Handbook: Engagement with State and Local Government

Robert M. Goerge

Acknowledgements

Support for the production of this document comes from the Laura and John Arnold Foundation. Lessons learned are drawn from experiences with the ADRF, Linking Federal Data to State and Local Data, and the HHS-ACF-OPRE funded Family Self-Sufficiency Data Center at the University of Chicago. The authors thank Leah Gjertson and Emily Wiegand of Chapin Hall at the University of Chicago for their astute comments.

Table of Contents

Table of Contents	ii
Introduction	1
Background and context	2
Scale and Scope.....	5
Approach	6
Due diligence.....	6
Working with political, policy, and technical government leaders and staff to understand the factors that hold back innovation (or simply to promote innovation)	7
Providing leaders with a vision that connects to their experience in an authentic way.....	9
Providing the agency with short-term, tangible benefits in terms of training and products	9
Providing the agency with a range of solutions to persisting challenges.....	11
Being flexible as a rigid approach would not fit all circumstances	12
References	13

Introduction

This handbook is intended to be a resource for both those who are providing data (government agency leaders) and those internal or external to those organizations who are requesting data for analytics projects particularly at the state or local level. Over the past 20 years, there has been a recognition that government administrative data (criminal justice schools, employment, human services, education, revenue) is an important resource for building evidence—from conducting rigorous evaluation to providing key descriptive analyses to improve quality, monitor caseloads, understand population composition (Commission on Evidence-based Policymaking 2017).

This has led to academic and research center-based investigators requesting administrative data from government agencies. Government agencies have been challenged to address these requests. Challenges include determining the legalities of providing data, the capacity to provide high-quality data, and the ability to review what is being done with the data provided to an external researcher. Of course, many of these challenges also exist internal to federal, state, county and city government. What are the tested approaches and learned truths to facilitating better use of the cornucopia of data available to us?

While there is an increasing number of federal agencies, states, and local governments committing to improve their use of the data they collect in the course of their operations, the United States is still not achieving the benefits of all that there is to offer from the innovations in computation and analysis to improve the well-being of poor and vulnerable individuals and families. In part, this is due to the elected and appointed leaders of government not having the motivation or means by which to take their agencies from the way their data has been analyzed in the past to the way that Amazon, Google, and Yelp, financial firms and the defense industry is using data to make strategic and tactical decisions. A culture shift recognizing the power of information and data in the policymaking process and in frontline decision-making is needed to take advantage of the data that state and local government collects (Mays 2018).

An example of this is that the government leaders are also wary of what the analysis of their data might bring them in terms of how well their programs are working, how cost efficient they are, and what they might be told to do with scarce resources. In the corporate world, executives are likely to be rewarded for identifying what does not work. Identifying a poorly operating program in government may lead to a requested resignation by a top official.

Its important to note that for the vast majority of state and county, there is absolutely no requirement to share data with anyone, regardless of the reason for wanting data. Therefore, it is necessary to engage state and county leaders to determine the conditions under which a researcher, external or internal to government, can access a rich set of data. While there are more and more datasets becoming publicly available, these datasets are typically de-identified or aggregated in ways that make them not optimal for many critical evidence-building activities.

There are individuals and organizations that are working tirelessly to bring more government bodies into the 21st century world of big data. Be it for-profit or non-profit, for academic or advocacy purposes, there are many visions and recipes for taking government data and turning it into knowledge that can improve services and make them more efficient. However, few are addressing the specific concerns and needs that leadership has. This document puts forth a set of activities(?) that together form an approach that has been successful in, at a minimum, improving the environment for the use of data, and, at the high end, producing actionable results that can change policy and practice.

Background and context

The content in this handbook draws on years of experience developing the LEHD at the U.S. Census Bureau and administrative data as a resource for social policy research at Chapin Hall at the University of Chicago. Currently, the experience in these two efforts are driving the Coleridge Initiative.

The Longitudinal Employer-Household Dynamics (LEHD) project is a data infrastructure project that linked unemployment insurance wage records for 49 states. These data were also linked to data from the Census Bureau, the Social Security Administration and the Internal Revenue Service. The result is a powerful data set that provides information on the dynamic interaction of workers and firms. It now serves as a vital tool to benchmark employment and earnings my area as a basis for development strategies, for planning transportation infrastructures, and more. It has been used to trace the effect of training programs, immigrant entrepreneurship, earnings inequality, and a host of other issues. Indeed, the micro data now represent the single most used labor market dataset for researchers.

The Integrated Database on Child and Family Programs in Illinois (IDB) is a unique resource for researchers and policymakers understanding of child and family experiences in the human service, health and education systems in Illinois.

The database includes the experiences and characteristics of the entire population of Illinois children who have had contact with the child protective services, income and nutritional support, health care, education, criminal and juvenile justice, and mental health systems in some cases back to the 1980's. Chapin Hall has worked with Illinois, Chicago and Cook County agencies to be a careful steward to government data. Chapin Hall's successful working relationship with the state of Illinois rests in part on the care taken to guard the confidentiality of records in its care by devising ways to link the disparate data from the public agencies and to guarantee confidentiality.

The IDB now provides a comprehensive picture of children's contact over time with publicly provided or financed service programs. The IDB makes it possible to produce an aggregate picture of the entire network of relationships linking children and public services, utilizing data routinely collected by the agencies for administrative purposes. With data going back from five to twenty-five years in the various service systems, the IDB makes it possible to do longitudinal analyses, tracking trends over time. Analyses using this Integrated Database on Child and

Family Programs in Illinois can track child and family experiences across otherwise unconnected service agencies to identify both redundancies and unmet needs. Data can also be used in combination with survey data with direct linkage of administrative data when consent is obtained. The database is continually updated and data on new services are added frequently. The project has been a springboard for dozens of research project both in Illinois and as a model for similar efforts in other states. It has also led to addressing the general utility of administrative data for research, including topics such as record-linkage, development of metadata, and spatial analysis.

The New York University (NYU) Administrative Data Research Facility (ADRF) offers a new approach to sharing data across agencies and across state lines as part of its Coleridge Initiative.¹ It enables multiple local, state, and federal agencies to keep data in separate secure areas, approved by [FedRAMP](#) protocols, and link data for approved projects.²

The ADRF allows agencies within the same state or different states to agree to share their data in a common area in the cloud for specific approved projects. If approved, staff from multiple agencies can jointly access the common area, so that they can work together to develop new integrated datasets, share information about coding differences or similarities, and develop common measures. No personally identifiable information is stored in the cloud – the data are hashed and deidentified (with a common hash algorithm) -- before being transferred to the secure area. Moreover, data stewardship modules can be deployed so that agencies can track use and work output. Over 200 agency staff, from about 90 agencies, have already accessed and used the ADRF. Sensitive Title 13 Census Bureau data have also been housed in the ADRF in a separate secure environment.

The NYU ADRF was set up by the Census Bureau to inform the decision making of the Commission on Evidence-based Policymaking (and was highlighted in their final [report](#)). It was

¹ <http://coleridgeinitiative.org/>

² <https://www.fedramp.gov/>

designed to build on the lessons learned from establishing the Census Bureau's Local Employment Dynamics (and Longitudinal Employer-Household Dynamics) program – mainly that state and local access to and use of linked data was essential to the creation of high value products. Access to the ADRF was made possible by the Laura and John Arnold Foundation and the Overdeck Family Foundation through a set of scholarships to government agency staff.

A set of courses enabling staff from state agencies to work in the sandbox environment, agency staff used exit and admission data on ex-offenders from the Illinois Department of Corrections, welfare recipient data from the Illinois Department of Human Services, and unemployment insurance wage records from the Illinois Department of Employment Security to answer questions including “what is the effect of neighborhood characteristics and transportation on the earnings and employment outcomes of ex-offenders and welfare recipients and their subsequent recidivism or exit from welfare?”

Scale and Scope

While some projects involve one or two datasets, in our experience, projects often expand as the use cases and research questions increase as a result of talking with policymakers. Taking the initial proposed data project to full implementation is an endeavor of considerable scale depending on the number of agencies involved, the nature and sensitivity of the data, the laws, rules and regulations around data use by external parties, the actual characteristics of the data, the context and familiarity of the agency relative to use of its data by external parties, and the needs of the leadership of the agency. To actually gather the information on these topics and then work through the request with the agency can take many months and often more than a year. The actual acquisition of the data and preparation of the data can take a similar amount of time. Any project must build in a sufficient amount of calendar time for the various steps in the process to be completed.

If there is any financial remuneration that will occur in either direction this could either ease or complicate any of the steps in the process. Specifically, if there is a contract for an external

party to produce a product for the agency and the agency is providing the resources, the regulations under which data can be shared will be more straightforward, because, presumably, the agency is securing a product it needs. Specific around how and what data are shared may be part of the contract. However, if government procurement is involved, there may be a whole different set of regulations that must be satisfied than if no exchange of funds occurs.

Approach

The important ingredients of our approach are:

- Due Diligence
- Working with political, policy, and technical government leaders and staff to understand the factors that hold back innovation (or simply to promote innovation)
- Providing leaders with a vision that connects to their experience in an authentic way
- Providing the agency with short-term, tangible benefits in terms of training and products
- Providing the agency with a range of solutions to persisting challenges
- Being flexible as a rigid approach would not fit all circumstances

In the sections below, we address each of these in greater detail.

Due diligence

Due diligence actually addresses most of the topics in this document. Prior to engaging a state, local government, agency or agencies, it is critical to have a sense of the political, legal, rules and regulations and bureaucratic structure. Getting to the point where mutual interests can be articulated depends on knowing the interests of the other party. It is probably the cases that others have taken a similar path and knowing the successes and failures of previous effort is critical to not making similar mistakes.

One fairly straightforward example is to know whether datasets have been shared with external parties in the past. Clearly, if they have been, the odds of acquiring that data are good. However, if datasets have not been recently shared, a new approach may be necessary. One example is Supplemental Nutritional Assistance Program data (SNAP). It is clear from federal regulation that this data can only be shared with external parties when the purpose is to improve the administration of the program. That being vague, federal and state administrators may have a range of opinions as to what analyses may improve the administration of the program. Knowing the types of projects that use SNAP data that have been approved and knowing the administrative challenges of the state SNAP program are a necessary, yet perhaps not sufficient piece of information collected as part of due diligence.

[Working with political, policy, and technical government leaders and staff to understand the factors that hold back innovation \(or simply to promote innovation\)](#)

Innovations to improve the use and application of data to the operations of an agency is not different from any other type of public sector innovation. Without a working knowledge of the institutional and social dynamics of the government agency or agencies of a particular initiative, it is more likely to fail.

The fact of government in the United States is that authority and discretion are widely distributed. While the three branches of government operate as a system of checks and balances, the administrative branch has been given significant authority to implement what might be needed in response to specific policy (Gormley 2014). Discretion is manifest throughout an organization from the chief executive to the frontline worker (Lipsky 1980). We have experienced that it is necessary to work with multiple levels of an agency in order to have both short-term and long-term success in changing practices and sustaining innovation.

As a consequence, this requires developing relationships with multiple staff at different levels across the agency. Clearly, it is important to be able to win over an agency chief executive as in most agencies, final approval of any major change is theirs'. However, engaging an agency can

begin wherever there is sufficient authority and discretion to begin an initiative. In fact, we would posit that long-term relationships with agency staff who are permanent civil servants is a necessary, although not sufficient, condition of successfully innovating within a government agency. More specifically, having an environment in which innovation can occur relies on government “investing in civil servants as the catalyst of innovation”(OECD 2017). The importance of the senior staff of an agency to move forward a particular innovation is a theme that is ubiquitous in our approach.

While it is certainly the case that some chief executives will be involved quite intensely in the development of new initiatives and be engaged on a daily basis, ultimately, they will have to delegate the tasks involved in innovation to others in the agency. Invariably, there are senior staff who are responsible for developing the strategy and tactics to innovate.

Similarly, once an initiative has begun, the technical information needed to properly use data—the metadata—is required. Given the current state of documentation of most information systems in government, gathering data about the data requires developing relationships with the internal experts who maintain the systems and databases on a daily basis. The interchange between these experts and the external innovators must be enabled by either the chief executive or the senior staff of the agency before it begins. However, once the exchange begins it must be done in a way that ultimately allows for the transfer of metadata and ongoing communication about data to allow for the proper use of the data.

If a prior relationship exists, it certainly is easier to introduce a new analytics initiative. Trust and credibility has already been established. However, leadership will not (and should not) blindly accept any proposal. Costs and benefits need to be articulated and understood at the outset. A clear, written agreement needs to be put into place that specifies the responsibilities of all parties.

If no prior relationship exists, it may be advantageous to connect with an organization that has had success collaborating with the specific government agencies to either learn what the process is or even get an introduction. A credible, trusted third party can help provide a head

start to developing a relationship with government. Chapin Hall provided such assistance to the many faculty and research centers at the University of Chicago.

[Providing leaders with a vision that connects to their experience in an authentic way](#)

Any innovation has to be communicated well so that leaders can understand how the innovation will address the needs of the agency and how the potential benefits will outweigh the costs. The diversity of prior experience of agency leadership is great. Governors may appoint individuals from the private sector, within agencies, or former elected leaders. Often, they may not have experience in the specific business of the agency. Innovators must provide a vision that the chief executive can sell to the governor, to the legislature, to advocates and special interest groups. The immediate payoff to the improved use of data are not obvious. The costs precede the benefits often by months or years.

This is also task where trusted internal senior staff are critical to bringing an idea to a chief executive and knowing how best to have the innovation accepted and promoted. In most cases, a chief executive will simply not have the many hours it takes to see how an innovation can be implemented within their agency. The ADRF is a many-faceted, complex technical and organizational tool. Written explanations are often inadequate.

[Providing the agency with short-term, tangible benefits in terms of training and products](#)

Tangible, more immediate benefits clearly hold more value in the course of engagement than the promise of research results that may or may be actionable. Given the capacity constraints of many state agencies, most states will value training for their current staff and assistance in providing research results that not only boost their capacity to do future work, but also provide initial results which can be built upon. We mention two efforts here.

The Administrative Data Research Facility (ADRF) provides an agency with a turnkey solution to many challenges of initiating an aggressive analytics agenda (Lane 2018). Housed at NYU, the ADRF provides a secure platform to host confidential micro-data. The ADRF system has been granted [FedRAMP](#) moderate pre-authorization and is designed to scale to different amounts

and types of use. (FedRAMP is the government-wide program that provides a standardized approach to security assessment, authorization, and monitoring of cloud-based products and services.) The ADRF is not just a single configuration of computing resources. Instead, it provides a set of data analysis components that can be combined in different ways, within the secure ADRF boundary, to meet a wide range of analytical needs. From traditional statistical analysis of data, to configurations that could be used to combine heterogeneous data. This technical flexibility, combined with our user, project, and data set management tools, allow for a given instance of the ADRF to securely and nimbly serve diverse data, analytical, and security needs.

It is important to build human capacity as well as technical capacity when developing and implementing new systems for working with data. The ADRF is a research and analysis sandbox that demonstrates the value of linking data by bringing together data and analysts from different administrative agencies to collaborate on shared problems through hands-on projects defined by senior management. The development of training programs delivered within the ADRF: (i) promotes system and tool adoption by cultivating teams of skilled practitioners who can demonstrate the value of new ways of linking and analyzing data for solving real world problems; (ii) promotes adoption of these practices to the benefit of the practitioners' own organizations; (iii) creates a pipeline of new product prototypes central to agency missions; and (iv) generates a growing set of linked data available as an ongoing asset for use in other projects.

A second way to build capacity is to work closely with state staff on specific research or analyses. An example of this was the OPRE-funded Family Self-Sufficiency Data Center (FSSDC) at the University of Chicago. The FSSDC embarked on demonstrating use cases for family self-sufficiency data, defining metrics for benchmarking across states and agencies, and providing opportunities to develop further models and code that we could share as guidance for agencies seeking to later replicate these analyses.³ There was significant interest from state agency

³ <https://harris.uchicago.edu/research-impact/centers/family-self-sufficiency-data-center>

partners in research activities focused on the self-sufficiency trajectories of SNAP and TANF cases, including questions of churn, recidivism, and long-term earnings. This research can be addressed simply using only SNAP/TANF administrative data, or explored in more detail with linked wage records.

Activities included designing an analytic file and sharing guidance on preparing that file from the state's administrative data, piloting a generalized spell script with states to simplify their work in data preparation, and bringing together a group of states to complete a share definitions and metrics and complete a common analysis. Key to the success of this project was adapting activities to the specific needs and technical capacity of each state agency. In some cases, data was shared with the FSSDC and in others the state analysts analyzed their data. A few states also requested assistance in navigating data sharing processes.

These two examples show how a relationship with state agencies can begin with research that is of use to the agency. Often, continuing the relationship after providing such opportunities

- requires a flexible approach that adapts to changing agency circumstances, especially data system transitions
- involves developing new relationships as agency staff turnover – and – getting buy-in from new staff
- navigating administration changes that bring new leadership, new priorities, and new approaches to data use
- being available to address state questions/needs in a timely manner

[Providing the agency with a range of solutions to persisting challenges](#)

Invariably, there will be multiple challenges that arise even when there is the will to provide access to for a specific analytic effort. If a particular agency does not have a formal process for preparing data sharing agreements, the data requestor may have to provide options that range from a formal contract to a memo of understanding to a data sharing agreement—all of which may have different meanings in a specific context. An issue that often comes up is that of

indemnification—who bears the cost if something goes wrong? Clearly neither party wants to take the burden so that it has to be clear under what circumstances either party might take blame. Often, the organizations receiving the data will be required to take the lion's share of the burden and will have to decide whether it can afford it. Taking all of the burden for a data breach may be one component of an agreement to convince a government agency to share data with another party.

A major concern among state SNAP agencies is whether the law allows them to share SNAP data externally for research or evaluation. Many leaders, in an uncertain environment, may find it not worth the risk for federal government sanction for mis-use of specific datasets. A requestor may need to determine what the options are in such a situation and orient the leader to what their options might be. In such cases, a contract may be necessary to address concerns. A formal opinion from the federal government may be necessary.

On the technical end, there may be a question of where the data will reside? Can it reside only on one computer disconnected to the internet? Can it reside on a server with limited access? Are cloud-based resources, such as the ADRE, acceptable to IT security officers? Is it only possible to analyze the data inside the agency that holds it? Being ready with a range of solutions is necessary.

Finally, it may be necessary to pay for data if there are no other ways to provide clear benefits for the data providing agency (or even when there are!). Increasingly government are asking external requestors to reimburse them for costs for producing the data.

[Flexibility: a rigid approach will not fit all circumstances](#)

Invariably, even when there is a clear process to engage with state and local agencies around beginning an analytics project, there will be obstacles that will arise. This will nearly always affect timelines for projects. Being able to adjust timelines to address the possible obstacles or changes in process is necessary. This may require extensions to grants, or at least a notification to funders.

An example of flexibility is around what is the minimum data set that would satisfy the needs of an analytic effort? Balancing asking for “everything” versus asking for specific fields is often a choice between the agency having to do more or less work to produce the data, but also have more or less risk around releasing more data.

Ultimately, the organization that holds the data is the gatekeeper. Without providing that agency with clear benefits of providing data access to a third party, a long-term productive relationship is unlikely to occur. Identifying the benefits is the challenge and without good intelligence about the current issues with which the agency is either struggling or understanding where they want to innovate, making a compelling case for accessing data to build evidence will be difficult.

References

Gormley, William T. 2014. *Taming the Bureaucracy : Muscles, Prayers, and Other Strategies*.

Princeton Legacy Library. Princeton: Princeton University Press.

<http://proxy.uchicago.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=791249&site=ehost-live&scope=site>.

Lane, Julia. 2018. “Building an Infrastructure to Support the Use of Government Administrative Data for Program Performance and Social Science Research.” *ANNALS OF THE AMERICAN ACADEMY OF POLITICAL AND SOCIAL SCIENCE* 675 (1): 240–52.

doi:10.1177/0002716217746652.

Lipsky, M. 1980. *Street-Level Bureaucracy. Classics of Public Administration*. doi:10.1007/978-3-319-21272-2.

Mays, Jeffrey. 2018. “Building an Infrastructure for Evidence-Based Policymaking: A View from a State Administrator.” *ANNALS OF THE AMERICAN ACADEMY OF POLITICAL AND SOCIAL SCIENCE* 675 (1): 41–43. doi:10.1177/0002716217739275.

OECD. 2017. *Fostering Innovation in the Public Sector*.

doi:<https://doi.org/http://dx.doi.org/10.1787/9789264270879-en>.

Policy-Making, Commission Evidence-based. 2017. "The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking."