UNIVERSITY of VIRGINIA

SCHOOL of DATA SCIENCE

Elson Building 189B
400 Brandon Avenue | PO Box 400249
Charlottesville, VA 22904

P 434.924.5261
E reia@virginia.edu
datascience.virginia.edu

October 3, 2022

NSTC Subcommittee on Equitable Data
Office of Science and Technology Policy
Eisenhower Executive Office Building
1650 Pennsylvania Ave. NW
Washington, DC 20504

Re: Federal Evidence Agenda on LGBTQI+ Equity RFI

The School of Data Science (SDS) at the University of Virginia (UVA) appreciates the opportunity to submit comments to the White House Office of Science and Technology Policy (OSTP) Request for Information (RFI) on the Federal Evidence Agenda on LGBTQI+ Equity (87 FR 52083).[1]

Introduction

This document will focus mostly on data collection and its impacts on gender diverse communities, and it is based on my work as a Responsible Data Science scholar and my lived experience as a non-binary member of the transgender community.

Gender diverse communities – including but not limited to trans, non-binary, gender non-conforming, and Two-Spirit people – face many disparities in accessing adequate, inclusive public services and have systematically experienced greater obstacles to access proper healthcare, labor market opportunities, and inclusive learning environments because of their gender identity and expression. These disparities can be exacerbated depending on intersectionality of other social identities such as race, ethnicity, sexual orientation, legal status, age, class, and cognitive, sensory, or physical disability. To reflect upon disparities among transgender individuals, a critical approach to data protection and privacy is crucial. As Federal agencies, scholars, advocates, and practitioners look to LGBTQI+ data[2] and official numbers[3] to improve equity and inclusion, invisibility and misrepresentation issues need to be addressed. This invisibility makes it harder to understand the dimensions of the community and to design better public policies; however, visibility of sensitive data without adequate safeguarding of human rights[4] might be harmful and exacerbate inequalities. Data privacy and protection as well as ethical approaches to data use must be at the forefront of LGBTQI+ data equity discussions. These discussions can be used to address the lack of preparedness of Federal agencies and civil servants to collect and share data of gender minorities, which often result in a lack of trust in the LGBTQI+ communities and flawed visibility mechanisms: from

---

[1] Federal Register. (2022). Request for Information; Federal Evidence Agenda on LGBTQI+ Equity. https://www.federalregister.gov/documents/2022/08/24/2022-18219/request-for-information-federal-evidence-agenda-on-lgbtqi-equity

[2] Guyan, K. (2022). Queer Data: Using Gender, Sex and Sexuality Data for Action. London and New York: Bloomsbury.

[3] Bouk, D., Ackermann, K., & boyd d. (2022). A Primer on Powerful Numbers: Selected Readings in the Social Study of Public Data and Official Numbers. Data & Society. https://datasociety.net/wp-content/uploads/2022/03/APrimerOnPowerfulNumbers_032022.pdf

[4] Wood, C. et al. (2022). The Role of Data Protection in Safeguarding Sexual Orientation and Gender Identity Information. Future of Privacy Forum and LGBT Tech. https://fpf.org/wp-content/uploads/2022/06/FPF-SOGI-Report-R2-singles-1.pdf

misrepresentation of gender markers, identities, and expressions in surveys to human rights violations (e.g., deadnaming, misgendering, and data privacy and protection violations).

Advancing data equity for gender minorities requires a holistic approach to how these individuals are perceived, seen, categorized, and excluded. From the perspective of data science, there is a momentum to define the boundaries of the discipline. Issues of data ethics, policy, governance, and digital rights are essential to advance the responsibility and equity within and beyond the discipline. Invisibility in official numbers, data use, and regulatory frameworks generates knowledge gaps. These issues are also related to a broader context in which public policies and politics play a central role. From general data protection laws and official data collection mechanisms to policies for (or against) transgender communities, data equity is shaped by different sets of sociocultural and political conditions.

Currently, there is a major opportunity to help shape the U.S. policy and regulatory agenda, especially around official numbers, and census data. In this context, the development of the Federal Evidence Agenda on LGBTQI+ Equity is extremely relevant and welcomed.

I - Describing Disparities

- **Response to 2:**

In the absence of official numbers, the efforts to collect data on gender minorities led by civil society and community-based organizations should not be overlooked. There are several groups, organizations, and scholars dedicated to making gender minorities visible and telling the stories that are so often not told by the census data[5] – to the point that it is believed that most data on LGBTQI+ populations "are captured by surveys [...] largely conducted by academic institutions and universities, think tanks, nonprofits, and advocacy organizations."[6]

Moving forward, the best practices and guidelines carefully designed (and thoroughly debated) by these organizations should be incorporated, in different stages, to establish an agenda that advances data equity for LGBTQI+ communities. The invisibility faced by so many members of gender diverse groups along the last several decades is exacerbated by other social and cultural identities, making certain disparities only visible to those engaged in assessing the situation, understanding shortcomings, and learning with these people.[7] Being invisible to the government can mean lack of proper care at the institutional level or protection against institutional violences. For those who can see through the invisibility, in numbers and lived experience, the perception of certain disparities are probably meaningful to the process of creating the Federal Agenda on LGBTQI+ Equity.

Including voices of those who have been invisible for so long and rarely have a seat at the tables where decisions are made is fundamental. Safeguarding human rights while doing so is extremely necessary. For this reason, learning from and with civil society and community-based organizations and responsible research efforts (in academia and beyond) can help us develop an environment in which data collection is critically assessed and data-informed policy can thrive. There is an urge to consider a multistakeholder approach that includes government, civil society, academia, and industry representatives to these processes., Inviting people and organizations working on topics that are significant for this Agenda, and also bring a fresh perspective on transversal issues, such as digital rights, data justice, racial equity, Indigenous sovereignty, among others.

---

[5] See, for instance, the U.S. Trans Survey (https://www.ustranssurvey.org/about) conducted by the National Center for Transgender Equality (NCTE), the National Black Trans Advocacy Coalition, the National Queer Asian Pacific Islander Alliance (NQAPIA), and TransLatin@ Coalition (TLC). Another example is the TransPop Survey (https://williamsinstitute.law.ucla.edu/projects/transpop/) developed by the Williams Institute at UCLA School of Law.

[6] Medina, C., Mahowald, L. (2022). Collecting Data About LGBTQI+ and Other Sexual and Gender-Diverse Communities. Center for American Progress. https://www.americanprogress.org/article/collecting-data-about-lgbtqi-and-other-sexual-and-gender-diverse-communities/

[7] A few examples of knowledge production by and for gender diverse organizations and individuals are the Center for Applied Transgender Studies (https://www.appliedtransstudies.org), the Southern Equality Research and Policy Center (https://southernequality.org/southern-equality-research-policy-center/), and the projects created by Gayta Science (https://www.gaytascience.com/).

At the forefront of this process is designing public consultations[8] that are committed to receive inputs in multiple platforms and venues and engage meaningfully with the actors involved while willing to admit (and correct) mistakes.

- **Response to 3:**

Currently, the rampant attacks on transgender rights in the U.S. puts gender diverse communities across many states at risk.[9] Combined with the increasing surveillance and online data collection, safeguarding human rights online is crucial. Thus, the promotion of an evidence-based policy agenda for gender minorities must take into account the protection and promotion of their digital rights, which means defending civil liberties and human rights in the digital environment. By protecting digital rights – mainly data privacy and protection – it is possible to ensure that digital technologies will support the safety of gender diverse communities and allow them to exercise their freedom of expression.

A recent example that shows the importance of protecting transgender digital rights is the public exposure led by an anti-trans organization, through GoogleMaps, of clinics that offered access to gender affirming healthcare.[10] If, on the one hand, these platforms can be used to promote access to information and healthcare, on the other hand, they may become instrumental in online gender-based violence and violence facilitated by information and communication technologies (ICTs).

Public policies that center digital rights must ensure that ICTs like online platforms avoid this type of instrumentalization. Implementing mechanisms that allow for the exercise of the right to be forgotten[11], through which any person may decide that some of their personal data should be excluded from certain databases, can help transgender individuals to protect themselves from future harms, for example, after they had access to a certain service that needed that data.

Ensuring that companies and public institutions that offer digital services include clauses that prohibit human rights abuses, especially based on gender, sexuality and gender identity in their terms of service (TOS) may reduce user vulnerability by making companies and institutions responsible for their users' well-being[12].

Inviting civil society and community-based organizations working on the promotion of digital rights in the US and across the Americas to join the conversation about LGBTQI+ data equity could generate a productive exchange of challenges and opportunities. Local and transnational perspectives on safeguarding human rights in a data-centric world can shed light on disparities or criteria to consider when reflecting on policy priorities from a public interest approach.

II- Informing Data Collections

- **Response to 2**

Combined data might be useful to understand how the community's self-identification exist in relation to other categories, such as in current US census survey questions on gender identity (Male, Female,

---

[8] Statistics Canada. (2019). The 2021 Census of Population Consultation Results: What we heard from Canadians. https://www12.statcan.gc.ca/census-recensement/2021/consultation/92-137-x/92-137-x2019001-eng.cfm

[9] Conron, K. J., et al. (2022). Prohibiting Gender-Affirming Medical Care for Youth. Williams Institute, UCLA School of Law. Brief. https://williamsinstitute.law.ucla.edu/wp-content/uploads/Trans-Youth-Health-Bans-Mar-2022.pdf

[10] Factora, J. (2022). TERFs Are Using Google Maps to Track and Target Trans Healthcare Providers. Them. https://www.them.us/story/terfs-google-maps-hospitals-community-centers

[11] Correia, M., Rêgo, G. & Nunes, R. (2021). Gender Transition: Is There a Right to Be Forgotten?. Health Care Analysis, 29, 283–300. https://doi.org/10.1007/s10728-021-00433-1

[12] Association for Progressive Communications. (2018). Providing a gender lens in the digital age: APC Submission to the Office of the High Commissioner for Human Rights' Working Group on Business and Human Rights. https://www.ohchr.org/sites/default/files/Documents/Issues/Business/Gender/APC.pdf

Transgender, None of these).[13] These single categories are useful as an overview of people whose gender identity is not contemplated by the gender binary, and can help us to consider characteristics of LGBTQI+ adult population more broadly.

However, combined data in the case of gender diverse communities might render relevant details invisible because gender is experienced in a myriad of ways. Given the fluidity and diversity of gender identities and expressions, fitting them into pre-established categories available for multiple choice responses is complicated and may hinder access to valuable knowledge that could be used to inform better policies. We still need consistent, open data[14] to comprehend generational gaps, regional differences and similarities, the urban and rural divide, access to education, employment and labor market, housing, health (especially mental health), veteran status, access to public service and public infrastructure – always keeping in mind safeguarding data privacy and protection, balancing visibility and a human rights-based approach (HRBA)[15] to shed light onto gender minorities. Careful consideration should be put into issues around immigration and legal status, for example, and other contexts where de-anonymization is possible.

Learning from the experience of other countries is important – despite differences in context, population size, and policy and regulatory frameworks, some of the challenges are similar and the public debate around sexual orientation and gender identity (SOGI) data collection is currently unfolding in multiple places.[16] Canada, for instance, became one of the first countries to provide census data on transgender and non-binary people following extensive consultation and countrywide engagement with the Canadian population to change the census in 2021.[17] According to the Canadian government, “the precision of ‘at birth’ was added to the sex question on the census questionnaire, and a new question on gender was included. As a result, the historical continuity of information on sex was maintained while allowing all cisgender, transgender and non-binary individuals to report their gender.”[18]

The Canadian government addresses the “diversity within gender diversity” for non-binary people via a write-in response so that they can indicate the correct term most relevant to them – with a response rate of over two-thirds of the people aged 15 and older. Words like “fluid,” “agender,” “queer,” “gender neutral,” “Two-Spirit,” and “gender-nonconforming” were prevalent, showing the various subcategories to the non-binary gender identity.[19] Some of these terms, such as “Two-Spirit,” are specific to Indigenous peoples and territories, proving that a more comprehensible set of categories and terms can help us understand the intersectionality of gender identity and, hopefully, advance equity and inclusion for systematically marginalized communities. As stated in responses III-2 and 3 below, considering Indigenous-designed principles for data use is crucial in this process.

A last remark concerns the question of whether collecting data on sex assigned at birth in nonclinical settings is necessary.[20] This question might be considered insensitive or harmful, especially when one considers the many barriers transgender people can face to have their gender markers and legal names changed – a process also deeply connected to legal status, class, support networks, and financial

---

[13] Anderson, L. et al. (2021). New Household Pulse Survey Data Reveals Differences between LGBT and Non-LGBT Respondents During COVID-19 Pandemic. Census Bureau. https://www.census.gov/library/stories/2021/11/census-bureau-survey-explores-sexual-orientation-and-gender-identity.html

[14] The Open Definition defines open data as data that “anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).” http://opendefinition.org/

[15] HRBA is one of the six Guiding Principles of the United Nations Sustainable Development Cooperation Framework. https://unsdg.un.org/2030-agenda/universal-values/human-rights-based-approach

[16] Guyan, K. (2022). Queer Data: Using Gender, Sex and Sexuality Data for Action. London and New York: Bloomsbury.

[17] Statistics Canada. (2022). Canada is the first country to provide census data on transgender and non-binary people. The Daily. https://www150.statcan.gc.ca/n1/daily-quotidien/220427/dq220427b-eng.htm

[18] Statistics Canada. (2022). Filling the gaps: Information on gender in the 2021 Census. Reference materials, 2021 Census. https://www12.statcan.gc.ca/census-recensement/2021/ref/98-20-0001/982000012021001-eng.cfm

[19] Statistics Canada. (2022). Canada is the first country to provide census data on transgender and non-binary people. The Daily. https://www150.statcan.gc.ca/n1/daily-quotidien/220427/dq220427b-eng.htm

[20] Medina, C., Mahowald, L. (2022). Collecting Data About LGBTQI+ and Other Sexual and Gender-Diverse Communities. Center for American Progress. https://www.americanprogress.org/article/collecting-data-about-lgbtqi-and-other-sexual-and-gender-diverse-communities/

resources. Additionally, tying responses to court orders, legal processes, and the medicalization of gender diverse communities is harmful, limiting the accuracy of the data collected. In this case, a focus on combined data about gender identity might be more relevant.

- **Response to 4**

Agencies should encourage public response to questions about gender identity by (1) safeguarding rights and ensuring that visibility in datasets is not risky or threatening, especially in contexts where the human rights violations directed at transgender individuals are on the rise; (2) working with local community-based organizations and advocates to build trust and enforce meaningful engagement with the data collection process; (3) designing questions and sets of responses that respect the complexity of gender identity and expression as much as possible, leaving space for nuances and contradictions; (4) investing in capacity building for people using data of gender minorities; and (5) creating strategies to prevent harm, assess risks and mitigate bias that are widely debated with the communities and publicly available. In order to better communicate with the public about methodological constraints to collecting and publishing SOGI data, agencies must have open channels of communication that are welcoming to transgender people (using preferred names, respecting gender markers, accessible to people with disabilities, available in more than one platform), and work consistently to improve the awareness, sensibilization and training of civil servants working with gender minorities.

There is valuable knowledge being produced and collected by civil society organizations, grassroots movements, and academia over the last decades. Following best practices and guidelines for gender identity, questions established in community-based surveys can be an important step to encourage public response to SOGI questions. For instance, the Center for American Progress states that "response options must be adapted to meet the needs of the specific LGBTQI+ population being surveyed;" some of the practices mentioned include accurately translating questions and materials for populations whose primary language is not English, as well as using "culturally specific terms." [21]

As mentioned before, the questions and set of responses must be able to acknowledge and encompass ever-changing terminology, and the fact that gender identity can change over time. For example, the fluidity and processes of self-discovery, coming out publicly (or not), and self-perception are relevant aspects to have in mind when collecting data on gender diverse communities.

- **Response to 5**

To address missing data and bias, it is important to consider broader sociocultural frameworks that can impact data collection. From distrust in government agencies to threats to one's safety, disclosing gender identity to official data collection efforts is not always a straightforward process. Qualitative information can become a helpful tool, in addition to statistical techniques, to connect dots, figure out missing links, and build trust with communities.

It is important to let gender diverse individuals tell their own stories and point out the limitations and potentialities of data collection. Combining statistical techniques with rigorous qualitative research (such as focus groups, in-depth interviews, and case studies) enhance chances to ask good and adequate questions that cover the various intricacies of gender identity. It also has the potential to involve people in the process, discuss the relevance of the data, map the needs of communities, and even create mechanisms to disseminate the findings later on.

Another aspect to be considered is engagement with community-based and civil society organizations. To listen to the public and advocates, the government must consider following a multistakeholder approach to the SOGI data collection, meaning individuals and organizations from different sectors – who are interested, impacted or have a role in this process – are invited to the table. Certain groups can be

---

[21] *Ibidem.*

inaccessible to government agencies, but reachable through other networks centred in providing care, support, and resources to gender diverse people facing various disparities. Working together with a willingness to learn and collaborate is key.

For an example of extensive multistakeholder engagement for collecting census data from gender minorities, the Canadian consultation process details the stages and findings.[22] Other examples of international multistakeholder engagement and policy guidance include United Nations documents.[23] Mapping the terrain, holding grasp of information needs, and brainstorming solutions can be good starting points in data collection and analyses for agencies in the U.S. However, it is important to keep in mind the limitations of multistakeholderism related to resources, meaningful participation, and skewed power dynamics.[24]

III- Privacy, Security, and Civil Rights

- **Response to 2 and 3:**

Federal agencies must adopt measures that prevent data de-anonymization relating to respondents' gender identities. Research shows that it is possible to de-anonymize and identify individuals who have gone through gender transition using only data available in public databases.[25]

In addition, it is necessary to make it clear to respondents about the intended uses of the collected data, as well as the conditions for sharing it with other agencies. Individuals with gender non-conforming identities may feel comfortable providing this type of data under certain circumstances, but they certainly would hesitate if they knew their data would be available to other federal or state agencies that could put them at risk.

Mechanisms that would protect the private information of people receiving gender-affirming care, such as the concept of "data sanctuaries" in Bill SB-107 in California[26] (and, subsequently, in other states) might be useful to consider in this context. Efforts to cut off at least some of the digital trails that civil liberties and privacy advocates fear could be used against transgender people, especially minors.

The above-mentioned right to be forgotten is also a mechanism that may improve transgender safety. The mechanism allows individuals who previously shared personal data to later exclude that data, after having a provided service. In other words, when there is no further reason to retain that data, it may be excluded to improve the privacy and safety of individuals. The right to be forgotten may also be exercised to remove gender information from ID systems, after any gender transition procedure – given the role of ID systems in the U.S. and how they overlap with Federal agencies and data collection, it is important to be mindful of the shortcomings of ID data for gender diverse communities.[27]

---

[22] Statistics Canada. (2019). The 2021 Census of Population Consultation Results: What we heard from Canadians. https://www12.statcan.gc.ca/census-recensement/2021/consultation/92-137-x/92-137-x2019001-eng.cfm

[23] United Nations. (2018). The Role of the United Nations in Combatting Discrimination and Violence against Lesbian, Gay, Bisexual, Transgender and Intersex People – A Programmatic Overview. https://www.ohchr.org/sites/default/files/Documents/Issues/Discrimination/LGBT/UN_LGBTI_Summary.pdf

[24] Sambuli, N. (2021). Five Challenges with Multistakeholder Initiatives on AI. Carnegie Council for Ethics in International Affairs. https://www.carnegiecouncil.org/media/article/five-challenges-with-multistakeholder-initiatives-on-ai

[25] Keyes, O. & Flaxman, A.D. (2022). How Census Data Put Trans Children at Risk. Scientific American. https://www.scientificamerican.com/article/how-census-data-put-trans-children-at-risk/

[26] Asiedu, K. G. (2022). California's bill for trans kids and abortion could make it 'data sanctuary'. Protocol. https://www.protocol.com/policy/california-data-sanctuary

[27] Privacy International (2021). My ID, my identity? The impact of ID systems on transgender people in Argentina, France and the Philippines. https://privacyinternational.org/long-read/4372/my-id-my-identity-impact-id-systems-transgender-people-argentina-france-and

In the federal context, the American Data Privacy and Protection Act (ADPPA)[28] does not mention gender identity in its definitions of sensitive covered data. In its current text, it only covers "sexual orientation" and "sexual behavior". Even though health and healthcare conditions are covered as sensitive data, the ADPPA should make it clear that any data regarding SOGI, as itself, is listed as sensitive covered data.

When collecting and using Indigenous data, it is crucial to consider issues of sovereignty. According to the CARE Principles, from the Global Indigenous Data Alliance,[29] "Indigenous data sovereignty reinforces the rights to engage in decision-making in accordance with Indigenous values and collective interests." Thus, building upon the CARE principles – Collective Benefits, Authority Control, Responsibility, and Ethics – or using them for guidance can strengthen trust and mitigate risks while inquiring about gender identity among Indigenous peoples in the U.S.

- **Response to 4:**

Promising practices for the collection of gender minorities' data must be evidence-based and informed by LGBTQI+ communities, combined with initiatives of capacity building and training. Gender identity and expression can be hard to grasp, and research studies show that Americans are "not paying close attention to news about bills related to transgender people."[30]

In this context, it is worth developing a culture of inclusion among civil servants in charge of data collection, data use, and data sharing. Investment in capacity building of gender diverse communities can help to embrace "diversity and inclusion as a foundational principle" and develop "comprehensive education programs on the needs and rights of the LGBTQ+ community"[31] via partnering with or subcontracting LGBTQI+ communities to develop trust. Learning from data stewardship[32] efforts can also be useful when determining promising practices for the collection of data on gender minorities.

Lastly, developing policy recommendations and guidelines for data collection alongside gender diverse communities and advocates for their rights in an inclusive multistakeholder setting is crucial to consolidate a robust framework.

Respectfully submitted,

Jess Reia, Ph.D. [they/them]
Assistant Professor of Data Science
School of Data Science, University of Virginia

---

[28] American Data Privacy and Protection Act, H.R. 8152, 117th Cong. (2022). https://www.congress.gov/bill/117th-congress/house-bill/8152/text

[29] Global Indigenous Data Alliance. (n.d.). CARE Principles for Indigenous Data Governance. https://www.gida-global.org/care

[30] Parker, K., Horowitz, J.M. & Brown, A. (2022). Americans' Complex Views on Gender Identity and Transgender Issues. Pew Research. https://www.pewresearch.org/social-trends/2022/06/28/americans-complex-views-on-gender-identity-and-transgender-issues/#h-most-say-they-re-not-paying-close-attention-to-news-about-bills-related-to-transgender-people; Blazina, C. & Baronavski, C. (2022). How Americans view policy proposals on transgender and gender identity issues, and where such policies exist. Pew Research. https://www.pewresearch.org/fact-tank/2022/09/15/how-americans-view-policy-proposals-on-transgender-and-gender-identity-issues-and-where-such-policies-exist/

[31] Marney, H. L., et al.(2022). Overcoming technical and cultural challenges to delivering equitable care for LGBTQ+ individuals in a rural, underserved area. Journal of the American Medical Informatics Association, JAMIA, 29(2), 372–378. https://doi.org/10.1093/jamia/ocab227

[32] Ada Lovelace Institute (2021). Participatory data stewardship. https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/