



6-2018

Data Sharing Governance and Management

ADRF Network Working Group Participants

Follow this and additional works at: https://repository.upenn.edu/admindata_reports

Recommended Citation

ADRF Network Working Group Participants, "Data Sharing Governance and Management" (2018). *ADRF Network Working Group Reports*. 2.
https://repository.upenn.edu/admindata_reports/2

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/admindata_reports/2
For more information, please contact repository@pobox.upenn.edu.

Data Sharing Governance and Management

Description

The Data Sharing Governance and Management Working Group focused on data intermediaries that help to expand access to administrative data for a broad range of researchers. The group interviewed 11 data intermediaries and identified 9 unique functions that they serve to facilitate the data sharing process between data providers and researchers.

Chair: Ken Poole (Center for Regional and Economic Competitiveness)

Lead Author: Monica King (ADRF Network)

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Data Sharing Governance and Management

JUNE 2018

By Working Group Participants



ADMINISTRATIVE DATA
Research Facilities **NETWORK**

Acknowledgments

The following people served as Working Group participants. They worked collaboratively to develop the working group charge, generate insights on the findings, and contribute to the writing and editing of the report.

- Ken Poole, Center for Regional Economic Competitiveness (chair)
- Monica King, ADRF Network (lead author)
- Trent Alexander, ICPSR
- Richard Clinch, Jacob France Institute
- Dennis Culhane, Actionable Intelligence for Social Policy
- Matthew Gee, University of Chicago
- Amy Hawn Nelson, Actionable Intelligence for Social Policy
- Kimberly Korejko, Delaware Valley Regional Planning Commission
- Graham MacDonald, Urban Institute
- Mallory O'Brien, Medical College of Wisconsin
- Andrew Reamer, George Washington University
- Maryanne Schretzman, NYC Center for Innovation through Data Intelligence
- Lars Vilhuber, Cornell University

We are grateful to our interviewees for their participation in the interview process. We would also like to thank the Alfred P. Sloan Foundation for making this work possible.

Contents

Acknowledgments 1

Summary 3

Introduction and Background 3

Timeline and Process 4

Inclusion Criteria 5

About the Data Intermediaries..... 6

Motivations for Data Sharing and Data Use 6

Data Intermediary Functions 8

Organizational Structure and Governance 12

Data Intermediary Maturation Process 13

Challenges and Recommendations 14

Conclusion..... 16

References 17

Appendix 1: Interview Protocol..... 18

Appendix 2: Data Intermediaries and List of Other Interviewees..... 20

Appendix 3: “Call for Examples” 21

Summary

This report summarizes findings from the Administrative Data Research Facilities (ADRF) Network’s working group on Data Sharing Governance and Management. Over the course of six months, we interviewed 17 organizations and identified nine functions unique to administrative data intermediaries. These findings advance our understanding of how data intermediaries facilitate secure, ethical, and efficient data sharing between data providers and a broad range of researchers.

Introduction and Background

Launched in June 2017 with funding from the Alfred P. Sloan Foundation, the ADRF Network is comprised of researchers, practitioners, and other stakeholders working to improve how administrative data are accessed and used for social science research and policy. One of the ADRF Network’s early initiatives was to form three working groups around high priority issues and questions in the social science research space. The three working group topics are 1) “Data Quality Standards”, 2) “Communicating about Data Privacy and Security”, and 3) “Data Sharing Governance and Management,” the focus of this report.

The topic of data sharing quickly emerged as a high priority issue. Currently, many sources of high-value administrative data are shared through a one-to-one relationship between senior researchers and data providers. These relationships usually take many years to develop, and the legal negotiations that ensue further add time before data is even accessed. The status quo is inefficient for both parties and hampers knowledge creation in the social sciences. However, data intermediaries, or trusted third party entities, have emerged as a model to share data more efficiently and to a broader range of researchers while addressing the concerns and protecting the interests of data providers.

In November 2017, the working group on Data Sharing Governance and Management formed under the leadership of Ken Poole. At our first meeting, we were motivated by the question, “**How do data intermediaries help share proprietary, restricted-access, or private administrative data to a broader set of researchers?**”

We first defined *administrative data* as data that are collected by governments, businesses, and other organizations for recordkeeping or operational — but not research — purposes. Moreover, we were interested in administrative data that are proprietary, restricted-access, and/or private, since those present significantly more barriers to access compared to public and open-access administrative data.

On the other hand, we entered the working group without a set definition of *data intermediaries*. During the initial group discussions, it became clear that we broadly conceived of data intermediaries as organizations that help facilitate a data sharing relationship between researchers and data providers, but the specific roles and responsibilities of data intermediaries remain unclear. We therefore decided to use the working group process to learn more about administrative data intermediaries and characterize their unique functions.

This report represents the working group’s initial effort to study and document the unique functions of data intermediaries in the sharing of proprietary, restricted-access, or private administrative data. In its charter, the working group participants agreed to the following:

- 1) Identify organizations that might serve as administrative data intermediaries
- 2) Interview these organizations and identify intermediary functions
- 3) Provide recommendations for the ADRF Network in advancing the work of data intermediaries

Timeline and Process

In December 2017, the working group began populating a list of academic institutions, government agencies, and other nonprofit organizations in the social sciences that serve as administrative data intermediaries. Monica King, the director of the ADRF Network, worked with the group members to draft a semi-structured interview protocol, which evolved over the course of the interview process (Appendix 1). Broadly speaking, interviewees were asked to describe their organization and walk through how they worked with data providers and data users.

From January to May 2018, Monica conducted interviews with the organizations, typically with an executive director or an operations manager (Appendix 2). The interviews lasted between 30 minutes and one hour. After each interview, Monica followed up with the interviewee and asked for referrals to other similarly-positioned organizations. This snowball sampling process helped to expand the list of data intermediaries generated by the working group members. To recruit additional interview participants, we also distributed a “Call for Examples” in March 2018 to help attract organizations not yet connected with this effort that self-identified as administrative data intermediaries (Appendix 3).

Inclusion Criteria

Although almost all of the interviewees self-identified as administrative data intermediaries, our report focuses only on those that ultimately met the following criteria:

- 1) Serve data users who are primarily researchers¹ (academic or otherwise)
- 2) Have policies and procedures in place for data access and use (as opposed to making ad-hoc decisions)
- 3) Support virtual or physical location where data are managed, queried, or used for analysis
- 4) Serve a broad range of researchers across multiple projects (not limited to those that are collaborating with data providers on a project-basis)

We developed these four criteria after the interview process concluded in order to identify data intermediaries that facilitate the sharing of administrative data and are equipped to accommodate a large number of researchers. A number of organizations that we interviewed met some but not all of the criteria above, indicating that there may be a maturation process with data intermediary who might start facilitating data sharing on a project-by-project basis before scaling up to accommodating a broad range of researchers and research projects. We developed a

¹ For the purposes of this report, we defined researchers as data users who use data to advance scientific knowledge. We recognize that administrative data intermediaries are valuable for other populations of data users.

preliminary maturation process for administrative data intermediaries toward the end of the report.

About the Data Intermediaries

Of the 17 organizations interviewed, 11 fulfilled all of the above criteria and are included in our findings below (Appendix 2). Eight of the 11 are based at a university, two are nonprofits, and one is part of the government. The 11 organizations are from three different countries – the United States, Canada, and the United Kingdom. Many of the organizations were established in the last ten years, reflecting the increased capacity and interest in data sharing that has emerged in recent years.

These newer data intermediaries were often established as a result of legislative changes that either directly or indirectly created demand for data intermediaries. As an example, the Institute for Research on Innovation and Science (IRIS) expanded on a previous initiative, STARMETRICS, that helped universities comply with new federal reporting requirements that came with the post-recession stimulus package for research grants. IRIS creates reports for the universities on metrics such as how many jobs were created by their grants while also sharing de-identified data securely with qualified researchers. Similarly, the Texas Education Research Center was created as a result of state legislature to establish longitudinal data infrastructure that connect education and workforce outcomes in Texas. NORC Data Enclave is yet another example that was developed in response to the Confidential Information Protection and Statistical Efficiencies Act (CIPSEA) of 2002. We did not get a clear picture of how older (those that began prior to 2000) data intermediaries were established, though there is evidence that they gradually grew from the successes of smaller projects and partnerships.

Motivations for Data Sharing and Data Use

Data intermediaries exist because they bring a value-add to the data sharing process between willing data providers and data users. There has been a lot of interest in why data providers, particularly private companies, share data (Future of Privacy Forum, 2017). But here we build on that question and ask, “why do data providers share data with researchers through data intermediaries?” During the course of the interview process, we documented a number of

benefits of sharing data for research through data intermediaries. We briefly highlight them below:

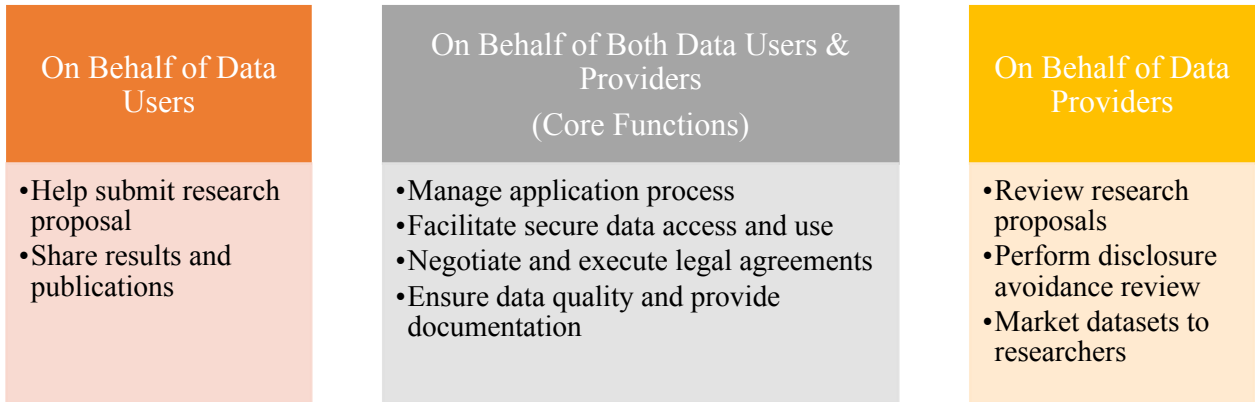
- **Revenue stream.** Depending on the business model, data intermediaries may be able to open up a new revenue stream for data providers without the amount of administrative burden and overhead. We see this benefit particularly valuable for private sector companies who have not been in the business of selling data to researchers.
- **Insights from Data.** By giving researchers access to valuable data, data providers are able to learn from the results generated from the studies. Data intermediaries connect data providers to more findings that can be transformed into business intelligence or insights that help them run their programs better.
- **Data management.** Data providers may work with data intermediaries to perform data management and reporting duties. In our interviews, we noticed that this is more common with data providers from nonprofits, universities, and government agencies.
- **Data philanthropy.** Sometimes companies will donate data to be used for academic research purposes as part of their data philanthropy strategy. These companies want to make their data available so that the broader researcher and the data science communities can use the data for public good. Data intermediaries can help connect these two parties.
- **Branding and marketing.** Private sector companies may share data to improve their brand image. By aligning their company with rigorous academic research and a trusted data intermediary, companies can establish themselves as thought leaders in the field.

From the data users' perspective, data intermediaries are a one-stop-shop to potentially a wide range of high-value administrative datasets in research-ready form. Importantly, data intermediaries provide transparency for data access and use. Rather than developing a one-to-one relationship with a data holder and negotiating the terms of a data use agreement over a long period of time, researchers can access the data through data intermediaries in less time and often at a lower cost.

Data Intermediary Functions

After interviewing the data intermediaries, the working group identified nine functions that data intermediaries perform to shift the burden away from data users and data providers². We categorize these functions into three groups: 1) those performed on behalf of data users, 2) those performed on behalf of both data users and data providers, and 3) those performed on behalf of data providers (Figure 1). We identified the four functions that are performed on behalf of both data users and data providers as “core functions” because they are essential to facilitating the relationship in the interest of both parties.

Figure 1. Data Intermediary Functions: What do they do and on whose behalf?



Functions performed on behalf of the data users

1. **Help submit research proposals.** Data intermediaries help researchers write more successful research proposals. For example, in order to access highly-restricted Census Bureau data through the Federal Statistical Research Data Centers (FSRDC) network, researchers must demonstrate that their proposal will benefit the Census Bureau programs in one of 13 predetermined ways. The FSRDC administrators have the experience to help researchers identify these benefits and make a compelling case to the Census Bureau. In that same vein, the Dartmouth Data Analytic Core, which functions as an intermediary between Dartmouth researchers and Centers for Medicaid & Medicare Services, reviews

² Throughout the report, we use the term “data providers” to describe the suppliers of the data. We acknowledge that some organizations refer to these parties as “data partners” or “data vendors”.

project proposals for researchers to ensure that the project is budgeted correctly and that the research agenda falls under what is appropriate within their Data Use Agreement.

2. **Share results and publications.** Several data intermediaries that we interviewed have policies in place to disseminate research results to the data providers and the greater public. For instance, the Texas Education Research Center requires all researchers to submit a policy brief that summarizes the implications of their research findings in a digestible manner for policymakers and data providers. These policy briefs are published on their website as well. Other organizations such as the Inter-university Consortium for Political and Social Research (ICPSR) and the Kilts Center also list publications and working papers using data from them.

Functions performed on behalf of both the data users and data providers

3. **Manage application process.** Data intermediaries perform the administrative and operational duties to manage the application process for researchers to access and use data. They often serve as the initial point-of-contact for researchers, whose first step is to identify a research question and the necessary data. Data intermediaries help researchers understand whether their research questions can be appropriately answered with the data available. They consult with potential data users and help them navigate the data inventory and variable list, ensuring that data users are aware of what is possible and what is not. At the same time, this function allows data intermediaries to act as gatekeepers on behalf of the data providers.
4. **Facilitate secure data access and use.** Data intermediaries provide researchers with secure data access and use. They generally approach this function in two ways – data can be directly downloaded for authorized users or, when greater security measures are warranted, data are accessed through a secure data enclave after research proposals are reviewed and approved. Wharton Research Data Services (WRDS) is an example of a data intermediary where all available datasets are only accessible by direct download for authorized users. Most of the other data intermediaries that we interviewed only allow data to be accessed through a physical or virtual enclave. The Consumer Data Research Centre and ICPSR are examples of data intermediaries with tiered access levels that

includes both approaches depending on the dataset. Regardless of the approach, we find that data intermediaries assume the role of developing the technology solution and infrastructure that enable data access.

5. **Negotiate and execute legal agreements.** Legal agreements are foundational to the existence of data intermediaries. Data intermediaries negotiate and execute legal agreements between data providers and data users. Although we did not perform an in-depth review of the legal agreements, we note that data licensing agreements are often in place between the data provider and the data intermediary. Additionally, data intermediaries ask data users (or their universities) to sign data use agreements, which may be modified from a master template or can vary across the user base depending on the data users or their organization. These agreements generally stipulate what data are shared, who can access the data, and under what conditions data can be used. We heard from almost every organization that setting up the necessary legal infrastructure is a time consuming, nebulous, and unpredictable process. We speak more to these challenges toward the end of the report.

6. **Ensure data quality and provide documentation.** Data intermediaries help ensure that the administrative data reach users in more research-ready conditions. Because the data that come in are often collected for different purposes, data accuracy and quality are not necessarily maintained to the same standards as they need to be for research purposes. Fields may be incomplete, inaccurately recorded, not fully validated, and not structured in the same way that a researcher would prefer but may be sufficient for internal use. Data intermediaries may perform this function by providing a common data structure, performing data quality checks, and ensuring that good documentation exists. The extent to which data providers are involved in this process varies across data intermediaries and depends on the mission of the data intermediary. For example, WRDS curates a collection of high-quality, research-ready datasets for finance and economics research. WRDS performs a rigorous data review and puts the onus on the data providers to ensure that the data are clean, without gaps (for longitudinal data), and have proper documentation before they are able to participate in WRDS. On the other hand, ICPSR's data archive is interested in making more data publicly available. ICPSR will work with

data providers to the extent that they can to help them properly structure and document their data. Yet another approach is from Consumer Data Research Centre. They acknowledge that data providers bring in administrative data that vary in quality. Rather than cleaning the data and potentially making inappropriate assumptions, the Centre provides users with detailed data profiles, variable distributions, and other metrics to help users understand the quality of the data.

Functions performed on behalf of the data providers

7. **Review research proposals.** Once research proposals are submitted, the data intermediary facilitates the review of research proposals. For most organizations that we interviewed, the data providers play a role in the proposal review process. At the Health Care Cost Institute, research proposals are sent to the Data Integrity Committee, which is comprised of actuaries from health plans who review the proposals to ensure feasibility and appropriate use. Similarly, the Consumer Data Research Centre directs research proposals to their Research Approvals Group, which includes data providers, and reviews for additional criteria such as ethical use and public benefit. There are also examples where data providers play a smaller role in the proposal review process. Because Nielsen has a pre-determined list of guidelines and criteria for data use, the Kilts Center reviews all of the proposals that come through for Nielsen data and only consults Nielsen if there is ambiguity. At ICPSR and NORC at the University of Chicago, the amount of data provider involvement in this process depends on the particular arrangement put in place for the dataset. But even when data providers do not review every proposal, the data intermediary represents the interests of data providers to ensure appropriate and ethical uses of their data.
8. **Perform disclosure avoidance review.** Data intermediaries that allow approved users to analyze restricted-access data in a secure setting will also perform a rigorous review of the results output. Typically, the results are reviewed to ensure that no individual-level or small cell counts are reported. The data intermediary may also ask for programming codes that were written to produce the statistical output. Many of the data intermediaries that we interviewed do not make available the list of criteria for disclosure review. This

review process usually takes one to two weeks. Although the data providers could be involved in this process, data intermediaries take the burden off the data providers to carry out this function. Of the 11 data intermediaries that we focus on in this report, we did not hear from the interviewees that the results are reviewed for reasons unrelated to scientific merit or privacy concerns (e.g. a company does not want a researcher to make a claim that could affect their reputation). Any potential inappropriate uses of the data are screened at the project approval phase.

9. **Market datasets to researchers.** Echoing a previous point on motivations for data sharing, we found that one important function of the data intermediary is to market datasets to the broader researcher population. Data intermediaries can speak the language of both the data providers and the researchers. They can help data providers communicate the value of their data to researchers as well as help them bring the data into research-ready form, which is usually vastly different from how the data are stored at data warehouses in companies. WRDS is one of the few organizations that explicitly brought this up, but it is clear that many data intermediaries perform this function on behalf of data providers.

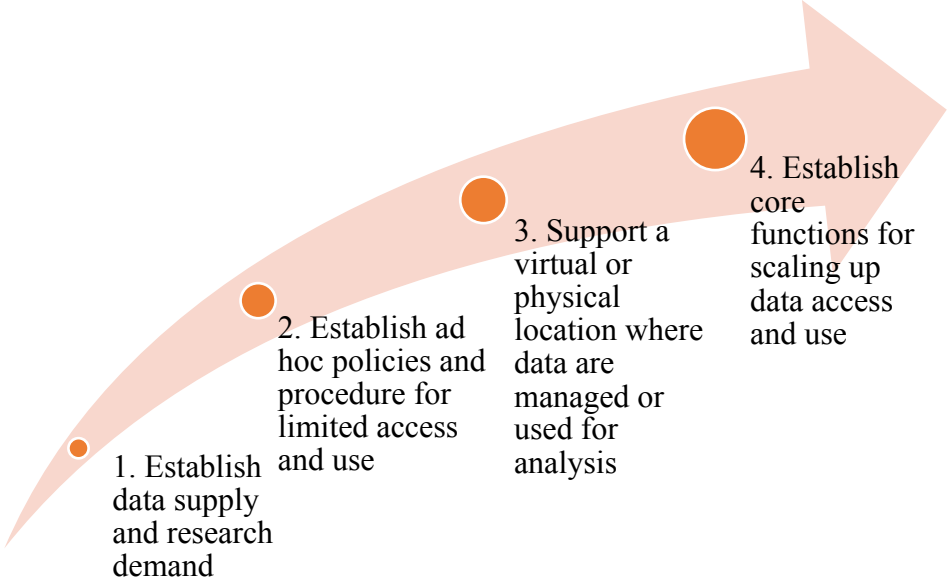
Organizational Structure and Governance

How are decisions made about how to approach the data intermediary functions? We bolster our findings above with a brief overview of the organizational structure and governance of the data intermediaries included in this report. Given that our time with the interviewees was limited, we provide a high-level overview and recommend that future work take a deeper dive into identifying the governance *models* of the data intermediaries.

As mentioned above, most of the data intermediaries that we interviewed are university-based. They are headed by an executive or managing director who may or may not have an academic appointment at the university. In some instances, they are run by both an academic and a non-academic director. Although the data intermediaries are hosted at an academic institution, we found that most of them are expected to be self-sustaining through user fees or project fees. The host universities typically do not provide financial investments beyond the space and limited infrastructure for the data intermediaries to operate.

In general, we found that the data intermediaries have governing boards or committees that support the development of their policies and procedures. There does not appear to be one common governance model or structure that the data intermediaries follow. But we did find that the governing body typically includes academic data users and representatives of the data providers. To the extent that we spoke to the interviewees about the roles and responsibilities of the board, we also found that the data intermediaries will consult the board or other governing committee on concerns related to business and fundraising needs, privacy and confidentiality, and legal agreements. Again, we recommend that future work dig deeper into the governing models and support structure for the longer-term sustainability of data intermediaries.

Figure 2. Proposed Model for Data Intermediary Maturation Process



Data Intermediary Maturation Process

As mentioned above, we found that a number of interviewees met some but not all of our inclusion criteria, suggesting that data intermediaries may undergo a maturation process that allows them to scale up over time. Given what we learned from all 17 interviews, we propose a preliminary model of a data intermediary maturation process described in Figure 2 above.

The data intermediary first matches data supply and research demand. This could look like hearing consistent requests from researchers for a certain dataset and hearing from providers of those data that they are interested in data sharing.

Once supply and demand are established, the data intermediary works with both parties to establish ad hoc policies and procedures that will enable limited data access and use. For instance, the data might only be allowed to be used for a particular research project by faculty at the university that is facilitating the relationship.

At the same time, the data intermediary will work to identify or build a technology solution that will enable data to be managed and used for research. These first three steps can lead to an initial pilot phase where trust is built and success is demonstrated. For example, we learned from the Quebec inter-University Centre for Social Statistics (QICSS) that Statistics Canada issued a call for proposals to demonstrate secure, research uses of business data. Currently, Statistics Canada allows access to a range of individual and household administrative data through its network of over 30 restricted data centers. The success of pilot projects can help make the case for expanding access to business data.

After the first three steps are established, the data intermediary can then establish its core functions for scaling up data access and use. More robust policies and procedures are put into place, legal agreements are renegotiated to now allow for broader use, administrative and operational capacity is bolstered, and the technology solution is capable of handling larger amounts of data transaction and analysis.

Challenges and Recommendations

During the interviews, the data intermediaries identified a number of challenges that they face. We discuss these below and provide recommendations for addressing these challenges.

- **Technology and security.** Data intermediaries strive to adopt technology that stores and delivers data efficiently. However, many grapple with the best way to accomplish that and feel that they lag behind with industry standards. For instance, a couple of interviewees mentioned that they want to transition from flat files to relational databases or other higher performing systems, but do not have the resources to make the transition

at this time. In addition to securing sufficient funding, the interviewees also found it difficult to identify best practices and standards for these approaches. One interviewee even found resistance from talking to other data intermediaries who are not willing to share information on how their data architecture is set up. Similarly, there is ambiguity in what data security practices to adopt and what security requirements are most appropriate for data enclaves. We recommend establishing efforts to build trust among data intermediaries so that they can openly discuss these challenges and share approaches. Developing shared standards for technology and data security will help advance the field for all data intermediaries and should not be seen potential competition.

- **Legal barriers.** As discussed above, a vast majority of data intermediaries found negotiating legal agreements to be a lengthy and confusing process. There are no clear standards on how to approach data use agreements under different circumstances, and legal counsel often cannot agree on interpretation of data protection laws that often hamper data sharing efforts. Legal guides coming out of Actionable Intelligence for Social Policy and State Data Sharing Initiative serve as good examples of tools to help government data intermediaries build legal capacity and develop a sound legal framework (Petrila et al., 2017; State Data Sharing Initiative, 2018). Similar efforts catered to administrative data intermediaries, especially those that work with private sector data, would be instrumental.
- **Business models.** Many of the newer data intermediaries were established in the past few years with seed funding from foundations. As seed funding runs out, they are concerned about finding a business model that will enable them to be self-sustaining. We learned from our interviews that self-sustaining data intermediaries are usually funded by user fees on a project basis or university subscriptions. These intermediaries can be very lean in its administration. As an example, the Texas Education Research Center, which is funded entirely on user fees and accommodations about 100 researchers a year, has a small staff comprised of a director, an IT coordinator, and a part-time administrator. We recommend that foundations consider funding the development of business plans for new data intermediaries and to make that a part of their seed funding. We also recommend that the ADRF Network conduct an in-depth survey of the business models of

administrative data intermediaries, much like the report that the National Neighborhood Indicators Partnership published on local data intermediaries (McTarnaghan and Hendey, 2017). These efforts would help increase our understanding of business practices that work in delivering these valuable data sharing services.

Conclusion

Data intermediaries are critical to the future of social science research. Through this working group, we interviewed administrative data intermediaries and identified nine unique functions that they serve to facilitate the sharing of restricted-access, proprietary, and private administrative data between data providers and a broad range of researchers. These functions lower the transaction cost of data sharing while speeding up the process of knowledge creation. We also highlighted a few key challenges that existing data intermediaries face. Our recommendations focused on facilitating dialogue among data intermediaries to ensure collaboration in sharing best practices and working toward developing standards for approaching their functions. As more and more data continue to be generated, we believe that there is space for more data intermediaries to be established. Organizations such as the ADRF Network should build on this current effort and continue to support existing data intermediaries and accelerate the growth of new data intermediaries.

References

- McTarnaghan, S., & Hendey, S. (2017). *The Business of a Local Data Intermediary*. Retrieved from Urban Institute Website: <https://www.neighborhoodindicators.org/sites/default/files/publications/The%20Business%20of%20a%20Local%20Data%20Intermediary%2012-4-17.pdf>
- Petrila, J., Cohn, B., Pritchett, W., Stiles, Stodden, V., Vagle, J., Humowiecki, M., Rozario, N. (2017). *Legal Issues for IDS Use: Finding a Way Forward*. Retrieved from Actionable Intelligence for Social Policy Website: <https://www.aisp.upenn.edu/wp-content/uploads/2016/07/Legal-Issues.pdf>
- State Data Sharing. (2018). *Legal Guide to Administrative Data Sharing for Economic and Workforce Development*. Retrieved from: http://www.statedatasharing.org/data-sharing/2018-03_-_SDS_Legal_Guide_to_Administrative_Data_Sharing_for_Economic_and_Workforce_Development.pdf
- Future of Privacy Forum. (2018). *Understanding Corporate Data Sharing Decisions: Practices, Challenges, and Opportunities for Sharing Corporate Data with Researchers*. Retrieved from: https://fpf.org/wp-content/uploads/2017/11/FPF_Data_Sharing_Report_FINAL.pdf

Appendix 1: Interview Protocol

Interview Protocol for the ADRF Network Data Sharing Governance Working Group

Introductions and Background (5 minutes)

Monica King to describe:

- The ADRF Network and the goal of working group on Data Sharing Governance and Management
- Informed consent:
- We will use the interviews to develop a written report that documents the various data intermediary functions. I will include what I learn about your organization from this interview in our written report to be published in June. I will circulate a draft of the report to all the interviewees in May, so you will have a chance to review.
- Do you have any questions for me before we proceed?

About the organization, tailor based on what we know (10 minutes):

- Give me an overview of [your org].
- Tell me more about the organizational structure of [your org].
- Who are your primary data providers and data users?

Questions on intermediary functions (40 minutes)

Now I'd like to turn to the specific data intermediary functions of [your org]. Walk me through how you work with data providers to bring their data into [your org].

Listen and probe for:

- How are decisions made about which data to include?
- What kind of guidance does [your org] provide to data providers on data quality, documentation, and metadata?
- What are the concerns data providers have? (i.e. privacy, misuse of data)
- How do you handle data use agreements? (e.g. Does the intermediary negotiate? Do you have templated agreements?)
- Tell me more about [your org's] role in linking and integrating data.
- What value does [your org] provide back to the data providers? (i.e. revenue sharing, data management)
- How often are data updated?

Turning to the data users, walk me through how you work with data users to gain access to the data as well as how they actually access the data.

Listen and probe for:

- How are research projects reviewed and approved?
- Tell me more about the user authentication process.
- Tell me more about the secure data environment. How are decisions made about the technical aspects for ensuring data security?
- How are results reviewed? Are results shared with data providers?

- Do you provide any training for data users?

Can you talk a little about how [your org] is funded? (e.g. User fees, federal funding)

Are there other functions that are important to what you do that we haven't talked about yet?

Closing (5 minutes):

- Are there services you'd want to provide but can't? If so, what's the barrier?
- What were some of the challenges you've faced as [your org] has grown over time? [or other challenge question.]

Appendix 2: Data Intermediaries and List of Other Interviewees

Data Intermediary Name	Type of Organization	Interviewee (role)	Year Est.	Brief Description of Data Provider(s)
Wharton Research Data Service	University	Robert Zarazowski (Managing Director)	1990s	Private sector including financial, banking, marketing
FSRDC Network (Census Bureau)	Government	Barbara Downs (Director)	1990s	Restricted Census Bureau data
Consumer Data Research Centre	University	Mark Birkin (Director)	2014	Private sector including retail, travel, energy
NORC Data Enclave	Nonprofit	Tim Mulcahy (Vice President)	2006	Various, including private sector, government, nonprofits, and foundation
Institute for Research on Innovation and Science	University	Jason Owen Smith (Executive Director)	2015	University research grants and outcomes
Dartmouth Data Analytic Core	University	Stephanie Tomlin (Operations Director)	1980s	Medicaid and Medicare
Quebec inter-University Centre for Social Statistics (QICSS)	University	Benoit Dostie (Academic Director) & Jean Poirier (Executive Director)	2001	Administrative and survey data from Statistics Canada and Quebec Statistical Institute
Texas Education Resource Center	University	Celeste Alexander (Director)	2006	State education, workforce
Kilts Center	University	Art Middlebrooks (Executive Director)	2013	Nielsen
Health Care Cost Institute	Nonprofit	Niall Brennan (President and Executive Director)	2010	Private health payer
ICPSR	University	Johanna Bleckman (Manager) & Trent Alexander (Associate Director)	1960s	Various related to social sciences

Organizations that participated in the interview process but are not included in the finding are:

- Jacob France Institute (Richard Clinch)
- Western Pennsylvania Regional Data Center (Bob Gradeck)
- National Student Clearinghouse (Joshua Leake)
- California Policy Lab (Evan White)
- JP Morgan Chase Institute (Fiona Greig & Gena Stern)
- Iowa Juvenile Data Warehouse (Laura Roeder-Grubb)

Call for Examples: Administrative Data Intermediaries

Are you an organization that facilitates the sharing and use of administrative data for research?

The Administrative Data Research Facilities Network hosts an active Working Group on Data Sharing Governance and Management led by Ken Poole, CEO of the Center for Regional Economic Competitiveness. The group is currently studying and documenting the unique functions that data intermediaries fulfill during the data sharing process. Our goal is to promote best practice standards among existing and emerging data intermediaries.

To that end, we are seeking to interview organizations that self-identify as administrative data intermediaries.

The organizations should fulfill most of the following characteristics:

- 1) Facilitate data access between multiple data providers and multiple data users, including academic researchers.
- 2) Do not necessarily have to house any data, but play an active role in connecting data users to data providers
- 3) Operate as a nonprofit (i.e. part of university, government, or other entity)
- 4) Work with data that are:
 - Collected by governments or businesses for non-research purposes
 - In the social sciences space, including economics and business
 - Restricted or sensitive in some manner (i.e. not open data)
 - Preferably not limited to a geographic area (i.e. national coverage preferred)
 - Able to be used by researchers at the individual-level (i.e. microdata)

We are conducting interviews now through May 2018. If you know of organizations that fit the criteria, please contact the Director of the ADRF Network below:

Monica King, PhD
moniking@upenn.edu
www.adrf.upenn.edu



Alfred P. Sloan
FOUNDATION



