INTRODUCTION TO

Data Sharing & Integration

VERSION 2.0

AISP

ACTIONABLE INTELLIGENCE FOR SOCIAL POLICY OCTOBER 2025

Contents

Suggested Citation: Hawn Nelson, A., Algrant, I., Jenkins, D., Rios Benitez, J., Kemp, D., Burnett, T.C., Zanti, S., Culhane, D. (2025). *Introduction to Data Sharing and Integration*. Actionable Intelligence for Social Policy. University of Pennsylvania.

Introduction

At Actionable Intelligence for Social Policy (AISP), we believe that the development of public data infrastructure will be the defining public works project of the 21st Century. It is critical that data infrastructure is constructed collaboratively and thoughtfully. Administrative data—the data collected and maintained by the government and other organizations during the routine process of administering programs—when repurposed, can provide valuable insights for improving programs, services, and ultimately the lives of residents. Despite the promise of cross sector data integration and some notable successes, strong coordination and data sharing across government agencies in the U.S. is still the exception rather than the norm.

With trust in government at record lows[1] and the current administration's attempt to rapidly consolidate data without strong safeguards in place,[2] this updated Introduction to Data Sharing and Integration is designed to support the intentional construction of infrastructure to share, integrate, and use administrative data to promote the public good.

This resource is intended to help partnerships, collaboratives, agencies, and community initiatives build a foundation for sustained administrative data sharing geared toward improving individual and community outcomes. We generally refer to these efforts as integrated data systems (IDS), but they have other names, including data hubs, longitudinal data systems, data collaboratives, and data intermediaries.

Whatever you choose to call your effort, development requires defining and articulating your purpose and identifying your partners and roadblocks. Once you have laid the groundwork, we recommend developing IDS quality across five components: governance, legal, technology, capacity, and impact. The work of designing an IDS is developmental and will change shape over time. As you move through this document, use what is helpful, leave what is not, and return to concepts you are not yet ready to address. Ideally, this will help you find a place to start that is aligned to your unique context.

[1] Pew Research Center. (2024). Public Trust in Government 1958-2024; Partnership for Public Service. (2025). The State of Public Trust in Government.

[2] American Civil Liberties Union. (2025). Data Silos, Dossiers, and Surveillance: The Unintended Risks of Federal Data Consolidation.

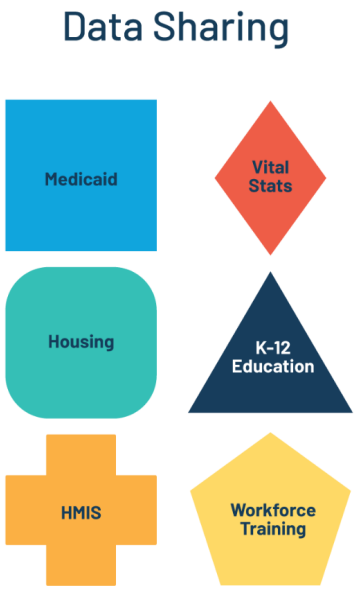
Data Sharing vs. Data Integration: What's the Difference?

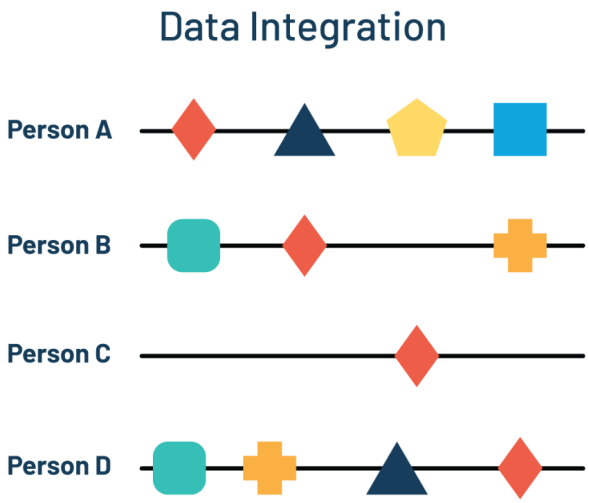
Data Sharing is the practice of providing partners with access to information (in this case, administrative data) that they cannot access in their own data systems. Data sharing allows interest holders to learn from each other and collaborate on shared priorities.

- **Example:** A group of early childhood providers agree to share information about their enrollment and subsidy waiting lists with a local advocacy group in order to better understand the community's early childhood capacity and unmet needs.

Data Integration is a more complex type of data sharing that involves record linkage, which refers to the joining of data based on common data fields. These data fields often include personal identifiers, such as name, birth date, social security number, or a common encrypted "unique ID" that is used to link or join records at the individual level.

- **Example:** Federally funded early childhood providers share identified information about the children they serve, including name, date of birth, and address, with their state early childhood agency. The state agency links these data with birth records and records from early learning programs funded by state and local sources to estimate how many children are attending publicly funded early learning programs, how many have had no formal early learning experience prior to kindergarten, and how child outcomes vary across these groups.





Find Your Purpose

Examples of Integrated Data Use

Data integration efforts can often seem lofty, technical, or intangible. As you work to define and articulate your purpose, take time to ground yourself and your potential partners in the actuality of what IDS are and what they can do. Below, we describe six real-world examples of how integrated data efforts were successfully developed and used across the AISP network to improve outcomes for those represented in the data.

EARLY CHILDHOOD IOWA

Early Childhood Iowa's integrated data system, I2D2, integrates public health, education, and human services data to better understand childhood service utilization and the early childhood workforce. Their initial effort focused on determining an unduplicated count of children enrolled in state programs from birth to age 5 across the state, and found that 73% of children had at least one center-based experience during the year before kindergarten entry. Importantly, the analysis also revealed significant gaps for underserved children, particularly those in rural counties. Results also showed shortages in both the quantity and quality of the early childhood workforce, with particularly acute staffing challenges in rural areas, which comprise 89% of Iowa counties. This project was an important precursor to I2D2 receiving a Preschool Development Birth-5 Grant in 2019. Since then, I2D2 has gone on to create Child Care Connects (C3), a tool which includes near-real-time updates directly from child care providers through various systems, integrates them, and then powers online dashboards and child care search applications that benefit families, child care providers, and state and local leaders across Iowa.

NEW YORK CITY'S MYCITY BENEFITS ENGINE

The New York City Mayor's Office for Economic Opportunity worked in conjunction with the Office of Technology and Innovation and the Office of Data Analytics to launch MyCity in March of 2023. Initially, this platform created a user-friendly experience for New Yorkers to apply for or recertify their childcare benefits and track application status. In November 2024, MyCity expanded to allow parents and caregivers to opt in to find out if they are eligible for other services such as utilities assistance or SNAP without completing a separate form. Within 48 hours, applicants received a one-time email containing a personalized list of benefit recommendations. Participation is completely optional, and the language used in the app was tested with New Yorkers to ensure readability and accessibility.

PHILADELPHIA IDEA

The City of Philadelphia, in partnership with the Penn Early Childhood and Family Research Center at the University of Pennsylvania, used its Integrated Data for Evidence and Action (IDEA) system to inform how Philadelphia could best leverage revenue from a newly established soda tax to expand universal pre-k offerings. The City of Philadelphia aggregated individual social services and neighborhood data to prioritize which communities would get access to new universal pre-school spots based on their exposure to risk factors for early childhood and their current access to high-quality pre-k programs. The city was able to build on this initial success and execute a separate legal agreement allowing trusted practitioners access to information about families who might benefit from the new pre-K slots, so that those most in need could be connected directly to services. Read the case study here.

NEW JERSEY CHCS AND THE CAMDEN COALITION

The Center for Health Care Strategies (CHCS) and Camden Coalition are collaborating on an effort to build new statewide infrastructure in New Jersey for collecting data on adverse childhood experiences (ACEs) and positive childhood experiences (PCEs). So far, CHCS and Camden Coalition have utilized the 2021 Youth Risk Behavioral Survey (YRBS), which includes national, state, and local data on high school-aged youth, as well as administrative data including birth and death certificates, child maltreatment records, clinical health data, and incarceration records, to study ACEs and PCEs for New Jersey children. Researchers also explored geographic variation, digging deeper into local conditions. Moving forward, CHCS and Camden Coalition will work to expand partner capacity to respond to findings and promote child and community well-being.

INDIANA MANAGEMENT AND PERFORMANCE HUB

The Indiana Management and Performance Hub (MPH) played an essential role in early COVID-19 prevalence and spread research, as part of a project led by the Indiana Department of Health and researchers at the Fairbanks School of Public Health at Indiana University–Purdue University Indianapolis. When the COVID-19 pandemic first reached the U.S., the understanding of infection rates was limited to positive test rates when what was needed was a full picture of the proportion of the population these positive tests accounted for. MPH created a population-level dataset that allowed researchers to understand the prevalence of COVID outside of symptomatic cases and hospitalizations. Using tax records, MPH and Fairbanks were able to randomly sample the population and build what became the "denominator" for COVID-19 analyses across the state and nationally. Indiana went on to test over 15,000 residents in April 2020, and MPH continued to serve as an integral partner, supporting vaccine distribution and accelerating public health response to mitigate disparate impacts of the pandemic. Read more in the AISP case study here.

ALLEGHENY COUNTY DATA WAREHOUSE

The Allegheny County Data Warehouse is hosted by the Department of Human Services, Offices of Analytics, Technology, and Planning. The Warehouse supports many use cases, both research and operational. For example, it is used to integrate human service and incarceration records to ensure continuity of mental health and substance use support during incarceration and in the transition afterwards. Upon release, formerly incarcerated individuals automatically retain the services they had previously without having to re-complete the intake process. This example illustrates how person-specific data allow for continuity of care and smoother reentry into the community.

Why do YOU want to share or integrate data?

Now that you have seen some potential use cases, it is important to identify your purpose. At AISP, we have seen some of the most successful efforts grow out of an attempt to answer one question, such as, "which service recipients are multi-agency involved?" Even if your goal evolves to answer other questions as well, building an IDS with a clear purpose will drive the design, and help engage and align partners. The following table shows some potential perspectives and entry points:

PERSPECTIVES AND ENTRY POINTS	
Perspectives	**Integrated data needed to**
Internal agency analyst	Better understand the overlap in the populations served by multiple programs and assess how best to maximize resources and impact
Community collaborative or organizer	Have access to information about residents' lived experience across a range of indicators, with timely updates
Project evaluator at a research firm	Evaluate a program's impact on individuals across multiple areas of interest (e.g., housing, health, workforce participation)
Director of a large regional nonprofit	Gain a better understanding of the backgrounds and experiences of the people served by the nonprofit, as well as how they fare after leaving the programs
University-based researcher	Better understand how outcomes vary for different populations based on characteristics or experiences
Deputy director of a state agency	Look across more information about the people they serve in order to coordinate referrals or offer targeted supports
Staffer at the governor's policy office	Draft a strategy to target new resources to support vulnerable populations, for example, new investments in early learning for disinvested neighborhoods

Questions for establishing your purpose:

- **Is there a question you need answered or a problem you need solved?**
- **How would having the capacity to share and link data drive action?**
- **What are the short- and long- term benefits?**

For more support identifying and articulating your purpose see Your Journey to Centering Racial Equity workbook activities I.iii & I.iv.

How will your purpose drive design?

We distill the purpose of data sharing and integration into three categories: Indicators and Reporting; Analytics, Research, and Evaluation; and Operations and Service Delivery.

	Indicators & Reporting	Analytics, Research, & Evaluation	Operations & Service Delivery
Purpose	Fulfills cross-agency planning and reporting requirements, and makes program outcomes more transparent	Enables detailed cross-agency analyses of long-term outcomes and impact of service utilization	Facilitates sharing of detailed cross agency data to improve service delivery and care coordination
Audience	Agencies, policymakers, the public	Researchers, evaluators, research and planning staff	Case workers, service providers
Example	A collective impact initiative wants to report on a set of common indicators through a publicly available dashboard	City agencies want to understand how outcomes vary for clients based on demographic characteristics and participation in multiple programs	County agencies want to link their data in near real-time to enable coordination for improved delivery and increased quality of care

The core purpose (or purposes) of your IDS will determine your approach to governance, legal, technology, capacity, and impact. The table below highlights considerations based on these purposes across our Quality Framework:

	Indicators & Reporting	Analytics, Research, & Evaluation	Operations & Service Delivery
Governance	Minimal	Shared processes and clear parameters around access and use are required	
Legal	Data may be publicly available already or access and use may require a simple agreement to receive data in de-identified or aggregate format	Data are protected and access generally requires multiple agreements to clearly outline permissible use	Data are protected and data sharing agreements must outline parameters for role-based, credentialed access. Access may also require client consent and non-disclosure agreements.
Technology	Data are linked, anonymized, and results are reported at the aggregate level	Data are linked, de-identified, and shared for a specific analytic purpose	Data are identifiable and may include case notes to support client-level services
Capacity	Costs and staffing are minimal	Costs and staffing are moderate	Costs and staffing are significant
Impact	Increased transparency and potential for data to drive public discourse, advocacy, and collective impact	Increased research capacity and potential for data to drive cross-agency coordination on policy and practice	Increased care coordination, streamlined referrals and potential for cross-program enrollment

The purpose for data sharing also impacts considerations related to privacy risks, the regularity of data sharing (data frequency), and data quality. The following table dives deeper into these distinct considerations:

	Indicators & Reporting	Analytics, Research, & Evaluation	Operations & Service Delivery
Privacy and Security	A lack of identifiers or small cell sizes means minimal risk of redisclosure, although data are potentially reidentifiable	Minimal access to identifiable data and a small group of approved users means that security requirements are essential but often procedural rather than technically advanced	Many users and identifiable data mean that staff training, complex permissions, and audit trails are necessary
Data Frequency	Data may be updated based on reporting cycles, often quarterly or annually	Data may be updated periodically depending on availability and analytic requirements for approved projects	Daily or real-time updates of entire client records may be required
Data Quality Standards and Documentation	Data are evaluated for correctness, missingness, accuracy, and stability over time	Data correctness, missingness, and accuracy are critical to support meaningful analysis	Clients should have the opportunity to correct errors

For more on moving from purpose to implementation in an SLDS-specific context, see an expanded discussion and version of this table in Defining Modern, User-Centered State Longitudinal Data System Design. For more information on legal frameworks, see Finding a Way Forward.

While it is possible to design infrastructure that combines several of the approaches described above, it is always important to differentiate them during design and planning. In most cases, we recommend starting with data sharing and integration work that can be reported in the aggregate and building on those successes as you approach the much more challenging work of coordinated operations and service delivery at the individual level.

Benefits, Risks, & Limitations

As you work to put your planning into action, it is important to consider the benefits, limitations, and risks of repurposing administrative data. These considerations will change depending on your context, but commonly include:

Benefits:

- **Whole-person, longitudinal view:** Using multiple sources allows a holistic view of individual experiences and outcomes across programs and, potentially, across time.
- **Whole-family view:** Administrative data linkages can enable us to better understand experiences and outcomes across a family or household unit.
- **Scale:** Administrative data can allow for a population view, rather than a sample, and is therefore less vulnerable to certain forms of bias, such as nonresponse.
- **Time & cost:** Data reuse is less time- and resource-intensive than collecting new data.

Limitations

- **Reliability:** Administrative data are collected to meet operational and reporting needs and may not accurately represent the concepts or outcomes an analytic plan aims to measure.
- **Quality:** Data quality issues are common, including missing data and insufficient data documentation, leading to issues of reliability and validity.
- **Depth:** These data are often one-dimensional and may need to be paired with qualitative and other forms of data in order to address deeper questions about causal relationships, client experiences, contextual factors, etc.
- **Access:** Gaining access is often difficult and time-consuming.

Risks

- **Unauthorized disclosure:** Any transfer of data beyond the scope of consent or legal authority that includes the risk of data being accessed improperly.
- **Misuse of data for research and evaluation:** Without sufficient data documentation, analysts may misuse or misinterpret data.
- **Replicating structural racism:** Administrative data include people who are disproportionately Black, Indigenous, and people of color, leading to oversurveillance. Seeing data as race-neutral is inaccurate, and such views could lead to system-level data usage that unintentionally replicates structural racism.
- **Harming individuals:** Certain uses of administrative data carry high risks of personal harm. This includes providing case workers, service providers, teachers, law enforcement, etc., with personal information that could lead to biased treatments, punitive action, or lengthened system involvement.
- **Harming communities:** Use of administrative data, especially when mapped or otherwise represented spatially, can create or deepen community stigma. When analysts fail to acknowledge the discriminatory practices and structural causes of disparate outcomes by race or geography, they risk using administrative data in ways that perpetuate harmful, deficit-based narratives.

The potential benefits, limitations, and risks of a data sharing effort should be considered carefully by a diverse group of partners, including those represented in the data. For guidance on weighing risks and benefits, see A Toolkit for Centering Racial Equity Throughout Data Integration and Your Journey to Centering Racial Equity activity II.iii.

Build your team

Data sharing and integration efforts are driven by people with distinct roles and motivations around data use. These individuals must come together to weigh the benefits, limitations, and risks of proposed data uses, and develop a shared agreement about if and how to move forward. It is helpful to start this process with an understanding of who is coming to the table and why, and then assess who might be missing from these initial conversations.

Establishing a clear purpose is critical before beginning any cross-sector data flow. **Data flow at the speed of trust**, and trust can be difficult to establish across agencies and organizations that lack a common language and are each focused on their own distinct duties and priorities. Therefore, it is important that agencies engaged in new data sharing efforts spend time up front discussing motivations, concerns, and expectations in order to build trust and document the rules of the sandbox and the needs of the community before beginning a data project.

There are three critical positions to take into account as you build your team: data owners,[3] data stewards, and data custodians:

The Role of Data Owners, Data Stewards, and Data Custodians

	Role in data sharing process	Role within agency
Data Owner	Accountable for the quality and security of the data and holds decision-making authority regarding access and use	Typically agency leadership that has signatory authority
Data Steward	Responsible for the governance of data, including metadata. Support established processes and policies for access and use	Typically the subject matter experts and data analysts that work with data
Data Custodian	Responsible for the technology used to store and transport data	Typically an IT person or team

[3] Note that we honor the principle of data sovereignty, and recognize that in the U.S. context, administrative data are legally owned by the agency that collects the data to administer programs and services. For more information about the concept of data sovereignty see Hummel, P., Braun, M. & Dabrock, P. (2021). Own Data? Ethical Reflections on Data Ownership. Philos. Technol. 34, 545–572; Haozous E.A., Lee, J., Soto, C. (2021). Urban American Indian and Alaska Native Data Sovereignty: Ethical Issues. Am Indian Alsk Native Ment Health Res. 28(2):77-97.; Tsosie, R. (2019). Tribal Data Governance and Informational Privacy: Constructing "Indigenous Data Sovereignty", 80 Mont. L. Rev. 229.

Incorporating external perspectives is also critical. Most importantly, this should include those who are represented in the data and those who have historically been excluded from these processes. As you gather partners, ask yourself: who's not at the table; what motivation is not being represented; and why are these voices not present? Below are some potential places to start:[4]

- **Core Partners:** Who is needed for the effort to achieve success?

  Examples:

  - **Data owners and stewards**
  - **Funding sources**
  - **Public agency leadership**
  - **Community members / individuals represented in the data**

- **Direct Partners:** Who else can help facilitate (or impede) IDS success?

  Examples:

  - **Data users (researchers, analysts)**
  - **Data custodians and technical experts**
  - **Privacy advocates**
  - **Community organizers**
  - **Funders**

- **Other Partners:** Who can broaden the interest of the integrated data and deepen its constituencies? What community groups are "in the data?" Who can this effort help? Who can it hurt?

  Examples:

  - **Business groups**
  - **Advocacy organizations**
  - **Public interest groups**
  - **Philanthropic organizations**

As you "build your table," take the time to ensure that power is acknowledged and decision-making is shared. For more specific guidance on convening partners, see Your Journey to Centering Equity activity II.ii.

[4] Adapted from IDS Governance: Setting Up for Ethical and Effective Use (2017).

What Data Will You Need?

(And What Do You Already Have?)

When considering your purpose and approach to data sharing, it is also important to begin with a realistic assessment of what data would be required to achieve your aims and how easy (or hard) access may be. It is helpful to first classify the accessibility of administrative data into one of three categories:

Open Data	Protected Data	Unavailable Data
Data that can be shared openly, either at the aggregate or individual level, based on state and federal law. These data often exist in open data portals	Data that can be shared, but only under specific circumstances with appropriate safeguards in place	Data that cannot or should not be shared, either because of state or federal law, lack of digital format (paper copies only), data quality, or other concerns

*Open Data: Data that can be freely used, re-used, and redistributed by anyone. For more information see What is Open Data?

> Cataloging data assets is an essential first step to determine legal and ethical use. See DataLinkCT's data dictionary as an example of publicly available metadata, including information on what data assets are not available through the IDS.

This is a list of protected data that AISP network sites are currently integrating:

- Vital records
- Health
- Child welfare & adult protection
- Early childhood
- Education
- Legal system & law enforcement
- Economic security
- Housing & homelessness
- Federal
- Nonprofit
- Private/corporate

Is an IDS Right for You?

You likely opened this document because you are interested in using data to inform decision-making and see an integrated data system as a solution, but that may not be the best way to achieve your goals. It may be helpful to first think through the following questions:

1. **Can these data be shared from one agency, de-identified?**
2. **Is there a collaborative / institution / agency / data intermediary that can provide access to these data?**
3. **Will these data need to be shared and integrated one time (often referred to as an ad hoc data request)? If so, ask the agency/ies about their process for data integration, access, and use.**

If you answered yes to the above questions, you likely do not need to embark on developing a new data sharing or integration effort. However, if your purpose requires data to be routinely shared across sectors and there is no other data system that has this capacity, then developing an IDS could advance your efforts.

Before beginning a data sharing and integration initiative, it is important to conduct a landscape scan of what efforts are in development and underway. Our experience is that every agency and site is sharing and integrating data in some form, typically through ad hoc projects, and conducting a scan can prevent duplicative efforts. Engaging with data owners, data stewards, and data custodians can be an effective way to better understand current data access and use practices, and will allow you to build upon pieces that are working.

Developing a Data Model

If you determine that an IDS would be beneficial in your context, you can begin to construct a data model. A data model is a framework used to organize data elements and standardize how they relate to one another. Data models describe the structure, integrity, and quality of the data stored in different data management tools and databases.

When you are selecting which data will be shared and integrated as part of your data model for a given analysis, we encourage you to consider the following broad principles:[5]

Organize around the life course: One of the advantages of administrative data is the ability to connect data on the same person longitudinally. Analysts can intentionally focus on key developmental periods and transitions which can be important for high-impact studies.

Include contextual factors: There are many frameworks that can define this phenomenon, but in general it is important to understand that individuals live within households and families and in neighborhoods; they attend schools in communities, cities, counties, states, and regions. Using individual-level records without broader context, like data on place, may limit insights and opportunities for action.

[5] Adapted from Establishing a Standard Data Model for Large-scale IDS Use (2017).

Ensure reliability and validity: Administrative data are collected during the routine administration of public service programs. When we reuse them in a way that was not originally intended (e.g., for analytics and insights), data quality can impact reliability and validity, and this is a primary consideration when developing a data model.

- Reliability refers to data that produce similar results under consistent circumstances (e.g., a record consistently links an event date to an event, such as starting a program).
- Validity refers to the extent to which conclusions drawn from analysis are accurate, e.g., an evaluation has a sufficient data model to measure outcomes related to a specific programmatic intervention, rather than other changes (such as improved funding or change in enrollment).

Example: If you are integrating data for an analysis that involves student outcomes in Pre K-12 education, then your data model should include information that allows the analysis to control for attendance. In this case, your data model would include absences, suspensions, and days in membership (number of days present during the attendance period). If possible, teacher attendance should also be included, so your data model also contains the teacher of record and how many days they were absent.

For further guidance on human service-oriented data models, see AISP's Establishing a Standard Data Model for Large-scale IDS Use (2017).

For an example of a conceptual data model, see Developmental Impacts of the COVID-19 Pandemic on Young Children: A Conceptual Model for Research with Integrated Administrative Data Systems (2020), pictured below.

How to Begin Data Sharing

The process of building sustainable, mission-driven data sharing and integration may at times be slow, but we firmly believe in moving slow to go fast. The more time you invest in planning, the better your resulting effort will be. This section will provide an overview of the five components of our IDS quality framework: governance, legal, technology, capacity, and impact. This overview provides resources for developing each of these components in your IDS.

Before we dive into the quality framework, we should start with four core questions to guide decision-making for data sharing and integration. You can read more about the four question framework in Four Questions to Guide Decision-Making for Data Sharing and Integration (2023):

1. Is it legal?	2. Is it ethical?	3. Is it a good idea?
What legal authority is in place to use these data? Are there federal or state statutes that prevent or constrain this data access or use? What are the specific state and federal law requirements enabling data sharing?	Do the benefits outweigh the risks, particularly for groups historically marginalized by discriminatory systems?	What action can be taken as a result of this data use? What can reasonably be changed or improved based upon this analysis? Is this a priority among marginalized populations and/or individuals included in the data system?
4. How do we know and who decides?		
This is typically determined by agency-involved legal counsel.	This is typically determined by a data governance group during the data request review process. The group should include a variety of partners, including those represented "in" the data and data users.	This is typically determined by a data governance group, including data stewards who have content expertise and data owners who will respond to insights that emerge from the analysis.

Governance

In order to decide if data sharing, integration, and use are legal, ethical, and a good idea, you need data governance. In other words, policies and procedures about how decisions are made and by whom should be clear, well-documented, and consistent. These policies along with established, routine meeting structures and well-documented proceedings, foster a culture of trust and transparency. As you work to establish sustainable governance, make sure to include diverse partners at your table.

> **Data Governance:** The people, policies, and procedures that determine how data are managed, used, and protected

Data governance for a cross-sector data sharing effort can draw upon existing data governance practices within one agency or involve a separate set of policies and procedures specific to the collaboration. Regardless of the approach, it is important that governance is explicit and collaboratively agreed upon, rather than implicit and driven by any one individual.

The particulars of the policies and procedures will vary widely based on the purpose for data sharing established by the data partners involved. For instance, a narrow goal of creating an academic research database to support indicators and reporting will suggest one governance approach, which will differ significantly from the approach needed to support an ambitious agenda to create access to real-time integrated data for credentialed users to support operations and service delivery within a health and human service agency. We recommend that a site devote time up front—both internally and with partner organizations—to build consensus around what data sharing and integration is intended to achieve. Taking the time to do this at the outset allows each site to establish tailored rules of engagement that best meet its needs and goals.

To learn more about key considerations for strong and inclusive data governance in cross-sector data sharing and integration, see our Quality Framework for Integrated Data Systems.

As you work on developing a strong governance framework, here are some additional resources with examples and activities:

- Four Questions to Guide Decision-Making for Data Sharing and Integration
- Survey Brief: Governance
- A Toolkit for Centering Racial Equity Throughout Data Integration
- Participatory Governance: Longform Work in Action
- IDS Governance: Setting Up for Ethical and Effective Use

Remember: data flow at the speed of trust. Cross-agency data governance is a way of institutionalizing trust to ensure that data use is legal, ethical, and a good idea.

For more resources on building strong and sustainable data governance, see https://aisp.upenn.edu/governance/.

Legal

Whether data can be shared legally depends on why you want to share, what type of information will be shared, who you want to share with, and how you will share the data. Legal agreements should reflect the purpose for sharing, document the legal authority of the host organization to serve that purpose, and ensure that data sharing complies with all federal and state statutes.

> **Legal:** the rules, regulations, and authority that determine why, when, what, how, and with whom data can be shared

Legal counsel, privacy security officers (PSO), and compliance officers are conditioned to hold data close to avoid risk, so negotiating data sharing can be time-consuming. These challenges are normal and surmountable and legal professionals play a critical role in ensuring that data sharing and integration is legally sound. Starting from a clear understanding about your data sharing purpose and being honest about the associated benefits and risks are critical to gaining the support of these partners.

When agencies and their partners bring data together safely and responsibly, policymakers and practitioners are better equipped to understand the complex needs of individuals and families. At the same time, data sharing and integration is not without risk, and clear legal frameworks are essential to mitigate those risks, protect privacy, and guide responsible data use. Designing the appropriate legal framework for the context can be a complex task. It is important to engage legal counsel early and often to ensure your data sharing effort has clear legal authority, and to design a framework to facilitate data sharing, spelling out the why, what, who, and how. We recommend a tiered framework for these agreements and a templated approach, whenever possible.

As you work to develop strong legal frameworks, here are some resources:

- Finding a Way Forward: How to Create a Strong Legal Framework for Data Integration
- Yes, No, Maybe?: Legal & Ethical Considerations for Informed Consent in Data Sharing and Integration
- Strong Legal Frameworks for Data Integration: Four Questions for Moving Forward
- A governance and legal framework for getting to "yes" with enterprise-level data integration

For more resources on designing legal frameworks, see https://aisp.upenn.edu/legal/.

Technology

Data sharing and integration is commonly approached as a technical project; however, we encourage sites to view technical components as tools to support analytic insights, allowing partners to ultimately improve policies, practices, and outcomes for the people they serve. A technical approach should not be the end goal, but a means to get there. Your technical approach can and should change as information needs develop and technological advances shift best practices. For this reason, we encourage starting small, with initial investments focused more on relationships and human capacity rather than large IT infrastructure. Technology should fit the purpose(s) of sharing and integration, so the tools and methods used may vary depending on your purpose.

> **Technology:** the infrastructure and tools that ensure secure data storage, linkage, and access while supporting analytic insights

For more resources on technology, see https://aisp.upenn.edu/technology/.

Capacity

Capacity is the human, relational, and material resources that enable a data sharing effort to implement governance, execute legal agreements, build technical infrastructure, and, above all else, demonstrate impact. Building shared capacity requires effective leadership and planning, as well as sustained commitment from data partners. It is an evolving and iterative process and often feels like a delicate dance. The goal is to keep growing the ability to meet partner goals without overpromising on what can be achieved with current data sharing capacity. Each successful proof of concept can help you advocate for more resources.

> **Capacity:** the staff, relationships, and resources that enable an effort to operate

For more resources on developing capacity, see https://aisp.upenn.edu/capacity/.

Impact

Impact is how we measure our progress toward our purpose and goals. All other components of the IDS quality framework exist to drive impact. We include impact in our quality framework because administrative data is a public good, and should be used to benefit the public. Data that are collected, safeguarded, and managed should be used to inform strategic insights that drive improvement. As you use integrated data to improve outcomes for the individuals you serve, it is important to document your impact and celebrate your wins. Ask yourself: what can we accomplish when we bring data together, and what can we do to more fully realize our intended impact?

> **Impact:** the outcomes realized when IDS create insights, drive action, and improve lives

For more resources on demonstrating impact, see https://aisp.upenn.edu/impact.

Call to Action

In our work, we frequently hear two misconceptions about government data: that the government cooperates, shares, and integrates data already; and that government data collection, sharing, and integration is, on the whole, problematic. This Introduction seeks to dispel both of these ideas by illustrating that administrative data are a public good that can be used to better understand and improve public services, but that these data are often stuck in siloes with minimal cross-sector collaboration. When these data are shared and integrated responsibly, they can have positive impacts on those we hope to serve. As you embark on your own data sharing and integration projects, remember the work is slow moving and, in many cases, completely new to your potential partners. We encourage you to work intentionally to incrementally build trust, and work together to integrate the data and the people needed to use these insights to improve lives.

Remember, you are not alone in this process. AISP works to support the field, and encourages developing sites to use our network as a guide. We look forward to learning more about what you create—now get going!

Actionable Intelligence for Social Policy
University of Pennsylvania
3701 Locust Walk, Philadelphia, PA 19104
www.aisp.upenn.edu