IDS Governance: Setting Up for Ethical and Effective Use

Actionable Intelligence for Social Policy,
Expert Panel Report

Prepared by

Linda Gibbs, Amy Hawn Nelson, Erin Dalton, Joel Cantor,
Stephanie Shipp, and Della Jenkins

MARCH 2017

ACTIONABLE INTELLIGENCE
FOR SOCIAL POLICY

Table of Contents

Actionable Intelligence for Social Policy

University of Pennsylvania

3701 Locust Walk, Philadelphia, PA 19104

215.573.5827 | www.aisp.upenn.edu

I. Integrating Data to Drive Sound Social Policy

Increasing consensus among policy-makers, practitioners, and researchers holds that good policy is evidence-based. Yet those seeking to solve social problems often do so with a limited picture of reality. The data they have available are limited in quality or scope. Without sufficient data access, data linkages, analytic capacity, or interpretive support, informed solutions can prove elusive. At the same time, more and more data are collected and held in administrative repositories, and now sit underappreciated or underutilized. Organizing and presenting that data for broader use can uncover knowledge and spur innovation in policy and human services delivery, offering the potential to improve life outcomes for families and communities.

Integrated data systems (IDS) are being used by a growing number of governments, universities, and non-profits to link administrative data across agencies in order to understand and improve programs and practices through evidence-based collaboration. IDS can help front-line human services workers holistically manage client needs more effectively, can assist policy-makers in identifying interventions that work—or don't—in the context of complex systems, and can help researchers decipher patterns out of the overlapping, often-confusing signals the data send.

But it's not easy. Data may be collected and stored in antiquated systems and formats that frequently don't talk to each other. Ownership of the data is also dispersed. Owners may guard the data closely, hesitant to share with others or unaware of the data's potential to inform policy and practice. Many data owners do not have existing communication and decision- making systems to connect them, and lack a familiar infrastructure around collaborative action. It's hard to work in a void. Creating an ethical and effective environment for an IDS can help to fill this void.

This guide shares key lessons from the field, compiled by practitioners who have spent years pioneering these efforts. In the following sections, we lay out considerations for putting together a highly effective IDS. But we begin with this caveat: an IDS will likely start small and grow over time. It is an iterative process that is not linear in development. Our recommendations here are based on an ideal end state of a well-functioning, well-resourced, mature IDS. Do not be daunted. Start with a clear vision and purpose, laying the groundwork for the healthy growth and development of your IDS over time. From there, work to build policies and a governing structure to accommodate these goals and promote ethical and effective use of the data and research as you grow.

II. The "Ethical" IDS

The ethical foundation of integrated data stems from the core belief that data should be gathered and used as a public asset to advance social good, making best use of public resources. Too often, conversations around data sharing start from the premise that releasing data is a risk. These concerns often stop a data-sharing effort in its tracks. **We suggest that an ethical imperative exists to respectfully share and use data to the best of our ability.** Knowledge about what works and what doesn't, as revealed by the sharing of data otherwise closely held, ensures that the public interest is served by shedding light on those facts and releasing them from, at best, ignorance, and, at worst, entrenched self-interest that would prefer to protect the status quo.

Of course, making data available responsibly and ensuring data quality and security takes a great deal of staff time and resources. There is no benefit from creating a data mausoleum, where data are collected and integrated but not fit for use. An ethical IDS should ensure that data are made available in a useable format and that incentives are created to ensure that they are used for the public good.

At the same time and of equal importance, the ethical use of IDS requires high standards of integrity around data quality and usage. An ethical IDS must ensure that data are de-identified **to protect personal privacy** and are not in danger of being re-identified. If identifiers are needed to match records for analysis, those should be hidden so as not to permit identification by inference on small data sets. While those of us with experience in building and operating IDS have not experienced such

a data breach, and acknowledge that they remain extremely rare, the potential damage of client re-identification is always a concern. In order to protect against such failures, it is imperative that IDS creators, partners, and staff are aware of the risks and build in technological and procedural safeguards to protect privacy and maintain data integrity.

IDS integrity also requires careful adherence to **human subject research standards, including Institutional Review Board (IRB) review** and reliance on informed consent when necessary. And, importantly, the ethical use of IDS requires policies that support strong data interpretation and presentation to protect against error or misrepresentation.

In addition, even clean and accurate data can carry biases or act as a repository of the legacy of over-surveillance of poor communities, particularly communities of color, subject to unequal treatment and discrimination. Because human services data have the burden of that legacy and are often used to support decision-making that deeply affects the lives of vulnerable people and groups, **a strong ethical consideration should be given to ensuring cultural competency and protecting against unintentional perpetuation of discriminatory patterns of behavior**.[1] For an example of best practices in this area, see Allegheny County's predictive analytics tool for child welfare (Packard, 2016).

All of these factors should be carefully considered in undertaking and managing an IDS.

III. Starting with a Vision, Mission, and Principles

To do all of this effectively, an IDS must start with a clear articulation of purpose in the form of collaboratively constructed vision and mission statements, and guiding principles. The governing structures and policies and procedures to be adopted will vary widely based on the precise nature of the defining vision. A narrow goal of creating an academic research database will suggest one governance approach, which will differ significantly from a format to support a broad and ambitious agenda to create open access use of real time integrated data for any user. Spending time internally and with partners to build consensus around what the IDS is intended to achieve will help guide you to the best rules of engagement.

A. Vision

A vision statement[2] defines **the end state or goal that you envision**. It is typically lofty and aspirational.

The vision is the start of a long collaboration that will become institutionalized over time. You will want participation of all your core partners from the start, and will also benefit by including others who may not be central to creating your IDS but can help to guide your initial thinking. Inclusiveness will not only help in getting the goals right from the start, but will also foster a broad sense of transparency and responsiveness.

The process of shaping the vision statement itself should start wide and end small, encourage creative brainstorming with lots of perspectives, but arrive at a succinct statement everyone can agree on. Stay away from "how" for now.

Instead, think about these foundational questions:

❖ Whom do we serve?

❖ Why do we want to integrate our data? What do we see as the ultimate end goal?

❖ Whom do we hope will benefit, in what way? Specifically, what benefits will the system provide to clients, citizens, communities, program managers, policy-makers, researchers, academics, and reporters?

Sample vision statement: Policy and programs are well informed by rigorous analysis of timely, comprehensive, and integrated data, available to experts, practitioners, and community residents.

B. Mission

A mission statement defines the project's motivation and primary objectives. It is more concrete than the vision statement; it explains **how your IDS will contribute to the end state or goal**. We recommend a short statement of purpose, followed by a list of desired outcomes.

Sample mission statement: Support best practice in integrating and disseminating de-identified, client-level, public and private human services data for program and analytic use, to:

❖ *Create an accurate picture of reality*

❖ *Open data to broader public use*

❖ *Promote analysis and program evaluation*

❖ *Strengthen broad outcome assessment*

❖ *Convene researchers and practitioners together*

❖ *Integrate people, systems, and knowledge in a continuous improvement cycle*

❖ *Facilitate relevant original policy research that addresses scholarly questions*

❖ *Contribute to more evidence-based policy and practice*

C. Principles

The manner in which the vision and mission will be achieved is of the utmost importance. Ideally, the shared values of the stakeholders involved will guide the work of creating the IDS, managing requests for data, and sharing the results of research and analysis. It is useful to put these values into words up front in order to establish trust between partners and lay the foundation on which future decisions will be made.

Principles can be articulated in a simple bullet format, or may be prioritized in a ranking.

Sample IDS Principles:

❖ *Act first and foremost in the service of improved public good and services.*

❖ *Ensure that all use of the data is ethical, non-partisan, and impartial.*

❖ *Ensure transparency in the process of creating and managing the data model.*

❖ *Engage people, community, and practitioners in the production and sharing of data.*

❖ *Err in favor of the release of data as a public good unless there is a clear public danger in doing so.*

❖ *Ensure appropriate standards for the production, review, and dissemination of the data while maintaining individual data-holders' privacy.*

---

1   American Public Human Services Association (2015) covers this topic thoroughly.
2   A good resource for developing mission and vision can be found in Van Korlaar (2012).

❖ *Build mechanisms to minimize potential harm or risk of re-identification of data.*

❖ *Create quality control procedures and processes to ensure rapid error correction.*

❖ *Maximize the frequency of data reporting to be as real time as possible.*

❖ *Produce the data in user-friendly format accessible to a range of users of varying skill levels.*

❖ *Minimize the costs of data production, while maximizing the benefits to be derived from the data.*

❖ *Build mechanisms to ensure that all data meet minimal standards for accuracy, validity, and professional collection methods.*

❖ *Ensure cultural competency from project approval through data analysis and research use.*

IV. Stakeholders in Integrated Data Systems

Engaging the full spectrum of stakeholders in designing, launching, and governing your IDS is essential for achieving its mission and vision. Stakeholder engagement is of both practical and ethical importance, and essential to adhering to the principles and values you have articulated for your IDS. Some stakeholders are on the "critical path" to creating an IDS in the first place, others will be needed to ensure success of IDS operations, and still others must be engaged to ensure that the IDS can be sustained over time. As important, active engagement of those whose lives will be affected by use of IDS data is an ethical imperative, both as a matter of fairness and to ensure that data are used to maximum benefit. Stakeholders can help ensure that the highest-priority questions are addressed using IDS data and that the interpretation of findings is well informed.

In this section, we begin by expanding on why stakeholder engagement should be an ongoing priority for your IDS; we then provide guidance on effective strategies for identifying, assessing, and engaging stakeholders; next. we turn to a description of categories of stakeholders you should consider in IDS development and operations; and finally, we enumerate strategies to consider for engaging stakeholders that are important to an ethical and successful IDS.

A. Why Engage Stakeholders

There are five essential reasons to engage IDS stakeholders at every stage of IDS development and operations:

1. Achieving high ethical standards

2. Ensuring compliance with applicable regulations and laws

3. Designing and maintaining governance structure and process

4. Identifying and overcoming barriers to implementation and successful operations

5. Promoting sustainability

Achieving high ethical standards demands that representatives of those agencies whose data are held by the IDS are aware of and have ample opportunity to influence IDS design and use. These individuals and groups can serve as a "moral compass" for the IDS, and if they are effectively engaged they can also counter resistant stakeholders, who may invoke privacy and confidentiality concerns as a way to impede data sharing.

Compliance with all applicable federal, state, and local laws and regulations is, of course, essential. But the web of rules and regulations can be confusing and sometimes conflicting. Not having strong legal and regulatory expertise available can leave the IDS vulnerable to arguments that data access and use is not permitted when, in fact, such access and uses are permissible and even encouraged.

Inviting the full range of IDS stakeholders to be part of formal governance can be unwieldy and is unnecessary. But understanding the perspectives and influence of your stakeholders is a critical first step to designing and maintaining effective governance.

Successful IDS implementation also depends on the cooperation of stakeholders who own or hold the data that you are seeking to integrate within the IDS, provide access to funding and other resources, and otherwise have power to advance or thwart critical IDS functions. Knowing who these stakeholders are, where they stand on IDS implementation, and how they can be persuaded to cooperate is essential. Whether stakeholders are part of formal governance or not, building strong relationships with them is imperative.

Finally, ongoing stakeholder engagement is vital to achieving not only the desired impact of the IDS on social policy, but also ongoing financial sustainability and access to important data.

B. Identifying and Mapping Stakeholders

A first step in stakeholder management is knowing who to engage, assessing their interests in the IDS, and working to engage them in productive ways. To achieve this, it is useful to proceed through a four-step process.

1. Brainstorm key stakeholders.

2. Identify their interests, formal positions, and implicit expectations.

3. Assess the ways that their interests, positions, and expectations can advance or impede IDS goals.

4. Prioritize stakeholders and develop plans to engage them.

Brainstorming should be comprehensive. It may be useful to consider stakeholders in three categories: (a) **core stakeholders**, without whose engagement the IDS cannot achieve success, (b) other **direct stakeholders,** whose engagement can help facilitate (or impede) IDS success but who are not in the core group, and (c) **other stakeholders,** who can broaden interest of the IDS and deepen its constituencies.

The stakeholder relationships that contribute to actionable intelligence produced by an IDS can also be conceptualized as a series of bidirectional relationships between executive leadership, researchers, practitioners, and the public, as described in *Figure 1*.
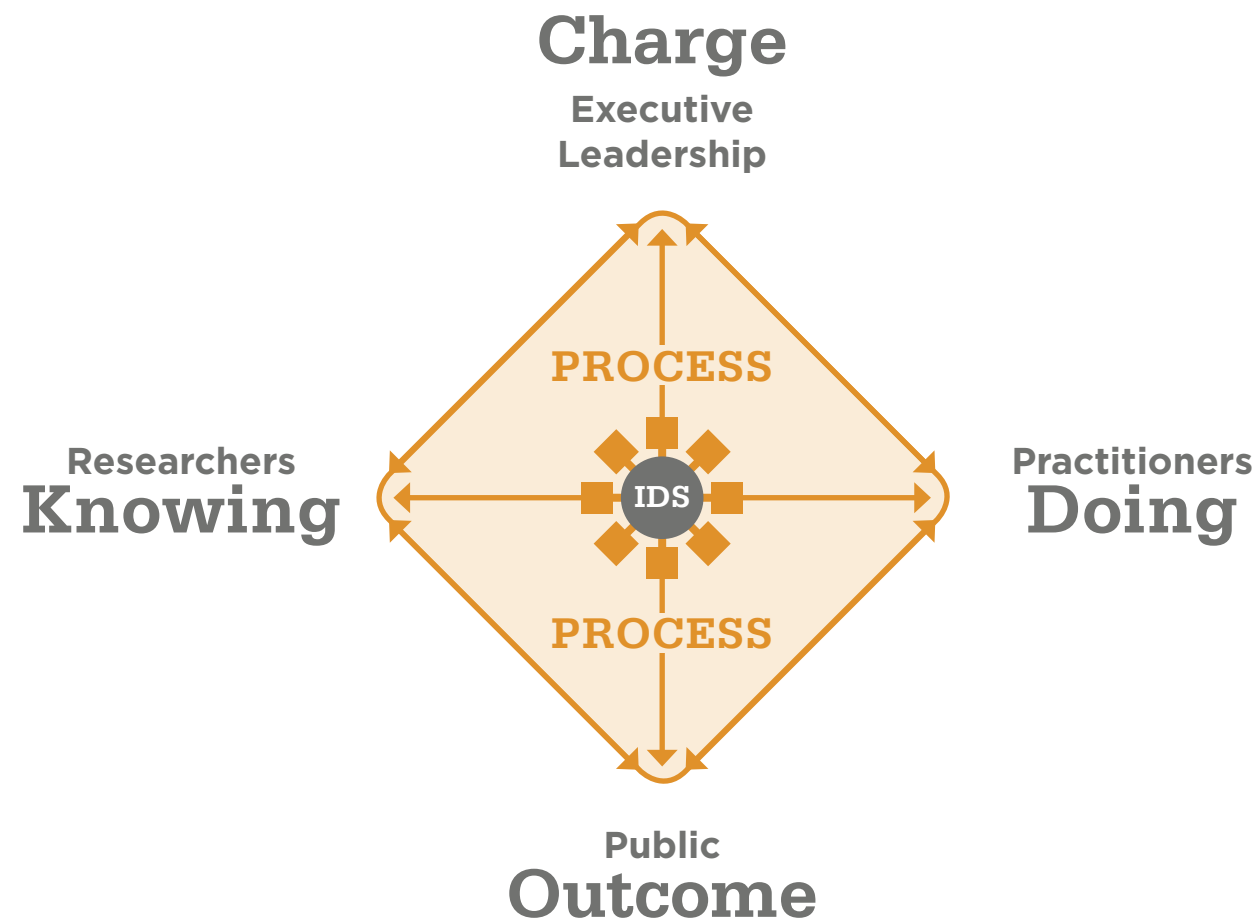
Charge
Executive Leadership

Researchers
Knowing

PROCESS

IDS

Practitioners
Doing

PROCESS

Public
Outcome

Figure 1: Configuration of Key Contributors to Actionable Intelligence.
Source: Adapted from Fantuzzo et al. (2015).

Once enumerated, the interests, positions, and expectations of each stakeholder group should be mapped. Next, assess whether or not they are likely to support IDS implementation and operation, or whether they are likely to be uninterested or neutral. Importantly, assess why or why not, and in what ways, each group is likely to advance or impede the IDS.

Consider the possible range of interests—both positive and negative—of each group, such as:

❖ Interest in improving service delivery, fostering research, or advancing policy goals

❖ Making the case for additional resources or identifying opportunities for savings

❖ Strengthening governmental administration, accountability, or efficiency

❖ Potential of being embarrassed about poor data quality, programmatic problems, or exposing unmet needs/new costs

❖ Potential burdens of cooperation

❖ Inertia and organizational culture

❖ Privacy and security

❖ Turf wars

❖ Legal compliance concerns

Next, prioritize the stakeholders. To do this, it is useful to think along two dimensions, as illustrated in *Figure 2*. Powerful stakeholders with strong interests (quadrant A) demand the most attention. Those who are likely to be advocates for the IDS should be engaged early and encouraged to help address the concerns of other groups that may be influential but less supportive. For example, the one business association, eyeing possible reductions in the cost of government operations, might be engaged to help persuade other pro-business constituents who fear that an IDS represents government overreach. Drawing this stakeholder mapping tool on a large whiteboard and placing sticky notes for each stakeholder in the quadrants can be a helpful group exercise to set outreach and engagement priorities.

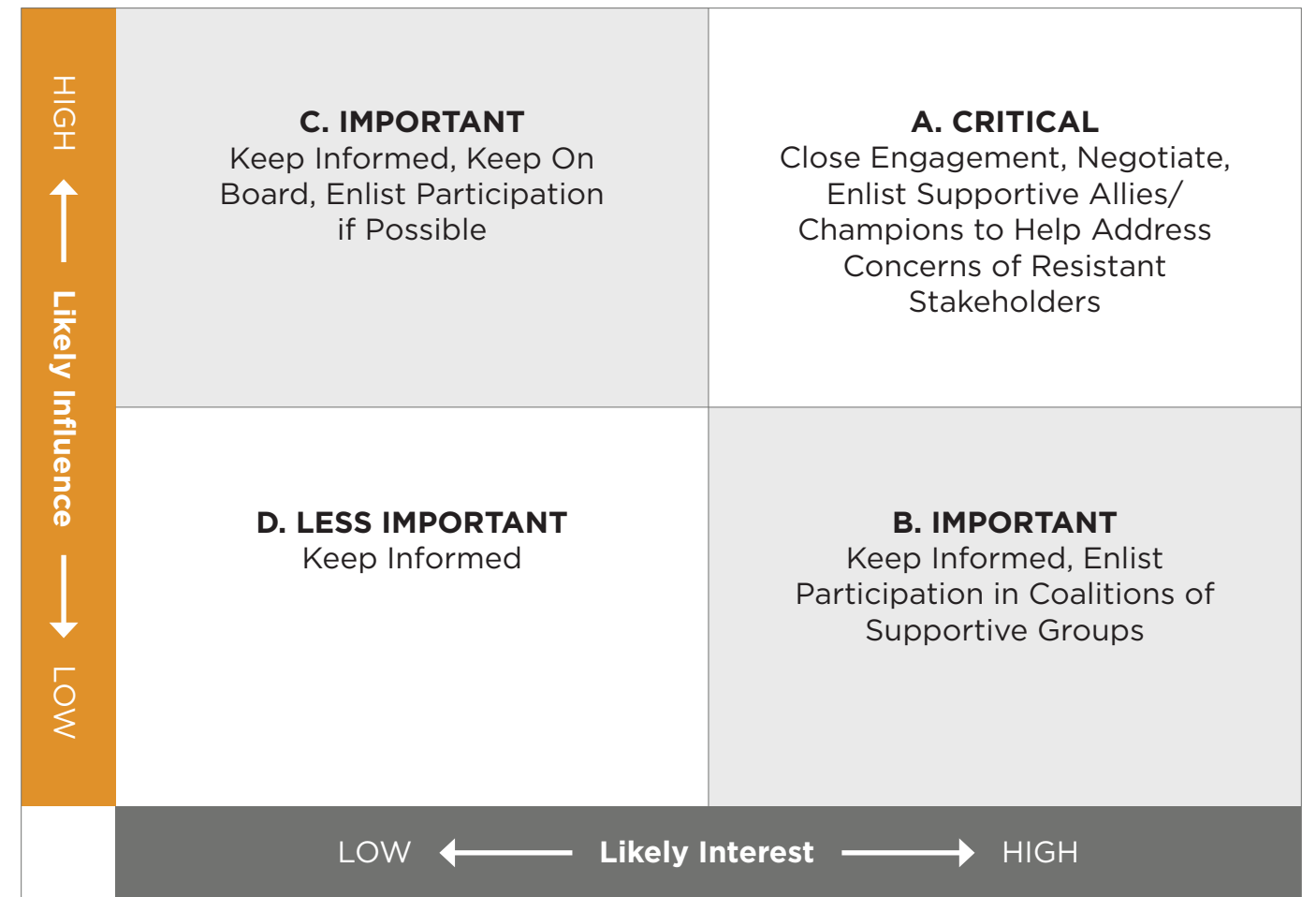|		
HIGH ↑ Likely Influence ↓ LOW	**C. IMPORTANT** Keep Informed, Keep On Board, Enlist Participation if Possible	**A. CRITICAL** Close Engagement, Negotiate, Enlist Supportive Allies/ Champions to Help Address Concerns of Resistant Stakeholders
	D. LESS IMPORTANT Keep Informed	**B. IMPORTANT** Keep Informed, Enlist Participation in Coalitions of Supportive Groups
	LOW ← **Likely Interest** → HIGH	

Figure 2: Prioritizing IDS Stakeholders.
Source: Adapted from Bryson (2004).

C. Categories of Stakeholders

The list of stakeholders who need to be engaged will vary depending on the particular context of each IDS, but there are types of stakeholders that should be considered in all cases. Drawing on the tripartite classification described above, a possible classification of stakeholders that should be considered in the identification and mapping process includes the following.

1. Core stakeholders

❖ Data owners and contributors (directly contributing, or facilitating access)

❖ Funding sources (government, private foundations, other)

❖ Public agency leadership and key elected officials

2. Other direct stakeholders

❖ Data users (researchers, advocacy groups)

❖ Technical experts (legal, data technology, security, research methods, fiscal)

❖ Privacy advocates

❖ Advocates for vulnerable populations and communities

3. Other stakeholders

❖ Business groups

❖ Good government groups

❖ Other citizen and public interest groups

D. Engagement Strategies

Once stakeholders are enumerated, their interests identified, and engagement prioritized, the next step is to develop an outreach and engagement plan. The most supportive and powerful stakeholders (quadrant A) can be enlisted to work with less supportive groups with which they have an affinity (as in the business group example above). Reaching out early to potential opponents and incorporating their input into IDS design and governance can be especially important tools for avoiding active opposition later in the process. More "wholesale" communications tools, such as e-newsletters and social media, can be employed for stakeholders who have fewer vested interests or are less influential (quadrants B and D). It may be helpful to try to stimulate engagement among groups that are influential but less interested (quadrant C), particularly if they have affinity with potentially powerful groups that may be positioned to impede IDS success. While a well-functioning IDS cannot satisfy all constituencies, it is vital to take all views seriously. A well-designed outreach and engagement plan will guide how best to use available resources. Regardless of the level of effort devoted to a given stakeholder group, it is most important to maintain strong communication with all.

E. Core Questions for Stakeholders

The questions to be addressed in stakeholder conversations should be driven, to a great extent, by the interests and concerns of the stakeholders themselves. However, the following are core questions that should be considered in formal and informal conversations with stakeholders:

❖ What are the specific reasons for creating an IDS? Research? Program development? Informing policy or budget decisions?

❖ Are all necessary parties part of the planning process? This is not simply a matter of having substantive experts at the table. There may be a political leader or influential staff person who is not a substantive expert on IDS issues but is an advocate of using integrated data.

❖ What ideas do people have regarding the structure of an IDS? Is there going to be a central repository in which data from different sources are pooled, or will data "stay where they are" while approved parties access them?

❖ What analytic priorities should the IDS pursue in the near term and longer run? Establishing an IDS is complex, and starting with a clear and manageable focus is essential for success.

❖ What data sets are required to address IDS research priorities? What data sources and elements are required to address those research priorities? Who "owns" the required data? Who currently has access to the data? Who potentially might have access? For more information on identifying data sets, see *Appendix A* and Wulczyn et al. (2017).

❖ What legal issues need to be addressed with the required data sources and elements? As the inventory of data sets is developed, it will create a foundation for identifying and discussing potential legal issues associated with each data set. As discussed, some of those legal issues will be real issues; others will result from a limited interpretation of the law. It is useful to spend time identifying as precisely as possible what legal issues exist and what issues might emerge. For more on the legal implications and requirements surrounding IDS development, see Petrila et al. (2017).

V. IDS Development

A well-functioning IDS will be supported by clear legal agreements, often in the form of Memoranda of Understanding (MOU) and data-sharing agreements among participating organizations, a well-defined governance structure, clear policies and procedures to support decision-making, routine meeting structures, and well-documented proceedings, all supporting a culture of trust, collaboration, and openness in managing the IDS.[3]

Many existing IDS have not created the formalized governance structures that we describe here as recommended best practice. All procedures will be iterative, as practice and need change. It is likely that you will start more informally and formalize structures over time. However, if you start informally, we recommend that you revisit this decision frequently as your IDS capacity and infrastructure develop to ensure that you adapt to organizational needs as they evolve.

Regardless of how formal or informal your IDS governance structure is, alignment between the many necessary policies and procedures is essential—ideally they should be conceptualized and designed together so as to build on one another. Key policy and governance documents are explained in more detail in *Appendix B,* and examples of each are provided for reference.

Be ready to tailor your approach to your local needs, and be aware that the IDS development process may happen in clear stages, as described below, or may be more organic and have overlapping stages. In either case, returning to the vision and mission, early and often, helps to improve chances of success. Though the evolution of each IDS is unique, it may be useful to think about the development of an effective IDS in three phases (Culhane et al., 2010).

---

3   Data-sharing agreements can take many forms and are sometimes called data use agreements or data use licenses. See Petrila et al., 2017.

❖ *Collaboration and planning phase*: This includes initial stakeholder engagement around vision, mission, and principles, as well as drafting and execution of MOU and data-sharing agreement(s) to ensure access and ethical use of the data.

❖ *Demonstration phase:* An infrastructure is built to store the data, and procedures are drafted and piloted for one or more research projects.

❖ *Institutionalization phase:* A regular flow of research projects is developed, and focus moves toward the iterative process of refining the IDS functionality, securing long-term funding, and disseminating findings.

As the IDS infrastructure is built and utilized, the culture of agencies will shift as data quality and data-driven practice is emphasized. Staff will operate within a changed environment, with shifting expectations. Attention to supporting and encouraging staff growth in the use of data, both as providers of data input and users of data reports for feedback and outcome improvement, can go a long way in overcoming resistance and accelerating adoption and ultimately improving outcomes.

VI. IDS Governance Functions

How various stakeholders are involved in IDS governance as your system develops will depend on context, but each IDS must identify governance structures that will best serve the needs of data partners and community stakeholders. These structures can be formal or informal, and may include an Executive Board, Data Subcommittee, full-time IDS Staff, Data Partner Group, and/or Advisory Committee or Workgroups. The creation and maintenance of these groups will be dependent upon the vision, mission, beliefs, and stakeholder groups supporting the development of the IDS.

A. Executive Board (Deciders)

The Executive Board is charged with providing oversight and direction. One of the most important functions of oversight and direction setting is staffing. The Executive Board should make staffing decisions with the governance model, vision, mission, beliefs, and key stakeholder groups in mind. While the work of IDS is technical, the core work involves relationship building with key stakeholders, and staffing decisions must reflect this. The work of IDS staff will be independent, and the Executive Board will need to provide guidance and approval of strategic decisions to ensure alignment with stakeholders.

Ethical use of the data is paramount, including confidentiality and data security. The Executive Board is charged with governing data use, which means championing, communicating, and committing resources to support these principles. A key component of governing data use is the board's role in setting the IDS research agenda. This agenda should be aligned with IDS goals and existing resources.

A key concern of any entity is sustainability, and the Executive Board is charged with seeking resources, developing a business model, and maintaining fiduciary oversight. Funding is critical, and funding often comes with commitments and obligations, some of which may be at odds with the vision, mission, and principles of the IDS. The foundation of any successful IDS is ethical, so it is important that the funding is aligned with priorities. Strategy for long-term sustainability should ensure that funding does not drive the direction of the IDS, but rather complements activities determined by stakeholders.

The Executive Board should be composed of a group of individuals who are committed to data sharing to benefit the community. This group should be made up of data partners, executive leadership, and community-level decision-makers such as funders and elected officials. As with the creation of any board, particular attention must go to the skills and assets that individuals, and the agencies they represent, bring to the table.

B. Data Subcommittee (Approvers)

The Data Subcommittee could fulfill an advisory function and be either a subcommittee of the board or a separate entity entirely. This group is charged with oversight of the technical policies, processes, and procedures of the IDS. The group oversees data security and confidentiality, and technical procedures governing acceptable use of the data, including data integrity and ensuring the use of appropriate methodological approaches to address stated research questions.

The most important role of the Data Subcommittee is to approve special case data and research requests for data beyond what is routinely and publicly available. This process should be multi-step and include independent and collective review.

The Data Subcommittee should also ensure that there is a sufficient internal audit function to check on the use of the IDS and determine whether any breaches have occurred. There should be strong standards for responding to any concerns raised about the misuse of data, and a proactive approach to verifying ongoing integrity of operations. For an overview of the Data Subcommittee's potential goals, roles, and policies, see *Appendix C*.

C. IDS Staff (Doers)

The IDS staff are engaged with the day-to-day work of making the IDS functional. Staff are charged with facilitating and completing work that is technical, procedural, and deeply relational. A variety of skill sets must be cultivated, with an emphasis on people who are adaptable, as the needs of the IDS will shift dramatically through the stages of development. Positions must be filled with individuals who possess a variety of skill sets and have a high aptitude for learning. The IDS landscape changes daily, and staff must be willing and able to make constant course corrections. Suggested positions can include Director, Assistant Director, Data and Research Coordinator, Research Analyst(s), Database Administrator(s), Data Scientist, Data Security Officer, and Legal Counsel. In the beginning stage of the IDS, some positions are mission critical (such as Director and Coordinator), while other positions can be shared, combined, or contracted as consultants as needed.

Care should be taken around potential conflicts of interest within the reporting structure. For example, the Database Administrator/Data Scientist/Data Security Officer should have autonomy and independence from research staff. The governance structure should encourage checks and balances, with a reliance on policies and procedures, rather than staffing/report structure.

IDS staff support the function and use of the data system. This includes managing all technical aspects of data sharing (technical assistance, data transfer, linking, cleaning, and pulls). The core of this work is not technical, but in developing relationships among stakeholders and data partners. Carefully cultivating and maintaining relationships supports the successful negotiation of data-sharing agreements, maintaining data quality and data integrity, supporting the reconciliation of inevitable conflicts among stakeholders, facilitating strong community engagement, and ultimately ensuring that the goals of the IDS are aligned with the needs of the larger community.

Staff are charged with the maintenance of multiple governance structures. This involves oversight of key functions of the Executive Board, such as supporting a long-term strategy for sustainability. In addition to supporting the creation and development of the research agenda to fit within community needs and available resources, IDS staff must lead the approval and review process of data and research requests. Regular and consistent communication among all stakeholders is critical.

All of these roles must be balanced with what many would consider the "so what" of an IDS. Now we have these data gathered, linked, and de-identified, so what? IDS staff must support researchers to ensure that the IDS is actually analyzing, interpreting, or using knowledge gained from research to guide its actions. IDS staff should also have a level of comfort, even expertise, with data holdings. This ensures that the idiosyncrasies of administrative data are communicated to researchers to support quality research. This knowledge can only be developed in conjunction with the data partners.

Staffing configurations will vary widely depending on the purpose, size, and capacity of an IDS. For more detail on particular staff roles and competencies to consider, see *Appendix D*.

D. Data Partner Group

Because an IDS typically does not "own" data, data partners are critical stakeholders. In simple terms, the IDS would not exist without the data provided by partners. Data partners support the technical aspects of data sharing, particularly data transfer. Agency-level data must be pulled from an enterprise system, and reconfigured for use in the IDS. This is a technical process that must be coordinated by staff and partners in a safe and secure manner.

Data partners not only provide the actual data, but also provide valuable expertise around data variables. This expertise is critical to ensure high-quality research. Researchers are typically trained to utilize data sets collected for research purposes, and because of this, researchers will make assumptions on the collection procedures and data quality of administrative data, particularly around the use of data longitudinally. Data partners are a critical part of the process to ensure that the variables being utilized to answer specific research questions are, in actuality, able to answer those questions. Because administrative data are collected for a specific purpose, to administer a program, these data often shift. These shifts can be due to a change in law, a change in agency-level data infrastructure, a change in personnel and/or training (example, a caseworker is directed by a new supervisor to not worry about inputting X field into a registration form), or a shift in service provision. The value of using administrative data is lost if the data and potential gaps in the data are not understood by the researcher. This insider knowledge will come primarily from those who regularly input the data: the programmatic staff at the data partner organization or agency.

E. Advisory Committees/Workgroups (Advisors)

The Advisory Committees and/or Workgroups will be dependent upon the vision, mission, beliefs, and stakeholder groups supporting the development of the IDS. Whatever the agreed-upon structure, the group of ad hoc advisors will help inform, advise, and support specific system or project needs. In many ways, these groups will complete the work of the board, in conjunction with staff. Advisors should represent a variety of institutions, disciplines, and stakeholder interests.

VII. IDS Policies and Procedures

Together, the stakeholders above must carry out their governance functions to ensure the success of the IDS. Policies and procedures will guide them each step of the way as they move each research project from an idea to a final product.

An IDS may have publicly facing data and prepared reports that are routinely supplied and made widely available. In other cases, the data needs of a user are specific and unique, and a special request would need to be made to produce the data. Depending on the degree of data integration, significant time and effort could be involved.

Effective and sustainable IDS must be flexible to respond to several different potential ways in which such research projects might be prioritized for action and approved for data usage.

A. Agenda Setting

The most thoughtful (and hopefully common) way that a project gets prioritized is through an agenda-setting process. This is primarily the charge and responsibility of executive leadership. The agenda-setting process can be fairly elaborate, inclusive, and time consuming. For example, an IDS Executive Board may agree to conduct an agenda-setting process every two to five years, with review on the off-years. This

process would begin by soliciting key stakeholders (researchers, practitioners, and community members) to provide insight regarding critical issues in the community. After this list is developed, IDS staff should conduct preliminary analysis of the problem statements, asking questions such as:

- ❖ Is this a problem that the community wants to solve?
- ❖ What data are needed to understand the problem?
- ❖ Can one agency or data provider solve the problem alone, or are multiple parties necessary?
- ❖ Does studying the problem align with our principles and have public benefit?
- ❖ What data-sharing/privacy issues need to be considered?
- ❖ Is this project feasible given our current capacity?

Early on in the development of an IDS, it will be important to consider projects that can publicly demonstrate the utility of integrated data. For example, a high-profile study by the Center for Innovation through Data Intelligence (2015) that showcased the adult outcomes of youth exiting the foster care and juvenile justice systems helped to solidify public support for their ongoing IDS work.

Ultimately, the IDS staff will score the projects by a set of decision rules. The IDS staff will likely present this rubric to the Executive Board, who will then prioritize the projects and set the agenda for the next two to five years.

While agenda setting might be the most common way projects get started, IDS leadership will likely take advantage of other opportunities. These could include the following:

- ❖ *A funder request:* If a critical partner such as a foundation comes to the IDS with a project, it should be seriously considered. The project will still be assessed by the decision rules suggested above, but if it fairs well, it could be prioritized over other projects and begin right away. Taking advantage of current priorities can elevate the profile of the IDS and make it part of the policy infrastructure of the community, increasing the likelihood of sustainability.

- ❖ *A policy-maker request:* Likewise, an IDS may wish to prioritize research that is of immediate interest to a policy-maker, relevant task force, or blue-ribbon panel in order to build on existing political momentum and public interest.

- ❖ *Support for a funded project:* Similarly, if a partner to the IDS with a funded grant or funding for a project, the project should still be assessed according to the IDS' decision rules and in light of all core IDS principles, but could be initiated faster than other projects if it meets the necessary standards.

- ❖ *Critical need from a data provider:* If a public agency (and data partner) has a dire need for data linking and analysis, perhaps to support analysis of an emerging problem, it will be in the best interest of the IDS to consider this request and perhaps even prioritize it to ensure timely delivery of actionable intelligence.

B. Request Refinement and Execution

Once a project is prioritized, it will go through a series of reviews to refine and execute the data request. IDS staff often shepherd the project through these reviews and use their relationships and persuasion skills to move the project along. If the project gets stuck, the Executive Committee might need to assist. Depending on the IDS context, a project may be subject to one or more of the following peer review processes:

❖ *Legal review:* Once the request is well understood and prioritized, it may be reviewed by an internal or external legal team. This process will vary widely from site to site. There may be a full committee convened to review together, or separate consultations with individual attorneys from each data provider. (See Petrila et al. (2017) for legal considerations.)

❖ *IRB review:* In a university setting, a formal IRB Committee will likely be required to approve the methodology and attend to human subjects concerns.

❖ *Technical review:* The project will be vetted by the technical team (Data Subcommittee). This group will work with the partners to further specify the data request, understand the variable definitions, execute the query along with IDS staff, and provide the data to support the project. This is typically a highly iterative process and requires both great knowledge of the data and strong working relationships with the partners who collect them.

❖ *Data supplier review:* In many IDS sites, the data partner or provider will review and approve the project. This review varies from site to site. In some sites, the data provider can raise objections with the goals/purpose of the project and seek clarification or veto use of data. In other sites, the review is more pro forma, a check to make sure that the legal and technical issues are in order.

Once the project has cleared the refinement process, the project will be approved by senior IDS staff. In some instances, this approval will also be signed off on by the Executive Board (or the Chair of the Executive Board).

C. Preparing the Data

Once the project has been approved, the data can be prepared and shared with the requestor. This process can be relatively straightforward, or it can be time consuming, depending on the complexity of the request and the researchers' familiarity with the data they are requesting. It will help immensely if data requests are standardized and specific. For example, an IDS may wish to require that all data requests include a clear articulation of the research question or questions to be answered, specific time frames for each data set requested, and a list of the data elements of interest within each larger data set.

Several consultation sessions with the database committee and/or subject matter experts may be required before IDS staff and researchers are both clear that the data requested are available with adequate detail and sufficient to run the desired analysis. Ultimately, it is the job of the IDS staff to ensure that the data requestor has the requisite understanding to conduct the analysis.

However, there may be innovative ways to approach this task. For example, data preparation responsibilities could be distributed or shared between partners in organized communities of practice to minimize the burden of programming code or cleaning and recoding of data each time, especially if several concurrent projects rely on all or some of the same administrative data.

D. Monitoring Data Security

From the moment data are shared, their security and integrity must be maintained. Whether staffed internally or convened as a workgroup or board subcommittee, a data security group should be responsible for ongoing monitoring of IDS data. The IDS may even wish to set up an external audit of their practices (similar to a HIPAA audit) on a routine basis to make sure that this group is not only compliant, but operating in a way that meets standards agreed upon by all data partners.

E. Analysis and Interpretation

Once the data are provided, the research team (which could be made up of IDS staff, data partners, independent academics, or some combination thereof) will conduct the analysis. Whether they are part of the research team or not, IDS staff will play an important role in the interpretation of results. Ideally, multiple times in the process, the IDS team will convene researchers and practitioners (and sometimes impacted communities) to review and translate findings, discuss policy and practice implications, and raise new questions. This process can range from informal to formal and can vary greatly in size depending on the project and the partners.

The product of the analysis can take multiple forms, from a journal article to a less formal report or briefing. Regardless of the format, the product should be reviewed by the data providers, executive committee members, or data subcommittee. The work may also need to be peer reviewed by academic experts and community experts. While the agency that is the subject of the work has a right to review the final products, they do not have editorial rights. Unless the work cannot pass peer review or is of substandard quality, it should be released. The IDS staff will often facilitate the peer review and dissemination processes.

F. Translation and Use

IDS research results can take many forms. Some research is published by the agency or IDS itself; for example, see *Education and Well-Being of Children in Assisted Housing Programs* (Patton et al., 2016), a paper that presents findings from a quasi-experimental study conducted by staff at Washington State's Department of Social and Health Services Research and Data Analysis Division using their integrated client database. Other results may be published in academic journals by university-based researchers granted access to IDS data and approved to distribute their findings independently. For an example, see Fantuzzo et al.'s paper published in *Educational Researcher* (2013) entitled "An Investigation of the Relations Between School Concentrations of Student Risk Factors and Student Educational Well-being," which utilized data from Philadelphia's KIDS. IDS research results can also take the form of briefings to key stakeholders or data visualizations that allow the public to interact with the data set; see, for example, the data brief and accompanying visualization, *Suicides in Allegheny County, 2002-2014* (Allegheny County Analytics, 2017).

Some IDS sites are involved in working with public agencies and other partners to implement policy changes based the results of the analyses. They might assist with an action plan that stems from an evaluation or work with partners to design new interventions based on analysis conducted with integrated data. This sort of engaged interaction between data owners and researchers helps to support the ethical use of the IDS, by ensuring that the findings are used in practice.

For some IDS, it may also be appropriate to aggregate common data fields and make them available to the public through basic search functions on a data dashboard or in selected tables or graphics. This must be done with the consent of all data partners involved. Presenting data from an IDS in a digestible format has potential to inform and engage stakeholders, inspire and inform new research, and demonstrate the value of an IDS infrastructure to the larger community. In this way, it may also help build IDS visibility, and contribute to the relevance and financial sustainability of the system. Still, it is very important to exercise caution when allowing users to access integrated data, especially to build tables with small sample sizes, as this could lead to data security concerns around risks of re-identification. For this reason, making even limited IDS data public in a user-friendly format is a task that requires significant expertise, resources, and capacity, and should only be attempted by a mature IDS with the governance and security infrastructure to support responsible translation and use.

VIII. Potential IDS Governance Approaches

A. Characteristics of Governance Models

A variety of governance models can engage stakeholders and accomplish the functions detailed above in an ethical manner. The most effective of these models manage the data, relationships, and processes.

❖ Managing the data involves establishing processes to ensure that the data are used appropriately and ethically to align with the vision/mission/principles of the contributing agencies, and to meet the needs of those relying on the IDS to support policy decisions and actions.

❖ Managing relationships entails creating and nurturing relationships that ensure the political, economic, and technical sustainability of the IDS. This includes supplying information to relevant stakeholders, discussing ideas and findings using everyday language and easy-to-understand visuals, providing context through examples, indicating the uncertainties associated with the data, and providing the range of possible results (Kettl, 2016).

❖ Managing processes occurs through social learning that combines content development and management with social involvement and education to achieve broad community outcomes. In social learning, the communities and agencies learn by participating, and the objective is not solely to achieve an optimal solution to one problem but also an ongoing learning and negotiation process that can resolve many problems through communication, perspective sharing, and a progression of adaptive group strategies (Pahl Wostl and Hare, 2004).

B. Governance Models

The following typologies and examples of potential IDS governance models are categorized in *Summary Table 1* as executive-led, agency-led, university-led, and hybrid models. Each case study provides a history of IDS development, identifies key players, and outlines their funding model, as well as examples of projects and evaluations they have conducted. They may prove a useful starting point both for envisioning the remarkable potential of a functioning IDS and understanding some of the challenges to implementation and sustainability.

Summary Table 1: Examples of IDS grouped by governmental and leadership type

Coverage	IDS Project Name	IDS Lead Organization	IDS Leadership			
			Executive	Agency	University	Non-Profit
Allegheny County, PA	Allegheny County Data Warehouse*	Allegheny County Department of Human Services		X		
Mecklenburg County, NC	Institute for Social Capital Community Database*	Institute for Social Capital, Inc., University of North Carolina Charlotte			X	X
Florida	Policy and Services Research and Data Center	Florida Mental Health Institute, Department of Mental Health Law & Policy, University of South Florida			X	
Illinois	Chapin Hall Integrated Database on Children and Family Programs	Chapin Hall, University of Chicago			X	
Cuyahoga County, OH	ChildHood Integrated Longitudinal Data System	Center on Urban Poverty & Community Development, Case Western Reserve University			X	
Los Angeles County, CA	Enterprise Linkages Project	Los Angeles County (Exec Office and Department Public Social Services)	X			
Los Angeles and state of California	Children's Data Network	School of Social Work, University of Southern California			X	
New Jersey	Integrated Population Health Data Project*	Center for State Health Policy, Rutgers University			X	
New York, NY	Center for Innovation through Data Intelligence (CIDI)*	Office of the Deputy Mayor for Health and Human Services (HHS)	X			
Rhode Island	DataSpark	The Providence Plan				X
San Mateo, Santa Cruz, and Santa Clara Counties, CA	Silicon Valley Regional Data Trust	Silicon Valley Regional Data Trust, University of California, Santa Cruz			X	
South Carolina	South Carolina Integrated Data Warehouse	South Carolina Office of Revenue and Fiscal Affairs	X			
Washington	DSHS Integrated Client Database	Washington State Department of Social and Health Services, Research and Analysis Division	X			

* Detailed example presented below.

1. Executive-led models

Definition

In executive-led models, a strong executive (or group of executives) enables an organization to continue operations over several political administrations and provides the resources for creating and maintaining an IDS. The strong leadership or support is often able to create reliable streams of revenues and resources necessary to sustain the organization's operating functions and to shield the organization from many of the political and economic threats to its existence. Agency executives provide a vision and find resources for creating a data-driven organization and culture.

Example: New York City Center for Innovation through Data Intelligence

In 2002, New York City executive-level officials created a "One City" coordinated approach to providing agency services and reducing bureaucracy for families requiring multiple services (Kitzmiller, 2013). The One City strategy brought together several agencies—Department of Homeless Services (DHS) (the initial lead), Human Resources Administration, Administration for Children's Services, Department of Probation, Department of Youth and Community Development, Department of Correction, NYC Housing Authority, the Department for the Aging, and Housing Preservation and Development.

DHS successfully led the research to conduct case studies of homeless families that required multiple services for different agencies. By coordinating their services, case managers learned about and overcame issues where policies conflicted, identified where changes were needed, and created best practices that cut across agencies.

By 2008, New York City created the Center for Innovation through Data Intelligence (CIDI) to provide the legal and analytical infrastructure that allowed the sharing of data across agencies. The CIDI IDS uses mapping and other visualization tools to quickly identify problem areas. CIDI was instrumental in analyzing data from Superstorm Sandy in 2012 to identify where residents did not have heat, electricity, or telephone service, and then used this information to more efficiently deploy resources. In another example, CIDI data and tools were used to assess whether foster children were at a higher risk of being involved in the juvenile justice system.

CIDI has an Executive Director, a Director of Research, and a Director of Technology and Innovation. Analysts report directly to the Executive Director and are tasked with leading projects. CIDI has a data analytics process, management, and infrastructure to ensure that the data are securely maintained and ethically used. CIDI initially planned to build a large IT infrastructure and then determine need, but quickly realized that it was more beneficial to build the IT infrastructure based on need, driven by the social service programs demanding their services. The Mayor's Office provides direct support for CIDI's work, and they receive additional funding from foundations and non-profit organizations to conduct specific evaluations.

Advantages

❖ The executive-led model provides a champion who finds resources, sustains the will to create data agreements, and is driven by the desire for evidence-based decision-making.

❖ Multiple sources of data are available to support program decision-making and improve delivery of services.

Challenges

❖ The strong association with a leader or executive may make the model overly susceptible to leadership influences.

❖ Continuous effort is needed to ensure that the use of data covers a broad range of issues and multiple data sources.

2. Agency-led models

Definition

Agency-led models typically reside in organizations such as the Department of Health and Human Services where the systems were originally designed to help workers manage their workload and create a comprehensive picture of the needs and service consumption of clients utilizing multiple programs and services. These systems evolved to include executive offices to manage the data-sharing agreements and legal procedures and to promote utilization of the data for research (Kitzmiller and Burnett, 2015). Examples include Allegheny County's Department of Human Services (DHS) (see description below) and the Washington State Integrated Client Database.

Example: Allegheny County Department of Human Services

Allegheny County's Department of Human Services (DHS) created a data warehouse in 1999 with funds from local foundations (Kitzmiller, 2014). The DHS data warehouse is a repository of administrative data from DHS programs, including behavioral health, child welfare and homeless services, and external sources such as the Pennsylvania Department of Human Services, 20 local school districts, the criminal justice system, and the housing authorities of the City of Pittsburgh and Allegheny County. These data, initially collected for programmatic and payment reasons, are also valuable for conducting research that drives decision-making, service quality improvement, and system planning. At the time of publication, the data warehouse had more than 1.4 billion records from approximately 1.2 million clients, including demographic characteristics and provider information. Data-sharing agreements are in place with participating agencies to ensure that the data are used in legal and ethical ways that comply with the requirements for each agency as well as state and federal laws.[4] The data warehouse connects DHS data with data from its data-sharing partners.

The data have proved extremely useful to practitioners and researchers. One project has identified and analyzed system involvement of students in the Pittsburgh Public Schools (PPS). Integrated PPS-DHS data have shown that about half of all students access DHS services during their K-12 education, and those students tend to have below-average school achievement. By using integrated data systems from multiple agencies, analysts can study these potential factors and provide evidence to help shape policy and practice.

Advantages

❖ The agency-led model offers increased transparency and collaboration as program leaders share, integrate, and analyze their data with the common goal of improving the efficiency and effectiveness in the delivery of services.

Challenges

❖ As agencies tackle the problems within their boundaries, additional effort may be necessary to ensure a continuous broad coverage of data, the involvement of a range of participants, and the development of strong interagency relationships.

---

4   See http://www.alleghenycountyanalytics.us/index.php/dhs-data-warehouse/

3. University-led model

Definition

Unlike executive- and agency-led models, university-led models are generally created in response to research or policy interest in cross-cutting social issues, and some lack formal governing boards. The university-led IDS described below was born from a legislative mandate, is governed by a diverse board appointed by key stakeholders, and relies on a university research team to conduct agreed-upon projects. However, university-led models may also arise from the specific interests of academics. In these cases, it is important that planning and implementation is done in conjunction with practitioners and community members to optimize the utility and public benefit of the IDS.

Example: New Jersey Integrated Population Health Data Project

Growing out of efforts to address social determinants of health, a coalition of organizations advocated for the creation of an IDS in New Jersey, resulting in the enactment in January 2016 of the Integrated Population Health Data (iPHD) Project[5] into state law. The iPHD law spells out governance structure and process and assigns responsibility for implementation and operation of the project to the Center for State Health Policy (CSHP) within Rutgers, the state university. The governing board has 10 members, including ex officio appointments of the State Commissioner of Health (who chairs the board), Commissioner of Human Services, Attorney General, and Treasurer. Implementation of the iPHD legislation is still under way, but it is anticipated that CSHP will draft policies and procedures for approval by the governing board, raise funding necessary to establish and operate the iPHD (the law did not include a state appropriation), and support iPHD operations.

Advantages

❖ Establishing the IDS in statute can help make it more sustainable and less vulnerable to state leadership turnover.

❖ Formalizing governance can add transparency and accountability.

Challenges

❖ Detailing governing board processes and structure in state law could create rigidity and make it more difficult to respond to unforeseen challenges or needs in the future.

❖ Embedding governance across state agencies and involving the legislature can slow the implementation and governance process, as state officials face many competing demands on their time and attention.

4. Hybrid model

The governance models described above are more common, but they are also modified as stakeholders adjust to circumstances, needs, and capacities.

Example: Hybrid University and Non-profit Model—Institute for Social Capital, Inc.

The Institute for Social Capital, Inc. (ISC), founded in 2004, is located at the University of North Carolina (UNC) at Charlotte and was created "to foster university research and to increase the community's capacity for data-informed decision-making."[6] ISC is a 501(c)(3) organization situated within the UNC Charlotte Foundation that provides oversight for the ISC Community Database. ISC's operational support comes from UNC Charlotte, with some funds available specifically for the database through ISC.

5   See http://www.cshp.rutgers.edu/content/nj-iphd
6   See https://ui.uncc.edu/programs/isc/faq.

ISC collaborates with non-profit organizations, governmental agencies, and other organizations in the Charlotte region to support data-informed decision-making, with a focus on vulnerable populations. The core work of ISC is to facilitate data sharing across agencies; this includes fulfilling data requests, supporting efforts around data quality and standardization, and providing analytical support to address questions asked by agencies and organizations in the region.

ISC has multiple data request channels. A partner agency can submit a data request, but a research request requires IRB approval before individual-level, de-identified data linked across one or more agency is allowed to be released. The researcher submits a Data License Request to the ISC Data and Research Oversight Committee (DAROC). DAROC is composed of University researchers and members of the community, including a representative from each ISC data partner. The agency that owns the data being requested must approve the use of its data during the approval process. The ISC's Data Scientist links and matches data; then data is de-identified and sent to DAROC for review, prior to being released to the researcher.

The ISC currently has data-sharing agreements with over 40 government agencies and non-profits, including Charlotte-Mecklenburg Schools, the Mecklenburg County Department of Social Services, the Mecklenburg County Sheriff's Office, and agencies within the Homeless Management Information System.

The ISC Community Database has been used for a variety of projects including academic studies and program evaluations, and the majority of projects seek to better understand outcomes related to education, criminal justice involvement, housing, and homelessness. ISC is situated within the UNC Charlotte Urban Institute, an applied research institute that has served the Charlotte region for 45 years. Many data license requests originate from UNC Charlotte Urban Institute staff.

Advantages

❖ A university provides infrastructure and capacity to create and maintain an IDS, particularly in regards to technical needs.

❖ A non-profit Board of Directors is well positioned to convene stakeholders, facilitate data sharing, and balance research requests with community needs.

❖ A separate 501(c)(3) may provide an additional layer of data security in certain states with specific requirements around Freedom of Information Act requests, as well as greater fiscal flexibility (e.g., receiving donations and coordinating independent contracts).

Challenges

❖ There is the potential for competing priorities of a university, a non-profit board, community stakeholders, and data partners.

IX. Conclusions

Creating an IDS is both challenging and rewarding. Addressing the inevitable challenges requires constant attention to creating and maintaining an information infrastructure and refining optimal policies and procedures to support it. This discussion attempts to share advanced-stage best practices for IDS governance, fully recognizing that each IDS is unique and that most will start small and formalize over time.

The success of an IDS depends heavily on strategic leadership to solicit the cooperation of key partner agencies and to demonstrate the potential of safely leveraging integrated data to answer key questions. Success also depends on well-trained analysts and skilled researchers to produce, analyze, translate, and present information to inform decision-makers and action in the policy realm.

Well-developed IDS should aim to demonstrate and amplify their hard-earned successes, ideally using an experimental design to evaluate their own impact on data-driven policy making. Of course, this will require additional resources and sufficient time to ensure that effects can be measured. All partners must work together to embrace the lifecycle of a project, from planning and design to implementation and evaluation to redesign and modification. Partnering is also critical in developing a sustainable funding model and continuously building, maintaining, and updating the tools and capacities that make information actionable.

Ultimately, when an IDS is operating efficiently and ethically, data connects practitioners and policy-makers to create and refine policies and programs to better serve the community. Data can create the capacity and will needed to recognize interdependencies and synergies across agencies, to collaborate to identify and solve intersecting problems, and to learn to work together to build trusted relationships to improve access to services, reduce bureaucracy and redundancies, and create a more unified and transparent government.

References

Allegheny County Analytics. (2017). Suicide in Allegheny County 2002 through 2014, accessed January 26, 2017. http://www.alleghenycountyanalytics.us/index.php/2016/02/01/suicide-in-allegheny-county-2002-through-2014/

American Public Human Services Association (APHSA). (2015). *Roadmap to Capacity Building in analytics.* aphsa.org/content/dam/aphsa/pdfs/NWI/APHSA%20Roadmap%20to%20Capacity%20Building%20in%20Analytics_Dec%202015.pdf

Bryson, John. (2004). *Strategic Planning for Public and Nonprofit Organizations* (3rd ed.). San Francisco, CA: Jossey-Bass.

Center for Innovation through Data Intelligence (CIDI), New York City Office of the Mayor. (2015). *Young Adult Outcomes of Foster Care, Justice, and Dually Involved Youth in New York City.* http://www.nyc.gov/html/cidi/downloads/pdf/foster-care-justice-dually-involved-report.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.* Actionable Intelligence for Social Policy, University of Pennsylvania.

Fantuzzo, John, Whitney LaBoeuf, and Heather Rouse. (2013). An investigation of the relations between school concentrations of student risk factors and student educational well-being. *Educational Researcher*, *43*(1): 25-36. http://www.aisp.upenn.edu/wp-content/uploads/2015/06/Investigation_Relations_bw_School_Concentrations_PrevPhilly.pdf

Fantuzzo, John, Dennis Culhane, Heather Rouse, and Cassandra Henderson. (2015). Introduction to the Actionable Intelligence Model. In John Fantuzzo and Dennis P. Culhane, Eds., *Actionable Intelligence: Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government,* pp. 2-38. New York: Palgrave Macmillan US.

Kettl, Donald F. (2016). Making data speak: Lessons for using numbers for solving public policy puzzles. *Governance, 29*(4): 573-579.

Kitzmiller, Erika M. (2013). *IDS Case Study: New York. Creating "One City": New York City's Center for Innovation through Data Intelligence.* Actionable Intelligence for Social Policy, University of Pennsylvania. http://www.aisp.upenn.edu/wp-content/uploads/2015/08/CIDI_CaseStudy.pdf

Kitzmiller, Erika. (2014). *IDS Case Study: Allegheny County's Data Warehouse: Leveraging Data to Enhance Human Service Programs and Policies.* Actionable Intelligence for Social Policy, University of Pennsylvania. http://www.aisp.upenn.edu/wp-content/uploads/2015/08/AlleghenyCounty-_CaseStudy.pdf

Kitzmiller, Erika, and T. C. Burnett. (2015). The AISP Network: Three organizational models for building, using, and sustaining integrated data systems. In John Fantuzzo and Dennis P. Culhane, Eds., *Actionable Intelligence: Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government*, pp. 169-190. New York: Palgrave Macmillan US.

Packard, Thomas. (2016). Literature review: Predictive analytics in human services. Academy for Professional Excellence, San Diego State University School of Social Work. https://theacademy.sdsu.edu/wp-content/uploads/2016/03/sachs-predictive-analytics-report-feb-2016.pdf

Pahl-Wostl, Claudia, and Matt Hare. (2004). Processes of social learning in integrated resources management. *Journal of Community & Applied Social Psychology, 14*(3): 193-206.

Patton, Deleena, Jim Mayfield, Callie Black, Barbara Felver, and Irina Sharkova. (2016). *Education and Well-Bring of Children in Assisted Housing Programs: Findings from Washington State.* https://www.dshs.wa.gov/sites/default/files/SESA/rda/documents/research-11-238_0.pdf

Petrila, John, Barb Cohn, Wendell Pritchett, Paul Stiles, Victoria Stodden, Jeffrey Vagle, Mark Humowiecki, and Natassia Rozario. (2017). *Legal Issues for IDS Use: Finding a Way Forward.* Actionable Intelligence for Social Policy. University of Pennsylvania.

Van Korlaar, Craig. (2012). Guide to creating mission & vision statements. *Top Nonprofits*, September 24. https://topnonprofits.com/vision-mission/

Wulczyn, Fred, Richard Clinch, Claudia Coulton, Sallie Keller, James Moore, Clara Muschkin, Andrew Nicklin, Whitney LeBoeuf, Katie Barghaus. (2017). *Recommended Minimum Data for Integrated Data Systems*. Actionable Intelligence for Social Policy, University of Pennsylvania.

Problem	
Problem Definition	What is the problem you are trying to solve?
	What does success look like/how much does the needle need to move?
Interventions	What interventions do you have available to solve the problem?
Impact	If this is successful, what impact will this project have?
	Will it encourage future projects/goodwill?
Available Data	What data sets do you have access to that are relevant to the problem?

Data Governance	
Ownership	For the data sets that you have access to—do you own the data? Do you have permission to use the data? If you do not own the data, do you have the relationships with the data owner?
Accessibility	Are the data accessible outside the department/agency?
Security Policy	What security policies and legal considerations need to be in place for each of the data sources? (HIPPA, FERPA)

Implementation and Maintenance	
Technical Implementation	Do you have people in house who can implement/deploy the solution?
Data Infrastructure	Do you have the internal technical and data infrastructure to provide a continuous data feed from all the systems, and integrate the results/recommendations back into the agency systems?
Maintenance	Can you update, maintain, and support the implemented solution?

Data Readiness	
Relevance and Sufficiency	Do you have data that are both relevant and sufficient to solve the problem?
Quality	How is the data quality?
Collection Frequency	How often are the data collected?
Granularity	What is the level of granularity for the data sources?
History	How much history is stored, and how are updates handled?
Privacy	What data privacy policies do you have in place?
Documentation	How well documented are the data?

Organizational Readiness	
Staff Buy-in	How bought in are staff throughout the organization? What percentage of the staff are involved in data collection? Data analysis?
Data Collector Buy-in	How bought in are the people on the ground doing the data collection? Do they understand the importance and nuance of data collection? Do they get direct benefit from collection data?
Leadership Buy-in	How does leadership value data? Do they require data to be presented in order to make decisions?
People Resources	Do the people who will act on the results buy in?
Data Use Policy	Are there policies in place around who can use data, how they can use data, which parts can they use, and for what purposes?
Intervener Buy-in	Do the people who will act on the results buy in?
Funder Buy-in	How do your funders consider data? What kind of data do they require? What support for technology and personnel do they give you?

(Adapted from Center of Data Science & Public Policy, University of Chicago)

Policy Area	Document	Sample document
Recipient/Researcher Requirements or **Data Use Agreement (DUA)**	This document outlines the duties of any approved data recipient or researcher, which likely include the protection of confidential data, use limited only to what is outlined in the data-sharing agreement, and immediate notification of the IDS if data privacy is breached. This might also include requirements for citation, peer review, or advance notification prior to publication of any research findings.	Confidential Data Sharing– Recipient Requirements (Allegheny County Department of Human Services) Standard Terms for Data Use Agreements under FERPA (University of Pennsylvania, KIDS Integrated Data System)
MOU or Data-Sharing Agreement	The IDS will need a Memorandum of Understanding (MOU) or data-sharing agreement in place with participating agencies or data providers to ensure that the data are used in legal and ethical ways that comply with the requirements for each agency as well as state and federal law. Agreements may be entered into with individual participants or collectively.	Memorandum of Understanding for Participating School Districts (Allegheny County Department of Human Services) Memorandum of Understanding (City of Philadelphia, KIDS Integrated Data System)
Bylaws/Board Policies/ IDS Protocols	Bylaws may lay out the IDS objectives (vision, mission, principles), as well as the roles and responsibilities of each member party. This document may also establish timelines for regular meetings, decision rules for project approval, and requirements for board composition. Bylaws should be regularly revisited and updated. Alternatively, in the absence of a formal governing board, objectives, roles, and responsibilities of member parties may be outlined in IDS protocols agreed to by leaders of all data-sharing partners, much as in an MOU or DUA.	Bylaws (Institute for Social Capital, Inc., University of North Carolina at Charlotte) CIDI Data Hive Protocols (Center for Innovation through Data Intelligence, New York)
Policies of Board Subcommittees	Board committees and subcommittees are often charged with the creation, management, and oversight of more specific policies related to legal agreements, data quality, data security, data use, and research oversight.	Data and Research Oversight Committee Policies and Procedures (Institute for Social Capital, Inc., University of North Carolina at Charlotte)

Goals and Roles	Common Policy Elements
Defining stakeholder representation	Clear guidelines around committee representation and participation; should include a mix of specific partners, agencies, and content expertise (technologists, methodologists, etc.)
Setting privacy policies	Determine formats for sharing data, including de-identification guidelines; set requirements around access to and release of data
Reviewing data requests	Determine required elements of data requests, decision rules on requests, and data request prioritization
Ensuring compliance with established policies	Flexibility to foster an iterative process of policy refinement, including oversight of requirements for and compliance with data license request process, such as IRB approval, peer review, data use and dissemination
Overseeing technical supports	Database development, data quality control

Role	Skills/Competencies	Training/Experience
Director	Nuanced understanding of human subjects research, social science research, and a wide range of research methodologies Familiarity with IDS stakeholder groups Comfort in navigating bureaucracy Ability to establish, cultivate, and maintain a wide range of relationships Excellent communication skills; ability to communicate effectively among a range of competing interests Ability to keep abreast of IDS field with dynamic legal parameters, best practices, and research methodologies	PhD, JD, or master's in relevant field Management experience, particularly with managing staff and budgets Experience in non-profit and/or government sector Experience in academic institution recommended Content expertise with one or more data holdings 10+ years of relevant work experience
Data/Research Coordinator	Advanced project management skills Strong communication skills Knowledge of research methodologies Experience with data set/data holdings management, including managing codebook	Master's in relevant field Experience in non-profit, government sector, or academic institution recommended Content expertise with one or more data holdings 3+ years of relevant work experience
Research Analyst	Demonstrated ability to support database management, including data cleaning, data set management, coding Demonstrated ability to support analysis and reporting Nuanced understanding of statistics	Master's in relevant field Content expertise with one or more data holdings

Role	Skills/Competencies	Training/Experience
Data Scientist	Demonstrated experience with algorithm development and data integration Strong technical documentation skills Demonstrated understanding of best practices for data security, data transfer, data cleaning, and data import Strong communication skills; ability to translate technical concepts into accessible language for a variety of stakeholders Basic understanding of social science methodologies	PhD or master's in relevant field 5+ years of relevant work experience
Database Administrator	Experience with design and development of flexible database solutions Strong technical documentation skills Demonstrated understanding of a variety of operating systems, memory and storage requirements, firewall restrictions and ports, and data security best practices	*Could be staffed part-time or on a consultant basis Master's in relevant field 3+ years of relevant work experience
Data Security Officer	Nuanced understanding of data security best practices across data transfer, storage, and export	*Could be staffed part-time or on a consultant basis
Legal Counsel	Nuanced understanding of legal issues pertaining to data sharing	*Could be staffed part-time or on a consultant basis