THE UNIVERSITY OF CHICAGO

Navigating Roles, Responsibilities, and Relationships in the Governance of Administrative Data in U.S. Public Sector Statistical Systems

Tia-Mae Surtees
August 2026

A paper submitted in partial fulfilment of the requirements for the
Master of Arts degree in the
Master of Arts Program in the Social Sciences

Advisor: Sara Lafia
Preceptor: Marshall Ryan Jean

Table of Contents

Abstract ***3***

Acknowledgments ***4***

Introduction ***5***

Background ***6***

Administrative Data, the U.S. Statistical System, and Those Within It *6*

Defining Administrative Data Governance and Exploring Challenges *9*

Theorizing Administrative Data Governance *11*

Data and Methods ***12***

Sampling and Recruitment *13*

Data Collection and Analysis *14*

Findings ***16***

Administrative Data Origins **16**

Data Collection and Challenges *16*

Defining Administrative Data Governance *17*

Data Sharing **19**

Federal Data Sharing Perspectives *19*

State and Local Data Sharing Perspectives *21*

Legal Challenges as Barriers *22*

Political and Relational Aspects of Data Sharing *24*

Data Access **26**

State Reflections on Different Approaches to Access *26*

Relational Data Access *27*

Data Use **29**

Data and Analytical Challenges *29*

Data Access Platforms *30*

Data Dissemination *31*

Influence and Impact of External Forces **31**

Political Forces *32*

Technological Forces *36*

Discussion ***38***

The Challenges of Administrative Data Governance *38*

Navigating Governance: Technical and Relational Strategies *39*

Governance as a Socio-Technical Process *41*

Limitations and Future Research *41*

Appendices ***43***

Appendix A: Interview Guide *43*

Appendix B: Index Codes and Definitions *46*

References ***47***

Abstract

Administrative data are an increasingly important resource for policy, research, and evaluation as survey response rates decline, and public sector and big data use expand. These data are not collected for research purposes, yet they have significant analytical value, so their secondary use in statistical contexts comes with challenges and tensions. Governing administrative data in the U.S. also means navigating a decentralized and fragmented statistical system while safeguarding the public who are represented within these data. This study examines how professionals navigate data governance, sharing, access, and use by drawing on in-depth interviews with 13 participants from government agencies, research organizations, and universities who work with administrative data at the federal, state, and local levels of government. Recurring challenges across each stage of the administrative data lifecycle were navigated primarily through relational work, which involved engaging stakeholders, building trust, and demonstrating value and mutual benefit. Political disruption and AI advancement introduced new uncertainties and pressures to this work. These findings show administrative data governance as a socio-technical and person-centered process, offering an insight into how this statistical system functions and how some of its challenges can be approached.

Acknowledgments

I am extremely grateful to Sara Lafia for supporting me with this project and being an amazing advisor, and to my preceptor Marshall Ryan Jean for his ongoing support throughout this academic year. I would also like to thank Jenny Trinitapoli whose guidance has been invaluable, especially in the early development of my project. A special thanks to Bob Goerge and Emily Wiegand for being receptive to my ideas and kindly offering their expertise. Finally, a huge thank you to my interview participants who I appreciate for being so generous with their time. I enjoyed every conversation and learned a lot from their insights.

I was awarded the Esther Johnson Student Research Award which supported me, as a MAPSS student, with my MA thesis research. I am also grateful to the US-UK Fulbright Commission for awarding me this opportunity to study in the U.S.

For my late Nana and Grandad, Pat and Pete Surtees.

Introduction

Administrative data, generated by government agencies delivering public services, are increasingly recognized as a valuable resource for research and policymaking. When data collected to administer programs within healthcare, education, and social services, for example, are combined and analyzed, they can provide large-scale, granular evidence about the lives of citizens and the systems that serve them, with the potential to be used to inform more effective policies (Fantuzzo et al. 2015). However, these data are not collected for research purposes which can result in limitations like measurement error, a lack of quality control, and high processing costs (Liao et al. 2020), meaning repurposing them is not straight forward.

Professionals that produce, manage, govern, and use administrative data do the everyday work of mobilizing data while safeguarding citizens, who are data subjects. This involves negotiation to prepare and share data for purposes beyond their original collection, balancing the benefits of data use with the risks of privacy, and navigating the competing interests of many different stakeholders. Understanding how these professionals carry out their work in this complicated context is essential to understand how administrative data governance functions in practice.

Existing scholarship documents the challenges faced when governing, sharing, accessing, and using administrative data to improve policies and practice (Allard et al. 2017; O'Hara, Shattuck, and Goerge 2017). This study seeks to add to this knowledge by exploring how professionals within these systems navigate challenges in practice. By drawing on 13 first-hand, in-depth interviews with professionals across the U.S. statistical system and the full administrative data lifecycle, this paper examines both the challenges they face and how they work through them. It also pays particular attention to the political and technological environments that professionals operate in, for example, addressing disruption at the federal level of government and the rise of AI. Overall, it offers a timely analysis into what pressures

those working with administrative data are facing and what strategies may be effective in response.

This study is guided by the question: *How do professionals working with administrative data understand and navigate their roles, responsibilities, and relationships in everyday governance practices within public sector data systems in the U.S.?* The paper proceeds by reviewing existing literature, to establish the study's contextual and theoretical background, and setting out the methods used. It then draws on the accounts of the interviewed professionals, structured by each stage of the data lifecycle, and external forces, before concluding with a discussion on what these findings mean together, and acknowledging their limitations and suggestions for further research.

Background

Administrative Data, the U.S. Statistical System, and Those Within It

Administrative data "derive from the operation of administrative systems, typically by public sector agencies" (Elias 2014, 47). These systems cover sectors such as health, tax, social security, housing, and education, where agencies are often responsible for the administration of social benefits, for example, "income maintenance, food stamps, Medicaid, nutritional programs, child support, child protective services, childcare subsidies, [and] Social Security programs" (Goerge and Lee 2002, 197). Crucially, these data are generated for "nonresearch activities," making them "a secondary data source in contrast to surveys that are conducted solely for research purposes" (Goerge and Lee 2002, 197). As put by Connelly et al. (2016, 2), administrative data are "found" data that require particular organization, management, and cleaning before analysis, in contrast to "made" data that are produced for defined research purposes.

Given these origins, administrative data offer considerable value for research[1]. In a bureaucracy where government is divided into functions to serve citizens, data from human and public service agencies provide comprehensive evidence on the needs of individual citizens that can inform policy and improve practice (Fantuzzo et al. 2015). Such data also have methodological advantages over traditional sources, such as surveys, generally offering "much larger sample sizes" and "far fewer problems with attrition, non-response, and measurement error than traditional survey data" (Card et al. 2011). The growing availability of large and granular administrative datasets has, in turn, expanded the scope for empirical research and methodological innovation (Einav and Levin 2014). Much of this potential is realized through linkage and integration, where data from across sectors can sometimes be joined at the individual level (Jutte, Roos, and Brownell 2011). This can provide comprehensive information about how policies and programs impact individuals and families, for example, for social policy evaluation (Zanti et al. 2022). The "increasing computerization of administrative records, has facilitated more extensive linking of previously disconnected administrative databases, to create more comprehensive and extensive information" on populations (Abowd, Schmutte, and Vilhuber 2021, 26). In addition, the linkage of administrative data to survey data sources can identify and improve problems associated with data quality and coverage (National Academies of Sciences, Engineering, and Medicine 2023).

These origins, however, give rise to ethical and legal concerns around consent, data sharing, and appropriate data use. Because administrative data are generated through the delivery of public services, the people represented within them, often those receiving social benefits and services, typically have not consented to their information being used for further

[1] The Abdul Latif Jameel Poverty Action Lab provide a large catalog of administrative datasets available for research purposes, and procedures and timelines on accessing these.
See https://www.povertyactionlab.org/catalog-administrative-data-sets (retrieved July 3, 2026).

research (Woollard 2014). The tensions that the secondary use of administrative data create frame the responsibilities and governance practices that this paper explores.

These dynamics play out within the U.S. federal statistical system, which "generally lags far behind other countries in making data available to researchers" (Card et al. 2011). The system has been described as "highly decentralized, with statistical activities spread across approximately 125 agencies," where principles of independence, credibility, and confidentiality can constrain innovations in data use (National Academies of Sciences, Engineering, and Medicine 2017).

This decentralization is compounded by differences across levels of government. Through administering the tax system and social programs, for example, the federal government collects vast amounts of granular administrative data, held in rich microlevel datasets maintained by federal agencies such as the Social Security Administration, the Internal Revenue Service, and the Centers for Medicare and Medicaid (Einav and Levin 2014, 7). State and local governments "similarly generate large amounts of administrative data," but with "less uniformity" than at the federal level (Einav and Levin 2014, 8). This results in a fragmented landscape in which data are shared and made available, so the systems governing it differ substantially across federal, state, and local governments.

Given these differences, there are a broad range of stakeholders involved in making administrative data available for research. Stiles and Boothroyd (2015, 125) distinguish between data owners, "who have legal or intellectual property rights or control" over data, and data users, "who use the data to create new intelligence." Data intermediaries are also involved in processes, and these individuals create and implement management and governance processes that connect owners and users to enable data use (Powar et al. 2025). The relationships between these groups are central to how administrative data move from the systems that produce them to those who seek to use them.

Defining Administrative Data Governance and Exploring Challenges

Across the collection, sharing, access, and use of administrative data, governance is of high importance. Administrative data governance is defined by Rosenbaum (2010, 1444) as the establishment of policies on the "access, management and permissible uses of data," and "the process by which stewardship responsibilities are conceptualized and carried out." Rosenbaum (2010) also frames data stewards in terms of relationships, emphasizing their responsibility not only to data owners but also those represented within data, extending beyond formal rules and procedures. Understood this way, governance relates to both directly and indirectly involved stakeholders and spans the entire data lifecycle, shaping how data are handled at each stage, which is the central focus of this study.

Much of the literature on administrative data use identifies challenges within these governance processes and across the data lifecycle, beginning at the point of collection. Because administrative systems collect data primarily to support service delivery rather than research, issues such as internal inconsistencies and measurement error can undermine data quality (Woollard 2014). From a user perspective, researchers often have limited control over the data quality and documentation, which can constrain analytical utility (Woollard 2014).

Governance is also shaped by legislation. Policies governing data can restrict opportunities for linkage and use across agencies beyond the scope for which data were originally collected, even when doing so could generate public benefit, for example, by enabling new research (Johnson, Massey, and O'Hara 2015). In the context of public health data, for instance, Jutte, Roos, and Brownell (2011) note that privacy legislation is often interpreted conservatively which can limit opportunities for new research and advancement.

Beyond formal legislation, organizational capacity and priorities affect administrative data access and use. In a report on federal data access, limited resources in statistical agencies are noted as a constraint to developing research partnerships with data users (Card et al. 2011). Accessing and linking data are frequently made difficult due to being "trapped in silos" and

dependent on cooperation from data owners, lawyers, and staff within agencies who have the knowledge and technical skills to mobilize data (O'Hara, Shattuck, and Goerge 2017, p. 71).

Data owners often express concern about how data are used. When interviewing Medicaid data owners, Stiles et al. (2011) found that their biggest concern in granting access to external researchers was the misinterpretation of the data, with many requiring that research outputs are reviewed prior to release. This is also common practice in some integrated data systems[2], where researchers share findings with agencies before dissemination for comment rather than approval, which preserves "academic freedom" by not allowing agencies to suppress results (Kitzmiller and Burnett 2015, 177).

For data users, challenges arise both from accessing data and limitations within the data themselves. Qualitative studies offer insight into these constraints. Stiles and Boothroyd (2015) emphasize that ensuring appropriate use depends on relationships and dialogue between data owners and researchers, where owners can facilitate this by offering data expertise to support analysis. Similarly, Wiegand and Goerge (2019) find that that having contextual information on data sources can help with understanding and mitigating challenges associated with being able to link data accurately. Allard et al. (2017) explain that administrative data users in human service agencies report limited resources and lengthy data-sharing agreements as barriers, alongside privacy and data quality concerns. They argue that further research is required on how organizational "workflows, structure, and culture" shape data use (Allard et al. 2017, 10). This highlights the value of this paper that examines how institutional and relational dynamics influence administrative data access and appropriate use in practice.

Finally, because research using administrative data directly affects citizens by improving social policies and interventions (Fantuzzo et al. 2015), and because citizens

[2] Integrated data systems bring together and provide secure access to public administration data on citizens to "provide a comprehensive picture of the whole person" (Fantuzzo et al. 2015: 19). These are often at the local or state level, and can be executive-based, agency-based, and university-based (Kitzmiller and Burnett 2015).

themselves are the subjects of these data, they have a clear stake in how the data are used. In a review of qualitative studies on public responses to the sharing and linkage of health data for research, Aitken et al. (2016) find that concerns center on confidentiality, consent, surveillance, and the potential for misuse when research is not in the public interest. Noting that many studies demonstrate low public awareness of research practices, they conclude that transparency towards how data are used, and meaningful public engagement, can strengthen trust and support for data linkage. This is also noted by Kitzmiller and Burnett (2015) where communicating the results of research strengthens the responsibilities that integrated data systems have to the public.

Theorizing Administrative Data Governance

The reviewed literature on this topic identifies a range of challenges across the governance, sharing, access, and use of administrative data, but pays less attention to how the professionals involved understand and navigate these challenges in their everyday work. This study draws on sociological theories to make sense of this, including social constructionism, bureaucracy, and relational data governance.

Berger and Luckmann's (1967) perspective on the social construction of knowledge states that knowledge is not objective but is produced through social interactions and institutional processes. Applied to administrative data, Zanti (2024) draws on Berger and Luckmann's theory to demonstrate how stakeholders involved in, or excluded from, integrated data systems shape the data themselves, and subsequent research using these data. Ultimately, administrative data are not neutral records and are influenced by how they are collected and the processes through which they are used. This frames governance practices as socially constructed, being interpreted and negotiated by the individuals who carry them out as they navigate their roles, responsibilities, and relationships.

In this study, a recurring theme from participants was the bureaucratic nature of the institutions they worked in, relating to formal procedures, rules, and hierarchies that shaped how data were governed. To conceptualize this, I draw on sociological theories of bureaucracy (Weber [1922] 1968; Lipsky 2010) by framing data professionals as bureaucratic actors operating within a rational-legal structure, exercising authority and navigating relationships to govern data on public citizens. Alongside this, a relational lens considers the interactions between these formal structures and roles. Viljoen's (2021) "Relational Theory of Data Governance" conceptualizes data governance as a set of institutional arrangements and complex interactions among actors. It reframes governance as relational rather than individual, where collective, population-level interests and harms arise from the production, sharing, and use of data.

Together, these perspectives frame administrative data governance not as a fixed set of rules or technical procedures, but rather something that professionals must interpret and navigate in their everyday work. It is shaped by the bureaucratic structures they work within, the relationships between them, and the wider political context. Exploring how professionals make sense of this and how it influences their day-to-day work with administrative data calls for a qualitative, interview-based method which is set out in the next section.

Data and Methods

The central question of this study is: How do professionals working with administrative data navigate their roles, responsibilities, and relationships in preparing, sharing, and using data for purposes beyond their original collection? This paper's aim is to explore how participants' priorities and objectives are pursued within complex and bureaucratic public sector systems, and what challenges, tensions, and trade-offs arise when balancing the benefits of data use with the risks of privacy.

Sampling and Recruitment

Between February and April 2026, I conducted 13 in-depth interviews with participants involved in the preparation, sharing, and use of administrative data for reasons beyond their original purpose. Participants worked at, or had previously worked at, government and statistical agencies, research organizations, and universities. As a recruitment framework, I considered three types of professionals: *data owners*, *data users*[3], and *data intermediaries*[4], presented in Table 1. These groups are not clear-cut, but recruiting across them supported a purposive sampling strategy designed to capture a range of perspectives.

Professional type	**Definition**	**Participants**
Data owners	Individuals working within federal, state, or local government departments or agencies who govern administrative data, including making decisions about how data are used and by who.	N = 4
Data intermediaries	Individuals involved in the management of data infrastructures or processes that facilitate data preparation, access, linkage, and use. These may work within government or academic statistical organizations, including integrated data systems and research institutes at the national, state, or local level.	N = 7
Data users	Individuals who use federal, state, or local government data for research and analytical purposes, often accessing through a data infrastructure. This group includes researchers and data analysts working in academia and research organizations.	N = 2
		Total N = 13

Table 1: Professional types, definitions and sample counts.

In practice, there was more overlap between these groups than the framework anticipated. Data intermediaries frequently had experience as data users, and those that were expected to be data owners had a more intermediary role. Rather than working within data producing agencies, they acquired and controlled administrative data across different federal and state agencies and made decisions about their use. As a result, data owners who are defined as those producing and controlling internal agency data are not represented in this sample. The

[3] Stiles and Boothroyd (2015, 125) define administrative data contributors as *data owners* "who have legal or intellectual property rights or control" and *data users* "who use the data to create new intelligence".

[4] Data intermediaries are defined by Powar et al. (2025) as those who create and implement management and governance processes that connect data owners and users to enable data use. In this study, they represent most participants. They are in a distinctive position, engaging with both owners and users, while at times acting as owners or users themselves, often holding roles that cut across these groups.

fact that participants did not fit neatly into these three types of professionals is consistent with this study's findings which show that roles within administrative data processes are fluid and relational.

The sample reflects federal, state, and local government perspectives, including participants from five federal agencies and national organizations, and representation from six U.S. states. To maintain participant confidentiality, names and organizational affiliations are not reported. Instead, participants are identified by the characteristics of their role and organization, and the level of government they work or worked at. Identifying details within quotations have also been replaced with bracketed generic descriptors.

My recruitment strategy began with contacting participants through my academic network. From these initial contacts, I utilized snowball sampling to be put in contact with additional participants. This method is appropriate given the specialized nature of these professionals. I then used the professional types and levels of government to guide purposive sampling, specifically recruiting participants where representation was limited. While deliberate steps were taken to increase variation in the sample, this is a convenience sample and is not representative of the entire U.S. public sector statistical system. The aim of this qualitative approach is to explore the richness and depth of a select number of cases to identify patterns and relationships, not to make causal claims (Gerson and Damaske 2020).

Data Collection and Analysis

I prepared an interview guide with core questions on participants' roles, responsibilities, priorities, the data they work with, and their perspectives on administrative data governance (see Appendix A). Prior to data collection, I piloted the interview guide with a subject matter expert and revised it based on their feedback. The guide included specific modules tailored to different roles in data governance, sharing, access, and use processes. Questions for *data owners* were about data sharing, *data intermediaries* about managing and

facilitating data use, and *data users* about experiences accessing and using data. The interviews were semi-structured which meant questions could be drawn from more than one module as participants' multiple roles often emerged throughout discussions. This approach also allowed participants to guide the conversation and for topics to emerge organically, beyond the interview guide questions. When topics did not arise naturally, I asked probing questions. An example demonstrating this flexible approach is that the interview guide did not include specific questions about relationships in data governance processes. Instead, I had a list of probes to ask when participants described working with a stakeholder. Not asking directly about important relationships and instead allowing participants to describe the significance of their interactions with stakeholders helped to reduce social desirability bias.

Interviews lasted between 43 and 70 minutes long, with an average duration of 59 minutes. They were digitally recorded with the consent of participants[5] and transcribed verbatim[6]. Within 24 hours of each interview, I wrote reflective and analytical memos. Transcripts were then imported into MAXQDA (26.1.0) for coding and qualitative data analysis. I employed the flexible coding approach outlined by Deterding and Waters (2021), which involved two rounds of coding. In the first round, I applied index codes that anchored to the interview guide questions and other broad themes. The index codes covered different elements of the administrative data governance process: priorities and responsibilities; current challenges; data sharing; data preparation and support; data access; data use; relationships; and future challenges (see Appendix B for definitions). In the second round, I applied fine-grained analytical codes to the indexed sections and examined how the distribution of these codes varied both between and within professional types. Following this approach and applying a

[5] This study was approved and marked exempt by the Social and Behavioral Sciences IRB in February 2026, and verbal consent was sought as per the IRB regulations.
[6] Transcription was conducted using the University of Chicago "SSD Scribe" service.
See https://sscs.uchicago.edu/transcription/ (retrieved July 2, 2026).

"uniform qualitative formula across the sample" resulted in increased consistency and reliability in my coding and analysis (Deterding and Waters 2021, 731).

Findings

Administrative Data Origins

This section draws on how participants described administrative data and understood its governance to establish the context for the analysis that follows.

Data Collection and Challenges

Participants mostly had experience with administrative data from public sector agencies at the federal, state, and local levels of government in the U.S. Through discussing administrative data governance, many acknowledged the context around which data are collected. A state Chief Data Officer (CDO) described the collection as agencies "doing the business of government, providing services, regulating industries," collecting data as a "byproduct." This secondary nature is beneficial, offering cost savings and the potential to improve government operations, as noted by a state integrated data system (IDS) manager:

> If data is an important asset now in the world and if there's ways that we can use data that are collected anyway for multiple purposes, that's just getting more value out of things that we're spending money on as taxpayers anyway. … If we can use that to help improve how these services or programs are administered, then even better, right?

However, utilizing these data beyond their original purpose resulted in challenges and tensions. When acquiring state data, one federal statistical professional noted that state agencies were not technically equipped to extract and share data because their roles were primarily operational, involving the management of programs and data entry. One academic data user noted that, because agencies collect data to meet their own needs, data are often not structured for analysis. The secondary nature of these data offer both strengths and limitations, being a cost-effective byproduct of government activity, but not designed or resourced for further use.

The overall challenges that data professionals faced in using administrative data for research stemmed from the operational context in which data are collected, and the possibilities

and limits that this creates for their secondary use. This reflection from a data intermediary, supporting states with administrative data use, raises a question central to data governance:

> Administrative data is messy and imperfect, and not collected for research, and so how do we document it, work with the people creating the data to improve the quality over time, and make smart decisions about which data we curate to actually use for research projects?

Their question on how administrative data are documented, and how decisions are made for future use, align closely with the governance challenges that this analysis seeks to explore.

Defining Administrative Data Governance

Participants were asked what administrative data governance meant to them, and responses generally related to policies and decision-making processes around how data are collected, managed, and used. For those working to acquire data from federal and state agencies, they frequently referred to obligations that they felt they had to agencies, like adhering to terms that data providers had set. Another theme was how governance processes varied significantly. A participant in a data acquisition role noted a lack of standardization, as agencies "all have their unique governance requirements to acquire and to use the data" for approved research projects.

Data users also felt they had obligations, and these related to security, privacy, and mitigating the risk of disclosure. Administrative data were frequently described as "public data" that needed to be protected since, as one participant put it, "these are real people that these numbers relate to." Therefore, many participants also emphasized obligations to citizens. A data intermediary explained that their team practices good governance by thinking carefully about who they consult to understand data, not just data providers or those directly involved in research, but also the people represented in the data. An IDS manager emphasized that being "good stewards" by "protecting the confidentiality" of data subjects was essential to encourage citizens to continue to give their data to the government, thus enabling further research. Another participant had similar thoughts, describing a responsibility to safely handle the public's data, while also using it to "make their experiences better."

In general, participants noted that proper governance is required so the government can use the data it collects to improve how it operates. A data intermediary believed that good governance supports "cross-sector collaboration," and that enabling administrative data use can improve "the siloed nature of public program data" so public sector employees can use their data to inform policy and practice. Collaboration was a central theme and governance processes were described as "people centered." One state-level data intermediary saw data governance as more than a "data management" definition, but as involving "people and structures that make decisions about how data are used." This suggests that governance in practice is structural and highly human-centered. One university researcher offered an analogy that captures the complexity of administrative data governance:

> It's like the sandbox everyone plays in where it's the container of all the rules and processes and things that help everyone function together. … More specifically, it's the policies and procedures, who the stakeholders are that have to be involved in each step of determining what happens with data, what we do with it, how we access it, collect it, analyze it, interpret it, share it back out – go through that whole data lifecycle.

This reflection foregrounds the entirety of the data lifecycle, from collection through to dissemination, and I adopt this as an organizing framework for this analysis.

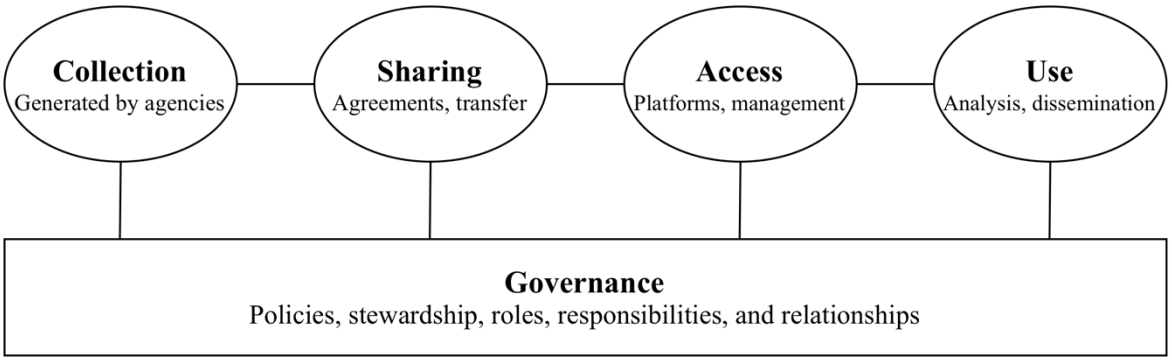


Figure 1: Administrative data lifecycle.

Figure 1 sets out the main stages of this data lifecycle and how governance underpins each one. The figure focuses on the secondary use of administrative data by those external to the agencies that generate data, rather than internal use by agencies themselves. It is intended as a

visualization of the lifecycle's main stages, which simplifies a nuanced and often non-linear process that involves much back-and-forth and other stages that may not be presented. Having already considered administrative data collection, I will explore how these data are shared, accessed, and used for research, before findings are then disseminated.

Data Sharing

After administrative data are collected, they may be used internally by the agencies that collect them or shared and acquired by external parties for secondary use. This typically involves establishing data sharing agreements between agencies and the intermediaries or users seeking access. In their responses, data professionals noted two factors as central to navigating challenges within the data sharing process: technological advancement and the building of positive relationships with data owners. These dynamics varied across different levels of government, which I examine first at the federal level, then at state and local levels.

Federal Data Sharing Perspectives

Data sharing processes at the federal level were varied and lacked standardization. To illustrate this, I draw on the accounts of two officials with experience in federal statistical agencies, both responsible for acquiring administrative data from elsewhere in the public sector to be used for research purposes. Firstly, they noted how technical progress had eased data sharing over time, where manual transfers were replaced with better tools and skills for handling large datasets. Describing the broader structural context, one noted that the U.S. is "decentralized with their data sources." The other described how federal agencies each had a "unique set of hurdles that need to be cleared" to acquire data. They summarized that, "the standard is, it's not standard," demonstrating how federal data acquisition was negotiated on a case-by-case basis.

The two officials emphasized how establishing a data sharing agreement could be lengthy, where completing one in less than six months "was probably a miracle," and more typically the process took somewhere between 12 and 18 months, while one had been in

negotiations that took around five years. When discussing what slowed these processes down, both used examples of acquiring state-level data. They pointing to resourcing challenges, cuts to spending, the absence of large legal departments, and a lack of staff or time within states to dedicate to this work. One official said that the challenge was related to capacity rather than unwillingness:

> Unless you have an advocate in the state, the whole process would die, or delay for years because no one wanted, not that they didn't want to, but they just didn't have the bandwidth to put those resources in place to keep an agreement going.

However, having a state advocate seemed to ease the process, and the official explained that negotiating effectively with advocates meant offering a "quid pro quo" for acquiring data, since agencies were not required to share data. A dedicated point of contact also helped by accommodating discussions "between the [the statistical agency] and the source agency." Again, this highlights how central people were to the everyday work of administrative data governance, involving much communication and negotiation. This framing of data sharing as "quid pro quo" suggests mutual benefit, a theme that recurred throughout the data lifecycle.

Resourcing constraints also shaped how data sharing agreements were sustained. When an agreement was in place, one official suggested that states often deferred to the federal agency: "Going back to how busy and resource strapped the states are, sometimes they just trusted us that it was OK and moved on." Part of this trust was down to the federal statistical agency's reputation and ability to anonymize and protect data. On building this trust, the official referred to an example, getting buy-in from agencies for their administrative data to be used for determining citizenship:

> The program takes a while to build up. … Many agencies didn't agree with that in the first place, and so that took some negotiation and building of trust that we really didn't get everybody's buy-in till very late … and even some of that buy-in was reluctant, only based on executive orders and [agency officials] enforcing that at a lower level. So, it does take time. The relationships don't happen overnight; they have to be built upon.

While executive orders could be applied to enforce sharing, this example demonstrates how trust with data owning agencies was built over time. Overall, these relational elements were pivotal, where advocacy, negotiation, and trust determined whether data sharing agreements were reached and sustained.

State and Local Data Sharing Perspectives

Data sharing processes at the state and local level also lacked standardization. To illustrate how these played out in practice, I draw on the accounts of two state data officials, an IDS manager based at a university and a Chief Data Officer (CDO) within a state agency, both leading in acquiring administrative data from other public sector agencies.

For the IDS manager, acquisition was not straight forward because "every agency is at a little bit of a different point in their data sharing journey." Their accounts also underscored relationship building with agencies as a way of navigating data sharing. For example, rather than seeking to acquire the entirety of the agency's data from the outset, they approached this beginning with one-off research projects as a way of "easing people into" the process:

> If we were to just jump in the deep end and say, 'OK, we want to link all your data' and suck you into the system, that's a really big ask of someone, especially if we don't have any history with them. So, it's a way for both sides to see if it works.

Like the federal official's example, relationships were built upon, and the IDS manager explained that a "slow process is exactly to build up that trust." Then, once a data sharing agreement was reached, relationships had to be sustained by continuing to "provide value" and prove that sharing was "worth it." Access relied on mutual benefit where both sides gained from the relationship, where data sharing also met internal agency goals. The CDO described a similar approach, ensuring that agencies in their state remained involved and in control of their own data. They described their framework as "decentralized and devolved back to the agencies," where they worked with "a point person for each agency" and provided support that enabled them to follow their own processes and governance structures. This echoes the federal

example, where a dedicated point of contact eased the process. In both cases, governance depended on specific individuals, rather than processes and policies alone.

The decentralized nature of this work and differences across agencies resulted in varying data sharing processes. To demonstrate this, the CDO ran a survey with state agencies that revealed "a couple hundred" forms of written agreements for sharing data. In response, they developed an "interagency data sharing framework" and "common shared templates." These were optional to adopt, which ensured agencies maintained the control, but offered a solution to making data sharing processes more consistent.

Overall, ways to ease the challenges of state and local data sharing involved building trust over time, modifying data sharing structures and frameworks, and investing resources to ease processes and work better with people in data owning agencies.

Legal Challenges as Barriers

In negotiating these agreements, participants repeatedly encountered legal challenges. For an academic user on the receiving end of data sharing, legal processes were described as a "bottleneck," though they recognized their importance for the "protection of the data and the privacy of the people who participate." For one federal official, legal risk created barriers within agreement negotiations as agencies feared being held liable for data used beyond their original purpose:

> That's why I say the agreements become the deal breakers. … It's the lack of trust because it's something where all these agencies are like, 'I collected it for one purpose, you're now using it for something else. What if I get sued by somebody who says you weren't supposed to be sharing that?'

These legal concerns had relational effects, where the fear of being sued broke down levels of trust. Participants that worked with data owners had privacy and security concerns, such as not wanting to violate privacy laws. This was compounded by legal complexities, where some

participants described U.S. privacy laws as “confusing” and a “complex web,” particularly at the state level where laws were more specific and individual.

These difficulties were also apparent when working with lawyers themselves. Participants understood that lawyers wanted to mitigate risk, which resulted in tensions when negotiating agreements, and sometimes led to sharing and research not happening altogether. As one state official put it, agency lawyers were “great in protecting their own data, and that's a benefit, and that can be a challenge if you're trying to put a data sharing agreement in place,” making the process “a constant search to try to find middle ground.” Lawyers were often not data security or data sharing specialists which led to further difficulties. For example, two professionals that helped states to mobilize and integrate their administrative data gave similar accounts on working with lawyers. One explained that “it takes time to bring attorneys up to speed on what is legal” since it might be the first time that they encounter U.S. privacy laws. The other pointed to how law schools do not offer specific training in this area, so lawyers are learning how to approach legal research on data sharing and privacy on their own.

This same professional offered two solutions to help navigate legal processes. The first was to engage lawyers as “partners in the process” from the start, rather than bringing them in only for “a signature at the end.” Involving them early in data sharing design helped increase their understanding and made them “more likely to feel like they're part of solving a problem.” The second was to draw on outside expertise and precedence, pointing lawyers to existing agreements and successful cases of data sharing, which helped them to “understand that there are templates and ways to do this.”

Overall, the legal side of data sharing was complex and created barriers. What helped navigate these barriers was engaging those who work with the law, involving lawyers as partners, and drawing on templates and examples to demonstrate workable ways forward.

Political and Relational Aspects of Data Sharing

Participants pointed to political and relational dynamics that shaped agencies' willingness to share data, which produced challenges like resistance and a lack of institutional buy-in. To navigate these, a key strategy that they noted was to engage agencies and stakeholders directly.

Agencies' resistance to sharing data was often political, related to protectiveness over the public's data and concerns about exposure. Participants described agencies responding to potential data sharing with scrutiny, skepticism, and suspicion as they recognized the damage that could be done if the personal data they owned ended up in the "wrong hands." A data intermediary said this protectiveness stemmed from a lack of understanding of the safeguards in place, which was further complicated by the number of steps involved in processing and working with data, creating "a lot of places where there can be miscommunications." Protectiveness also related to more political forms of caution. One participant described agencies as sometimes acting "politically defensive" since sharing data left them "more vulnerable" to external scrutiny and questioning. Another provided an example where some agencies thought the risk of the public learning about poor outcomes for a population, through the findings of administrative data analysis, outweighed the benefits of data sharing to their organization. In their view, this highlighted a need for institutional buy-in.

Related to the importance of buy-in was the issue of political misalignment. An official at the federal level noted that misalignment particularly influences negotiations between "federal agencies and state governors." Sometimes agencies declined to share data because they did not want to be associated with a specific project or agenda. One university IDS manager gave an example of how political sensitivity could disrupt a project, even one an agency thought was worthwhile:

> In one case, [the project] was about the Dobbs decision, reversing the abortion rights, and the proposed research project wanted to look at the impact on women's health in general, not specifically about abortion. But the [state agency] was like, 'we cannot take this heat right now, ... it seems like a worthy project, but we just can't do it at this moment.' So, there are times where it's just a no, and we move on, but it's also important

> that agencies feel that we will hear that and respect that too. That helps build that relationship where they feel like they actually have some control over it, instead of just pushy academics going and sneaking around and doing stuff.

Their identity as an academic shaped this interaction, and respecting the agency's refusal meant they could build an equitable partnership to sustain a positive working relationship.

Solutions to overcome the resistance of agencies depended largely on communication. Some participants emphasized the importance of explaining processes and offering assurances in a clear and balanced way that was "not too complicated or overly simplified." This communication also involved demonstrating the value and mutual benefit that can come from sharing data to secure institutional buy-in. This value could be practical; for example, an IDS manager explained that agencies sharing their data reduced the administrative burden of having to deal with data access requests directly. Another participant, an intermediary with experience in an IDS, described how this practical value helped to develop relationships with agencies:

> One of the best ways we build relationships with [agencies] is being responsive to their kind of real-time needs around data and data strategy because that's not their expertise. So, we've kind of built a relationship where we provide something above and beyond what they can do with the staff that they have. … That ended up being critical to relationship building too. They know I am being very meticulous about how I'm treating the data, and I'm also trying to deliver something of value back to them along the way.

Here, the participant used capacity, resources, and expertise to add practical value to the agency.

Other participants noted that demonstrating financial value was a strategy for securing institutional buy-in to acquire data. Cost savings could come from administrative data being an existing data source that can be used strategically. For example, one professional working within an IDS explained how some agencies lacked "stability within the larger health and human services system" because of public scrutiny and the threat of their organization being restructured or losing funding. In this context, the IDS could help the agency use their data in a defensive way to "make a case for their value." Similarly, a federal data intermediary described how state agencies are required to meet many federal benchmarks, and that administrative data can help an agency to understand "the underlying dynamics of how that

benchmark is either met or not met" which, in their words, moves "the agency to a place where it can perform better, which means potentially more money." More bluntly, in persuading leadership for the institutional buy-in to share data, they said: "The sales part of it isn't about confidentiality ... or privacy. It's about cost savings. That's the only thing that gets the government's attention right now. We can save money, and we can reduce burden." Therefore, an effective lever was to frame data sharing in terms of cost saving and financial benefit.

Overall, navigating the tensions within administrative data sharing was technically, legally, and politically challenging. Across each of these dimensions, relational work eased processes through building trust, securing buy-in, and demonstrating value.

Data Access

The next stage in the administrative data lifecycle is access. Participants described users accessing data via federal and state agencies, national and academic research organizations, IDSs at the state and local level, and other access platforms like Federal Statistical Research Data Centers[7] (FSRDCs) and Administrative Data Research Facilities[8] (ADRFs). As with data sharing, access processes were not standardized. I will first draw on examples from state data professionals to explore how processes differed, then conclude by reflecting on the relational access requirements that participants navigated.

State Reflections on Different Approaches to Access

One state official, responsible for enabling access to other agencies' data for research, described how processes had shifted from manual and case-by-case to now happening within an Administrative Data Research Facility (ADRF). Previously, staff would develop and package

[7] Federal Statistical Research Data Centers (FSRDCs) are secure facilities that enable approved researchers access to restricted data, including censuses, surveys of businesses and households, linked data, and administrative records from federal and state agencies and other sources.
See https://www.census.gov/about/adrm/fsrdc.html (retrieved July 2, 2026).

[8] Administrative Data Research Facility (ADRF) systems, developed by the Coleridge Initiative, are secure platforms that enable administrative data access, linkage, and use across agencies.
See https://www.fedramp.gov/marketplace/products/FR1819057982/ (retrieved July 2, 2026).

extracts for each request, which carried security risks as users often still could access data when agreements expired. The ADRF addressed this by improving security and governance, where access and usage could be tracked more easily. Notably, platform-based access also enabled collaboration through the linkage of data from multiple agencies, which the participant described as a "synergy... that doesn't happen when you're just transmitting data."

Another participant, a user of local and state administrative data, noted having direct access to a case management database. An advantage to this approach was that it allowed them to resolve problems within the database quickly and independently, without "a bunch of different people who control different parts" of its access. They also experienced accessing data through an ADRF, which was more secure and eased "legal hurdles." However, they described that this process had considerably more stages and people involved, which made it "much more difficult to just get through the process" than when they handled data themselves. They thought the access platform was "useful for solving the problems that it exists to solve," but was not the only approach to access, demonstrating the lack of standard approaches.

Relational Data Access

Access to administrative data often depended on specific connections or knowledge, where relationships with data owners or intermediaries were necessary, and knowledge of data and processes were required to be successful in accessing data.

Firstly, participants described needing an "internal sponsor" or "trusted agent" to access data. Some systems did not allow external requests altogether, allowing only those affiliated with a given state or institution. As one state IDS worker put it, someone could be "the best researcher in the world, but if you don't have any connection to [the state] or the [university research institute], we're just not gonna open up the data to you." This was largely to ensure that data were used in ways that aligned with an agency's remit and served the public good.

The way that users presented their proposed research when requesting access was also important. For one IDS manager, much of their intermediary work between agencies and users involved explaining what data was required and how it would be used. To ease this back and forth, they emphasized that projects should provide a thorough explanation of intended data use and show a benefit to the agency to justify them "disclosing something that they're not obligated to disclose." For example, they would "tweak every request to highlight how the findings would be relevant to each agency's mission." Another data intermediary, drawing on their experience in an IDS, related access requests to the broader importance of relationships:

> Sometimes outside researchers who don't really know about the IDS or that much about working with policy partners. They might just fill out the form for a data request and skip the knowing people, relationship building part, or not really know where to start. So, they start with the data request. But it's more successful, in general, if you know the people and you can make sure your request is aligned with what they're interested in, what their needs are.

This emphasizes how building relationships and demonstrating value and public good were tangible strategies for being more successful in accessing data.

Where users had secured a relationship, a lack of internal knowledge on data themselves could remain a barrier. One academic researcher was attempting to access local administrative data and worked hard to find an internal partner who provided them with documentation to help specify the tables and columns they required. However, without firsthand experience with the data, navigating this documentation proved difficult as information was "vague" and required "guesswork." The user summarized: "I feel like I got probably 85% of the way to the level of precision that I needed, and the last 15% hinged on some internal knowledge."

Taking participants' accounts together, administrative data access was difficult to navigate, shaped by technology, the relationships and connections that a user had available to them, and their knowledge of data themselves. Platforms like ADRFs eased access by improving security and collaboration, but the relational aspects of data access meant that building trust and connections were vital. Interestingly, one federal official discussed the

National Secure Data Service[9] (NSDS), a proposed initiative for accessing administrative data, describing how the NSDS's "data concierge service"[10] could act as "that trusted agent recognized by at least all federal authorities and state authorities." They suggest that this approach may begin to bridge the relational gap, offering a solution where the "trusted agent" is embedded into the infrastructure itself, meaning users did not require specific connections or relationships, making administrative data access more equitable overall.

Data Use

Once administrative data were accessed, their use brought specific governance challenges, like needing contextual knowledge to analyze data, technical constraints within access platforms, and lengthy processes to disseminate results.

Data and Analytical Challenges

Using administrative data accurately and appropriately was considered a part of good governance by participants. As explored earlier, the secondary nature of administrative data that are collected operationally, rather than for answering research questions, led to challenges at the point of analysis. A common difficulty was the lack of documentation. Describing their experience with understanding "messy" hospital claims data, one data user explained how this relied on institutional knowledge: "There's basically one person's job in that research group [that] existed just because they knew the history of this one database at this medical system, and that knowledge was not written down anywhere." This knowledge was person-dependent and resided with individuals, a feature that appeared across many interviews. When documentation did exist, it was often limited. For example, one state data intermediary that supported users described how documentation often was, "here's the data, here's how it's

[9] The National Secure Data Service (NSDS) is a government wide project for strengthened data linkage and access infrastructures. See https://ncses.nsf.gov/initiatives/national-secure-data-service-demo (retrieved July 2, 2026).
[10] The NSDS data concierge service is intended to support users of the proposed NSDS infrastructures. See https://www.americasdatahub.org/wp-content/uploads/2025/03/DCS_23_N03-Environmental-Scan-and-Outreach-Final-Report_Public.pdf (retrieved July 2, 2026).

organized," rather than "here's what it means, here's what's reflected in it, here's what you can and can't know from it." This contextual knowledge on the meaning of administrative data was not captured within the basic documentation that often existed.

Solutions to this involved working with those who held the knowledge, likely within the data producing agency, and documenting it well for others to use. As one IDS intermediary described, this process depended on "a lot of communication and back-and-forth" between their IDS, or the researcher accessing the data, and the source partners who understood the data and how they had been collected. Many other participants also noted a "back-and-forth" element to understanding data. Enabling collaboration was often the role of data intermediaries, with one explaining that they do "a lot of connecting and relationship managing," like between individual researchers and data owning agencies. This collaboration between data owner and user again shows the person-to-person nature of this work, in this case, the sharing of institutional knowledge to use administrative data more accurately.

This in-depth knowledge could also be developed by data users themselves. One participant described an analytical team in a local IDS that had built "good historical knowledge" through "multiple iterations of data warehouse redesigns" in an agency they worked with. Their connections and longstanding expertise were an asset that helped others with questions on data use, and this knowledge exchange was informal, described as "crowdsource[d]" and "word of mouth." Generally, knowledge on how to use administrative data travelled through people and networks as well as documentation, demonstrating how the processes to make these data usable were largely person-centered.

Data Access Platforms

Platform-based access enabled appropriate data use, offering safeguards, security, and collaboration, but participants also experienced these controls as limiting. One user, working within a Federal Statistical Research Data Center (FSRDC), explained how getting new

statistical packages approved for use would have taken around a year, and that these restrictions resulted in "a long process in doing anything innovative." While approval may have been possible, they believed there was a "speed issue in the bureaucracy," where platform protections often limited their work at the cost of time.

Data Dissemination

Once users had produced research outputs, they were typically reviewed before results could be published. This was often a requirement in data sharing agreements, giving agencies a way of knowing what findings were being released that used their data. Participants expressed that these reviews did not give agencies the power to reject or veto findings, which preserved researcher independence and academic freedom. They were disclosure reviews to mitigate agency concerns, reflecting the legal worries that they had to protect data and data subjects.

Users understood the need for these reviews, but their lengthy nature proved frustrating. One described these reviews as "good," but wished they would "move faster." Another found that reviewers were "so particular" about what could be reported, and processes were "kind of painful" given how long they took. However, users largely accepted these obstacles as they safeguarded data and ultimately enabled access from the data owners themselves.

Influence and Impact of External Forces

The previous sections have examined the governance challenges that relate to the secondary use of administrative data within and across each lifecycle stage. This section looks outward, as a key idea that emerged within interviews was how external forces imposed change and shaped governance processes. Participants frequently described these forces that were out of their control and impacted how they navigated administrative data governance and use. External forces were driven by political disruption and technological development. Political and technological forces were also felt together. One participant made this connection, which provides context to this section that explores how professionals approached challenges:

> It's just the uncertainty associated with the larger [political] climate, … and there's lots of other technological things that … can cause disruption too, but you add all of this together and it's just so messy. It's like 'what do you do with this?' You just muddle through the dark.

Political Forces

An overwhelming majority of participants identified disruption at the federal level of government as a threat to their work with administrative data. The first challenge was navigating uncertainty, where government shutdowns, cuts to funding, and limited resources led to disappearing data, the loss of institutional knowledge, and difficulties in accessing data. A second challenge related to mistrust, among both agencies and the public, in how the federal government uses data. While solutions were explored, participants emphasized the knock-on effects of disruption, particularly the negative impact on data quality and its wider implications.

Uncertainty

The most noted sources of disruption were government shutdowns, funding cuts, and resourcing challenges across the federal statistical system. One manager of a national data archive captured the nature of this disruption:

> Neglect causes just as much harm as actually cutting funding for something or trying to harm something, and you lose people. People leave the public sector. People stop doing this sort of work. You don't invest in systems. So, things just kind of hobble along and it creates these unanticipated problems, and then it casts doubt on the quality of it.

They suggest that a lack of investment in data systems has knock-on effects. People leaving the public sector was a recurring concern. A data intermediary supporting federal data access pointed to staffing and funding constraints as a key challenge, particularly the loss of institutional knowledge as experienced staff leave the federal statistical system. They warned that these losses have the potential to "compound," with effects that might "feel very, very big in five years." This is especially pertinent with administrative data because, as discussed earlier in this paper, the knowledge required to be able to use it well is historical, specific, and person-dependent, with limited documentation. When individuals left, that knowledge left with them.

The same participant concluded that this disruption is "not a level of stability that's appropriate for consistent quality data production." Therefore, instability also had a negative impact on data quality, as well as the people and systems that sustained data use. Even within this disruption, the federal data intermediary emphasized relational solutions: "That's where we're trying to maintain positive relationships with data providers. … We are aware that they are facing some pressure from the other side too." This emphasizes the importance of recognizing the impact of wider constraints when building relationships with stakeholders.

Another consequence of this disruption was disappearing data, where data were no longer collected or able to be accessed, given cuts to funding and resources. One academic data user, a professor, had their access to restricted federal data revoked, and experienced uncertainty that slowed down their work. In response, they anticipated having to "pivot to other data sources," likely shifting their research questions entirely. Similarly, a pre-doctoral researcher explained how government shutdowns had limited their access to an in-person FSDRC, and how their data access was then revoked entirely. The federal agency owning the data cited "limited resources" as the reason for this, and that access might be restored "in the near future" through a "virtual data enclave." This instance of processes being replaced in response to limited resources also echoed a concern raised by the intermediary supporting federal data users:

> There are processes that break and the person who can fix them isn't there, and maybe the process was old and brittle, so it gets replaced with something more efficient, and that's great, but do you have time to do the A/B testing to make sure the new process is giving you what you expect? I think it's that staffing and funding constraint.

In this sense, uncertainty from resourcing and staffing pressures impacted researchers as well as the processes and systems meant to sustain data use.

Mistrust

Another political influence on administrative data governance was mistrust in government uses of data. As one state intermediary put it, navigating mistrust was "tricky in the U.S. context."

Concerns came from the agencies themselves, who were wary of how their data might be used by others in government. One state CDO identified the "current federal administration, and the real push for unfettered access to state data" as a challenge. They explained how this had changed the system's feelings towards data sharing:

> [It] creates an environment where there're more concerns about sharing and using data than there would have been in the past because people believe, probably correctly, that there's increased risk. … There're a few cases where [the federal government] have gone sort of directly to state agencies to access data so that's created both, I think, direct and indirect pressure … and creates real challenges.

These apprehensions about sharing data were based on concerns that data would be used for harm. An intermediary working with agencies warned that skepticism about data sharing would compound existing capacity constraints: "There's already a lot of little places where the process can get derailed as it is, and if there's even lower capacity and lower willingness to share data in the future, that's gonna complicate that further."

Data professionals also noted that mistrust in government uses of administrative data was a concern of the public. One IDS manager anticipated "greater questioning of the integrity of the federal statistical system from the people who actually represent the response rates." Many participants framed this mistrust in relation to the potential for harm. For example, a data intermediary described "a foundational shakeup of public trust" in the government's use of administrative data, attributing this to a federal government "doing things that are unprecedented and illegal with data." Several participants recognized how this mistrust threatened people's willingness to engage with public services. As an example, one state CDO referred to proposals for "a national elections database of people to use for immigration enforcement" and warned this could undermine "people's willingness to engage with services." Another participant considered the wider consequences of this as a "domino effect", where people engaging less with systems and programs affects "the state and local people who are working to protect that data and use it responsibly" and, in turn, impacts administrative data

quality. This knock-on effect was noted to negatively affect the representativeness, completeness, and quality of administrative data themselves.

Rather than offering solutions to resolve this public mistrust or its effects on data quality, participants focused on providing assurances to protect what trust remained. One state level intermediary saw the "long term erosion of what administrative data can tell us if we've got whole populations opting out of public services" as a "secondary challenge." Instead, their immediate priority was: "How do we stay trusted in our state and how do we ensure we can tell our constituents truthfully that their data won't be misused in that way, that we've got protections in place for that?" For several participants, these protections meant strengthening internal governance processes. An IDS researcher's priority was to answer the question: "How do we advance what we're doing for the public good while also not ... putting people at risk, or violating people's trust?" A data intermediary similarly reported that state data stewards were focused on strengthening and upscaling governance processes to ensure their data would be protected if requested by federal agencies. Interestingly, these protective behaviors arguably reversed what may once have been described as good governance. One intermediary gave an example of a state that had:

> pushed really towards full transparency and a lot of community engagement around their data work, and public data dictionaries of all of their linked data … and they, in the last year, have removed all of that, and you need a state credential to access the metadata about what's in the administrative data systems because they don't want federal agencies looking at that.

They said that transparency, often a key component of trustworthy data practices (Aitken et al. 2016), now seemed "more risky than beneficial." This may compound the existing lack of available information and documentation on administrative data, making them harder to access and use for public good. This clearly demonstrates how external political pressure in the U.S. has reshaped the meaning of responsible administrative data governance.

Technological Forces

As noted throughout the data lifecycle, technology plays a central role in the governance of administrative data. This section explores it as an external force which brought both new opportunities and challenges. The most prominent technological concern for the future of administrative data governance was the influence of artificial intelligence (AI).

AI Governance

Some participants noted that there were expectations that AI could solve problems which led to external pressures on how administrative data were produced and used. Because AI solutions require large volumes of data, one state CDO anticipated pressure on states to produce and publish AI-ready data. However, participants saw clear risks in this, particularly around disclosure and potential harm. An IDS manager raised the concern of data leakage, where agencies had "concerns about people using a chatbot where the data leaves the secure environment and then gets mixed in with the training data." However, they explained that the level of concern varied, where some agencies were more cautious, wanting data use agreements amended to prohibit AI use, while others wanted the IDS to do more with their data and AI. They concluded that navigating these differences will be a "challenge as everybody gets comfortable on what AI means and what their concerns are about it."

For several participants, they responded to the rise of AI by adapting their governance practices. One data intermediary believed that a way of responding to the pressure of using AI responsibly would be to treat it as a "tool that people will have to integrate into their workflow," with clear policies, approved tools, and defined uses so the public sector statistical system does not get left behind "while everyone else advances on it." The CDO framed the development of AI governance as a process, working with agencies to "foster AI innovation," but only after establishing "privacy protections and policies for data quality and data governance." In their words, they highlighted the need for establishing a "good policy framework" before entering "the wild west of AI."

AI and Administrative Data Use

Some data users described how access platforms constrained innovation. The pre-doctoral researcher felt that outdated technology within infrastructures and slow approval processes did not empower researchers using data. They also reinforced the earlier concern that the private sector drove innovation through technological advancement in data use, while the public sector statistical system risked being left behind.

Participants considered how AI could reshape uses of administrative data, such as linkage and integration. However, for one intermediary within the IDS space, new technologies did not change this:

> Doing data integration within strong governance and legal frameworks, ethical frameworks, engaging people, I think the core of that is the same. AI can't really change that. It could mean that places do really poor data integration quickly because they can [with AI] … but in the true IDS community … we have a set of values for how we operate.

Even with the introduction of new technologies, this participant believed that the work of mobilizing, linking, and using administrative data for the public good remained a relational process, maintaining the governance principles that this data community are founded upon.

A related concern was algorithmic bias. One local IDS manager emphasized that, "in this age of data technologies being used through machine learning models," it was essential to not "lose track of the context under which the data is generated." Their solution centered on people, ensuring that "stakeholders that are most impacted by the way in which we're going to portray the data … are involved at the very root, at the very beginning of the data analytic process." In other words, building responsible algorithms depended on understanding the lived experiences of those represented within data, or those impacted by decisions using it, suggesting that new technologies cannot be blindly automated without human context.

Across these accounts, technology and AI promised to accelerate administrative data work, but it remained fundamentally human, dependent on knowledge held by people, trust

built through relationships, and an understanding of data in context. Some participants believed that technology cannot replace this. This work was relational at its core, where the human context seemed to become more important as technologies and tools developed to guarantee that data and populations are represented and protected.

These external forces raised unique challenges for participants across administrative data governance, sharing, access, and use. These were largely outside of participants' control, for example, driven by pressure to share and make data available. Political disruption and technological change created a complicated and uncertain landscape, and solutions were not so clear cut. However, participants did note strategies to navigate these complexities, for example, updating or establishing frameworks to ensure data and data subjects were safeguarded.

Discussion

The Challenges of Administrative Data Governance

The insights of this qualitative study reveal that there are common patterns in the challenges faced by professionals working across the data lifecycle to collect, prepare, access, and use administrative data. These challenges were rooted in the secondary, operational nature of data themselves, because administrative data are collected to deliver public services rather than for research, so using them for research led to recurring tensions. This finding is consistent with existing literature which links administrative data's origins to concerns about quality and consent (Woollard 2014), resourcing constraints and lengthy agreements that constrained use (Allard et al. 2017; Card et al. 2011), and the difficulty of obtaining permission to access linked data (O'Hara, Shattuck, and Goerge 2017). By examining the full data lifecycle rather than a single stage, this study shows that these challenges are not unique to one point in the process, but recur at different points across data governance, access, and use.

These recurring challenges include differences between agencies and processes, resourcing and time constraints to properly mobilize data, and legal and political caution

around the risks of secondary use. The lack of standardization in processes meant that professionals were required to navigate a complex landscape. This relates to theories of bureaucracy (Weber [1922] 1968), particularly street-level bureaucracy (Lipsky 2010), where professionals working within a rational-legal structure of formal rules and laws had to exercise discretion to carry out their day-to-day work in mobilizing administrative data.

In contrast to the challenges within the data lifecycle, the external forces of political disruption and technological change that participants experienced resulted in uncertainty and mistrust that threatened administrative data use more widely. Particularly in relation to mistrust, participants expressed a growing concern among both agencies and the public about government uses of data, relating to literature on public concerns about administrative data use (Aitken et al. 2016). This had consequences concerning data themselves, where populations engaged less with public programs and services and data became less complete and representative. Viljoen's (2021) relational theory of data governance applies here as harms were not solely individual, but occurred at the population level, impacting whole communities.

These external forces threatened the day-to-day governance and uses of administrative data, where the relationships, trust, institutional knowledge, and resources that professionals utilized to navigate the data lifecycle were impacted by political disruption and funding constraints. Therefore, these findings emphasize how administrative data governance is not neutral or purely technical but is political and shaped by the wider context in which it takes place. This is consistent with the social construction perspective (Berger and Luckmann 1967) as governance depended on the institutions, practices, and relationships within systems.

Navigating Governance: Technical and Relational Strategies

Returning to the research question, this study finds that professionals navigate the governance of administrative data through two types of strategies: technical and relational. A common theme was how technical innovations helped to ease the challenges that professionals faced,

particularly the relational challenges of working with data owners and users. Data access platforms and integrated data systems often standardized processes, and governance in general, by centralizing access, which could make data available to those without the connections or resources to negotiate bespoke arrangements. These technical solutions may be more pertinent as the use of AI expands and makes rich administrative data more valuable. Some participants were clear that the public sector must invest in capabilities to not to fall behind other sectors.

While technology eased this work, relationships were often what made it possible. The most prominent theme across interviews was how central the relational dimensions of this work were. Even where technology helped, participants felt that it was not a solution to everything. For example, technical assurances alone did not generate the trust required for data to flow through the lifecycle. The people and relationships within these systems were critical to stewarding data responsibly, developing individual and context-dependent governance frameworks, and ensuring that data were not misused. This finding echoes Rosenbaum's (2010) framing of data governance as extending formal rules and processes, and Viljoen's (2021) relational data governance, where stewardship is about the management of relationships.

A central strategy that professionals emphasized for navigating the governance of administrative data was to build and sustain positive relationships. They noted using relationships to gain access to data, involving stakeholders across the whole data lifecycle, and cultivating relationships over time. This was slow and incremental work within a complex landscape, but ultimately helped to build trust and secure buy-in. Similarly, collaboration, clearly communicating intentions, and demonstrating mutual benefit repeatedly helped to move work forward, where data sharing was reframed as something both parties gained from. Also, much of the knowledge required to use administrative data was person-dependent; for example, contextual information is necessary to navigate technical challenges and use data appropriately

(Wiegand and Goerge 2019; Stiles and Boothroyd 2015). Relationships were vital to acquire this knowledge, but also to understand and document it for future use.

Governance as a Socio-Technical Process

To conclude, this study explores how administrative data systems function, and how professionals working within them approach the challenges and tensions that arise as they navigate their roles, responsibilities, and relationships. Ultimately, this study shows that governing and mobilizing administrative data is a socio-technical process, shaped by both relational and technological aspects. In Science and Technology Studies, technological systems are explored as networks of people, technologies, and institutions that interact in ways that each shape, and are shaped by, the systems themselves (Hughes, 1987). In this study, technical infrastructures that enable administrative data use depended on trust, collaboration, and human judgment to sustain them. Interpersonal relationships and political dynamics shaped these socio-technical systems just as significantly as more tangible and visible features like policies, legal requirements, and technologies. Therefore, to understand administrative data governance, attention must be paid to how people navigate technical data systems, specifically their judgements and personal connections that these systems depend on.

Limitations and Future Research

A key limitation of this study is that the professionals interviewed were mostly those who bring together or use external administrative data, rather than those in data owning agencies who work to collect data or prepare data to be shared. Many participants shared their experiences working with people in these roles and were able to give insight into the challenges they face, like resourcing and time constraints. However, the first-hand experiences of those within data owning agencies are under-represented here. Therefore, speaking directly with those individuals would give a more complete picture of these data processes. This study also focuses on public sector administrative data, but third-party, corporate, and private sector

administrative data were also referenced in a handful of interviews. While this study does not present those findings, it could be interesting to explore the differences in accessing and using these data, in contrast to public sector data.

This study also has an intentionally broad scope, intended to cover the entire data lifecycle with participants in a range of roles at different geographies and levels of government. This provides a strength as patterns across the system are identified. However, the diversity and complexity of participants' accounts suggest that there is value in conducting more focused research. For example, examining a single stage of the data lifecycle, a particular level of government, or region, or data type would likely provide more specific strategies and solutions on how professionals navigate challenges, and these could provide more tangible solutions to apply in particular circumstances.

Through interviewing individuals across the U.S. statistical system, it was clear that there is a community of professionals, connected through organizations, networks, and channels, who are deeply committed to using data for the public good to improve the lives of citizens. Administrative data governance is a complex and expansive topic, likely to become only more complicated in the future. Continued research that explores in detail the connected experiences of those within these systems will therefore be essential. Such findings could enable professionals to better share data and grant access to researchers and policymakers for more innovative uses of administrative data for the public benefit, in ways that also maintain the confidence of the public.

Appendices

Appendix A: Interview Guide

CORE QUESTIONS

Role, background, and governance as practice

1. To start, can you describe your current role?
    - How long have you worked in this role?
    - Have you had any other related roles?

2. When you think about administrative data "governance", what does that mean to you?
    - How does this look in your day-to-day work? What responsibilities do you have? What decisions are you involved in?
    - What rules, policies, or formal processes shape what you do?
    - To whom do you feel accountable?

3. Which types of data do you typically work with/use?
    - Which levels of government do these data come from (federal, state, or local)?
    - Which populations are represented in the data?
    - What makes the data particularly valuable, sensitive, or challenging to work with?

4. What is your organization's main priority or responsibility when working with administrative data? (I.e. to produce, share, or use data).
    - Who is typically involved in deciding these priorities?
    - What considerations are there when carrying out these priorities? (I.e. legal, regulatory, or policy requirements, external considerations like researchers, other agencies, the public).

PROFESSIONAL-SPECIFIC MODULES

A. Data Producers

1. Can you describe how decisions are made about whether your agency/department's data can be shared or reused?
    - What factors are typically considered in these decisions?
    - Who is usually involved in these decisions, and who has the most influence?
    - How do you try to get buy-in from stakeholders?
    - In what ways do political, legal, or resourcing considerations or constraints shape these decisions?
        - Can you give a recent example?
        - How do these considerations affect opportunities for positive data use?
        - If [**insert consideration that they mention**] didn't exist, how do you think data sharing and use would be different?

2. Can you give an example of a typical data sharing process?

3. What visibility do you have into how data are used by others once shared?
    - Can you give a recent example?
    - Who are the typical data users? Are they internal or external?
    - Are there aspects of this data reuse that raise concerns for you?
    - How do you ensure that data users are able to use the data appropriately?
    - How do you navigate any tensions between data quality, confidentiality, and reuse?

B. Data Intermediaries

1. Can you walk me through your experience acquiring/facilitating/managing administrative data?
    - Which stakeholders are involved in this process?
    - Are there any points in this process that tend to take longer or be more challenging? Why do you think this is?
    - Has there ever been a time when you were unable to access certain data?
        - What implications did this have on your work/organization?
        - If access to that data had been possible, what difference do you think it would have made to the work that you do?

2. Do you play a role in mediating between data producers and data users? If so, can you walk me through what this process looks like?
    - How do decisions about appropriate data use get made?
    - What factors do you think make a data user more likely to be successful in accessing data?
    - How do you ensure that data users are able to use the data properly?
    - Does this process change depending on the type or sensitivity of the data (i.e. public-use files vs more restricted data)?
    - Do you know of any examples where data has been misused?

C. Data Users

1. Can you walk me through your experience accessing administrative data?
    - Think about a time when you were able to successfully access the data you needed. What do you think contributed to your success?
    - Are there any points in this process that tend to take longer or be more challenging? How do you try to combat this?

2. Has there ever been a time when you were unable to access data that you needed?
    - What implications did this have on your work?
    - If access to that data had been possible, what difference do you think it would have made to your work?
    - What do you think would have made getting access to this data easier?
    - Have specific data infrastructures or technologies made access easier?

3. How do you ensure you are using administrative data appropriately?
    - Are there agreements, guidelines, or ethical standards that you must follow? Who sets these?
    - Do you receive materials from data owners or data stewards, like documentation, code books, or support from people themselves?
    - Do you have to report findings back to data owners?

Navigating relationships

[Ask as probes to the above questions when participants describe working with stakeholders, navigating processes and relationships.]

I.e. stakeholders across agencies or systems (data owners/users/intermediaries, researchers, policy makers, lawyers).

- How did your relationship with this stakeholder develop?
- How important has building or maintaining relationships with this stakeholder been to your work?
- What impact does this relationship have on your work, or meeting your priorities, or for general data access/use?
- Can you give an example of a recent interaction with this stakeholder?
- What mechanisms help you to collaborate effectively with this stakeholder?
- Is trust important in this relationship?
- Has there ever been a time where your priorities or goals have differed? What was this likely due to? How did you navigate this? If not, what systems or practices do you think help ensure you and this stakeholder stay on the same page?
- Do you think your relationship with this stakeholder reveals anything about how the broader data system operates?

[Identify other important stakeholders throughout the interview and ask these questions.]

Reflections

5. What do you think will be the biggest challenges in governing administrative data over the next five years?
    - Potential answers:
        - Changes in data availability, linkage, or technology.
        - AI-readiness and the big data revolution.
        - Shifts in public attitudes or political contexts.
    - What solutions do you think there are to these challenges?
    - How prepared do you think your organization/agency is to address these challenges?
    - What do you think would help data professionals to navigate these challenges? (I.e., support, resources).

Appendix B: Index Codes and Definitions

Index Code	**Definition**
Priorities and responsibilities	Captures how participants described their professional roles and duties, including what they saw as most important to their work with administrative data.
Current challenges	Captures the difficulties and barriers that participants referenced facing in their work, including those relevant to the characteristics of administrative data, and institutional and political factors.
Data sharing	Captures discussions of processes and experiences related to the sharing of administrative data between agencies, organizations, and individuals. Includes establishing processes, data sharing agreements, and other legal factors.
Data preparation and support	Captures insights related to preparing administrative data for secondary use, including cleaning, linkage, documentation, and giving advice to data users.
Data access	Captures experiences related to who can access administrative data, how they are accessed (i.e. through access platforms), and what processes participants went through to gain access.
Data use	Captures how participants used administrative data, including analysis and processes involving controls and reviewing outputs.
Relationships	Captures different interactions, connections, and dynamics between participants and other stakeholders, including key themes like collaboration, trust, mutual benefit, and specific relationships (i.e. with owners and the public).
Future challenges	Captures participants' reflections on anticipated and emerging challenges and risks in the future of administrative data work, including political factors, public attitudes, logistical and technological issues, and AI.

References

Abowd, J. M., Schmutte, I. M., & Vilhuber, L. (2021). Disclosure limitation and confidentiality protection in linked data. In A. Y. Chun, M. D. Larsen, G. Durrant, & J. P. Reiter (Eds.), *Administrative records for survey methodology* (pp. 25–60). John Wiley & Sons.

Allard, S. W., Wiegand, E. R., Schlecht, C., Datta, A. R., Goerge, R. M., & Weigensberg, E. (2018). State agencies' use of administrative data for improved practice: Needs, challenges, and opportunities. *Public Administration Review, 78*(2), 240–250. https://doi.org/10.1111/puar.12883

Aitken, M., de St Jorre, J., Pagliari, C., Jepson, R., & Cunningham-Burley, S. (2016). Public responses to the sharing and linkage of health data for research purposes: A systematic review and thematic synthesis of qualitative studies. *BMC Medical Ethics, 17(1)*, Article 73. https://doi.org/10.1186/s12910-016-0153-x

Berger, P. L., & Luckmann, T. (1967). *The social construction of reality: A treatise in the sociology of knowledge*. Doubleday.

Card, D., Chetty, R., Feldstein, M., & Saez, E. (2011). *Expanding access to administrative data for research in the United States* [White paper]. National Science Foundation. https://eml.berkeley.edu/~saez/card-chetty-feldstein-saezNSF10dataaccess.pdf

Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research, 59*, 1–12. https://doi.org/10.1016/j.ssresearch.2016.04.015

Deterding, N. M., & Waters, M. C. (2021). Flexible coding of in-depth interviews: A twenty-first-century approach. *Sociological Methods & Research, 50*(2), 708–739. https://doi.org/10.1177/0049124118799377

Einav, L., & Levin, J. (2014). The data revolution and economic analysis. *Innovation Policy and the Economy, 14*, 1–24. https://doi.org/10.1086/674019

Elias, P. (2014). Administrative data. In A. Duşa, D. Nelle, G. Stock, & G. G. Wagner (Eds.), *Facing the future: European research infrastructures for the humanities and social sciences* (pp. 47–48). Scivero Verlag.

Fantuzzo, J., Culhane, D., Rouse, H., & Henderson, C. (2015). Introduction to the actionable intelligence model. In J. Fantuzzo & D. P. Culhane (Eds.), *Actionable intelligence: Using integrated data systems to achieve a more effective, efficient, and ethical government* (pp. 1–38). Palgrave Macmillan.

Gerson, K., & Damaske, S. (2020). *The science and art of interviewing*. Oxford University Press.

Goerge, R. M., & Lee, B. J. (2002). Matching and cleaning administrative data. In M. Ver Ploeg, R. A. Moffitt, & C. F. Citro (Eds.), *Studies of welfare populations: Data*

collection and research issues (pp. 197–219). National Academy Press. https://doi.org/10.17226/10206

Hughes, T. P. (1987). The evolution of large technological systems. In W. E. Bijker, T. P. Hughes, & T. J. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 51–82). MIT Press.

Johnson, D. S., Massey, C., & O'Hara, A. (2015). The opportunities and challenges of using administrative data linkages to evaluate mobility. *The ANNALS of the American Academy of Political and Social Science, 657*(1), 247–264. https://doi.org/10.1177/0002716214552780

Jutte, D. P., Roos, L. L., & Brownell, M. D. (2011). Administrative record linkage as a tool for public health research. *Annual Review of Public Health, 32*, 91–108. https://doi.org/10.1146/annurev-publhealth-031210-100700

Kitzmiller, E. M., & Burnett, T. C. (2015). The AISP network: Three organizational models for building, using and sustaining integrated data systems. In J. Fantuzzo & D. P. Culhane (Eds.), *Actionable intelligence: Using integrated data systems to achieve a more effective, efficient, and ethical government* (pp. 169–190). Palgrave Macmillan.

Liao, D., Berzofsky, M.E., Couzens, G.L., Thomas, I. & Cooper, A. (2020). Improving Quality of Administrative Data: A Case Study with FBI's National Incident-Based Reporting System Data. In Hill, C. A., Biemer, P. P., Buskirk, T. D., Japec, L., Kirchner, A., Kolenikov, S., & Lyberg, L. E. Hoboken (Eds.), *Big Data Meets Survey Science: A Collection of Innovative Methods* (pp. 217–244). NJ: John Wiley & Sons.

Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public services*. Russell Sage Foundation.

National Academies of Sciences, Engineering, and Medicine. (2017). *Innovations in federal statistics: Combining data sources while protecting privacy*. The National Academies Press. https://doi.org/10.17226/24652

National Academies of Sciences, Engineering, and Medicine. (2023). Blended Data: Implications for a New National Data Infrastructure and Its Organization. in *Toward a 21st Century National Data Infrastructure: Mobilizing Information for the Common Good* (pp. 75–108). Washington, DC: The National Academies Press. https://doi.org/10.17226/26688

O'Hara, A., Shattuck, R. M., & Goerge, R. M. (2017). Linking federal surveys with administrative data to improve research on families. *The ANNALS of the American Academy of Political and Social Science, 669*(1), 63–74. https://doi.org/10.1177/0002716216678391

Powar, J., Janssen, H., Cloete, R., & Singh, J. (2025). From policy to practice in data governance and responsible data stewardship: System design for data intermediaries. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2491–2504). ACM. https://doi.org/10.1145/3715275.3732164

Rosenbaum, S. (2010). Data governance and stewardship: Designing data stewardship entities and advancing data access. *Health Services Research, 45*(5, Pt. 2), 1442–1455. https://doi.org/10.1111/j.1475-6773.2010.01140.x

Stiles, P. G., & Boothroyd, R. A. (2015). Ethical use of administrative data for research purposes. In J. Fantuzzo & D. P. Culhane (Eds.), *Actionable intelligence: Using integrated data systems to achieve a more effective, efficient, and ethical government* (pp. 125–156). Palgrave Macmillan.

Stiles, P. G., Boothroyd, R. A., Robst, J., & Ray, J. V. (2011). Ethically using administrative data in research: Medicaid administrators' current practices and best practice recommendations. *Administration & Society, 43*(2), 171–192. https://doi.org/10.1177/0095399710390644

Viljoen, S. (2021). A relational theory of data governance. *Yale Law Journal, 131*(2), 573–654.

Weber, M. [1922] (1968). *The theory of social and economic organization*. Translated by Henderson, A. M. & Parsons, T. New York: Free Press.

Wiegand, E. R., & Goerge, R. M. (2019). *Using and linking administrative datasets for family self-sufficiency research*. Family Self-Sufficiency and Stability Research Consortium.

Woollard, M. (2014). Administrative data: Problems and benefits. A perspective from the United Kingdom. In A. Duşa, D. Nelle, G. Stock, & G. G. Wagner (Eds.), *Facing the future: European research infrastructures for the humanities and social sciences* (pp. 49–60). Scivero Verlag.

Zanti, S., Berkowitz, E., Katz M., Hawn Nelson, A., Burnett, T.C., Culhane, D., & Zhou, Y. (2022). Leveraging Integrated Data for Program Evaluation: Recommendations from the Field. *Evaluation and Program Planning* 95(102093). https://doi.org/10.1016/j.evalprogplan.2022.102093

Zanti, S. (2024). *The Theory, Methods, and Value of Integrating Data to Advance Social Policies and Human Services.* Doctoral dissertation, University of Pennsylvania. https://repository.upenn.edu/server/api/core/bitstreams/0d9962d0-5366-49e1-8386-3c3328e8163f/content