribed options. In sum, a number of beneficiaries expressed their doubts as to whether they were recognized in the data and whether this attempt at creating recognition was actually beneficial to them. One student, when asked about their experience with the new application, stated that “there’s a lot of big talk with not a lot of action” (Student Beneficiary, Interview) with “the amount of performativity that happens” (Student Beneficiary, Interview). “Are we doing this just to capitalize on diversity and say we’re being really inclusive when we’re not? Is it just [StateU]’s way of saying oh, we’re such an inclusive school because we have all these queer people on campus?” (Student Beneficiary, Interview), wondered another student.

A group of employees was also concerned about the risks of collecting these highly sensitive data. There was some perception that these constituted a “risky set of data to collect” (Technical Expert, Group Interview) due to a decision to associate SOGI data with StateU user ID numbers. More clarification was needed on “why we’re collecting this data” (Technical Expert, Group Interview). Some students felt at risk: “This is a list of all the trans students, but also ‘this is a list of all trans students’ [negative emphasis], it’s nice because it represents me, but also ‘it represents me’ [negative emphasis]” (Student Beneficiary, Interview).

Finally, there was a sense at StateU that, while the EBD project delivered the SOGI data application on time and within budget, some of the decisions made as part of the project to a certain extent eroded the mission and vision of providing recognition of gender identity and sexual orientation in SOGI data. As one of the EBD project team members talking about lessons learned reflected:

I know from [the EBD project] that we made some decisions early on that ended up being short sighted. ... So we’d love to learn from that and avoid it even though I do want to acknowledge that I understand that creating data [formats] in this way presents difficulty. But not doing it right in the first place, not doing it in a way that responds to the community’s needs and asks can be extremely negative. (Domain Expert, Meeting)

What happened during the EBD project that resulted in the ultimate perceived lack of recognition in data? Below we analyze the recognition of gender identity and sexual orientation in SOGI data at StateU.

Recognition in SOGI Data at StateU

The EBD project team was in charge of conceptualizing, developing, and implementing the SOGI data solution at StateU. In one of the first project meetings, the team conceived a web application with the aim of asking about legal sex, gender identity, and sexual orientation on one page, with a range of response options available (see Figure 1 for a representation of the application that was ultimately delivered). Such a solution required more work on the part of the project team, as it could not rely on the baseline version of its MIS and instead had to build a custom application. This challenge arose early in the project and made the team aware of the amount of work that needed to be completed in about six months.

While domain experts were in charge of specifying the new data points, technical experts focused

on the technical aspects of what changes need to be made to what tables, what data is going to be collected, who has permissions We were definitely not the driving factor of these decisions, only when we needed to weigh in on technical limitations, but usually leaning on [domain experts] to inform us of what’s needed. And then we might tweak that based on what’s technically possible. (Technical, Interview)

This dynamic was set early on in the project, whereby domain experts would work out the semantics of SOGI data, and then hear from the technical experts on the technical feasibility of the semantics in data formats, along with suggested modifications. Below, we show the details of how domain experts worked on providing recognition through data semantics while technical project members focused on providing recognition through data formats. Specifically, we analyze data semantics and data formats in terms of the four facets of gender identity and sexual orientation that emerged in the project: layered identities, non-binary gender, plural genders, and gender fluidity.

Figure 1. StateU’s Web Application to Disclose Legal Sex, Gender Identity, and Sexual Orientation Data

Table 2. Functional Definitions of SOGI Data at StateU (Internal Document)

Term	Functional definition
Legal sex designation	The sex marker that is on government-issued identification such as SS card, passport, or drivers license. Self-reported on admissions applications for students (all Admission and Non-Degree applications). Designation is verified for employees.
Gender identity	The internal sense of one's gender that may be different than their sex assigned at birth. Typically used with reference to social and cultural differences rather than biological ones. The behavioral, cultural, or psychological traits typically associated with one sex.
Sexual orientation	An individual's sexual identity, which may include the gender(s) to which they are attracted. Self-reported on all forms where biographical data (including gender, race, or ethnicity) are included.

Layered Identities

Recognition through data semantics: From the outset, the project charter acknowledged the role of the internal Commission on Gender Equality and the state bill in recognizing legal sex, gender identity, and sexual orientation as three separate data fields that, taken together, can build a more accurate picture of an individual’s identity. The Commission, in its memorandum, advocated for the addition of SOGI data, since just collecting legal sex was not a complete way to recognize identity. This was echoed by the state bill work group, which in one of its first recommendations “determined that sexual orientation data must be defined as sexual orientation, gender, and legal sex” (External Document). In further justifications, the work group report stated that

a single question ... is not sufficient to adequately represent the intent of the legislation, nor the makeup of the student body and faculty/staff. It was agreed that the

standard gender question asked on most applications should be expanded to incorporate more inclusive and accurate indicators of gender and gender identity. (External Document)

Recognition, as championed by the Commission on Gender Equality and the state bill work group, amounted to recognizing that an individual’s identity is composed of three layers: legal sex, gender identity, and sexual orientation. Domain experts on the EBD project were committed to the recognition of layered identities, as moving from one data field to three would reveal the full extent of diversity among beneficiaries, and they insisted on ensuring that the semantics of SOGI data reflected this. For example, they developed clear and concise functional definitions of SOGI data (Table 2) for the project’s “data cookbook” (Internal Document) that showed the project team how the three layers are distinct but need to be considered together.

Further, they drafted and refined a website that accompanied the launch of the application and included a clear specification of the three layers of identity. On the website, domain experts included a sentence stating that StateU “recognizes that gender is not limited to, nor always congruent with, the categories available in legal sex designation” (Internal Document), explaining why data on all three aspects were needed. For domain experts, data about legal sex, gender identity, and sexual orientation were complementary in allowing individuals self-identification and recognition. In other words, any of the three data points considered separately from the other two did not allow for complete recognition. These were the data semantics that domain experts insisted on in order to recognize layered identities in SOGI data.

Recognition through data formats: The task for technical experts was to figure out how to reflect the data semantics of layered identities established by domain experts in data formats. Technical experts reviewed various options to format SOGI data and presented them in a project meeting. At first, they suggested deploying unused columns in an existing table for biographic information, but this was disregarded by domain experts since the columns only allowed binary answers without any write-in responses or multiple-choice options. Instead, the team settled, upon the suggestion of another technical expert, on creating a new biographic data table that would import legal sex data from an existing table and store SOGI data in two additional tables, constituting a nested table format (see Figure 2). After this initial discussion, a technical expert was asked to work with a colleague to develop the technical specifications for the new tables and present data formats at the next meeting.

In the meeting, the nested table format was presented as a way to provide recognition of layered identities in the database. Technical experts explained that this format would allow data about an individual’s legal sex to be stored together with data about their gender identity and sexual orientation, thus conforming to the desire for recognition of layered identities. Despite this solution entailing more work for technical experts, the overall EBD project team decided that the nested table format was appropriate for ensuring the recognition of layered identities at the level required by domain experts. With SOGI data storage worked out, the EBD project team focused on how to ensure that the right data are disseminated through various systems. Technical experts were concerned that there were many other systems at StateU that relied on legal sex data and that the creation of the nested table format would both require updates to data feeds and integrations and present a good opportunity to review which systems needed the data.

Continuing to rely on legal sex data only was perceived by domain experts as going against the very recognition they were striving for; therefore, they requested that technical

experts compile a spreadsheet of all systems and reports pulling legal sex data. For example, class rosters routinely pulled legal sex data from the MIS, and StateU’s internal analytics software was configured to generate numerous reports based on legal sex. Over several meetings, the 32 identified systems were reviewed, with several decisions made to minimize the reliance on legal sex data in favor of gender identity data, wherein domain experts instructed their technical colleagues to “remove [legal sex] from all class rosters, make a new report [using gender identity data] for analytics” (Internal Document). In some cases, e.g., student profiles for advising teams, gender identity data were used instead. In the majority of cases, neither legal sex nor gender identity data were deemed necessary; therefore, “we should do our best to remove it [legal sex] from any reports that have it and do not need it, starting with public reports” (Internal Document). The team also discussed the need to report SOGI data, but a conclusion was noted that “right now we cannot think of any specific reports that need to be created, but we should bring the data into [the reporting system] so it’s ready when the need arises” (Internal Document).

Prioritizing data semantics for recognition: Paradoxically, the decision to remove legal sex and gender identity data from most systems and reports was presented by domain experts as ensuring the recognition of layered identities: they made a “conscious decision not to show the [gender identity] table because of its nature” (Stakeholder, Interview). Domain experts were concerned that individuals had complex identities, parts of which they may want to reveal (or not) themselves, depending on the context and system they were interacting with. Fully recognizing this complexity, in the view of domain experts, meant giving individuals the capacity for “self-authorship” (Domain Expert, Interview), and disseminating data about their legal sex, gender identity, and sexual orientation by default went against this value. Domain experts’ insistence on fully recognizing the many layers that an individual’s identity could have had implications for recognition through data formats. On the one hand, SOGI data were collected on the same page in a web application and appropriately stored in a nested table format (reflecting the recognition of layered identities). On the other hand, the level of protection and access limitations imposed on the new tables sparked discussion among the project members about a secret table or a “behind-the-scenes table that no one can see” (Domain Expert, Interview), with frequent references made to the SOGI data being “locked down” (Domain Expert, Interview), hampering the recognition of layered identities in wider StateU systems. Data semantics for recognition were, in this case, prioritized, as the application to collect SOGI data gave beneficiaries a sense of recognition with its interface and accompanying descriptions recognizing layered identities, while wider data dissemination into StateU systems was barred. Recognition through data formats was deprioritized.

<p>zgbbioid -- Biographic Data table</p> <ul style="list-style-type: none"> a. zgbbioid_pidm number (8) b. Zgbbioid_aidm number (8) c. zgbbioid_gi_code varchar2 (char 1) d. zgbbioid_gi_other varchar2 (char 30) <ul style="list-style-type: none"> i. populated if zgbbioid_gi_code = 'OR' ("identity or identities not listed") e. zgbbioid_so_code varchar2 (char 2) f. zgbbioid_so_other varchar2 (char 30) g. Zgbbioid_activity_date date h. Zgbbioid_id number(19) i. Zgbbioid_version number(19) 	
<p>zgtvbdgi -- Bio Data Gender Identity Validation Table</p> <ul style="list-style-type: none"> a. zgtvbdgi_gi_code varchar2 (char 2) <ul style="list-style-type: none"> i. 'AG' = Agender ii. 'GQ' = Genderqueer iii. 'MN' = Man iv. 'NB' = Non-Binary, including gender fluid and gender non-conforming v. 'QU' = Questioning or Unsure vi. 'TM' = Trans Man vii. 'TW' = Trans Woman viii. 'TG' = Transgender ix. 'WM' = Woman x. 'OR' = Identity or Identities not listed (Please specify) xi. 'NR' = Prefer Not to Answer b. Zgtvbdgi_description varchar2(30) c. zgtvbdgi_eff_date d. zgtvbdgi_nchg_date e. Zgtvbdgi_activity_date f. Zgtvbdgi_id number(19) g. Zgtvbdgi_version number(19) 	<p>zgtvbdso -- Bio Data Sexual Orientation Validation Table</p> <ul style="list-style-type: none"> a. zgtvbdso_so_code varchar2 (char 2) <ul style="list-style-type: none"> i. 'AS' = Asexual ii. 'BS' = Bisexual iii. 'GA' = Gay iv. 'HS' = Heterosexual/Straight v. 'LE' = Lesbian vi. 'PS' = Pansexual vii. 'QR' = Queer viii. 'QU' = Questioning or Unsure ix. 'SG' = Same Gender Loving x. 'OR' = Identity or Identities not listed (Please specify) xi. 'NR' = Prefer Not to Answer b. zgtvbdso_description varchar2(30) c. zgtvbdso_eff_date d. zgtvbdso_nchg_date e. zgtvbdso_activity_date f. Zgtvbdso_id number(19) g. Zgtvbdso_version number(19)

Figure 2. Biographic Data Table (Top) with the Gender Identity Table (Bottom Left) and Sexual Orientation Table (Bottom Right)

Yet, as a result, StateU was not able to provide full recognition to its LGBTQ+ population. The reports and dashboards that included data on legal sex were devoid of SOGI data, depriving their users of insights into the actual LGBTQ+ populations that StateU wanted to serve. Relying only on legal sex data (limited to “male,” “female,” or unknown), StateU users, for example department chairs, could not recognize transgender or non-binary individuals in their communications or even their department statistics. For all intents and purposes, identities were still equated to legal sex, which nullified the lived experience of many StateU students, staff, and faculty. This deficit in recognition was further aggravated by the fact that beneficiaries could disclose SOGI data, hinting towards the recognition of layered identities, which created an unmet expectation that StateU would recognize them as such.

Non-binary Gender

Recognition through data semantics: The state bill work group stated clearly in its report that “gender can no longer be limited to the traditional binary choices of male and female” (External Document). Developing expansive systems for SOGI data was intended to illustrate “the point that sexuality

is a spectrum with each point on the spectrum worthy of notice and recognition” (External Document). At the same time, there was a movement in the state to add “X” as the official non-binary legal sex designation on state ID cards. Domain experts on the EBD project team were strongly informed by the idea of non-binary gender and advocated for the inclusion of a number of gender identity options in the field: “I remember there was definitely a lot of conversation about what selection options we should provide in the validation table, and we definitely look to [the Commission on Gender Equity], [the Queer Center], our stakeholders to define those” (Technical Expert, Interview). In fact, one of the domain experts on the EBD project was involved in the state bill work group and kept the project team informed about the work surrounding data semantics. The work group engaged in extensive discussions on this point: “There was significant discussion regarding the most appropriate phrasing of the questions and whether or not institutions would have the expertise or capacity to ensure that the proposed language is inclusive and free of any potentially offensive phrasing” (External Document). Finally, the work group put together proposed questions and response options for SOGI data fields which were recommended to all institutions. StateU also created a glossary of gender identity and sexual orientation

terms that was published on its website (Internal Document). As a result, beneficiaries were glad that “they [StateU] recognize it’s more than just two or three options, where it’s female, male, and other, or just female and male. So I’m glad that there were more options” (Student Beneficiary, Interview). Working out the language around gender identity became an important part of the domain experts’ role. Following the report of the work group, domain experts advocated in project meetings for including the proposed 11 different response options for gender identity and sexual orientation (see Figure 2). They also insisted on the need to build the capacity to edit and add response options over time as other non-binary genders evolve. In this way, domain experts wanted to ensure that data semantics would contribute to the recognition of non-binary gender.

Recognition through data formats: Both gender identity and sexual orientation responses entered by beneficiaries were to be mapped in a validation table and recorded appropriately as a variable character field (varchar) of 2 characters’ length (see Figure 2), as designed by technical experts. This allowed a possible expansion of the values available, i.e., adjusting existing SOGI data values as well as adding new ones, in line with the desired data semantics. Technical experts considered the inclusion of response options fairly straightforward to implement once domain experts decided on them. The developer responsible for the new application recalled:

I remember when I was developing, I was just like: for the options, you could pick a color, could be blue, white, and red. And at some point, somebody went in and defined what those options could actually be. But I wasn’t involved in that at all. I made a place where they could define what could be selected, but the actual options were outside of the scope of what I was working on. (Technical Expert, Interview)

For example, in one of the meetings, domain experts decided to “remove ‘Gender non-conforming’ since this is already included in the ‘Non-binary’ choice” (Internal Document), and this change was easily implemented by the developer. Mapping the various responses to values in a validation table was meant to ensure that all non-binary genders could be captured in the database and that each individual’s data would be properly recorded. Recognition through data formats was, initially, seen as easy to achieve, and the implementation of response options was quickly resolved.

Prioritizing data formats for recognition: A few months into the project, however, some domain experts began to express doubts about the solution developed for non-binary gender. It was best captured retrospectively by one domain expert:

that’s something that I just have an issue with in general, but in terms of [SOGI data], it would be cool to have a field where folks could write in what they like. It gets more complicated about data collection, but to have an open text field feels important. Knowing that we don’t have a full list, or people use words that we might not even know about. So I’d love to see an open text field. (Domain Expert, Interview)

Similar concerns were expressed by students: “When you have a list to choose from, it puts things in neater categories for data collection, [but] if I wanted to say I identify as trans and gender queer, when I’m asked to explain further, I’m like, you know, some sort of third, or in-between, or other gender” (Student Beneficiary, Interview). Mapping response options to the variable character format in a validation table was not seen as sufficient, because the options available still binarized identities into being one thing but not another, with some controversy over the separation of “man” from “trans man.” Further, a drop-down list of choices created a sense of a hierarchy of options and impacted readability.

In other words, domain experts began to realize that due recognition of non-binary genders would involve having a write-in text box instead of a drop-down list of response options. In a crucial meeting regarding this aspect, one of the domain experts raised this question, but technical colleagues were quick to explain the challenge this would pose for collecting, analyzing, and disseminating the data. One of the senior technical experts explained that a write-in text box would require a change of the data format to a string, which would make it virtually impossible to actually access and report on the data. This would, in turn, mean that individuals would be illegible to the application and wider StateU systems, as string data could not be disseminated and consumed by other systems and reports. It was suggested that if a write-in text box was used, the majority of systems pulling gender data would not be able to work with string data and would default to legal sex instead. Prioritizing data formats for recognition, in this case, was necessary to actually recognize the existence of non-binary genders in SOGI data.

Plural Genders

Recognition through data semantics: Domain experts on the EBD project strongly believed that an individual may identify with more than one gender identity or sexual orientation at any given time. For them, recognizing the plurality of genders was essential, in response to the words of actual beneficiaries who expressed the need to identify with a number of different options. This need was also spelled out in the recommendation of the internal Commission on Gender Equality and indicated in the state bill and the report of the

work group (“per [the Commission’s] proposal and [bill] doc, one may select more than one [gender identity]” (Internal Document)). Domain experts therefore equated recognition with plural genders—that is, the possibility of selecting more than one gender identity or sexual orientation response option—and worked to imbue the data with these semantics.

Recognition through data formats: A few weeks into the project, when the application was first demonstrated to domain experts in a project meeting, they were concerned that only one response option could be selected. Domain experts insisted that beneficiaries should be able to select more than one response option. Yet they faced resistance from technical colleagues. One domain expert wanted “folks to be able to select more than one field if they identify with more than one term, but I was told we would break [MIS] if we did that” (Domain Expert, Interview). Technical experts explained that it was not feasible to implement multiple-choice options in SOGI data because of the knock-on effect on the other systems and reports. A senior technical expert clarified that although collecting multiple-choice responses would be possible in the database, “reporting will be skewed if we allow more than one choice” (Internal Document). Effectively, data formats did not recognize the plurality of genders.

Compromising on data semantics and formats: The issue sparked an intense debate in a project meeting. One domain expert probed: “Is it a matter of difficulty [to implement the preferred data format] or is it a matter of ‘absolutely cannot’? Is it a lot more work but still possible or ‘absolutely will fail,’ ‘cannot do’ type of thing?” (Domain Expert, Meeting). With the technical team’s insistence on the infeasibility of the solution, domain experts then proposed a write-in text box where beneficiaries could specify “Identity or identities not listed” as a way of recognizing plurality, as beneficiaries who wanted to identify as more than one identity could write both of them in: “They said they wanted another box where if they pick ‘other,’ they can type in whatever they wanted to” (Technical Expert, Interview). Similarly to the discussion on the write-in text box to deal with non-binary gender, a technical expert warned again that although the write-in box was technically possible, this solution created an obstacle to potential data analysis because it was difficult to process string data, and data cleaning was required beforehand. In the discussion that ensued, domain experts insisted on including the write-in text box as a solution to ensure recognition of the plurality of genders. Technical experts finally conceded that the use of the write-in text box would not cause major issues, and, if needed, the string data could be processed manually to be included in other systems and reports, as the usage of the box was not estimated to be significant. Both sides decided to compromise on this solution and worked out the details in a meeting. The write-in text box was implemented as a 30-

character string, with this constraint being based on the technical experts’ analysis of the capacity of other systems.

The chosen solution was deemed insufficient by beneficiaries, who emphasized that allowing multiple-choice answers was preferred. First, they explained that they either did not see the option to specify their identity because of unclear user interface design or they did not think to use it to write two options from among those provided. Second, the limit of 30 characters was not enough to capture many responses, as a number of truncated descriptions of individuals’ gender identities were identified, as summarized by a beneficiary: “If my sexual orientation or gender were not listed on there, I wouldn’t think, honestly, to put in. I might just hit prefer not to answer. Yeah, it wouldn’t include me. And 30 characters is definitely not enough. There’s just not enough room” (Staff Beneficiary, Interview). Third, data in the string format proved indeed to be very difficult to access and analyze. In fact, StateU had not looked at the data for a number of years due to a lack of resources and sanctioned uses. When the data were finally accessed more than six years after the launch of the application, it became apparent that relying on a string as the data format was inadequate. Many responses were variations of one another with alternate spellings, or indeed two identities from the existing options (Internal Document); thus, individuals who gave these responses were not recognized for their identities in the database.

In hindsight, domain experts reflected that these decisions did not account for the overlapping of individual identities, proceeding largely as if these were not data about complex, intersecting, malleable aspects of people. In retrospect, a domain expert stated that there should have been more attempts to “disrupt some ways that [the project team] thought about how data affects its intended recipients—people” (Domain Expert, Meeting). Further reflections pointed to the nature of the EBD project that made it about people, which was not fully appreciated by everybody on the team at the time: “This particular subject is something that is really tied to disrupting systems that haven’t been built for people” (Domain Expert, Meeting). SOGI data beneficiaries did not feel reflected and properly recognized in the data: “I’m a trans man but put in gender [identity data field] that I’m just a man because I am a man. It doesn’t tell the complete story of what it is. The story isn’t accurate” (Beneficiary Staff, Interview).

Gender Fluidity

Recognition through data semantics: From the outset, domain experts on the team insisted on fluidity in two forms: an individual’s gender identity and sexual orientation may change at any point in time, and the terms used to describe these two aspects of identity evolve over time as well.

Acknowledging these two elements was necessary to provide recognition through data semantics. This understanding was grounded in the state bill work group: “The original [work group] that included myself and the current director of the Queer Center determined that it [not accounting for fluidity] is detrimental to ways in which people describe themselves. Definitions are fluid and evolve over time, and they are very personal” (Domain Expert, Interview). Therefore, SOGI data had to account for the fact that gender identity and sexual orientation are fluid and can change and evolve. From the domain experts’ perspective, this meant that beneficiaries should be able to change their response to gender identity and sexual orientation at will and at any point; at the same time, a solution was needed to allow modifying the response options in the system: “It’s a constantly changing landscape of technology, and [we] want to make sure that we’re using appropriate terms” (Technical Expert, Interview). Beneficiaries demanded that “all this data should be dynamic data that can be changed and updated rather easily. Once you change it, it updates across the board really quickly” (Student Beneficiary, Interview). Nonetheless, recognizing gender fluidity meant that there should be no history of previous responses kept.

Recognition through data formats: In the case of evolving terms, the expansive design of the table described earlier made it possible to modify and add new terms as and when needed, thus allowing authorized administrators to make appropriate changes to the response options. Even during the project, domain experts identified a need to separate “Non-binary, including gender fluid and gender non-conforming” into three response options. Such changes were described as relatively simple by technical experts, who explained that the option to make changes was given to a handful of authorized users at StateU: “There’s an administrative side of the [application] where certain people would have access ..., who could go in and see all the current options, and they can edit them or add new ones, they can actually add from 15 to 20 options without needing my help” (Technical Expert, Interview).

In terms of individuals’ fluid identity, several discussions were held about data formats. First, domain experts stated that beneficiaries should be able to update their SOGI data whenever and as often as needed. This led to pushback from technical experts, who explained that databases have certain update cycles and that it would be challenging to ensure that all of them consumed the data when changed. Ultimately, technical experts audited all systems to check their update times, and domain experts were made aware of the time it took to disseminate the data across systems. Further, domain experts decided that no history of SOGI data should be kept,

and this was communicated to the technical colleague in charge: “That was actually part of the planning discussions, they [domain experts] didn’t want a history” (Technical Expert, Interview). “From a technical perspective it’s easier to just not track [history] and just whatever’s there today is what’s accurate or not, tracking [history] over time becomes challenging” (Technical Expert, Interview).

Forfeiting data semantics and formats: These decisions had conflicting impacts on recognition. Domain experts concluded that the process of adding or changing existing options was too cumbersome to be executed in practice due to the need to enter the application and possess the required skills to insert modifications. They felt that this process should have been designed in a more approachable way. In hindsight, this approach was seen as symptomatic of “patriarchal and supremacist thinking” (Domain Expert, Meeting), whereby someone in an administrative position of power decides on which options of gender identity and sexual orientation are available and allowed in the application instead of beneficiaries themselves.

The fluctuation of individuals’ identities over time was also varyingly recognized. On the one hand, not storing historical data allowed for fluidity in that it avoided assigning prominence to previous gender identities or sexual orientations while privileging an individual’s current identity, recognizing that gender can be fluid. On the other hand, without historical data showing fluidity, it was difficult to acknowledge and recognize that individuals’ identities were, in fact, fluid and could change over time. StateU had no way of identifying the fluidity of identities from the data it was storing and therefore could not recognize such identities. A domain expert acknowledged that this caused recognition to suffer:

Looking at the decision now, through the lens of data collection, research, and aggregate data usefulness, I can see that we are losing some information that could be useful for understanding these data. I can imagine how it could be useful on the aggregate level to quantify how many of our students change this information, or how often. As a new type of bio-demo data, is studying its fluidity useful? Since we are not recording changes, does that limit our understanding of it? (Domain Expert, Interview)

Thus, not using data semantics and data formats to provide recognition of individuals’ gender fluidity resulted in low overall recognition of gender fluidity. We summarize these findings in Table 3 below and discuss them in the next section.

Table 3. Summary of Recognition in SOGI Data at StateU		
	Recognition through data semantics	Recognition through data formats
Actors	<p>External to StateU: State senate, state bill work group, U.S. Department of Justice and U.S. Department of Education</p> <p>Internal at StateU: Commission on Gender Equality, Queer Center, Registrar's Office, stakeholders on the EBD project team (representatives of Office of Diversity and Inclusion, Institutional Reporting, Registrar's Office, Human Resources, Queer Center)</p>	<p>External to StateU: MIS software vendor, other software vendors</p> <p>Internal at StateU: IT employees, including analysts and developers (e.g., manager of business intelligence team, director of human resources information systems, MIS team manager), and the project manager</p>
Actors' objectives and intentions	<p>Providing due recognition for minoritized and underrepresented LGBTQ+ individuals through SOGI data collected about them "to create a system that focuses on recognition, respect, and safety" (External Document) so students feel seen and validated</p>	<p>Providing the resources and expertise "on the technical aspects of what changes need to be made to what tables, what data is going to be collected, who has permissions" (Technical Expert, Interview) to ensure feasibility of the project and integrity of the systems</p>
Scope of influence	<p>Definitions, terms, language, processes, policies, procedures, website text, requirements</p>	<p>Data formats, databases, reports, interoperability and integration with other systems, dashboards</p>
Aspect	Layered identities	
Locus of recognition	<p>Recognition through data semantics: Legal sex, gender identity, and sexual orientation data, as the layers of individuals' identities must be considered together to provide recognition of an individual's identity.</p>	<p>Recognition through data formats: Legal sex, gender identity, and sexual orientation data must be recorded in a nested table with separate tables for each value to ensure SOGI data are stored and displayed alongside legal sex data, with the required updates to data feeds and integrations to show all values.</p>
Configuration of aspects of data, resulting data warping, and its outcome	<p>Prioritizing data semantics for recognition: Displaying legal sex, gender identity, and sexual orientation data in all systems goes against the capacity for self-authorship and limits the layers through which individuals may construct their identities. Thus, data should be removed from as many systems as possible to allow individuals to construct layered identities.</p>	<p>Data warping: While SOGI data were collected on the same page as legal sex data and appropriately stored in a nested table format (recognizing layered identities), limited access and near non-use of the SOGI data effectively hampered the recognition of layered identities in StateU systems.</p> <p>Outcome: Collecting SOGI data alongside legal sex data created an expectation among beneficiaries that their layered identities would be recognized, while StateU users of SOGI data were unable to provide due recognition in StateU's communications, statistics, or reports.</p>
Aspect	Non-binary gender	
Locus of recognition	<p>Recognition through data semantics: An individual's identity lies on a spectrum outside of the binary choices of 'male' or 'female'; therefore, other response options have to be provided to ensure non-binary genders are recognized in data.</p>	<p>Recognition through data formats: Legal sex with a binary choice; gender identity and sexual orientation in a variable- character field of length 2 corresponding to 10 different options as attributes in a validation table (to ensure values can be appropriately recorded and disseminated in the database and further systems); and a 30-character length string available for write-in responses to allow for the recognition of non-binary gender.</p>

<p>Configuration of aspects of data, resulting data warping, and its outcome</p>	<p>Prioritizing data formats for recognition: The validation table ensured that the databases and systems find individuals' responses legible and can process them (e.g., count, or show in reports) which allowed the recognition of non-binary gender in the systems. This came at the expense of recognizing that individuals may not fit the response options provided.</p>	<p>Data warping: Providing recognition by mapping the responses to a validation table still binarizes identities into falling into one category but not the other, forcing beneficiaries into selecting binarizing responses.</p>	<p>Outcome: StateU was able to collect data with an expansive definition of gender that recognizes the existence of 10 genders and provides the option of write-in responses. This provides recognition of some non-binary genders, but not all, and some beneficiaries feel they are not recognized.</p>
<p>Aspect <i>Plural genders</i></p>			
<p>Locus of recognition</p>	<p>Recognition through data semantics: An individual may identify with more than one gender identity or sexual orientation at a time and thus needs to be able to select one or more response options in SOGI data.</p>		<p>Recognition through data formats: Only one response option for each data field can be selected and stored in order to ensure the data can be disseminated into further systems and reports. The write-in text box can serve as a catch-all for those with plural genders.</p>
<p>Configuration of aspects of data, resulting data warping, and its outcome</p>	<p>Compromising on data semantics and formats: Allowing less recognition through data semantics (inability to select multiple response options does not recognize plural genders) and less recognition through data formats (write-in text box to accommodate plural genders renders responses illegible to the database) resulted in a compromise solution that allowed some recognition.</p>	<p>Data warping: The plurality of genders is not fully recognized in the data, as the responses of those who opted to use the write-in text box to declare their plural genders are not accessed, analyzed, or used in any way.</p>	<p>Outcome: Individuals with plural genders are not recognized at StateU, as even the adopted write-in text box solution resulted in string data that were not usable for the organization.</p>
<p>Aspect <i>Gender fluidity</i></p>			
<p>Locus of recognition</p>	<p>Recognition through data semantics: An individual's legal sex, as well as their gender identity and sexual orientation may fluctuate and change over time, thus the need to change SOGI data at will. Gender identities, sexual orientations, and legal sex response options themselves may evolve over time.</p>		<p>Recognition through data formats: An individual's legal sex, gender identity, and sexual orientation are recorded only at a given point in time, it is not possible to record its change over time as a history table does not exist. Administrators are able to edit and add response options.</p>
<p>Configuration of aspects of data, resulting data warping, and its outcome</p>	<p>Forfeiting data semantics and formats: Deprioritizing data semantics for recognition (by not wanting to keep past responses) and deprioritizing the recognition of fluidity in data formats (no past data were recorded) resulted in an overall lower recognition of gender fluidity in data.</p>	<p>Data warping: Data collected without tracking changes only reflect the current gender identity or sexual orientation.</p>	<p>Outcome: StateU is unable to record individuals whose legal sex, gender identity, or sexual orientation change over time, denying them full recognition in terms of gender fluidity, resulting in a recognized gap in organizational data.</p>

Discussion

In this paper, we set out to uncover how data semantics and data formats are involved in providing recognition in personal data. In the case we studied, the external bodies, the organization, and those involved in creating SOGI data all shared a commitment to providing recognition to LGBTQ+ individuals through the data. Yet the data that were conceptualized as part of a well-resourced project intended to achieve greater recognition still misrecognized the intended beneficiaries. Therefore, despite the recognition-oriented intentions of the actors involved in conceptualizing data semantics and formats, personal data still ended up misrecognizing the individuals to whom they pertain, a phenomenon that we term *data warping*. Data warping, a mechanism leading to a lack of recognition, is a source of injustice in and of itself (Honneth, 2003; Taylor, 1994), denying individuals a voice, visibility, and respect. It also contributes to further injustice by impeding the distribution of resources in a more equitable manner instead of contributing to redistribution (Fraser, 2003) to support individuals with additional resources or opportunities. In summary, despite the best of intentions and conscious efforts to achieve recognition in personal data, injustice occurred as an outcome of data warping.

Our findings indicate that providing recognition in data is a complex endeavor involving diverse actors with different objectives, intentions, and varying scopes of influence. Domain experts, including representatives of the Office of Diversity and Inclusion, the Registrar's Office, and the Queer Center, had a specific understanding of recognition in mind when working to conceptualize SOGI data semantics that acknowledged layered identities, non-binary gender, plural genders, and gender fluidity. Their objectives were aligned with StateU's mission to provide due recognition for LGBTQ+ individuals, and their intentions were focused on ensuring SOGI data were a source of recognition, respect, and validation. Through their involvement in definitions, terms, language, and website text, as well as requirements, processes, policies, and procedures, domain experts worked toward what we call *social recognition*—that is, the acknowledgment of diverse individuals' intrinsic value by others who see such individuals as equal and worthy of respect. In the case of recognition in data, this entails imbuing data with semantics that allow data users to recognize others. The semantics that data carry (Alaimo & Kallinikos, 2022; Kallinikos, 2006; Kallinikos et al., 2013; Tuomi, 1999), contained in, e.g., wording, definitions, or response options, allow others, such as those who look at, analyze, process, or otherwise work with the data, to recognize, acknowledge, and respect those to whom the data pertain for the person they wish to be perceived as (Leidner & Tona, 2021).

By contrast, technical experts, including developers, the business intelligence team manager, and the project manager, worked under a different form of recognition. They saw themselves as providing the technical resources and expertise necessary to ensure the feasibility of the project and the integrity of StateU's IT systems. Their scope of influence extended to data formats, databases, reports, and integrations with other systems, and, as such, they were working toward what we term *systems recognition*, that is, the legibility of diverse individuals in systems that can register differences and diversity. For recognition in data, this means data need to take specific formats that allow them to function as carriers of semantics (Alaimo & Kallinikos, 2022; Faulkner & Runde, 2019; Hirschheim et al., 1995; Recker et al., 2021). Data formats, e.g., tables or types, allow systems, such as databases, dashboards, or reports, to record, register, and disseminate differences between individuals to whom the data pertain. In other words, systems recognition, through data formats, ensures that individuals are legible to the systems underpinning the circulation of personal data in organizations.

Thus, recognition in data is possible only if both social recognition and systems recognition exist. Social recognition is embedded in data semantics, and systems recognition ensures that social recognition is translated into data formats. However, the presence of data warping indicates that the coexistence of social and systems recognition is far from straightforward. In fact, these two forms of recognition in data are often at odds. For example, the social recognition of plural genders at StateU, i.e., recognition that an individual may identify with more than one gender identity or sexual orientation at a time, contrasted with systems recognition where disseminating data into wider systems and reports required that only one response option from a preset list be selected. In other words, while social recognition dictated a multiple-choice data format, opting for such a format would render individuals illegible to systems, as they would not be able to process more than one response. Therefore, overall recognition in data would not be achieved: opting for complete social recognition would entail very little to no systems recognition at all, failing to grant any recognition in data.

Recognition in data, thus, requires managing both social and systems recognition that often work in opposite directions, which we term *recognition concessions*. Recognition concessions are the trade-offs between social recognition and systems recognition that are necessary to achieve an overall state of some recognition in data that are manifested in different configurations of data semantics and data formats. In the case we studied, we observed four types of recognition concessions: (1) prioritizing social recognition through data semantics at the expense of systems recognition through data formats to achieve higher recognition in data

and (2) vice versa; (3) compromising by allowing less social recognition through semantics and less systems recognition through formats to achieve some recognition in data; and (4) forfeiting recognition, whereby deprioritizing both forms of recognition through semantics and formats results in lower overall recognition in data. Our findings indicate that regardless of the recognition concession, data warping still occurs because the trade-offs are both necessary to achieve some recognition in data and accessory to some misrecognition in data.

Aiming for recognition in data is therefore a fraught undertaking, where the degree of recognition or misrecognition achieved is a complex and dynamic condition (rather than just recognition or a lack thereof) based on recognition concessions between social and systems recognition. Recognition in data is not a finite state; rather, it constantly evolves as recognition concessions change. This is to say, recognition in data is not something that can be achieved once by an organization that wants to foster social justice through recognizing individuals; rather, it is a constantly changing, temporary accomplishment. This is because recognition concessions themselves are not permanent and change along three dimensions. First, the exact balance between social and systems recognition achieved in a recognition concession to obtain the most recognition in data possible may change as the state of recognition evolves. For example, StateU had to introduce changes to SOGI data as a non-binary legal sex designation was implemented by the state, allowing any resident to obtain an ID card with “X” as their official legal sex designation. Recognizing that legal sex can also be non-binary in data required changes to both social and systems recognition. Second, recognition concessions may also change as the understanding of recognition shifts—for example, when StateU’s stakeholders recognized that keeping historical SOGI data changes would have indeed provided more recognition in data. Finally, recognition concessions may change as the impact of the obtained state of recognition in data becomes clearer. In other words, as the consequences of recognition in data on social justice become more evident, a new state of recognition in data may be sought, with a different set of recognition concessions in place.

Theoretical Implications

Our findings have several implications for existing theory. First, we contribute to the recognition literature (Benhabib, 2018; Fraser, 1997, 2000, 2008; Fraser & Honneth, 2003), as we demonstrate “the struggle for recognition” (Honneth, 1996) in practice from the perspective of a group of actors who want to foster recognition. While the current literature focuses on arguing for the importance of recognition as a

necessary element of social justice (Fraser & Honneth, 2003) and teases out how individuals are negatively impacted by the lack of recognition when they could benefit from due recognition, political and philosophical thought does not capture the complexity involved in *providing* recognition. We show that agreeing on the need for recognition of certain individuals or populations opens up a difficult path of working toward achieving such recognition among the variety of actors, objectives, interests, and scopes of influence. Recognizing, in practice, involves making recognition concessions, and even with the best intentions, the outcomes may be fraught: constant managing of recognition may be required to ensure that due recognition is indeed provided. Further research might explore how this constant managing is approached, tracing specific impacts on recognition, as our case is limited to a snapshot of a certain state of recognition. Longitudinal studies of organizations striving for recognition over time may provide more insight into managing recognition under changing dimensions. Additionally, while our study concentrates on recognition as social justice (or, perhaps more correctly, misrecognition as social injustice) (Honneth, 2003), future studies could delve into the relationship between recognition and redistribution (Fraser, 2003). As we focus on recognition, we collected data most closely connected to the EBD project, which explains the relatively narrow scope of interview participants. Further research encompassing data beneficiaries and wider organizational actors is needed to trace subsequent redistribution.

At the same time, our paper adds an ingredient to the current theorizing of recognition, predominantly focusing on the cultural (Honneth, 2003) and political (Fraser, 2003) aspects of recognition. While the former argues that recognition is based on accepting and valuing diversity and difference, and the latter emphasizes that giving a voice and representation serves recognition, we show that the systemic aspects of recognition may be just as important. In our case, the materiality of digital technology (Kallinikos et al., 2013; Leonardi et al., 2012), manifesting itself in data formats, databases, dashboards, reports, and wider management information systems, became a force in the struggle for recognition. Bringing the systems aspect to theorizing recognition may reveal new ways in which recognition and misrecognition can be achieved in a world where recognition is increasingly mediated by data and associated digital technologies (Baskerville et al., 2020; Recker et al., 2021). Only a handful of studies have investigated this aspect so far (Waelen, 2023; Waelen & Wiczorek, 2022). Increased scrutiny of the material systems, including digital technology, can be taken up further to uncover how systems are involved with recognition, studying recognition not only by digital technology but also in digital technology. IS research traditions can further the theorization of recognition.

Second, we contribute to the growing field of data justice by bringing in IS research on data as artifacts to extend the study of data-related harm and injustice. We bring to light the phenomenon of data warping, where data injustice occurs despite the best intentions to achieve the opposite, which so far has not been studied in depth. By rooting our study in the theory of recognition, we explicate how injustice can unfold when data are first thought of, before they even begin to serve as evidence for distributing resources or before they begin to be used in decisions. In this vein, we continue and expand on the efforts of data justice scholars who begin to draw on recognition as a lens (Braun & Hummel, 2022; Cinnamon, 2020; Dencik et al., 2022), and we enrich this line of research by providing an as yet missing, sharp focus on the systems involved in data (Bruhn Jensen, 2021; Dencik et al., 2022). In short, our study shows that data-related harm and injustice can happen as early as when data semantics and formats are conceptualized, as this very process can result in misrecognition or the outright lack of recognition. This shows how IS can contribute to a better understanding of data justice and, by extension, social justice issues that can be explained through the lenses that our field can offer. At the same time, IS research can benefit from adopting recognition as a lens on social justice, especially as increasing numbers of digital technologies attempt to digitize human beings.

Importantly, our study contributes to IS literature investigating data as human-made artifacts (Aaltonen et al., 2021; Kallinikos et al., 2013) and specifically unpacks the relationship between data semantics and formats: both semantics and formats are in a mutually constitutive relationship that impacts what data can ultimately convey. While the prior literature acknowledges that data are made up of the semantics they carry and the formats they are put into (Aaltonen & Penttinen, 2021; Alaimo & Kallinikos, 2022; Beynon-Davies & Wang, 2019; Eriksson & Ågerfalk, 2010; Faulkner & Runde, 2019; Hirschheim et al., 1995; Tuomi, 1999), we provide a detailed analysis of the emergence of both semantics and formats of data in practice. Our study also adds to the literature investigating the impact of data formats on organizational realities (Beynon-Davies, 2016; Bowen et al., 2006; Eriksson & Ågerfalk, 2010; Tempini, 2020) by theorizing this impact on data themselves: formats contribute to the constitution of organizational facts not only by affecting what operations can be carried out on data by their organizational users but also at a more foundational level by shaping what can be contained in data. Of course, our study is limited by its context of highly novel and, to a certain extent, sensitive SOGI data. This type of data provided us with a setting that made the decisions around data semantics and formats particularly visible; however, we agree that other contexts may become sites of research to further our contribution.

Finally, by bringing recognition as a lens to IS research, we provide a sharp focus on personal data. While personal data have an inherent link to human dignity (Leidner & Tona, 2021), they are rarely recognized and studied in IS as intrinsically connected with issues pertaining to social justice. Our findings demonstrate the difficulties in trying to capture individual identities or preferences in personal data. People's identities (not only gender but also racial and ethnic, as well as their preferences, likings, needs, and perceptions of self) are layered, non-binary, plural, and fluid, and for this reason, not straightforward to capture in data. Transporting the known approaches and theories about data to personal data without accounting for their personal aspect does not allow for a full appreciation of the complexity of this context. While we studied the conceptualization of personal data—and, thus, the evidence we collected is focused deliberately on this aspect—further research might investigate personal data through the lens of recognition by expanding the scope more broadly to data beneficiaries.

Practical Implications

Based on our findings, organizations and actors involved in similar laudable IS projects that involve personal data should start from the understanding that providing full recognition in data is an unachievable and, as such, inviable goal. Rather, accepting the existence of trade-offs between social and systems recognition allows for the framing of the project objectives in terms of the balance sought. The following guidelines are useful discussion points for data conceptualization projects.

First, projects should not be based on good intentions, but rather on carefully assessed impacts (see also Wagner & Piccoli, 2007). If the overall impact on the intended data beneficiaries is unlikely to be net positive, it is worthwhile to carefully consider whether to proceed with the project. Second, as recognition in data requires managing both social and systems recognition, which often require trade-offs and sometimes work in opposite directions, project teams should discuss the recognition concessions they will need to manage and their outcomes for the intended beneficiaries. Third, projects should ensure that all stakeholders take seriously the materiality of digital technology and the systemic aspects of recognition at play: seemingly minute and detailed technical choices will have consequences for the overall outcome. Data-related harm and injustice can happen as early as when data semantics and formats are conceptualized, and this very process can result in misrecognition or an outright lack of recognition. Fourth, recognition in data is not something that can be achieved once by an organization that wants to foster

social justice through recognizing individuals. Rather, it is a constantly changing, temporary accomplishment. Therefore, organizations need to be prepared to iterate over time and dedicate resources to ongoing work. Constant managing may be required to ensure that due recognition is indeed continuously given.

Conclusion

In 2021, the U.S. government published the first-ever National Strategy on Gender Equity and Equality extending to “LGBTQI+ persons” in order to strive for a world in which equal opportunity is afforded to all people, regardless of gender or any other factor. With respect to gender, 12% of people in the 18-34 age group in the U.S. “do not identify with the sex they were assigned at birth, or their gender expression is different from conventional expectations of masculinity and femininity” (GLAAD, 2017, p. 1). The government emphasizes the need for “adequate collection of gender data,” and encourages “more analysis and dissemination of the gender data we do collect,” stating that “we will support efforts to close gender data gaps” (The White House, 2021, p. 40). This is just one example of SOGI data now gaining momentum in many parts of the world (Guyan, 2022). Yet “a society with more data about LGBTQ people is not automatically a society that is better for LGBTQ people” (Guyan, 2022, p. 32). More data, through data warping, can cause harm (Wilson, 2018), lead to additional work to ensure data are correct across multiple systems (Spiel, 2021), result in unreliable credit histories as credit rating systems fail to acknowledge the continuity in a person’s identity (Mackenzie, 2017), or deny critical healthcare when medical and insurance systems do not deal properly with gender identity (Costelloe & Hepburn, 2021; Dunne et al., 2017; Faye, 2021; Whitley & Greene, 2017). The lack of adequate recognition caused by data warping can lead to the inability to seize opportunities to gain greater market share by serving the unmet needs of parts of the population: “If our products don’t [recognize non-binary gender], we will continue to leave out entire populations of people” (Bennett, 2020). Our research contributes to the intersection of data and social justice by warning against the potential for inadequate recognition of individuals in warped data and by proposing the concept of recognition concessions as an explanation for why personal data warping happens. We agree that the collection of personal data can, in many ways, support the urgent fight against social injustice, the opposite of justice, equality, diversity, and inclusion. We hope that this work, by revealing the complexity of providing recognition in data, will contribute to fostering social justice.

Acknowledgments

We would like to thank the special issue senior editors, as well as the associate editor and reviewers who provided constructive and developmental feedback on previous versions of this manuscript. We received helpful comments from several colleagues, including Aleksi Aaltonen, Cristina Alaimo, Sarah Lebovitz, Silvia Masiero, and Katherine Wyers. The paper benefitted from a presentation and comments received in a research seminar at the London School of Economics and Political Science.

References

- Aaltonen, A., Alaimo, C., & Kallinikos, J. (2021). The making of data commodities: data analytics as an embedded process. *Journal of Management Information Systems*, 38(2), 401-429. <https://doi.org/10.1080/07421222.2021.1912928>
- Aaltonen, A., & Penttinen, E. (2021). What makes data possible? A sociotechnical view on structured data innovations. In *Proceedings of the 54th Hawaii International Conference on System Sciences* (pp. 5922-5931). <https://doi.org/10.24251/hicss.2021.716>
- Alaimo, C., & Kallinikos, J. (2022). Organizations decentered: Data objects, technology and knowledge. *Organization Science*, 33(1), 19-37. <https://doi.org/10.1287/orsc.2021.1552>
- Baskerville, R. L., Myers, M. D., & Yoo, Y. (2020). Digital first: The ontological reversal and new challenges for information systems research. *MIS Quarterly*, 44(2), 509-523. <https://doi.org/10.25300/MISQ/2020/14418>
- Bauer, M. (1996). *The narrative interview: Comments on a technique for qualitative data collection* (Papers in social research methods—Qualitative series). London School of Economics & Political Science, Methodology Institute.
- Bauer, M., & Gaskell, G. (2007). *Qualitative researching with text, image and sound: a practical handbook*. SAGE.
- Belcher, W. L. (2019). *Writing your journal article in twelve weeks: A guide to academic publishing success*. University of Chicago Press.
- Bendl, R., Fleischmann, A., & Hofmann, R. (2009). Queer theory and diversity management: Reading codes of conduct from a queer perspective. *Journal of Management and Organization*, 15(5), 625-638. <https://doi.org/10.5172/jmo.15.5.625>
- Bendl, R., & Hofmann, R. (2015). Queer perspectives fuelling diversity management discourse. In R. Bendl, I. Bleijenbergh, E. Henttonen, Elina, & A. J. Mills (Eds.), *The Oxford handbook of diversity in organizations* (pp. 195-217). Oxford University Press.
- Benhabib, S. (2018). From redistribution to recognition? The paradigm change of contemporary politics. In *The claims of culture* (pp. 49-81). Princeton University Press.
- Bennett, M. (2020). *Beyond the binary: 5 steps to designing gender inclusive fields in your product*. UX Collective. <https://uxdesign.cc/beyond-the-binary-5-steps-to-designing-gender-inclusive-fields-in-your-product-ff9230337b4f>
- Bera, P., Soffer, P., & Parsons, J. (2019). Using eye tracking to expose cognitive processes in understanding conceptual models. *MIS Quarterly*, 43(4), 1105-1126. <https://doi.org/10.25300/MISQ/2019/14163>
- Beynon-Davies, P. (2016). Information and organization instituting facts: Data structures and institutional order. *Information and*

- Organization*, 26(1-2), 28-44. <https://doi.org/10.1016/j.infoandorg.2016.04.001>
- Beynon-Davies, P., & Wang, Y. (2019). Deconstructing information sharing. *Journal of the Association for Information Systems*, 20(4), 476-498. <https://doi.org/10.17705/1.jais.00541>
- Bowen, P. L., O'Farrell, R., & Rohde, F. (2006). Analysis of competing data structures: Does ontological clarity produce better end user query performance. *Journal of the Association for Information Systems*, 7(8), 514-544. <https://doi.org/10.17705/1.jais.00098>
- Braun, M., & Hummel, P. (2022). Data justice and data solidarity. *Patterns*, 3(3), Article 100427. <https://doi.org/10.1016/j.patter.2021.100427>
- Bruhn Jensen, K. (2021). *A theory of communication and justice*. Routledge.
- Chua, C., Indulska, M., Lukyanenko, R., Maass, W., & Storey, V. C. (2022). MISQ research curation on data management. *MIS Quarterly*. <https://www.misqresearchcurations.org/blog/2022/2/11/data-management>
- Cinnamon, J. (2017). Social injustice in surveillance capitalism. *Surveillance and Society*, 15(5), 609-625. <https://doi.org/10.24908/ss.v15i5.6433>
- Cinnamon, J. (2020). Data inequalities and why they matter for development. *Information Technology for Development*, 26(2), 214-233. <https://doi.org/10.1080/02681102.2019.1650244>
- Costelloe, S. J., & Hepburn, S. (2021). Management of transgender patients in Laboratory Information Management Systems—Moving on from binary and ternary logic. *Annals of Clinical Biochemistry*, 58(4), 264-266. <https://doi.org/10.1177/0004563220984825>
- D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.
- Dencik, L., Hintz, A., Redden, J., & Treré, E. (2022). *Data justice*. SAGE.
- Dencik, Lina, Hintz, A., Redden, J., & Treré, E. (2019). Exploring data justice: Conceptions, applications and directions. *Information Communication and Society*, 22(7), 873-881. <https://doi.org/10.1080/1369118X.2019.1606268>
- Dhoest, A., & Szulc, L. (2016). Navigating online selves: Social, cultural, and material contexts of social media use by diasporic gay men. *Social Media and Society*, 2(4). <https://doi.org/10.1177/2056305116672485>
- Dunne, M. J., Raynor, L. A., Cottrell, E. K., & Pinnock, W. J. A. (2017). Interviews with patients and providers on transgender and gender nonconforming health data collection in the electronic health record. *Transgender Health*, 2(1), 1-7. <https://doi.org/10.1089/trgh.2016.0041>
- Eisenhardt, K. M. (1989). Building theories from case study research. *The Academy of Management Review*, 14(4), 532-550. <https://doi.org/10.5465/AMR.1989.4308385>
- Emerson, M. E., & Lehman, L. G. (2022). Who are we missing? Conducting a diversity audit in a liberal arts college library. *The Journal of Academic Librarianship*, 48(3), Article 102517. <https://doi.org/10.1016/j.acalib.2022.102517>
- Eriksson, O., & Ågerfalk, P. J. (2010). Rethinking the meaning of identifiers in information infrastructures. *Journal of the Association for Information Systems*, 11(8), 433-454. <https://doi.org/10.17705/1.jais.00234>
- Faulkner, P., and Runde, J. (2019). Theorizing the digital object. *MIS Quarterly*, 43(4), 1279-1302. <https://doi.org/10.25300/MISQ/2019/13136>
- Faye, S. (2021). *The transgender issue: An argument for justice*. Penguin.
- Fraser, N. (1995). From redistribution to recognition? Dilemmas of justice in a "post-socialist" age. *New Left Review*, 212, 68-93.
- Fraser, N. (1997). *Justice interruptus. Critical reflections on the "postsocialist" condition*. Routledge.
- Fraser, N. (2000). Rethinking recognition. *New Left Review*, 3, 107-120.
- Fraser, N. (2003). Social justice in the age of identity politics: Redistribution, recognition, and participation. In N. Fraser & A. Honneth (Eds.), *Redistribution or recognition? A political-philosophical exchange* (pp. 7-88). Verso.
- Fraser, N. (2008). Abnormal justice. *Critical Inquiry*, 34(3), 393-422.
- Fraser, N., & Honneth, A. (2003). *Redistribution or recognition? A political-philosophical exchange*. Verso.
- Ghaziani, A. (2019). Methodological problems and possibilities in gayborhood studies. In A. Ghaziani & M. Brim (Eds.), *Imagining queer methods* (pp. 103-118). NYU Press.
- GLAAD. (2017). *Accelerating acceptance 2017*. https://www.glaad.org/files/aa/2017_GLAAD_Accelerating_Acceptance.pdf
- Gray, K. L. (2015). #CiteHerWork: Marginalizing women in academic and journalistic writing. Retrieved October 12, 2021 from <http://www.kishonnagrays.com/manifestmy-reality/citeherwork-marginalizing-women-in-academic-and-journalistic-writing/>
- Guyan, K. (2022). *Queer data*. Bloomsbury Publishing.
- Heeks, R., & Renken, J. (2018). Data justice for development: What would it mean? *Information Development*, 34(1), 90-102. <https://doi.org/10.1177/0266666916678282>
- Hirschheim, R., Klein, H. K., & Lyytinen, K. (1995). *Information systems development and data modeling conceptual and philosophical foundations*. Cambridge University Press.
- Holvino, E. (2010). Intersections: The simultaneity of race, gender and class in organization studies. *Gender, Work and Organization*, 17(3), 248-277. <https://doi.org/10.1111/j.1468-0432.2008.00400.x>
- Honneth, A. (1996). *The struggle for recognition: The moral grammar of social conflicts*. MIT Press.
- Honneth, A. (2003). Redistribution as recognition: A response to Nancy Fraser. In N. Fraser & A. Honneth (Eds.), *Redistribution or recognition? A political-philosophical exchange* (pp. 110-197). Verso.
- Honneth, Axel. (2004). Recognition and justice: Outline of a plural theory of justice. *Acta Sociologica*, 47(4), 351-364. <https://doi.org/10.1177/0001699304048668>
- Johnson, J. A. (2015). Information systems and the translation of transgender. *TSQ: Transgender Studies Quarterly*, 2(1), 160-165. <https://doi.org/10.1215/23289252-2848940>
- Kallinikos, J. (2006). Information out of information: On the self-referential dynamics of information growth. *Information Technology and People*, 19(1), 98-115. <https://doi.org/10.1108/09593840610649989>
- Kallinikos, J., Aaltonen, A., & Marton, A. (2013). The ambivalent ontology of digital artifacts. *MIS Quarterly*, 37(2), 357-370. <https://doi.org/10.25300/MISQ/2013/37.2.02>
- Kelemen, M., & Rumens, N. (2008). Queer theory. In G. Haggerty & B. Zimmerman (Eds.), *An introduction to critical management research* (pp. 1-15). SAGE. <https://doi.org/10.4324/9780203487884-134>
- Kitchin, R. (2022). *The Data Revolution. A critical analysis of big data, open data and data infrastructures*. SAGE.
- Kitchin, Rob, & Lauriault, T. (2014). *Towards critical data studies: Charting and unpacking data assemblages and their work*. SSRN.

- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112
- Labuski, C., & Keo-Meier, C. (2015). The (mis)measure of trans. *TSQ: Transgender Studies Quarterly*, 2(1), 13-33. <https://doi.org/10.1215/23289252-2848868>
- Leidner, D. E., & Tona, O. (2021). The CARE theory of dignity amid personal data digitalization. *MIS Quarterly*, 45(1), 343-370. <https://doi.org/10.25300/MISQ/2021/15941>
- Leonardi, P., Nardi, B. A., & Kallinikos, J. (2012). *Materiality and organizing: Social interaction in a technological world*. Oxford University Press.
- Love, H. (2019). How the other half thinks. In A. Ghaziani & M. Brim (Eds.), *Imagining queer methods* (pp. 28-42). NYU Press.
- Lukyanenko, R., Parsons, J., Wiersma, Y. F., & Maddah, M. (2019). Expecting the unexpected: Effects of data collection design choices on the quality of crowdsourced user-generated content. *MIS Quarterly*, 43(2), 634-647. <https://doi.org/10.25300/MISQ/2019/14439>
- Mackenzie, L. Z. (2017). The afterlife of data. *TSQ: Transgender Studies Quarterly*, 4(1), 45-60. <https://doi.org/10.1215/23289252-3711529>
- Masiero, S., & Das, S. (2019). Datafying anti-poverty programmes: implications for data justice. *Information Communication and Society*, 22(7), 916-933. <https://doi.org/10.1080/1369118X.2019.1575448>
- McCann, H., & Monaghan, W. (2019). *Queer theory now: From foundations to futures*. Red Globe Press.
- Mearns, G. W., Bonner-Thompson, C., & Hopkins, P. (2020). Trans experiences of a university campus in northern England. *Area*, 52(3), 488-494. <https://doi.org/10.1111/area.12595>
- Miles, M. B., & Huberman, M. A. (1994). *Qualitative data analysis: An expanded sourcebook*. SAGE.
- Monteiro, E., Constantinides, P., Scott, S., Shaikh, M., and Burton-Jones, A. (2022). Qualitative research methods in information systems: A call for phenomenon-focused problematization. *MIS Quarterly*, 46(4), iii-xix.
- Recker, J., Lukyanenko, R., Jabbari, M., Samuel, B. M., & Castellanos, A. (2021). From representation to mediation: A new agenda for conceptual modeling research in a digital world. *MIS Quarterly*, 45(1), 269-300. <https://doi.org/10.25300/MISQ/2021/16027>
- Robinson, C., & Franklin, R. S. (2021). The sensor desert quandary: What does it mean (not) to count in the smart city? *Transactions of the Institute of British Geographers*, 46(2), 238-254. <https://doi.org/10.1111/tran.12415>
- Rumens, N., de Souza, E. M., & Brewis, J. (2019). Queering queer theory in management and organization studies: Notes toward queering heterosexuality. *Organization Studies*, 40(4), 593-612. <https://doi.org/10.1177/0170840617748904>
- Sarker, S., Xiao, X., Beaulieu, T., & Lee, A. S. (2018). Learning from first-generation qualitative approaches in the IS discipline: An evolutionary view and some implications for authors and evaluators (part 1/2). *Journal of the Association for Information Systems*, 19(8), 752-774. <https://doi.org/10.17705/1jais.00508>
- Schilt, K., Meadow, T., & Compton, D. (2018). Introduction: Queer work in a straight discipline. In D. Compton, T. Meadow, & K. Schilt (Eds.), *Other, please specify* (pp. 1-34). University of California Press.
- Schucan Bird, K., & Pitman, L. (2020). How diverse is your reading list? Exploring issues of representation and decolonisation in the UK. *Higher Education*, 79(5), 903-920. <https://doi.org/10.1007/s10734-019-00446-9>
- Spiel, K. (2021). Why are they all obsessed with gender? (Non)binary navigations through technological infrastructures. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference: Nowhere and Everywhere* (pp. 478-494). <https://doi.org/10.1145/3461778.3462033>
- Szulc, L. (2019). Profiles, identities, data: Making abundant and anchored selves in a platform society. *Communication Theory*, 29(3), 169-188. <https://doi.org/10.1093/ct/qty031>
- Szulc, L. (2020). Digital gender disidentifications: Beyond the subversion versus hegemony dichotomy and toward everyday gender practices. *International Journal of Communication*, 14(699745), 5436-5454.
- Tarafdar, M., Rets, I., & Hu, Y. (2023). Can ICT enhance workplace inclusion? ICT-enabled workplace inclusion practices and a new agenda for inclusion research in Information Systems. *The Journal of Strategic Information Systems*, 32(2), Article 101773. <https://doi.org/10.1016/j.jsis.2023.101773>
- Taylor, C. (1994). The politics of recognition. In Amy Gutman (Ed.), *Multiculturalism: Examining the politics of recognition* (pp. 25-73). Princeton University Press.
- Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data and Society*, 4(2), 1-14. <https://doi.org/10.1177/2053951717736335>
- Tempini, N. (2020). The reuse of digital computer data: Transformation, recombination and generation of data mixes in big data science. In S. Leonelli & N. Tempini (Eds.), *Data Journeys in the Sciences* (pp. 239-264). Springer Open.
- Tuomi, I. (1999). Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. *Journal of Management Information Systems*, 16(3), 103-117. <https://doi.org/10.1080/07421222.1999.11518258>
- Wade, L., & Ferree, M. M. (2018). *Gender: Ideas, interactions, institutions*. Norton.
- Waelen, R. A. (2023). The struggle for recognition in the age of facial recognition technology. *AI and Ethics*, 3(1), 215-222. <https://doi.org/10.1007/s43681-022-00146-8>
- Waelen, R., & Wiczorek, M. (2022). The struggle for AI's recognition: Understanding the normative implications of gender bias in AI with Honneth's theory of recognition. *Philosophy and Technology*, 35(2), 1-17. <https://doi.org/10.1007/s13347-022-00548-w>
- Wagner, E. L., & Piccoli, G. (2007). Moving beyond user participation to achieve successful IS design. *Communications of the ACM*, 50(12), 51-55. <https://doi.org/10.1145/1323688.1323694>
- Wand, Y., & Weber, R. (1995). On the deep structure of information systems. *Information Systems Journal*, 5, 203-223. <https://doi.org/10.1111/j.1365-2575.1995.tb00108.x>
- Wand, Yair, & Weber, R. (1990). Toward a theory of the deep structure of information systems. In *Proceedings of the 11th International Conference on Information Systems* (pp. 61-71).
- West, C., & Zimmerman, D. (1987). Doing gender. *Gender and Society*, 1(2), 125-151.
- The White House (2021). *National strategy on gender equity and equality*. <https://www.whitehouse.gov/wp-content/uploads/2021/10/National-Strategy-on-Gender-Equity-and-Equality.pdf>
- Whitley, C. T., & Greene, D. N. (2017). Transgender man being evaluated for a kidney transplant. *Clinical Chemistry*, 63(11), 1680-1683. <https://doi.org/10.1373/clinchem.2016.268839>
- Wilson, M. (2018). *How tech giants design for transgender users-or don't*. Fast Company. <https://www.fastcompany.com/90165071/how-tech-giants-design-for-transgender-users-or-dont>
- Yin, R. K. (2018). *Case study research and applications: Design and methods*. SAGE.

About the Authors

Marta Stelmaszak is an assistant professor of information systems in the Isenberg School of Management at the University of Massachusetts Amherst. She holds a Ph.D. from the London School of Economics and Political Science and an M.Sc. in data science from Birkbeck, University of London. Marta's research interests concern data science and analytics as well as their responsible, sustainable, and ethical management in organizations. Her research has appeared in, among other outlets, *Information Systems Research* and *Journal of the Association for Information Systems*. She serves as a reviewer for information systems journals and an associate editor for leading conferences. ORCID: <https://orcid.org/0000-0003-3173-7470>.

Erica L. Wagner is a professor of information systems in The School of Business at Portland State University. Erica formerly served as PSU's vice provost of Student Success and also served as an associate dean of Undergraduate Programs in The School of Business. Prior to her employment at PSU in 2009, Erica was an assistant professor at Cornell University. She earned her Ph.D. from the London School of Economics. Her research has been published in a variety of outlets including *Information & Organization*, *Journal of Management Studies*, *Journal of the Association of Information Systems*, and *MIS Quarterly*.

Nicolle Nixon DuPont is an associate vice provost and university registrar at Portland State University and has served in student support positions since 2001. Nicolle is a student-centered professional with a rich history of leadership and advocacy for improving the educational experience by examining and reimagining how students interact with systems.

Appendix

Citation Values and Metrics

As we began to work on the manuscript, we created a table of citation values and corresponding metrics inspired by similar approaches pioneered in other fields, including the publication of values statements and principles by the University of Maryland’s African American History, Culture, and Digital Humanities Initiative, and the University of Delaware’s Colored Conventions Project. Notably, we drew inspiration from D’Ignazio and Klein’s (2020) book, which provides a comprehensive explanation of how citation values and metrics can be established and audited, as well as other diversity audits (Emerson & Lehman, 2022; Schucan Bird & Pitman, 2020). We proceeded as follows: We discussed the citation values adopted by D’Ignazio and Klein and agreed that we subscribe to them—that is, we also insist on intersectionality, advocate for equity, prioritize proximity, acknowledge the humanity of data, and aim to be reflective, transparent, and accountable (2020, p. 215). To reflect these values in our work, we established aspirational citation metrics at the beginning of the project. Then, we contracted a research assistant with a background in anthropology and qualitative methods to independently audit the citations and subsequently calculated the metrics in the manuscript submitted. We updated the metrics throughout the review process.

Table A1. Aspirational, Manuscript, and Final Metrics					
Structural problem	Aspirational metrics	Submitted manuscript metrics	First revision metrics	Second revision metrics	Published manuscript metrics
Lack of ethnic diversity	<ul style="list-style-type: none"> • 40% of citations include a scholar identifying as non-White (at least one included) 	<ul style="list-style-type: none"> • 9.9% of citations include a scholar identifying as non-White 	<ul style="list-style-type: none"> • 5.2% of citations include a scholar identifying as non-White 	<ul style="list-style-type: none"> • 5.7% of citations include a scholar identifying as non-White 	<ul style="list-style-type: none"> • 5.6% of citations include a scholar identifying as non-White
Lack of gender-diverse scholarship	<ul style="list-style-type: none"> • 50% of citations from scholars identifying as non-male, including at least 25% of citations from scholars identifying as outside of gender binary (fractionally) 	<ul style="list-style-type: none"> • 54.7% of citations from scholars identifying as non-male, including 10.1% of citations from scholars identifying as outside of gender binary 	<ul style="list-style-type: none"> • 46.2% of citations from scholars identifying as non-male, including 6.8% of citations from scholars identifying as outside of gender binary 	<ul style="list-style-type: none"> • 42.8% of citations from scholars identifying as non-male, including 4.9% of citations from scholars identifying as outside of gender binary 	<ul style="list-style-type: none"> • 42.8% of citations from scholars identifying as non-male, including 4.8% of citations from scholars identifying as outside of gender binary
Lack of non-heteronormative scholarship	<ul style="list-style-type: none"> • 25% of citations from scholars identifying as non-heteronormative (fractionally) 	<ul style="list-style-type: none"> • 7.0% of citations from scholars identifying as non-heteronormative 	<ul style="list-style-type: none"> • 6.0% of citations from scholars identifying as non-heteronormative 	<ul style="list-style-type: none"> • 2.1% of citations from scholars identifying as non-heteronormative 	<ul style="list-style-type: none"> • 2% of citations from scholars identifying as non-heteronormative
Lack of voice given to marginalized populations in research	<ul style="list-style-type: none"> • 50% of citations in literature review feature or quote people directly impacted by an issue • Use of LGBTQ+ inclusive language throughout the article 	<ul style="list-style-type: none"> • 47.8% of citations in literature review feature or quote people directly impacted by an issue • Use of LGBTQ+ inclusive language throughout the article 	<ul style="list-style-type: none"> • 46.6% of citations in literature review feature or quote people directly impacted by an issue • Use of LGBTQ+ inclusive language throughout the article 	<ul style="list-style-type: none"> • 28.6% of citations in literature review feature or quote people directly impacted by an issue • Use of LGBTQ+ inclusive language throughout the article 	<ul style="list-style-type: none"> • 28.6% of citations in literature review feature or quote people directly impacted by an issue • Use of LGBTQ+ inclusive language throughout the article