Teaching Case Study: Gender Data Trouble in a Student Information System

Marta Stelmaszak Portland State University stmarta@pdx.edu Erica Wagner
Portland State University
elwagner@pdx.edu

Abstract

In 2022, StateU, a large public university in the United States, embarked on a project to collect and use personal pronouns in its information systems. The project lead and functional expert was StateU's Administrative Leader. As she prepared for the first project meeting, she reflected on lessons learned from a past project she led to expand the collection of student gender data to record legal sex, gender identity, and sexual orientation. That project involved navigating challenging decisions about user interface design, underlying databases, data privacy and security, and reporting, underpinned by the desire to best serve minoritized and vulnerable populations. She recalled that: "A society with more data about LGBTQ people is not automatically a society that is better for LGBTQ people¹". She wondered if collecting pronoun data was the right choice in the first place.

Keywords: Data ethics, gender, LGBTQ, database design, decision making.

1. Introduction²

Back in her office for only one day a week, Stephanie Owens³, Administrative Leader at StateU, reflected on the difficulties that the two years of the pandemic brought for the whole university. As an urban, public university comprising ten academic units, employing approximately 1,700 faculty members and enrolling about 24,000 students, StateU was looking at a long path towards recovery. Even in 2022, the dire situation of many students in terms of basic needs, mental health, and finances made it challenging for them to continue with their studies, and faculty and staff felt increased pressure. Many investments and hiring had been suspended for the past two years, but with the numbers of COVID-19

One of the projects that has just received a green light from the university was the Pronoun Collection and Usage Project. In 2018, the university started investigating the possibility of collecting personal pronouns⁴ from students (e.g., she/her, they/them, he/him, and many others) and using them across various university systems as part of the drive towards more diversity, equity, and inclusion on campus. On the heels of the Expanded Bio-Demographic Data Collection Project, the Pronoun Collection and Usage Project was intended to centrally collect and display students' personal pronouns on class rosters, in the learning management system, on Zoom, on ID cards, and in various information systems.

StateU has traditionally ranked highly in LGBTQ+ inclusivity rankings in the United States, and it prided itself on being a diverse, equitable, and inclusive institution. StateU has always been committed to supporting LGBTQ+ students and employees as part of its mission. The university wanted to embark on the Pronoun Collection and Usage Project to foster a community that recognizes and celebrates individual identities. The basis of this approach was the belief that people should have the right to self-identify and denying them this right through university systems may be causing harm, for example by referring to them using wrong pronouns. StateU also believed that universities should provide students with an opportunity to reflect and support shifting identities. Further, personal pronoun data was seen as useful to direct funding to initiatives fostering diversity, equity, and inclusion. These goals were important for Stephanie Owens, as in her role as Administrative Leader she has always been a champion of student success and an advocate for potentially marginalized student communities.

https://www.portlandoregon.gov/article/730061.



cases decreasing, some projects were added to the agenda.

¹ Guyan, K. (2022). Queer Data. Bloomsbury Publishing.

² The teaching note accompanying this teaching case is available from the authors.

³ All names of people, institutions, and organizational units in this teaching case are pseudonyms.

⁴ For a glossary of terms related to personal pronouns, legal or biological sex, gender identity, and sexual orientation used in this teaching case study refer to "Glossary of LGBTQ+ and Gender Terms" available at

Now StateU was catching up with other universities in the country that have already implemented pronoun collection and usage⁵. StateU was also learning from its peers. The University of Washington, for example, had published a comprehensive guide related to the handling of personal pronoun data, including governance, definition, collection, integration, and use. The guide included instructions for displaying pronoun data: "Display the pronoun data as you receive it, unchanged. Additional data processing is discouraged in order to preserve the integrity of what each person has expressed for pronouns"⁶.

With the funds unfrozen, Stephanie was looking at the first pronoun meeting invitation in her email inbox. The Pronoun Collection and Usage Discovery meeting was, as per the agenda, aimed at completing the project discovery questionnaire, including project background and goals, project impact and risks, stakeholders, team members, communication, and timeline and costs. As an Administrative Leader, Stephanie was a vital member of the project team, serving as the project lead and a functional expert, because she has previously been involved in the Expanded Bio-Demographic Data Collection Project.

For this reason, Richard Penn, the IT project manager who set up the meeting, sent a quick follow-up email to Stephanie.

"Stephanie, as we're preparing for the meeting, it would be great if we could hear from you first about lessons learned from the expanded biodemo project. Many colleagues on the team were not here back then, so it would be helpful for them to get background on that project. Could you provide a general overview and also highlight the challenges, risks, and concerns that emerged from that project that we should keep in mind with the pronouns? Thanks!"

Stephanie opened a web browser and navigated to the Student Information System (SIS), and then clicked on the link to the expanded bio-demographic data collection web page. She scrolled through the options available and started to recall how StateU got to this point.

https://www.insidehighered.com/news/2022/01/25/colleges-update-systems-include-preferred-pronouns-names. Accessed on: 5 May 2022.

2. Expanded Bio-Demographic Data Collection

2.1. Project Background

The Expanded Bio-Demographic Data Project (EBD) started within the StateU Office of Information Technology in January 2016. StateU undertook this project for two main reasons. First, to implement recommendations from the internal Commission on Gender Equity. Second, StateU sought to align with legal provisions from the state's Senate bill requiring public universities to capture additional biographical information from students. No longer would universities capture just legal sex; they would also capture sexual orientation and gender identity data in student information systems. The bill empowered the state's Higher Education Directorate to work with representatives of higher education institutions to determine the best method to collect additional data and to identify potential barriers to implementing the legal provisions, including legal issues, cost issues, and data systems limitations.

The Directorate convened a work group of all institutions, including StateU. The work group acknowledged the need for expanded biodemographic data "in an effort to better serve vulnerable or underrepresented students and staff"7. The work group agreed to add a question on sexual orientation and gender identity to each student record to accompany the existing and required legal sex designation, and developed three questions and possible answers (see Exhibit A). The work group identified several barriers to data collection, most notably around costs in time and money, different systems used across institutions, lack of relevant education and training and project prioritization. The main issue revolved around the costs of making changes to existing SIS - systems that are predominantly off-the-shelf products that differ between institutions. It has been noted that such changes would impact not only the SIS, but other university systems, reports, and pages that involve gender data, while the work group was concerned that not all dependencies were known. The work group sent a survey to all institutions to estimate these costs and establish whether they could be borne by each institution. As per the summary of the survey, the

https://itconnect.uw.edu/connect/identity/pronoun-data-governance-and-integration/. Accessed on: 10 May 2022.

⁵ Smalley, S. (2022). "Gender Identity Norms Shift, and Institutions Move to Reflect Them", *Inside Higher Ed.* Available online:

⁶ University of Washington (2022). Pronoun data governance and integration. Available online:

⁷ All direct quotations are taken from confidential case materials unless otherwise credited.

average number of hours needed to make the change was 131, and many institutions reported small numbers of IT staff available. Outsourcing this work at the recommended rate of 200 USD per hour could result in significant costs for each institution. The report of the work group to the Directorate stated:

"This survey illustrates that a major barrier to successful implementation of the bill will be the cost associated with the needed system upgrades. Given the complication of a wide variety of systems and software versions used across the colleges, the ability to cross share upgrades or programming changes, while perhaps more cost effective, may prove problematic. As illustrated above, the cost in dollars is not the only expense. Many institutions have very small IT staff and some may not have programming expertise for their particular software in-house. The necessary amount of hours to update computer systems could top 200 and may be spread across a very small number of staff. This may be considered a more significant impact on an institution's resources than just the monetary value itself".

The work group also highlighted a challenge around training and educating staff and students about collecting expanded data: "Students and staff who are unaware of the reasons behind the questions may fear answering them or believe the data may be used for some other purpose".

In its recommendations, the work group suggested that universities that can bear the costs of updating their systems should collaborate to create and implement the necessary modifications. In the meantime, an anonymous survey should be sent out to students as a temporary solution to comply with the bill's regulations. Internal recommendations of StateU's Commission on Gender Equity were in line with legal requirements. The recommendations were based on almost 10 years of work by the Commission to extend social justice towards marginalized sexual orientation and gender identity communities. Some past initiatives of the Commission included queer student outreach, conducting diversity, equity and inclusion training, and participation in shaping university-wide policies, for example by suggesting standards for collecting bio-demographic data on campus which will be discussed in the next section.

2.2. Expanded Bio-Demographic Data Collection Project

In order to follow both internal and external recommendations, StateU decided to draw on its

sizable IT department to modify its SIS. StateU's Office of Information Technology launched the EBD project to achieve the following goals: 1) compile an inventory of all university forms and reports that collect bio-demographic data, 2) develop a technical solution to store the three data points, 3) modify the interface of all forms and reports to allow expanded data capture, 4) work with the external provider to modify the SIS to allow students to enter, update, and modify data, 5) update the master database with expanded data, and 6) develop policies for acceptable use and collection of the data. The project goals would be achieved by August 2016.

The project charter identified the following stakeholders: StateU's students, StateU's employees, Queer Center, Office of Diversity and Inclusion, Human Resources, Commission on Gender Equity, Office of the Registrar, Institutional Research, Admissions, Graduate Studies, Student Financial Services, and Information Technology. The project team consisted of a sponsor from the Office of Diversity and Inclusion, five representatives from Information Technology, a representative from Institutional Research, and a representative from the Office of the Registrar, Human Resources, and Queer Center each.

Over the course of the ensuing months, the EBD project team worked on realizing the project goals. The questions for expanded data collection from the Directorate working group (see Exhibit A) were designed to be stored in the database in two new tables (see Exhibit B). To acknowledge the plurality of gender identities and sexual orientation, meaning that a person can identify with more than one, StateU 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 the [SIS] if we did that", as reported by a project team member, so the open text box with a limit of 30 characters was a middle-ground solution to acknowledge this diversity.

As work progressed, some issues were identified. For example:

"While working with the new biographic information in [the] SIS my team noticed there are some students with records in the table that have not specified a gender identity or sexual orientation, so their records seems to have no content. Is this intentional or in need of cleanup? Here is the SOL to identify these records:

select * from zgbbiod where zgbbiod_gi_code is null and zgbbiod_so_code is null;"

2.3. Legal Sex Data

Since gender identity and sexual orientation questions were optional, missing data was less problematic as it was not required in every record, but this became an issue with legal sex. Historically, this field has not always been required on various applications, thus some student records contained a NULL response, rather than Male or Female. Additionally, as the state within which StateU is located established a nonbinary legal sex destination, StateU included Nonbinary as a legal sex destination. A project member described the complexity of this change back then:

"As part of the Expanded Biographic Data project, we renamed Gender to Legal Sex Designation and added two new fields: Gender Identity and Sexual Orientation. As part of an earlier project, we added a third option, Nonbinary, to the Legal Sex Designation field, which previously only allowed for Male and Female (and null). It's my understanding that federal processes require a Male or Female value. Stephanie Owens would know more about the federal reporting requirements and why adding nonbinary had historically been complex. In the Expanded Biographic Data project, we were creating brand new fields with brand new values, so we did not have to take historic reporting into account. In the Nonbinary Legal Sex designation, we were adding a third value to an existing field, so we had to consider how doing so would impact existing reports and reporting requirements."

Indeed, the presence of Nonbinary and NULL responses to legal sex created the need to transform the data before reporting it to the federal system, the Integrated Postsecondary Education Data System (IPEDS). IPEDS is a large-scale survey collecting institution-level data from post-secondary institutions in the U.S. Using a web-based collection system all higher education institutions have to report, including many other fields, the gender of their students. The usage of the term gender in the system was seen as a source of confusion, as reported by a StateU employee: "Much would be solved if we had clearer definitions around exactly what we're supposed to report in those fields. Are we supposed to report legal sex, are we supposed to report gender, if we're supposed to report sex is this sex [that] was assigned at birth or sex on a state ID?".

The IPEDS system requires universities to report on whole populations rather than individual students

and collects only legal sex (erroneously termed gender) as a binary response. Thus, the only allowed responses are Male or Female. Reports to IPEDS are required several times throughout each year and are prepared by the Institutional Research Office. As such, various institutions including StateU (see Exhibit C), developed ways to report unknown or NULL legal sex data. It is common practice to reassign students whose answers are other than Male or Female to one of those two categories: "the federal government recommends dealing with students who you do not know the legal sex of by report[ing] them all as one or the other, or to split the numbers between male and female based on the known ratio of males to females". At StateU, such students are reported as Male for historical reasons motivated by the desire not to artificially overstate the size of non-male student population when increasing this student population was one of the university, state, and nation-level goals. In order to align with a more common practice of reassigning students based on the last digit of their student ID number, StateU was planning to change its approach.

2.4. Gender Identity Data

The separation of legal sex and gender identity reflected "shifting language from what we previously called the gender field which really was legal sex field to have a more clear distinction between legal sex that's on your birth certificate versus gender identity, so that was part of the project as well", as reported by a project team member. This required the creation of a new data table to store data that previously were not collected by StateU, with significant implications, as the project team had to make sure "we're not breaking any of the existing processes or federal reporting requirements or financial aid reporting requirements based on adding a new [table] which seems simple but there's a lot of downstream impact".

The gender identity field was optional and allowed selecting only one response from 11 answers, all stored as two-character variables in the data table. An additional open text box enabled users, including those identifying as more than one gender identity, to specify an identity or identities not listed in up to 30 characters. While the 11 answers provided did not require any processing, the open text box "makes it a little bit of an obstacle because there are some people obviously joking around, and [are] not real results", as reported by an employee of StateU. This meant that before any of that data could be used, they would have to be cleaned and since "there's a lot of variations on the same theme" the data would also need to be categorized. However, there has been very little need for the use of these data in light of other priorities

concerning community building and taking direct action; therefore, none of the cleaning as part of processing was actually taking place. Changes to the answers, which users could introduce themselves at any time, were not tracked; that is, a history table did not exist. This decision was clarified by Stephanie Owens as follows:

"Hi Richard.

Thank you for asking for clarification on this.

I can confirm that yes - it was the decision of the implementation team (including QC representatives) not to keep track of changes. And yes - one of the stated reasons having to do with privacy, that 'no one would need to know' what a previous gender is/was. I agree with this on an individual level (I cannot articulate a use-case where the institution needs that information).

In 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?

I don't know that this would have been enough of a priority to change the decision of how it was collected, and I don't know of it would have outweighed the concern for personal privacy. The element of personal privacy was really strong during the implementation discussions.

I hope this is helpful. Best, Stephanie"

Gender identity data were classified as highly sensitive, according to StateU's data privacy policies. These data were therefore not used in any internal or external reporting. Access to them required the approval of the Office of Diversity and Inclusion, granted only for justified educational and research objectives.

2.5. Sexual Orientation Data

As with gender identity data, the intention was to make sexual orientation data available to the Queer Center to better identify students, faculty, and staff in need of appropriate support and resources. As an employee of StateU explained, some of the data "helped to quantify LGBTQ+ students which helped with receiving a fair allocation of funds" and "the first full-time queer students of color coordinator, one of the first in the nation, [was hired] because the data that showed we had over 1000 students on campus who identify as queer and BIPOC [Black, Indigenous, and People of Color]".

The collection of sexual orientation data specifically surfaced the use versus protection trade-off with which StateU struggled. Existing regulations stated that "nonconsensual disclosure of personally identifiable information (PII), such as a student's birth name or sex assigned at birth, could be harmful to, or invade the privacy of, transgender students and may also violate the Family Educational Rights and Privacy Act (FERPA) A school may maintain records of this information, but such records should be kept confidential". At StateU, FERPA was also extended to sexual orientation data collected.

StateU's information security employees were not involved in the EBD project. A few years after the project's completion, security team members expressed concerns about collecting individually identifiable sexual orientation and gender identity data that could be subject to potential hacking attacks or, down the road, required for federal reporting. These unintended consequences of collecting sensitive personal data may expose individuals to danger: "the creation of the record in itself is a source of danger", explained an information security team member. The employees acknowledged the usefulness and need for these data in aggregate but suggested that StateU should "stop collecting it the way we do it today" and instead separate their collection from student identities: "student support goals are important but should take a back seat with respect to safety". For these reasons, StateU took several steps to ensure that the data were secure, kept private, and with limited access. With no clear indications of how data could be used, there was some perception that these constituted a "risky set of data to collect" and that more clarification was needed about "why we're collecting these data", as explained by the information security team. Sexual orientation data were effectively not being used at the individual or collective level, as their use was seen as a potential risk to privacy, security, and safety.

3. After EBD Project Launch

A few years after the EBD project was completed, a handful of LGBTQ+ students on campus were asked about their views on the expanded data collection. One student commented: "I'm really glad that it [expanded bio-demographic data] is not just two or three options", and that having these expanded data collection options is "in itself sort of a political statement, a kind of allowing people to exist". At the same time, students shared their worries about the data being collected. In terms of why the data were being collected, one student expressed a worry that "there's a lot of big talk with not a lot of action" with an attitude to "collecting it [data] and then figure out the rest". No students were directly involved in the EBD project.

It was true that StateU started the EBD project hoping that the collection of additional gender identity and sexual orientation data would bring several advantages to the whole community. One of them relied on StateU's ability to analyze and report on these data to be able to better support its queer students. For example, StateU wanted to use the reporting on gender identities and sexual orientation to allocate resources better and improve its decision-making. However, the many difficulties around collecting, recording, and processing the data meant that StateU never realized this benefit.

As Stephanie found herself reviewing the materials related to the EBD project, she came across some resources she used back then (see also Exhibit D):

- Reachable: Data collection methods for sexual orientation and gender identity
- <u>Current Measures of Sexual Orientation and</u> Gender Identity in Federal Surveys
- Dear Colleague Letter on Transgender Students

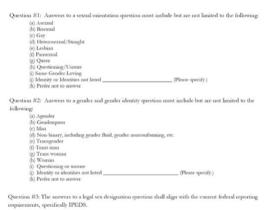
Stephanie revisited these source documents to prepare for the upcoming meeting that would launch the pronoun project and recalled a conversation with a colleague who stated that many universities faced difficulties making meaningful use of the expanded data the institution collected. Her colleague had said, "they'd been collecting it [expanded bio-demographic data] because they did want to be inclusive, but they just didn't know what to do with that information once they had it or where the value was [in the collected data]". In hindsight, she began to wonder if this was true at StateU. Having thought about the EBD project now. Stephanie wondered how to share her experience and lessons learned with the new Pronoun Collection and Usage Project team. She also began to reflect on whether pronoun data should even be collected and used in the first place. After all, expanded biodemographic data were used internally and for external reporting purposes, while pronouns were meant to be visible to others in academic spaces and used when addressing students.

4. Case Questions

- What are the most important elements of the Expanded Bio-Demographic Data Collection Project relevant to the Pronoun Collection and Usage Project? Identify five similarities between the projects that you consider the most relevant for Stephanie to include in her update in the meeting. Highlight a crucial difference between the projects.
- 2. What challenges and risks emerge from the Expanded Bio-Demographic Data Collection Project relevant to the Pronoun Collection and Usage Project? Consider at least technological, privacy and security, and ethical issues.
- 3. Despite these challenges and risks, should StateU go ahead with the Pronoun Collection and Usage Project? Evaluate this question using ethics theories or decision-making frameworks and present your reasoning and an answer that would represent Stephanie Owen's perspective in the meeting.
- 4. "A society with more data about LGBTQ people is not automatically a society that is better for LGBTQ people". This quote from Guyan's book "Queer Data" points to the fact that just collecting more data about vulnerable or marginalized populations does not necessarily improve their conditions. Under which conditions should data such as personal pronouns be collected and used? Based on the case and resources mentioned and your own research, devise a set of principles around ethical sensitive data collection that Stephanie could present in the meeting.

5. Exhibits

Exhibit A Proposed questions and answers to be included in university systems



Source: Confidential case materials

Exhibit B Database structure for the collection of expanded bio-demographic data

```
1. zgtvbdso - Bio Data Sexual Orientation Validation Table
a. zgtvbdso_so_code varchar2 (char 2)
i. 'AS' = Asexual
iii. 'GA' = Gay
iv. 'HS' = Heletrosexual/Straight
v. 'LE' = Lesbian
vi. 'PS' = Pansexual
vii. 'QR' = Queer
viii. 'QR' = Queer
viii. 'QR' = Queer
viii. 'QR' = Gueer
```

Source: Confidential case materials

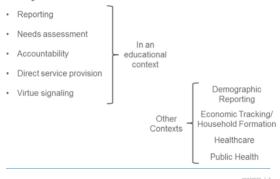
Exhibit C Proposed reassignment of Gender Identity to binary gender markers for federal reporting

Gender Identity		Assigned Binary Gender Marker
if Male	then assign	Male
<i>if</i> Trans Male/Trans Man		
if Female	then assign	Female
<i>if</i> Trans Female/Trans Woman		
if Genderqueer or Nonbinary Gender	then assign	Gender assigned based on the last dig of student/employee identification cod even digits assigned female and odd digits assigned male
if Different Identity		
if No		
Response/Decline to State		

Source: Confidential case materials

Exhibit D Practices shared around collecting expanded bio-demographic data

Why we collect data



Slide 1

How to move forward at your school/campus

- · Identify and communicate with stakeholders
- · Identify and communicate purpose(s) of data collection
- · Create a data collection/use policy
 - Collection
 - Specify wording of question(s)
 - · Determine when and where data is collected
 - Communicate how data can be changed
 - Data storage
 - · Data use/disclosure
 - · Data retention/destruction

11/13/2018 | 2

Slide 2

Other important considerations

- Placement of questions (Williams institute)
 - Place questions with other demographic data or in a separate section. Do not place question near questions about sexual assault or sexual abuse.
- · Be explicit about data use (Consortium
 - How the data will be used
 - Which offices or individuals will have access to the data
 - That the information will not affect admissions, enrollment, academic status, etc.
 - Be clear with campus partners about how and when data can be used (Van Matre)
- Tell students how they can change their selections in the future (Consortium)
- Have an action plan for responses that indicate that a student/applicant is troubled or emotionally distressed. (Allen and Trimble 1993)

11/13/2018 | 1

Slide 3

Source: Van Matre, J. C. (2018, November). Collecting Sexual Orientation and Gender Identity Data in an Educational Context [Conference presentation]. Annual Conference of the California Educational Research Association 2018, Los Angeles, CA, United States.

https://doi.org/10.17605/OSF.IO/U4CX8