

THE GIFT OF FEDERAL STATISTICS: REFRAMING THE GIVENS IN U.S. PUBLIC SECTOR DATA ACCESS

Jayshree Sarathy and danah boyd

Northeastern University and Cornell University

Over the last decade, Congress has introduced a flurry of legislation in the U.S. to democratize access to public-sector data. One important channel of access is through Federal Statistical Research Data Centers (RDCs), which are secure enclaves across the country that offer external data users access to sensitive data from multiple government agencies. The RDCs serve as critical infrastructure for public science, yet the process of accessing data through these centers remains unpredictable, bureaucratic, and inequitable. Extensive research has been conducted *using* the RDCs, and scholars are just beginning to examine the role of RDCs in advancing empirical academic research [6]. Our contribution goes one step further by examining the dynamics around data access within the RDCs themselves. In this work, we seek to understand the RDCs' competing constraints around risk, trust, and equity, which complicates efforts to enact scalable open data for the public.

The struggle for more access to federal data has spanned over the last half century. In the 1960s, business economists put forth a controversial proposal to create a centralized warehouse for federal data, a National Data Center, but this resulted in a fierce Congressional battle and a renewed concern about privacy [4]. Congress rejected any proposal to centralize federal data, but many data users still believed they should have the right to access “the public’s data” that were collected using taxpayer’s money. In 1994, the Census Bureau created an alternate structure to support economic research—RDCs—which evolved to include data from the National Center for Health Statistics, Agency for Healthcare Research and Quality, Bureau of Labor Statistics, and the Census Bureau. As of this writing, a network of 33 RDC facilities operate across the country, enabling researchers at different geographic locations to apply for access and conduct multi-year analyses on federal microdata [1].

Congress has shifted its orientation towards data sharing over the decades, creating new data-forward policies like the Evidence-Based Policymaking Act,¹ the OPEN Government Data Act,² Federal Data Strategy,³ and the CHIPS Act⁴. Collectively, these proposals aim to democratize and streamline access to public-sector data in the US. Given this, we ask: How do RDCs maximize rewards while managing risks of data sharing? What are the impacts of these risk-management strategies on equity of access, transparency, and knowledge production? And what lessons can we learn about creating effective systems of public-sector data access?

Our study draws on documents, interviews, and observations. We initially analyzed historical documents and contemporary public materials such as federal notices, proposal guidelines, white papers, and legislation to understand the challenges and visions around federal data access. We then conducted 15 interviews with a diverse set of Census Bureau staff, RDC directors and administrators, academic researchers, and funders. To supplement the interviews, we attended several workshops and conferences related to federal data access, including the annual RDC business meeting.

We found that researchers’ increasing demands for access to data have complicated the relationship between government and data users. For the US Census Bureau and other agencies that contribute data to the RDCs,

¹ <https://www.congress.gov/bill/115th-congress/house-bill/4174>

² <https://www.congress.gov/bill/115th-congress/house-bill/1770>

³ <https://strategy.data.gov/overview/>

⁴ <https://www.congress.gov/bill/117th-congress/house-bill/4346>

providing data access to external researchers is critical for improving data quality and building research capacity inside government. At the same time, sharing data comes with risks that data users might de-anonymize sensitive data or use their access to undermine the legitimacy of the data collectors. To manage these tensions, RDCs have relied on bureaucratic processes, trusted networks, and limited visibility to moderate the risks of data sharing within a highly politicized and underfunded federal system. However, these strategies have been critiqued for exacerbating inequities and opacity in scientific research. More recently, legislation and proposals have failed to address these long-standing inequities in who gets access to federal data and have additionally undermined the social arrangements that enable data sharing.

To theoretically ground our discussion of the social dimensions of federal data exchange, we turn to the canonical anthropological argument presented by Marcel Mauss in his book “The Gift” [5] and subsequent work analyzing how data exchanges can be understood as data gifts [2,3]. We extend these theoretical arguments to show how public data intermediaries – like the RDCs – complicate the social exchange of data. We highlight how the RDCs must actively attend to the obligations embedded in the social exchange in order to maintain trust among all who are involved. By reframing federal data as a complex social exchange rather than simply a public resource, our work develops a more nuanced vision around open data projects and raises new possibilities for modernizing federal data access.

REFERENCES

- [1] Census Bureau. (2007). *The Center for Economic Studies (CES) at 25: A Short History*.
- [2] Ferryman, K. (2017). *Reframing Data as a Gift*. <http://dx.doi.org/10.2139/ssrn.3000631>
- [3] Fourcade, M. & Klutzz, D. (2020). A Maussian bargain: Accumulation by gift in the digital economy. *Big Data & Society*. 7. 205395171989709. 10.1177/2053951719897092.
- [4] Kraus, Rebecca. 2013. “Statistical Déjà Vu: The National Data Center Proposal of 1965 and Its Descendants.” *Journal of Privacy and Confidentiality* 5 (1). <https://doi.org/10.29012/jpc.v5i1.624>.
- [5] Mauss, M. (1925). An essay on the gift: the form and reason of exchange in archaic societies. *L'Année Sociologique*.
- [6] Nagaraj, A., Stipanovic, F., & Tranchero, M. (2025). Restricted-Access Data in Economics: Adoption, Diffusion, and Impact of U.S. Census Bureau’s Microdata. *Harvard Data Science Review* (Special Issue 6). <https://doi.org/10.1162/99608f92.19a0e1c7>