

## REFLECTIONS ON THE SUCCESSES AND CHALLENGES OF RESEARCH DATA CENTERS IN CANADA AND THE U.S.

RON S. JARMIN\*

\* Deputy Director and Chief Operating Officer, U.S. Census Bureau  
*e-mail address:* ron.s.jarmin@census.gov

**ABSTRACT.** Ron Jarmin reflects on the progress in the past twenty years of research data centers in the US and Canada. Based on his contribution to the panel on “Microdata Access: Perspectives from Canada, France, the UK and the US” at the October 2020 Canadian Research Data Centre Network (CRDCN) conference celebrating the 20th anniversary of the network.

This short paper is based on my remarks at CDRDCN20 “Public Policy & The Data Revolution: Building on 20 years of CRDCN research impact.” They reflect my perspective as a researcher and an executive who has seen both the U.S. and Canadian Research Data Center (RDC) networks from their initial foundation to their current structure and status. These remarks are, by design, brief and do not constitute an adequate history of either network, nor a complete analysis of their accomplishments and shortcomings. For interested readers, I have provided references throughout for further reading.

Let me start by saying how proud and pleased I am to be celebrating the 20th anniversary of the Canadian Research Data Center Network (CRDCN) and its partnership with Statistics Canada. While the U.S. Research Data Center network is slightly older, both systems experienced many of their growing pains together. We learned much from Gustave Goldmann and many other colleagues in those early days, and I believe they gleaned a thing or two from us as well. Although differing legal and institutional foundations impact the nature and operations of the two networks, both have become indispensable assets for research, evidenced-based policy and statistical innovation.

One of the key differences between the Canadian Research Data Center Network (CRDCN) and the Federal Statistical Research Data Centers (FSRDC) is how they were established and the consequent differences in research focus. Whereas the CRDCN was

---

*Key words and phrases:* confidentiality; privacy; research data centers.

I thank Randy Becker, Lucia Foster and an anonymous referee for helpful suggestions and Barbara Downs for Figure 1. All errors are my own. Any opinions and conclusions expressed herein are my own and do not necessarily reflect the views of the U.S. Census Bureau.

*Editorial note:* This article is an edited version of the author’s talk at the October 2020 Canadian Research Data Centre Network (CRDCN) conference. Information on the conference can be found at <https://www.crdcn20.ca/crdcn20/program>. Articles in the Perspectives series reflect the author’s opinions, and do not necessarily reflect the opinions of the journal’s editorial board.

established as a network of nine RDCs (Currie and Fortin, 2015), the U.S. RDC network started small and grew over time. U.S. RDCs originated in the mid-1990s as an expansion of the research program at the Census Bureau’s Center for Economic Studies (CES). CES itself was established in the early 1980s to develop and support research on longitudinal establishment data (see Atrostic, 2008). Public use microdata from the Decennial Census and household surveys such as the Current Population Survey had been available to researchers since the 1960s, but it was not possible to release similar datasets for business surveys due to confidentiality concerns. The only way for external researchers to access Census Bureau business establishment-level data was to collaborate directly with its staff. Moreover, there was no infrastructure in place at the Census Bureau to support such activities. Much of the early research featured collaborations between researchers at the Census Bureau and academics who were often on leave as Census Fellows.

Thus, major role for CES was to gather, curate, document and deploy business microdata to researchers. Such infrastructure already existed for Household censuses and surveys with Public Use Micro Samples (PUMS). Similar investments were not made for business data (nor, as it turned out, for the components of household surveys not surfaced in PUMS files). Neither did CES have the resources to fully deploy all relevant data initially. Early efforts focused on key manufacturing surveys where there was strong external interest among academic researchers. A norm that developed early and persists to this day is engaging the assistance of external researchers to assist documenting and improving microdata for future researchers.

The success of these early efforts increased interest in these new data resources among researchers in both academic and government settings. Recognizing the scientific value of increasing access and seeing the inefficiency of external researchers traveling to Suitland, MD, the late Bob McGuckin, the former visionary Chief of the CES—who happened to hire me (certainly not why he was a visionary)—pushed to establish the first Research Data Center located in the Census Bureau’s Boston regional office in 1994 (see McGuckin, 1992, 1995).

In 1998, the Census Bureau partnered with the National Science Foundation to develop a set of procedures to expand the Research Data Center Network even further. This model was broadly followed to establish the CRDCN (Currie and Fortin, 2015). While the network was expanding physically over these years, the data available to researchers grew even faster. Both dimensions of expansion, as well as IT issues, contributed to the growing pains for both networks.

The original, and still a leading motivation for the Census Bureau’s use of Research Data Centers was to support research access to business data. This is a key difference between the U.S. and Canadian networks. In both countries, business data are constructed from a combination of survey and administrative sources, which complicates research access. While business and economic data are not available through the CRDCN, Statistics Canada established the Centre for Data Development and Economic Research (CDER) in 2011 to permit access in Ottawa (Goldmann, 2009). CDER’s mission is similar in many respects to the early mission of CES—to develop and promote the use of business (especially longitudinal) microdata.

An early challenge to the network in the U.S. stemmed from Internal Revenue Service (IRS) concerns about RDC researchers accessing Federal Tax Information (FTI) on Census Bureau projects. These concerns not only cast doubt on the viability of RDC projects using Census Bureau business datasets, which virtually all contained commingled FTI, but also on

access to FTI by the Census Bureau more generally. The latter was a grave threat to many key Census Bureau programs and resulted in a crisis between the Census Bureau and the IRS and their associated stakeholders. Fortunately, an agreement was reached that allowed the use of FTI on Census Bureau projects, conditional on establishing that the primary purpose was to benefit Census Bureau programs (see [Greenia, 2004](#)). While adding an extra hurdle for researchers, this agreement made agencies with rich administrative data, like the IRS, important and active partners in the RDC Network.

Unlike Canada's unified statistical system, the statistical system in the United States is spread across multiple agencies. Whereas the CRDCN was able to host diverse data across several domains of Statistics Canada's broad portfolio, the U.S. RDC system was initially limited to only the social and economic data held by the Census Bureau. Starting in the mid 2000s, the Agency for Health Research and Quality (AHRQ), and the National Center for Health Statistics (NCHS) began to permit researchers to access their data at Census Bureau's RDCs under reimbursable agreements with Census. More recently, the Census Bureau's RDC system was rebranded as the Federal Statistical Research Data Centers (FSRDC) with several other federal statistical agencies joining the network. The FSRDCs are now jointly governed as a partnership between the agencies and the RDC partner institutions. In an important early initiative that follows the guidance of the Foundations for Evidenced-Based Policymaking Act to streamline access, the FSRDCs are piloting a single application portal.

The scientific value of providing secure access to confidential statistical agency microdata through RDCs has been more than adequately demonstrated by thousands of published articles and other scholarly outputs. While important, the value of the RDCs to the research community alone is probably not sufficient to gain the level of support for their operation that we have seen from agencies such as Statistics Canada and the Census Bureau. Rather, the value of the RDCs to the statistical agencies stems from three broad classes of benefits that are more direct than general research.

First, and most obviously, RDC projects extract more value from data that are collected and processed at public expense. The agencies do not have the time or resources to examine all relevant policy and research questions that could be addressed with their data, so RDCs are a way to leverage that initial public investment, derive more value added from existing data, and build support for agency programs.

Second, RDCs help the agencies improve their research capacity. Importantly, this is not just by gaining access to academic researchers working in the labs. RDCs provide a network of researchers working both within and outside the agencies, who innovate and improve measurement. As junior economists at the CES, my colleagues and I assumed the Census Bureau would no longer need us if Harvard and MIT researchers could access the same data after the opening of the Boston RDC. But rather than seeing the research function of CES diminish as the RDC network grew, we saw CES researchers thrive as they became part of expanding network of researchers using similar data and tools to explore issues in economic measurement in new ways.

Figure 1 summarizes the rich collaboration across nodes on the U.S. FSRDC network on projects where Census staff researchers are principal investigators. The "Census HQ" node is not an RDC, but represents Census Bureau staff researchers. The Census Bureau itself is the naturally the central hub as pertains to this important subset of projects using its data. Figure 1 highlights that Census staff collaborate with researchers across most of the other nodes across the network. The ties to Washington (DC), Michigan, Chicago and Boston (NBER) are particularly strong and include many productive and long-term collaborations.



innovation. An excellent example is the collaboration between academic and Census Bureau researchers to utilize new data sources, tools and improved methods to build prototype integrated price and quantity measures for the retail trade sector (Ehrlich et al., 2020). This work points the way to how to deliver on users' needs for more timely and more granular data. Projects such as this demonstrate that to be successful, we will need to find ways to deploy new data sources, tools and computing power to the RDC networks and to bring in new collaborators including those from the private sector, to move the 21st century measurement agenda forward.

## REFERENCES

- Abowd, John, John Haltiwanger, and Julia Lane.** 2004. "Integrated Longitudinal Employee-Employer Data for the United States." *American Economic Review*, 94: 224–229. <https://doi.org/10.1257/0002828041301812>.
- Abraham, Katharine G., Ron S. Jarmin, Brian Moyer, and Matthew D. Shapiro.** 2020. *Big Data for 21st Century Economic Statistics*. Vol. forthcoming.
- Atrostic, B.K.** 2008. "The Center for Economic Studies (CES) at 25: A Short History." U.S. Census Bureau. [https://www2.census.gov/ces/misc/CES\\_History\\_2007\\_Extract.pdf](https://www2.census.gov/ces/misc/CES_History_2007_Extract.pdf).
- Bayard, Kimberly, Emin Dinlersoz, Timothy Dunne, John Haltiwanger, Javier Miranda, and John Stevens.** 2018. "Early-Stage Business Formation: An Analysis of Applications for Employer Identification Numbers." <https://doi.org/10.3386/w24364>.
- Bean, Charles.** 2016. "Independent Review of UK Economic Statistics." [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/507081/2904936\\_Bean\\_Review\\_Web\\_Accessible.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/2904936_Bean_Review_Web_Accessible.pdf).
- Buffington, Catherine, Lucia Foster, Ron Jarmin, and Scott Ohlmacher.** 2017. "The Management and Organizational Practices Survey (MOPS): An Overview." *Journal of Economic and Social Measurement*, 42(1). <https://doi.org/10.3233/JEM-170441>.
- Currie, R.F., and S. Fortin.** 2015. *Social Statistics Matter: A History of the Canadian RDC Network*. Canadian Research Data Centre Network = Réseau canadien des Centres de données de recherche. <http://rdc-cdr.ca/sites/default/files/social-statistics-matter-crdcn-history.pdf>.
- Ehrlich, Gabriel, John C. Haltiwanger, Ron S. Jarmin, David Johnson, and Matthew D. Shapiro.** 2020. "Re-Engineering Key National Economic Indicators." In *Big Data for 21st Century Economic Statistics*. Vol. forthcoming, , ed. Katharine G. Abraham, Ron. S. Jarmin, Brian Moyer and Matthew D. Shapiro. <https://www.nber.org/books-and-chapters/big-data-21st-century-economic-statistics/re-engineering-key-national-economic-indicators>.
- Goldmann, Gustave.** 2009. "From a Seed to a Forest: Microdata Access at Statistics Canada." *Statistical Journal of the IAOS*, 26(3-4): 75–87. <https://doi.org/10.3233/SJI-2009-0703>.
- Greenia, Nicholas H.** 2004. "Developing Adoptable Disclosure Protection Techniques: Lessons Learned from a U.S. Experience." *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3050: 343–352. [https://doi.org/10.1007/978-3-540-25955-8\\_28](https://doi.org/10.1007/978-3-540-25955-8_28).

- Jarmin, Ron S.** 2019. “Evolving Measurement for an Evolving Economy: Thoughts on 21st Century US Economic Statistics.” Journal of Economic Perspectives, 33(1). <https://doi.org/10.1257/jep.33.1.165>.
- McGuckin, Robert H.** 1992. “Analytic Use of Economic Microdata; a Model for Researcher Access with Confidentiality Protection.” U.S. Census Bureau Working Paper 92-8. <https://ideas.repec.org/p/cen/wpaper/92-8.html>.
- McGuckin, Robert H.** 1995. “Establishment Microdata for Economic Research and Policy Analysis: Looking beyond the Aggregates.” Journal of Business and Economic Statistics, 13(1): 121–126. <https://doi.org/10.1080/07350015.1995.10524585>.