January, 2018

Maximizing the Use of Integrated Data Systems: Understanding the Challenges and Advancing Solutions

Dennis P Culhane
John Fantuzzo, *University of Pennsylvania*
Matthew Hill
TC Burnett

Maximizing the Use of Integrated Data Systems: Understanding the Challenges and Advancing Solutions

By
DENNIS CULHANE,
JOHN FANTUZZO,
MATTHEW HILL,
and
TC BURNETT

State and local governments are seeking to use their health, education, and human services administrative data to address major social problems. As they attempt to move forward to develop and use integrated data systems (IDS), they have encountered substantial challenges. The Actionable Intelligence for Social Policy (AISP) team at the University of Pennsylvania targeted four common sets of challenges to IDS use: IDS governance, legal agreements, technology and data security, and data standards. The purpose of this article is to present the results from a year-long convening of four panels of national experts in each of these key topic areas. The results specify the greatest IDS challenges in each of these areas, and develop state-of-the-art responses to innovate the IDS field. It discusses how these solutions promote more effective, efficient, and routine use of IDS that are scalable to advance the IDS field beyond these current limitations in practice.

Keywords: integrated data; governance; administrative data; social policy; public policy

Currently, the federal government spends nearly $4 trillion per year on behalf of its citizenry (U.S. Government Publishing Office 2016). And the United States' population is larger and more diverse than ever. There are now more than 323 million people living in the United States, who speak more than 350 languages (U.S. Census Bureau 2015). With the

Dennis Culhane is the Dana and Andrew Stone Professor of Social Policy at the School of Social Policy and Practice at the University of Pennsylvania. He is a co–principal investigator (PI) of Actionable Intelligence for Social Policy (AISP). His primary research expertise is in the field of homelessness.

John Fantuzzo is the Albert M. Greenfield Professor of Human Relations at the Graduate School of Education at the University of Pennsylvania. He is also a co-PI of AISP. His primary research expertise is in the areas of early childhood risk, early childhood education, and Head Start.

Correspondence: culhane@upenn.edu

DOI: 10.1177/0002716217743441

national debt growing to more than $20 trillion, there is pressure to address more complex social problems with less. Yet only 20 percent of Americans would describe government programs as well run, and just 19 percent of Americans trust the federal government most of the time (Pew Research Center 2017). Americans want a responsible government, one that delivers more effective and efficient services to its citizens and abides by ethical standards of conduct (Kettl 2009). But what hope is there to improve government within a context of growing need, limited resources, and low public confidence?

Responding to this challenge, the Office of Management and Budget (OMB) and the United States Congress have called for the cross-sector use of government-collected administrative data to inform social problem-solving processes that lead to evidence-based policy. This resulted in the passage of H.R. 1831 in March 2016, establishing the Commission on Evidence-Based Policymaking. The goal of the commission is to figure out how administrative data from federal programs can be integrated and made available to facilitate "program evaluation, continuous improvement, policy-relevant research, and cost-benefit analyses by qualified researchers" (U.S. Congress 2016). It also seeks to make recommendations on what type of "data infrastructure" and "database security" can best support these objectives.[1]

Fortunately, there is a robust national movement at the state and local levels to use integrated, individual-level administrative data across public service systems to address vexing social problems (Fantuzzo and Culhane 2015; Lane 2016; Heidbreder 2016; Jennings, Hall, and Zhang 2012). This is critical because the role of the federal government is largely to redistribute funds to state and local governments, which possess the actual assets required to deliver and tailor services to the needs of their constituencies (Perlman 2010).[2] Since many of the important decisions about government service provision are ultimately made by states and local jurisdictions, sustainable program evaluation, policy analysis, and planning processes are needed at this level. State and local integrated data systems (IDS) have demonstrated their ability to fulfill this function by engaging cross-sector stakeholders across administrative silos and creating the legal and collaborative infrastructure to make longitudinal, cross-system analyses possible on a routine basis. When education, health, and human service records are successfully linked at the individual level, a broader range of relevant factors and outcomes can be examined longitudinally for entire populations. For example, in 2016, Los Angeles County administrators from multiple agencies wanted to determine how they could better coordinate programs to address homelessness and to examine their spending on this population. They were able to quickly

Matthew Hill is the former executive director of AISP at the School of Social Policy and Practice at the University of Pennsylvania. He is currently a senior research fellow at the University of Massachusetts Amherst Center for Heritage and Society.

TC Burnett is the associate director of AISP at the School of Social Policy and Practice at the University of Pennsylvania. Her work covers a wide range of topics related to integrated data systems, as well as the administration and management of AISP activities.

accomplish this by using L.A. County's long-standing integrated data system to look at county service provision to the homeless population. County administrators found that, across just six agencies, almost $1 billion was spent in 2015 on services to individuals experiencing homelessness. More specifically, they were able to determine that just 5 percent of individuals experiencing homelessness accounted for 40 cents of every dollar spent by the county on homelessness services (Wu and Stevens 2016). Armed with this information, the county enacted strategic policy and programmatic changes to better focus outreach and services to these individuals to decrease both county spending and homelessness (County of Los Angeles, Chief Executive Office 2016). As evidenced through the example of L.A. County, integrated data systems provide state and local government with actionable evidence to inform decision-making. The use of established IDS offers a promising avenue for government leaders to improve decision-making and generate more effective data-driven solutions for policy and practice.

Recognizing the potential of these IDS to produce cross-sector actionable intelligence for government leaders, and the complexity they represent, the MacArthur Foundation provided funding to the University of Pennsylvania to establish a network of integrated data systems to study the best practices of well-established state and local IDS in the United States. University of Pennsylvania researchers identified high-functioning IDS with strong track records of using integrated cross-sector administrative data to address complex social problems in their jurisdiction. Sites were determined using a key-informant process. University of Pennsylvania researchers contacted administrators of known integrated data sites, a federal human services research sponsor, and other university researchers that were identified as having potential knowledge of integrated data system sites. These individuals were asked to identify integrated data systems that met the following three criteria: (1) contains data from multiple agencies, (2) the IDS was developed to serve a general purpose rather than to complete a one-off research project, and (3) the IDS links data at the individual level. Responses were then used to generate the convenience sample of sites (Culhane et al. 2010). These included state IDS sites (Florida, Michigan, and South Carolina), as well as county- or city-level IDS sites (Allegheny County [Pittsburgh], Cook County [Chicago], Cuyahoga County [Cleveland], Los Angeles County, and Philadelphia) that produce, in a sustainable real-time manner, actionable intelligence to advance social problem solving in government.[3] With a network in place, AISP researchers studied the best practices across these exemplary IDS sites. A principal finding in the AISP Network study was the necessary and sufficient contribution of four critical *core components* of effective IDS operations required for data-driven solutions: IDS governance, legal agreements, technology and data security, and data standards (Fantuzzo and Culhane 2015).

Discovering these critical, common components of effective IDS sites is a significant accomplishment, but it is just the beginning. There is more work to be done to fully develop the potential of effective and efficient IDS use to foster ongoing quality improvement of government public services. The next important step is to consider the greatest challenges or barriers in establishing each of these

essential IDS components, and to then search for existing state-of-the-art responses to these challenges to advance the field of IDS practice. With the support of the Laura and John Arnold Foundation, AISP established four panels of national experts for each of the IDS core components. Panel members were selected based on current public sector or academic leadership experience in the governance, legal, technology, or data-related aspects of IDS development. They were charged first with identifying the major barriers to IDS use related to each of these components. Once this had been done, each panel was tasked with recommending innovative solutions to effectively address these challenges. The purpose of this article is to present a summary of the expert panels' findings. This work presents viable next steps in IDS development by showing how to concretely address current challenges. Taken together, these recommendations have the potential to increase the speed, scalability, and sustainability of effective IDS.

IDS Governance

Governance is the foundation of IDS use. The governance of IDS refers to the people, policies, procedures, and technologies required to manage the operations of an IDS under the charge of an executive government leader such as a mayor or governor. A governing board, which consists of a group of key stakeholders, is typically appointed to oversee the operations of the IDS. They supervise how the IDS is used to accomplish high priority research and evaluation inquiries with the aim of improving public services.

The challenge: Public mistrust of data integration

The biggest challenge in the governance of an IDS is *public mistrust* of government's ability to safeguard the personal information found in an individual citizen's administrative records. The public fear is that the perceived risks of integrating the administrative data of individuals across public service agencies are far greater than the benefits. We live in an era of low public trust in government. This is particularly true regarding American's confidence in government's ability to protect their personal data. According to a recent Pew Research Center report on *Americans and Cybersecurity*, only 12 percent of respondents said they were "very confident" government agencies can keep their records private and secure, and half reported that they do not trust the federal government to protect their data (Pew Research Center 2017). Moreover, the public perception is that this situation is getting worse, with nearly half of those surveyed reporting that their personal data are less secure now than compared to five years ago.

This general fear and mistrust is fueled by news stories of government surveillance and "unmasking" of the identity of citizens, and media reports of increasing data breaches of private records stored by national retailers, insurance companies, and financial institutions. These fears are further intensified by anecdotal stories about the potential of combining government datasets stripped of

personal identifiers with publicly available data to reveal the identity of citizens in these de-identified government records. In one highly publicized case, a computer scientist was able to use zip code information contained in voter rolls to reidentify individuals (including the governor) in an "anonymized" dataset on state employee hospital visits released by the Massachusetts Group Insurance Commission (Anderson 2009).

These public fears about the unauthorized use of data records pose a real threat to using personally identifiable information in a government operated IDS to foster evidence-based improvements in public services. Public mistrust is a real challenge to IDS governance, and it calls us to directly respond to this challenge by intentionally prioritizing innovative ways to enhance the ethical uses of IDS (Stiles and Boothroyd 2013). The following section considers the advancements proposed by the expert panel on governance. These concrete strategies are aimed at making both the real benefits of IDS use and the safeguards that can be put in place to minimize risk of personal data use more evident to the public.

Opportunities

The adage "*the best defense is a good offense*" captures the spirit of the experts' recommendations for IDS governance. Our expert panel on governance recommended taking very proactive and transparent steps to build public trust and confidence in IDS use by establishing the ethical use of personal information to advance the social good (Gibbs et al. 2017). This starts with grounding all IDS operations in the bedrock ethical principles of using human participants' data in research—*beneficence*, *autonomy*, and *justice* (Fantuzzo and Culhane 2015).[4] First and foremost is the ethical principle of *beneficence*. This superordinate principle asserts that all research uses of an IDS must be high-priority uses that are in the best interest of the persons being served. This means that the benefits of participation are clear and that they exceed the risks. Next is the principle of *autonomy*. Autonomy ensures that the beneficent uses of personal information respect the dignity of all participants and give the public an active voice in IDS decision-making (governance). Finally, speaking directly to public mistrust is *justice*. This principle respects all the public laws that protect the rights of participants and demonstrates that the use of individuals' information is reasonable, nonexploitative, and identifies and minimizes risks. The following three subsections outline concrete recommendations from the governance expert panel on how to incorporate these tried-and-true principles of ethical conduct into the modus operandi of IDS governance.

Advancing beneficence. The AISP governance expert panel's report appropriately introduces concrete ways to advance the *beneficence* of IDS use. This starts with the realization that the IDS will only operate effectively if at its inception there is a clear articulation of its purpose and expected benefits to the public. This takes the form of collaboratively constructed *mission* and *vision* statements. So much of the existing public fear and mistrust about IDS is associated with ignorance and confusion that results from the lack of any direct communications

to the public. Developing a transparent and straightforward mission statement will communicate why the IDS exists, what the IDS does, and who does what for whom. A mission statement should be constructed with all the relevant stakeholders to provide a "we" consensus justification of the beneficence of the IDS's existence in a state or local jurisdiction. Such a statement must underscore and make visible the core purpose and mechanisms of the IDS to use public resources to achieve more effective and efficient public services. The mission statement is an assertive step toward dispelling myths and fears about IDS, but it is not sufficient on its own. It must be accompanied by a coconstructed vision statement. The vision statement is a simple declaration that describes in plain language the end goal of the IDS and points to the long-term expected benefits of IDS operations. The vision statement makes evident to all how the IDS will ultimately benefit the clients being served by the participating public agencies; the citizens of the community; the government leaders and policy-makers in the community that will use the actionable intelligence it provides; as well as other communities, both national and international, that can use the research and evaluation knowledge resulting from the IDS. The expert panel report underscores the importance of developing these clear and honest consensus statements of *What is it?* and *How it will benefit us?* as key to establishing *beneficence* as a cornerstone of ethical IDS use.

Advancing autonomy. The AISP governance expert panel's report also provides clear guidance to actualizing the ethical principal of *autonomy*, through engaging relevant stakeholders in the design, launch, and governance of an IDS. Autonomy emphasizes the significance of providing for and respecting the public *voice* and *choice* of relevant stakeholders as a means of concretely building trust among all the key participants in IDS use. All too often, mistrust, fear, and resistance are generated when individuals or groups feel that others are doing something important that impacts them that they have no ability to influence. Their resistance ("No!") may in fact be a negative expression of their autonomy—voicing reasonable concerns that have not been yet been considered. To effectively address resistance to IDS use in a state or local jurisdiction, we must *partner with resistance* (Fantuzzo, McWayne, and Childs 2006). Core to the principle of autonomy is the *practice of respect* for the distinctive perspectives of all relevant stakeholders (Fantuzzo 2015). As recommended by the panel, this first involves the recognition of key contributors and beneficiaries and then the creation of an inclusive process that allows for their points of view to influence the decision-making about the routine use of the IDS.

The panel's report provides guidance for how to identify and include key stakeholders. It lists four major stakeholder categories to consider that are involved distinctively in effective IDS operations: government executive leadership, frontline service providers, researchers and data analysts, and the public (i.e., both the direct beneficiaries of the services and the community at large). Identifying key stakeholders within these categories is essential to inviting the most interested and capable stakeholders to serve on the various boards and advisory groups that are necessary to the ongoing governance process of the IDS.

Once stakeholders across categories have been identified and prioritized with respect to their ability to contribute to the governance of the IDS, plans should be made to engage them. This is an important consideration of where their voice and distinctive contributions would most benefit the various activities of the IDS. The sum of these functions include the whole range of work executed by those involved in the governance process of the IDS: (1) inaugurating the IDS with mission and vision statements; (2) generating the various legal agreements that permit the integration of individual data across agencies for IDS use; (3) establishing the integrity of the data infrastructure and analytic capacity to conduct research projects; (4) determining the priorities for IDS research projects to improve services; (5) monitoring and overseeing the successful completion of those projects and ensuring that the results are translated into actionable intelligence in accord with the IDS mission and vision statements; and (6) communicating, in an open and transparent way, to the public what has been learned and its contribution to more effective and efficient public services. The aim here is to draw upon all those stakeholders who are interested and able to contribute to the robust "we" of IDS use. Here our aim is to achieve equality of respect by recognizing that respect is "not something we possess, but an ongoing ethical practice that requires a sincere effort" (Fantuzzo 2015, 85).

Advancing justice. The ethical principal of *justice* in IDS governance builds upon the principals of *beneficence* and *autonomy*. Establishing the purpose and benefits of IDS use and inclusion of all relevant stakeholders must be followed by a written agreement ensuring that there are ethical and legal safeguards in place governing all the concrete policies and procedures of IDS use. Here, justice is codified in the memorandum of understanding (MOU) agreement, which is signed by all the key contributors involved in IDS use. The MOU sets forth the core features of the IDS structure and conduct, and defines the legal rights and responsibilities of each party within the IDS in a just manner. The MOU provides the collaborative foundation for how the "we" of the IDS will achieve the benefits of the IDS. The MOU accomplishes three important objectives. First, protecting the private information of individuals being served by the respective service agencies participating in the IDS is made a top priority. Second, it respects the rights and responsibilities of the agencies that collect private information during the course of service provision to use these data to inform how they can improve the quality of the services they provide. Third, it affirms the "we's" commitment to beneficence and autonomy by making its policies, procedures, and accomplishments transparent and open to the public at large. In this way, the MOU is both an ethical and a legal document that upholds equal burden and equal benefit of IDS use.

IDS Legal Issues

The second major core component of IDS is legal issues. IDS utilize the personal information found in government administrative data records to improve public

services. These data are originally collected by government agencies through the routine provision of programs and services. The agencies hold and use this information in the context of existing laws. It is therefore essential to understand the legal issues related to IDS use.

The challenge: Too many red lights

Fundamentally, the purpose of law in a society is to govern and guide actions and relations among and between persons, organizations, and governments to protect the valued liberties and rights of members of that society from unreasonable intrusions by persons, organizations, or government. The law at its best provides *freedom within form*. It regulates transactions to protect liberties and can be equated to traffic lights in a big city, which use red and green lights to permit many individuals to move about the city freely with minimal harm. The red lights protect citizens from the impulses of other drivers, and the green lights permit citizens to get to their destinations while also regarding the rights of other drivers. A driver's license signifies that an individual knows the law and is willing and capable of abiding by it. Therefore, to a naïve, uninformed person the legal component of IDS should be equally simple—just identify the laws that govern the use of the personal data collected by the government, and give the government a license to integrate and use these data in accordance with the existing laws to regulate IDS use. This sounds like a simple, linear, and rational process. However, this does not reflect the reality of the complex and often irrational world of twenty-first-century America. Therefore, our contemporary context poses significant legal challenges to IDS use.

There are two prevailing forces that beset lawyers and generate legal challenges to state-of-the-art IDS use. They include (1) the unprecedented crisis of public mistrust surrounding government's use of personal data; and (2) the unprecedented, though complex, opportunities to use IDS to make substantial improvements in government health, education, and human services. As we mentioned, there is currently a substantial lack of public trust in government's ability to safeguard personal data. As a result, there is a predisposition to be cautious by legal counsel in government agencies. Fears of litigation; long-standing cultural trends; norms and policies within government agencies against sharing; as well as overly conservative interpretations of federal, state, and local laws all point to a "no" red-light, legal response (Petrila et al. 2017, 6). Unfortunately, this climate of "many reasons to say no" is a breeding ground for myths, misinterpretations, and half truths about the risks associated with IDS use. More importantly, it diverts attention from the benefits of how IDS can contribute to a more innovative, effective government.

To complicate matters further, the effectiveness and utility of an IDS is enhanced when there are more data partners and community stakeholders involved throughout the life of the IDS projects. This translates into "more opportunities mean more complexity." From a legal perspective, this makes the formulation of comprehensive MOUs very complex and time-consuming, with many moving parts to regulate—many complex actions and relations among

partners. This requires an experienced general counsel that understands the intricacies of the relatively new and burgeoning world of big data and legally sanctioned IDS uses. It also requires an extraordinary amount of time to negotiate and finalize these complex agreements. All this may be too much for the typically overburdened legal counsel in government service. The nature and amount of knowledge, experience, and time required to craft these complex agreements creates a sizable burden for existing legal counsel. This burden further challenges and thwarts effective IDS use and increases the likelihood of a "no," or red-light legal response.

Opportunities

As highlighted above, the value of the law within society is to promote liberties within a social contract that governs, guides, and regulates the expression of those liberties for the social good of all. This important freedom within form is actualized by both red light *and* green light applications of the law. The challenges presented above reflect primarily red-light applications of the law designed to prevent or minimize risk. Absent is the counsel and leadership to promote liberties and to pursue opportunities amid risks that have the potential to yield the greatest good for the greatest number of citizens. Heineman, Lee, and Wilkins (2013), in a paper titled, *Lawyers as Professionals and Citizens: Key Roles and Responsibilities in the 21st Century*, draw our attention to three distinct roles that lawyers should play in the twenty-first century. These roles are technical expert, wise counselor, and effective leader. A lawyer, for example, as legal technician (or technical expert) makes a specific application of existing law to a particular set of facts. Here the implication is that there is always a specific legal answer to a given situation. We want the technician's answers to be a simple yes or no, even though the problem may be immensely complex and convoluted. However, such a stance causes us to look less to the *wise counsel* or *effective leader* roles in the legal profession. To appropriately address the legal challenges of IDS and advance government innovation, we need to call upon these roles. IDS use in government has the potential to advance evidence-based decision-making needed to improve the quality of care and services received by millions of citizens. As such, it necessitates the highest levels of functioning from our legal professionals. The AISP legal issues expert panel report invokes all three roles to lay out concrete steps to address these contemporary legal challenges. Fortunately, in constructing the report, the expert panel exercised *effective leadership* and pointed to "green light" responses and pathways forward.

At the outset, the legal issues expert panel report asserts that the primary purpose of IDS use in federal, state, and local government is to achieve more effective, efficient, and responsive government by facilitating the core government functions of audit, evaluation, research, and evidence-based practice in public programs and policy. At this time, we are not proposing IDS use for the day-to-day operations and case management of individual clients in public service agencies. This specific type of individual, client-level information sharing across agencies almost always requires that the individual consent to having his or her

personal data shared for these purposes. In the context of this article, IDS use is focused on linking thousands of individual records across multiple agencies to achieve a broader view of a social problem and policy solutions for entire service populations. Here the primary aim is to study services to enhance them and generate more effective policies in an ongoing data-driven process. The AISP Legal Issues Expert Panel (LIEP) vision of the legal profession's contribution to IDS use is to identify appropriate legal regulation of IDS to maximize IDS as a means to improve evidence-based practice while minimizing privacy and data security risks. Therefore, the preamble of their wise counsel is that the legal issue is no longer *whether* we should integrate data to drive data-based decision-making, but *how* to integrate data such that we address existing legal barriers and concerns to realize the spirit of the law. To accomplish this charge, the legal issues expert panel report seeks to expose myths of IDS use, explicate permissible uses in federal law, and demystify the legal agreements that govern IDS use.

Expose myths and explicate permissible uses. Pervasive public mistrust and fear of litigation are the breeding grounds for misinterpretations and myths about IDS use. The LIEP report identifies the most common arguments posed by legal counsel in opposition to IDS data sharing across agencies and provides clear legal responses to refute these misconceptions. The first misconception is, "This is not legal." As the LIEP report explains, such an assertion is not true because IDS use is legal. All federal and most state laws authorize data sharing for appropriate governmental and research purposes. The LIEP report provides a detailed review of the existing federal laws to illustrate the legal pathways to legitimate use. Another set of common objections involves issues of individual rights, such as, "This (IDS use) requires obtaining individual consent to re-disclose data, which is not administratively feasible and it pits individual interests against societal interests." Again, this is not true because most data privacy laws allow the agency holding the data to use or share that data, including personal identifiers, for research or policymaking purposes without obtaining individual consent. The LIEP report argues that the perceived conflicting interests reflect a false dichotomy. While individuals have a strong interest in data privacy, they have an equally strong interest in effective and efficient government programs and policies. A well-constructed IDS preserves individual privacy through policies and procedures. At the same time, it helps to ensure that government carries out its functions to the highest quality.

Another set of legal oppositions involves misconceptions of how IDS is a threat to the participating public agencies. Examples include, "This (IDS use) exposes us to too much liability … we are going to get sued"; or "This is not a well-accepted practice; it is uncharted and unsanctioned territory placing us at risk." These often push the agency to a "no" response, but such objections are not true. First, major data privacy laws not only allow and encourage data sharing for these purposes, but they also do not contain private right of action for individuals to sue over a data breach or misuse of private data (Petrila et al. 2017). To say that IDS use in government is uncharted territory is also false. IDS exist throughout the United States, and are endorsed at the federal and state levels. The LIEP report points to the AISP national network of existing IDS sites, which consists

of local- and state-level IDS that include more than 26 percent of the U.S. population.[5] Adding further evidence to this is the fact that, in 2016, the Evidence-Based Policymaking Commission Act of 2016 (H.R. 1831) became law. Its goal is to promote IDS. Additionally, the National Conference of State Legislatures has prioritized opening government data for public use, including integrated data (Petrila et al. 2017).

Finally, the LIEP report cites two major legal misconceptions that directly exacerbate Americans' fears of government, cybersecurity, as well as governmental ability to manage the (perceived) overwhelming complexity of IDS use in government: "It (IDS use) is too big for government to handle, and it makes a serious data breach more likely." Again, the LIEP report points to the exemplary IDS sites in the AISP network, some of which have been operating for more than 30 years, that, to date, have zero security breaches. It can be done and done well if the IDS is well constructed and uses the best data security practices. High-quality IDS place a premium on data security and formulating legal agreements that maximize beneficial use while minimizing risks to personal data breaches. We devote the remaining sections of this article to addressing how to demystify IDS use and innovate IDS operations to lessen the burden on personnel (including legal counsel) and foster economically sustainable administrations of IDS in government.

Demystify foundational IDS agreements: MOU and DUL. Both the ethical and legal principles addressed in the governance and legal issues' reports require that legal counsel has an adequate understanding of the operations and laws related to an IDS, experience in negotiating and drafting IDS data sharing and data use agreements, and the time and resources to continually develop and monitor these agreements going forward. The LIEP report proposed three responses to legal challenges. It addresses the legal content, process, and structure of the foundational legal agreement documents of an IDS—the MOU and the data use license (DUL). The MOU is the bedrock agreement among the lead IDS agencies and data contributors. It is coconstructed by the IDS stakeholders, and it codifies both the legal rights and responsibilities of each party in the IDS and the procedures and policies that govern sanctioned IDS operations. The DUL is the other basic legal agreement in an IDS. The DUL details the terms and conditions under which a researcher, evaluator, or outside party may gain access to data from the IDS related to a project conducted in partnership with the governing body of the IDS.

The three advancements proposed by the LIEP report are designed to demystify the MOU and DUL legal agreement processes and reduce the burden on government legal counsel (Petrila et al. 2017). The report first specifies the content of these agreements by generating a checklist of information that legal counsel needs to obtain to craft the MOU and DUL. This checklist comprises a set of key questions that relate to laws and regulations governing IDS use. For example, "What are the legal, regulatory, and administrative policies governing the specific types of data involved in the IDS and provide applicable citations?" "What are the specific categories of data to be shared and with whom?" "Are there any

restrictions (legal, regulatory, administrative, or other) regarding who can be an authorized user of the data?" Next, the LIEP identifies the logical steps in the process of gathering information, negotiating agreements among those involved in IDS use, and finalizing the written agreements. This process involves major categories of work related to understanding the data sources to be included, identifying the specific data elements to be integrated, gathering all the laws and regulations associated with those elements, considering safeguards related to access and usage, identifying access restrictions and usage requirements, and composing agreements such that they can be comprehended by all partners before signing.

The LIEP report also provides annotated templates of MOU and DUL agreements, and points to online exemplars of these agreements from AISP IDS sites across the United Stata. These templates and examples of existing agreements provide legal counsel with a valuable framework to move forward. The final proposal of the LIEP report is to develop standardized legal agreements (i.e., these annotated templates of agreements) endorsed by recognized legal authorities and legal organizations. This would save an enormous amount of time and reduce the burden on state and local general counsel. The legal leadership of LIEP provides a clear green-light pathway forward to actualizing the benefits of IDS use in government.

IDS Technology and Security

At first glance, the dual charges of the Commission on Evidence-Based Policymaking seem incompatible: to increase access to administrative data for establishing the evidence base for social policy, and to increase the protection and security of these data. But, perhaps counterintuitively, the only way to increase access and use of administrative data will be through the adoption of increased security for these data. As noted previously, among the key objections that agency administrators and attorneys use to argue against data sharing is the fear that making data accessible will increase the risk of a data breach or the reidentification of protected personal information. Indeed, the relatively slow rate at which jurisdictions have been adopting integrated data systems reflects this basic fear. Through the implementation of all four panels' recommended innovations, the security and safety of data sharing can be greatly enhanced. A thorough and thoughtful adoption of standards can enable a community to provide appropriate assurances—both to agency leadership and to the public—that these data integration efforts can be undertaken in a way that safeguards private information.

As with the enhanced governance and legal standards, technology innovations are also making data sharing much more secure by greatly reducing, if not eliminating, the potential risk for data breaches and reidentification risks. Secure research platforms for data sharing have proliferated in several fields, specifically with respect to protected health, education, employment, and social services data. Many countries in Europe, provinces in Canada, states in Australia, and

New Zealand all provide authorized users with access to linked administrative data for approved projects. They have also adopted similar technological approaches and procedural safeguards that point a way forward for jurisdictions in the United States. To facilitate that roadmap, we surveyed these countries, convened an international conference in November 2016, and charged the AISP technology and data security expert panel (TDSEP) to recommend a set of technology-based solutions to greatly reduce the barriers to the implementation of integrated data systems in the United States.

Challenges

Beyond the governance and legal challenges that state and local agencies seeking to share data face, the practical challenges of sharing data in a secure and safe manner create additional barriers. While sharing data between two agencies under the same government auspice, and possibly using a shared secure platform, may seem relatively safe, sharing data with external evaluators and researchers is inherently more complex. Secure file transfer protocol (FTP) and encrypted file transfers can provide some increased security, but having the data travel outside the direct control of the government agencies responsible for these data creates an increased risk that the data can be stolen or otherwise shared with unauthorized people or for unauthorized purposes. Similarly, the risk of reidentification grows because data can be manipulated or linked with other data sources for this purpose.

States and local governments are also suspicious of big information technology (IT) projects, and with good reason. The typical IT project in government involves a complex and time-consuming procurement process. Costly consultants must be retained simply to draft the appropriate specifications for a procurement. Contractors propose and build highly customized solutions at enormous costs. Typically, government agencies are then tied to these contractors for the life of the system, and must reengage them at high costs to make even basic modifications. Legislatures are wary of such projects for all the apparent threats. From our survey of existing integrated data systems in education, health, and human services, sites report that the more sophisticated platforms cost between $2.5 and $4 million to develop. Only some states and a handful of localities have budgets that permit them to build such costly systems, and even jurisdictions with the capacity to fund them are reluctant to engage contractors to build highly customized solutions given the high future attendant service costs.

AISP's work with state and local governments has also revealed what is perhaps the most important resource constraint that they face in developing systems of data sharing and collaboration: workforce capacity. Governments are often working with threadbare staff. The last two decades have seen a shedding of government workers across the board, but especially in the areas of research and evaluation. Not only are attorneys and general counsel offices overwhelmed and overworked, and thus reluctant to take on new work related to IDS MOUs and DULs, but the agency staff to which the operational aspects of this work would fall are often equally overworked. Simply undertaking a single data-sharing agreement between two agencies—the simplest use case—can take nine months

to a year to execute. The prospect of having to process multiple requests across multiple departments, and to manage simultaneous transactions with both internal and external analysts, all the while maintaining data security and proper oversight of projects, is unimaginable with current staffing capacities. Technological innovations will be required to make the integrated data system process move at a vastly different scale and efficiency, and to offer genuinely actionable intelligence in a timely manner.

Opportunities: An open source, shared technology solution set

The TDSEP report proposes a shared set of technology solutions that simultaneously address data security, cost, and transaction management challenges that confront jurisdictions seeking to develop an integrated data system (Patterson et al. 2017). Recognizing that states and local governments have many data integration needs, and that no single system can or should be burdened with the responsibility of meeting the diversity of needs and possible uses, they propose a specific archival approach comprising a thin stream of data that would be updated periodically (quarterly or yearly) and that would be designed to meet the specific needs of program evaluators, policy analysts, and researchers. TDEP recommends an open-source set of solutions, consisting of two primary components. The first component, "DataHub," would standardize the technology and workflow processes for acquiring, storing, linking, and provisioning the data. The second component, "Clearinghouse," would manage the transactions associated with processing data requests, managing secure access, and providing oversight of approved projects and users.

The solution set proposed in the TDSEP report includes specific features designed to address the data security concerns of agency administrators. Through the adoption of state-of-the-art encryption methods, the data would be encrypted by source agencies in transit to the DataHub, and at rest within the system, so that personal identifiers are not attached to records. Tools would be available to the system administrator for creating customized linked research datasets to the specifications of an approved project and user. Research datasets would be made available to evaluators and analysts through a secure portal, through which queries could be sent and run against the data. The statistical output generated from these queries would be run through an automated disclosure filter, designed to ensure that only aggregate data with minimum cell size limits are returned to the analysts. Analyst queries would also be monitored to make sure that queries conform with approved purposes. Of course, a manual disclosure review for final output is also possible. Analysts would not be able to view record-level data; however, simulated, record-level data views could be generated to verify statistical output. Each jurisdiction would have its own instantiation of the DataHub, with physical and technical control over the data and server. While some sites may eventually seek cloud-based solutions, TDSEP's assessment of state and local governments' preferences has identified a robust consensus for site-based physical and technical controls over their data; DataHub is designed to provide that.

The Clearinghouse platform will operate as a website that will standardize the workflow processes associated with end user data requests and project execution.

The site would be configured to appear as the portal to a given DataHub installation, or, in a future phase of the initiative, to enable multisite, cross-jurisdiction data requests. The Clearinghouse would provide metadata through the jurisdictional DataHub. Evaluators and researchers would submit requests for projects using a standardized form. Once a designated governing board approved a project, a data use license would be generated for the signature of the end user and their sponsoring institution (countries surveyed only permit universities and approved research institutions access to microdata). The DUL would specify the data elements and research questions that have been approved, and the time period for which data access will be permitted, as well as other responsibilities and safeguards required of the end user and their institution to protect against unauthorized uses of the data and related sanctions for violations (usually lifetime refusal of future access, in addition to financial and legal penalties for the institution and end user). AISP is also planning to create a tutorial on the laws and ethics related to the analysis of administrative data, which end users may be required to complete. Once approved, the Clearinghouse would provide user authentication and electronic certification, permitting the analyst to access her or his designated research dataset. One important feature of the Clearinghouse is that it would also permit researchers to submit approved external datasets for linkage to the DataHub. Queries would be submitted via the Clearinghouse, and the statistical output would be provided via the secure portal at the specified DataHub. The automated and possible manual disclosure review would screen output for approval. The Clearinghouse would host a results forum so that evaluators, project sponsors, IDS administrators, and data source agencies could discuss results and data interpretation issues before any findings are made public. Similarly, public forums for specific topics would be hosted for the discussion of results and papers by subject matter experts, policymakers, and other interested parties. Last, the TDSEP has recommended that the Clearinghouse include a mechanism for jurisdictions and funders to post requests for proposals, or a notice of research priorities, to which evaluators could respond. The site would also manage financial charges for access to research datasets, standardize contracts, and invoice end users, thereby avoiding the bureaucratic site-specific procurement and payment processes that can often slow projects.

The solutions proposed in the TDSEP report would be commissioned and overseen by a governing board comprising jurisdictions seeking to adopt this common solution set, as well as experts in computer science and evidence-based policymaking. The governing board would also be responsible for developing a business plan.

IDS Data Standards and Minimum Datasets

IDS data challenges

Jurisdictions collect thousands of data elements across scores of datasets as part of their administrative duties. However, only a fraction of these data is of

sufficient quality for research and evaluation. Given that administrative data are not typically collected with these uses in mind, great caution must be exercised in selecting data elements that are reliable and valid. Moreover, tremendous variability exists in the data elements collected by state and local governments, making cross-site comparisons, let alone simultaneous multisite analysis, potentially problematic.

A further challenge is that the diverse data sources in an IDS need to be organized in a way that facilitates an understanding of what data are available. Viewed as a series of one-off, or program specific, data siloes, the data potentially available to an IDS can seem vast and overwhelming. The specific program uses addressed by a given dataset can also seem so particular as to require years of experience and domain-specific knowledge to understand their peculiarities. Moreover, state and local variations in data definitions and even different measures in varying datasets for some of the same variables create significant challenges for data managers and administrators seeking to provide meaningfully curated research datasets.

Opportunities

The data standards expert panel (DSEP) was charged with identifying the most promising data from across education, health, human services, justice, housing, and workforce programs that could be incorporated in a state or local government's IDS. Recognizing that the data holdings of these entities are large and complex, the DSEP was asked to consider which datasets are most likely to be common across jurisdictions, which data elements within them are most likely to be governed by federally mandated minimum data standards and definitions, and which data elements are most likely to be valid and reliable for research purposes. To fulfill this charge, the DSEP first developed a conceptual framework for organizing these diverse data holdings. They surveyed the data sources of existing IDS to create an inventory of the optimal candidates for inclusion in a robust IDS data model that would meet their criteria for universality (or near universality) across the United States, reliability, and validity. They also developed a data schematic to classify the types of data most commonly held in these datasets. Last, they considered some of the issues that should be addressed in the repurposing of administrative data for research.

DSEP adopted a life course conceptual framework to structure the recommendations provided in their report (Wulczyn et al. 2017). Data about citizens begin with the birth certificate and end with the death certificate. In between, there are data about infancy and early childhood, including immunizations, early intervention testing and screening, and early childhood education and enrichment programs. Data from school districts track entry and progression through school, including attendance, achievement, special education status, standardized test scores, and disciplinary actions. Social programs for children, including child welfare investigations and out-of-home placements, and juvenile justice placements, record special services to children and youth at risk. The transition to adulthood is recorded through higher education datasets, as well as workforce

training programs. Employment and earnings data are available through state labor department records. Special population data, for adults and children who experience homelessness, for example, are collected across the life course, as are public assistance receipt and assisted housing participation. Inpatient and emergency room services are tracked by state "all payer" datasets. Some programs for people with disabilities, including vocational rehabilitation programs, are tracked, as are placement in assisted living and nursing home care.

The DSEP inventoried all these data sources, surveyed the data holdings of existing IDS sites, and rated the data on their accessibility for a given IDS installation. The results of these efforts are listed in their final report, which also includes an appendix of the candidate data sources, the types of data held within them, and their likely utility for an IDS (Wulczyn et al. 2017). While a given agency may track hundreds of variables, the DSEP report identifies the relatively small subset that is likely to be nationally standardized (or approximately so), provided that it is subject to mandated federal minimum data requirements, and with prescribed data definitions. Not surprisingly, the data elements with the highest reliability tend to be those that are audited because they are associated with tracking service provision, billing, and payment. The DSEP report also provides a schema for classifying the types of data likely to be found in these datasets to improve the ease of understanding by potential data requestors (Wulczyn et al. 2017). These include distinctions for persons, types of service encounters, dates associated with services, places or providers for services, and exit codes or destinations.

Finally, the DSEP report considers how communities can address some of the data management issues associated with repurposing administrative data for research and evaluation. Data managers must consider the historical legacy of various data sources, how to reconcile conflicting pieces of information from one or more data sources, how to conduct and document record linkage approaches, and how to assess the quality and completeness of various data elements (Wulczyn et al. 2017). Each of these involves careful assessment by the data management staff of an IDS. The report presents some best practice guidance, and AISP hopes to create a community of experts in this area who can share their experiences to advance the collective understanding and appreciation for the data that can be most effectively used to generate actionable intelligence.

Conclusion

Our growing and diverse nation faces a myriad of social challenges. States and local governments administer dozens of major social programs intended to meet these challenges. However, they have limited knowledge about the people they serve, the impacts programs have, and the best ways to improve the effectiveness of these programs. In most cases, the lack of data is not the problem. Instead, there is a lack of collaborative dialogue among the public agencies that serve the population, the service providers who deliver the programs, the researchers and

subject matter experts with domain knowledge, and the public whose needs are to be addressed. That lack of collaboration extends to—and is partly the result of—the lack of data-sharing across agencies. As a network of advanced IDS practitioners, AISP has identified the four key domains for IDS development and operations. The expert panels commissioned in each of those domains have identified the most common challenges to institutionalized data sharing procedures and the recommended solutions to those challenges. The resulting reports, summarized here, provide a roadmap for states and local governments to more quickly and readily adopt best practices for IDS implementation, including guidance for how to address the complex legal and governance issues that set the framework for dialogue and collaboration. A recommended set of technology solutions would enable communities to adopt a low-cost and shared approach to doing this work, while maintaining site-specific governance, authority, and control over their data and how they are used. By adopting a national data model, jurisdictions can further engage in multisite collaborations with data that have known generalizability, reliability, and validity. With these reports, the path forward is clearer and the barriers reduced, and hopefully many more communities will be able to adopt IDS-based approaches to actionable intelligence, thereby improving the quality of life of their citizens through more effective and efficient public services.

Notes

1. The commission was also charged with determining the kinds of administrative data that are ultimately relevant for program evaluation and policymaking, and how to make these data available to researchers through a clearinghouse. See https://www.govtrack.us/congress/bills/114/hr1831/text.

2. States and counties spent more than $2.5 trillion in direct general expenditures for government services in 2012, the majority of which went to education, health care, and social safety net programs. When intergovernmental transfers are factored in, including federal funds, those expenditures increased to more than $3.5 trillion. See http://www.urban.org/policy-centers/cross-center-initiatives/state-local-finance-initiative/projects/state-and-local-backgrounders/state-and-local-expenditures.

3. Since 2008, nine additional jurisdictions have been added to the AISP network.

4. See the Belmont Report, https://videocast.nih.gov/pdf/ohrp_belmont_report.pdf.

5. See https://www.aisp.upenn.edu/aisp-network/.

References

Anderson, Nate. 2009. *"Anonymized" data really isn't—and here's why not*. Ars Technica. Available from https://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/.

County of Los Angeles, Chief Executive Office. 2016. *Homelessness initiative recommendations*. Los Angeles, CA: County of Los Angeles, Chief Executive Office. Available from http://file.lacounty.gov/SDSInter/bos/supdocs/101295.pdf.

Culhane, Dennis P., John Fantuzzo, Heather L. Rouse, Vicky Tam, and Jonathan Lukens. 2010. Connecting the dots: The promise of integrated data systems for policy analysis and systems reform. *Intelligence for Social Policy* Working Paper, Philadelphia, PA.

Fantuzzo, John. 2015. Towards a "what if" class: Practices of respect as the aim of teaching ethics to court-involved youth. *Teaching Ethics* 15 (1): 83–86.

Fantuzzo, John, and Dennis P. Culhane, eds. 2015. *Actionable intelligence: Using integrated data systems to achieve a more effective, efficient, and ethical government*. New York, NY: Palgrave Macmillan.

Fantuzzo, John, Christine McWayne, and Stephanie Childs. 2006. Scientist-community collaborations: A dynamic tension between rights and responsibilities. In *Handbook of ethical research with ethnocultural populations and communities*, eds. Joseph E. Trimble and Celia B. Fisher, 27–49. Thousand Oaks, CA: Sage Publications.

Gibbs, Linda, Amy Hawn Nelson, Erin Dalton, Joel Cantor, Stephanie Shipp, and Della Jenkins. 2017. *Principles and practices for ethical and effective ids governance: Setting up for success*. Philadelphia, PA: Actionable Intelligence for Social Policy, University of Pennsylvania.

Heidbreder, Brianne. 2016. Change and continuity in the study of state and local governance: A conversation with Ann Bowman. *State and Local Government Review* 48 (1): 63–71.

Heineman, Ben W., William Lee, and David B. Wilkins. 2013. *Lawyers as professionals and citizens: Key roles and responsibilities in the 21st century*, 1–84. Cambridge, MA: Center on the Legal Profession, Harvard Law School.

Jennings, Edward T., Jeremy F. Hall, and Zhiwei Zhang. 2012. The American Recovery and Reinvestment Act and State Accountability. *Public Performance and Management Review* 35 (3): 527–49.

Kettl, Donald F. 2009. *The next government of the United States: Why our institutions fail us and how to fix them*. New York, NY: W.W. Norton & Company.

Lane, Julia. 2016. Big data for public policy: The quadruple helix. *Journal of Policy Analysis and Management* 35 (3): 708–15.

Patterson, David, Ken Steif, Niall Brennan, Andreas Haeberlen, Aaron Schroeder, and Adam Smith. 2017. *Technology and data security expert panel report*. Philadelphia, PA: Actionable Intelligence for Social Policy, University of Pennsylvania.

Perlman, Bruce J. 2010. Governance challenges and options for state and local governments. *State & Local Government Review* 42 (3): 246–57.

Petrila, John, Barbara Cohn, Wendell Pritchett, Paul Stiles, Victoria Stodden, Jeffrey Vagle, Mark Humowiecki, and Natassia Rozario. 2017. *Legal issues and agreements expert panel report*. Philadelphia, PA: Actionable Intelligence for Social Policy. University of Pennsylvania.

Pew Research Center. 2017. *Americans and cybersecurity*. Washington, DC: Pew Research Center.

Stiles, Paul G., and Roger A. Boothroyd. 2013. *Ethical uses of administrative data for research purposes*. Philadelphia, PA: Actionable Intelligence for Social Policy. University of Pennsylvania.

U.S. Census Bureau. 2015. *Census Bureau reports at least 350 languages spoken in U.S. homes*. Available from https://www.census.gov/newsroom/press-releases/2015/cb15-185.html.

U.S. Congress. 2016. H.R. 1831. Available from https://www.govtrack.us/congress/bills/114/hr1831/text.

U.S. Government Publishing Office. 2016. *Budget of the United States government*. Available from https://www.gpo.gov/fdsys/browse/collection.action?collectionCode=BUDGET&browsePath=Fiscal+Year+2017&isCollapsed=false&leafLevelBrowse=false&isDocumentResults=true&ycord=931.

Wu, Fei, and Max Stevens. 2016. *The services homeless single adults use and their associated costs*. Los Angeles, CA: Chief Executive Office Service Integration Branch Research and Evaluation Services Unit. Available from http://www.aisp.upenn.edu/wp-content/uploads/2015/03/LACountyHomelessness2016.pdf.

Wulczyn, Fred, Richard Clinch, Claudia Coulton, Sallie Keller, James Moore, Clara Muschkin, Andrew Nicklin, Whitney LeBoeuf, and Katie Barghaus. 2017. *Recommended minimum data for integrated data systems expert panel report*. Philadelphia, PA: Actionable Intelligence for Social Policy, University of Pennsylvania.