

Four Questions to Guide Decision-Making for Data Sharing and Integration

Amy Hawn Nelson^{1*} and Sharon Zanti¹

Submission History

Submitted:	26/04/2023
Accepted:	04/08/2023
Published:	04/10/2023

¹University of Pennsylvania,
Actionable Intelligence for Social
Policy, Philadelphia, PA 19104

Abstract

Introduction

This paper presents a Four Question Framework to guide data integration partners in building a strong governance and legal foundation to support ethical data use.

Objectives

While this framework was developed based on work in the United States that routinely integrates public data, it is meant to be a simple, digestible tool that can be adapted to any context.

Methods

The framework was developed through a series of public deliberation workgroups and 15 years of field experience working with a diversity of data integration efforts across the United States.

Results

The Four Questions—*Is this legal? Is this ethical? Is this a good idea? How do we know (and who decides)?*—should be considered within an established data governance framework and alongside core partners to determine whether and how to move forward when building an Integrated Data System (IDS) and also at each stage of a specific data project. We discuss these questions in depth, with a particular focus on the role of governance in establishing legal and ethical data use. In addition, we provide example data governance structures from two IDS sites and hypothetical scenarios that illustrate key considerations for the Four Question Framework.

Conclusions

A robust governance process is essential for determining whether data sharing and integration is legal, ethical, and a good idea within the local context. This process is iterative and as relational as it is technical, which means authentic collaboration across partners should be prioritized at each stage of a data use project. The Four Questions serve as a guide for determining whether to undertake data sharing and integration and should be regularly revisited throughout the life of a project.

Highlights

- Strong data governance has five qualities: it is purpose-, value-, and principle-driven; strategically located; collaborative; iterative; and transparent.
- Through a series of public deliberation workgroups and 15 years of field experience, we developed a Four Question Framework to determine whether and how to move forward with building an IDS and at each stage of a data sharing and integration project.
- The Four Questions—*Is this legal? Is this ethical? Is this a good idea? How do we know (and who decides)?*—should be carefully considered within established data governance processes and among core partners.

Keywords

data governance; ethical data sharing; legal frameworks for data integration

*Corresponding Author:

Email Address: ahnelson@upenn.edu (Amy Hawn Nelson)

Introduction

As sharing and integrating administrative records becomes increasingly common across health, education, and human service agencies, it is imperative to build high-quality systems that safeguard individual-level data and ensure that these data are used for the public good and without doing harm. At Actionable Intelligence for Social Policy (AISP) we support state and local governments in the United States (U.S.) on their efforts to share data collaboratively and responsibly [1]. While data integration efforts vary widely and are driven by local context, we have identified five key components of quality for data integration efforts—Governance, Legal, Technology, Capacity, and Impact (see Table 1 for definitions of each component) [2]. These five components are interrelated and are all essential for building and maintaining high-quality Integrated Data Systems (IDS)ⁱ and in evaluating individual project requests using IDS data. However, the first two components—Governance and Legal—set the foundation and should be considered throughout decision making for any data integration effort. Although creating data governance and legal structures are distinct streams of work in IDS decision-making, and each requires different expertise at the decision-making table, they are inextricably linked and are both indispensable for high-quality and ethical data integration. For example, drawing on the lived experience of community members whose data are represented in an IDS supports more equitable governance processes and ethical data use, while concurrently engaging legal teams that can reconcile the respective needs and legal limitations of each data partner results in structures that provide the broadest potential for data sharing that is legal, ethical, and a good idea.

This article introduces a Four Question Framework to bridge key concepts in data governance and legal structures—*Is this legal? Is this ethical? Is this a good idea? How do we know (and who decides)?* (see Figure 1). These questions should be carefully considered within established governance processes and alongside core partners—specifically data owners, data stewards, and those represented within the data—to determine whether and how to move forward at each stage of data integration. The Four Questions provide an overarching framework to guide decisions around data sharing and integration; however, these questions can also be applied within existing decision-making frameworks, such as the Five Safes [3].

Below, we discuss the evolution of this framework over 15 years of working in the field. We then discuss the four questions in depth, with guidance for addressing each. In addition, we provide examples of how these questions can support decision-making around data use, with a focus on the role of data governance. While we developed the four questions within the U.S. context, and with a focus on public data, this framework is meant to be broadly applicable. Moreover, it offers a simple, digestible way to bring together partners with a diversity of skills and experience in service of building a strong governance and legal foundation to support ethical data use.

ⁱThroughout this paper we invoke multiple terms when referring to IDS, such as data sharing and integration efforts, cross-sector data integration, shared data infrastructure, and data collaborations.

Methods

The Four Question Framework has been workshopped over a period of 15 years, since AISP's inception in 2008. It is informed by hundreds of discussions and presentations with government agencies, university research partnerships, and policymakers; providing training and technical assistance to 55+ data integration efforts; and convening expert workgroups on data governance and legal issues in data sharing and integration. The workgroups employed public deliberation methods to assemble a diverse group of data integration practitioners who collectively have decades of experience developing strong data governance and legal frameworks to support cross-sector data integration.

Public deliberation

As a method, public deliberation requires assembling a diverse group of participants to consider wide-ranging viewpoints through a series of discussions, with the objective to come to an agreement on collective statements and recommendations that incorporate numerous perspectives [4, 5]. In 2016–2017 AISP convened four groups of experts to generate best practices in data sharing and integration across key topics: legal issues, governance, data standards, and technical considerations. All workgroup members were part of established IDS across the U.S. and had experience developing and maintaining cross-sector shared data infrastructure. The governance and legal workgroups supported the development of this Four Question Framework.

The expert workgroup tasked with creating best practices for IDS governance included six experts representing a range of professional roles, geographic locations, and data governance structures. Concurrently, AISP convened a similarly diverse group of eight experts in legal issues for data sharing. Both workgroups involved on-going deliberation events for seven months, including a two-day in person meeting, and bi-weekly virtual meetings.

For both workgroups, the in-person public deliberation session included an assigned facilitator and scribe. The lead facilitator created a discussion agenda in partnership with workgroup members. The agenda focused on enduring challenges to data sharing and integration posed by workgroup members. These members collaboratively distilled their combined experiences into practical problem-solving approaches for staff committed to strong governance and legal frameworks to support legal and ethical use of data. Dialogue was facilitated and ideas were captured through meeting notes and then synthesized and agreed upon by workgroup members. These public deliberations resulted in published reports with guidance for navigating the legal, political, and relational challenges of building an IDS, including guidance for effective governance approaches, and an overview of laws pertaining to administrative data reuse [6, 7].

Given substantive shifts in the IDS field since AISP first convened the 2016–2017 expert workgroups, in 2021 we sought to update our guidance and convened a second legal advisory workgroup. We invited all original workgroup members to participate as well as new members who could bring different perspectives and ended up with eight members. Over three public deliberation events, as well as extensive independent

Table 1: AISP’s quality framework for integrated data systems

Governance	Data governance is the people, policies, and procedures that support how data are used and protected.
Legal	A legal framework articulates how legal authority for data access and use is operationalized. 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 to serve that purpose, and ensure that data sharing complies with all applicable laws.
Technology	Technical components (e.g., technology to securely transfer/federate, link, validate, and analyze data; data standards and data management policies, etc.) are created to support analytics and insights that can help further improvements in policies, practice, and outcomes.
Capacity	Data sharing capacity refers to the staff, relationships, and resources that enable an effort to implement governance, establish legal authority, build technical infrastructure, ensure sustainability, and above all else, demonstrate impact.
Impact	All components of quality—governance, legal agreements, technical tools, and staff capacity—exist to drive impact. The extent to which an effort achieves its desired impacts depends on both how actionable the initial research questions are and how well they communicate findings to those who can take action.

Figure 1: Four questions for ensuring legal and ethical data use

1. IS IT LEGAL?	2. IS IT ETHICAL?	3. IS IT A GOOD IDEA?
<p>What legal authority permits the access and use of these data?</p> <p>What legal authority prevents or constrains the access and use of these data?</p> <p>What are the specific legal requirements that must be met to lawfully access and use these data?</p>	<p>Do the benefits outweigh the risks, particularly for groups historically marginalized by discriminatory systems?</p>	<p>What action(s) could be taken by creating integrated data capacity?</p> <p>What can reasonably be changed or improved based upon proposed analysis?</p> <p>Is this a priority among marginalized populations and/or individuals included in the data system?</p>
4. HOW DO WE KNOW? (AND WHO DECIDES)?		
<p>This is typically determined by agency-involved legal counsel.</p>	<p>This is typically determined by a data governance group, during the review process for data requests, that should include a variety of stakeholders, those “in” the data and users of the data.</p>	<p>This is typically determined by a data governance group, including data stewards who have deep expertise of the data, and data owners who will respond to insights that emerge from the analysis.</p>

review by workgroup members, the group considered the 2017 legal report and suggested updates to the guidance offered. The group was particularly interested in decision-making frameworks. A core focus was distilling their guidance for legal and ethical data sharing into a simple, accessible, and effective framework that could be used by new and seasoned legal counsel, policymakers, practitioners, and community members when building an IDS.

The Four Question Framework—which synthesizes the deliberations from three of AISP’s expert workgroups and 15 years of work in the field—was embedded into an updated report on legal frameworks for cross-sector data integration and released in 2022 [8].

Results

The guidance informed by the public deliberation process outlines key considerations and effective practices for finding a way forward amidst the challenges of building and sustaining an IDS. There are sections of the guidance that detail relevant

U.S. privacy laws and provide templates for legal agreements in that context, but the core of the document is the decision framework that workgroup members all agreed was universal. The group asserted that developing shared data infrastructure largely hinges upon asking the right questions, which we distilled down to a Four Question Framework to guide data sharing efforts in any context. These questions should be asked when developing an IDS and when considering data requests for specific projects.

The four questions

When establishing data flow in public sector agencies, the initial question partners often ask is, “Is this legal?” While legality is the first question, it is also the lowest bar for determining whether to move forward with a particular project. We strongly encourage agencies sharing data, along with their partners, to grapple with key questions about the ethical implications of data use. Figure 1 presents our Four Question Framework for establishing legal and ethical data use at all

stages of a project. The Appendix includes guiding questions to put the framework into practice. Below we explore each question in depth.

Is this legal?

While the legality of data sharing and integration is complex and specific to local context, it largely comes down to gaining clarity on two concepts: 1) legal authority, and 2) permissible data access parameters. Thinking through legal authority and data access while developing a proposal for IDS development or a specific use of data can help you to better understand the relevant legal parameters and craft a proposal that fulfills both the need to use data to inform policymaking and the need to adhere to privacy and confidentiality laws.

Legal authority

Though contracts are the most common legal mechanisms authorizing and facilitating data sharing, cross-sector data integration often relies on a combination of legal authorities and mechanisms. These may include: enabling legislation that grants authority to an agency or office to lead cross-agency data sharing [9]; statutes and regulations that specify how data can or will be used; program rules or policies; executive orders mandating data sharing in service of a specific policy priority or population; and/or contracts and other agreements, such as a Memoranda of Understanding, Data Sharing Agreement, Data Use License or Agreement, and Informed Consent [10]. Additionally, judicial interpretation reflected in case law, court orders, consent decrees, and administrative decisions can clarify legal basis for data sharing and integration.

Data access

Legality also depends on why, what, and how data will be accessed and by whom. We recommend classifying the data in question as either open, restricted, or unavailable, as defined in Table 2. Most often, data owners (and their legal counsel) determine data access parameters. For example, data may be classified as unavailable for sharing and integration if there are significant data quality concerns. Most data that include identifiers and are shared at the individual level are restricted, while almost any data source can be categorized as open data if aggregated at a large geography.

Other considerations for lawful data access are purpose of use, to whom, and how the data will be released. Releasing integrated then de-identified row-level data to a researcher for analysis and publication of aggregated results could be permissible use, as could releasing identifiable row-level data to a case worker that may only be disclosed for case management

purposes. While both instances require access to restricted data, the purpose changes data access determination, with research typically requiring de-identification and case management necessitating access to identifiable data.

Is this ethical?

Ethics considers what is good for individuals, communities, and society at large. Ethical data use must ensure that individual-level data are protected and not used for harm. At the same time, it is also ethical to make data available when it can provide actionable intelligence to benefit society. Data integration of public data requires consideration of the sometimes parallel and opposing principles that individuals have a right to privacy and that data are a public good.

It is also imperative to acknowledge that vulnerable and disenfranchised populations have been harmed by research and data use. Many ethical concerns around administrative data reuse and integration stem from this fraught history as well as current surveillance practices [11]. The Belmont Report is the conceptual underpinning for human subjects review of research in the U.S., as operationalized by Institutional Review Boards (IRB), and emphasizes respect for persons (privacy must be protected), justice (risks and benefits must be fairly distributed), and beneficence (benefits must outweigh risks) [12]. Although many uses of IDS do not require IRB approval, these principles offer a strong foundation for thinking through the ethics of a proposed IDS effort or specific data use project. Importantly, these principles are not hierarchical and must be weighed equally even when in conflict. For instance, consent is not typically required when collecting administrative data for routine, operational purposes in U.S. government agencies [10]. Obtaining individual consent may also be unnecessary when such data are de-identified for reuse in research. When consent is not required, its absence should still be considered when examining the ethics of data use, as it falls under “respect for persons.” Respecting a person can mean giving them choice in how their data are used. Yet *not* acting upon these data may go against the principle of beneficence, as there could be substantial benefits and limited privacy risks (assuming appropriate data security is in place) in leveraging data to inform policy.

Moving beyond legality and considering the ethical implications of data use makes the question of whether or not to share and integrate data less straightforward. Balancing oppositional values requires discernment across all relevant parties to consider potential benefits and risks, which is why data governance is central to this work. If executed well, a strong data governance process will rigorously address the ethical concerns of all involved parties and create social license for data use.

Table 2: Data classifications by access level

Open data	Restricted data	Unavailable data
Data that can be shared openly, either at the aggregate or individual level, based on state and federal law.	Data that can be shared, but only under specific circumstances with appropriate safeguards in place.	Data that cannot or should not be shared, because of legal restriction or another reason (e.g., data quality concerns).

Social license

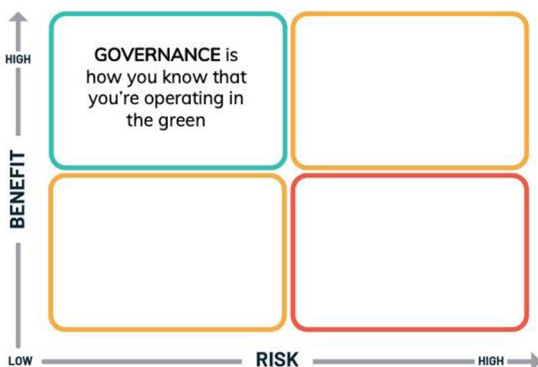
Generating public approval, or “social license,” to share and integrate data, is an important consideration for ethical data use. Social license is derived from perceived credibility, legitimacy, compliance with legal and privacy rules, and overall public trust in how data are accessed and used. It is earned by dedicating time and resources to building relationships, seeking out and incorporating feedback, and regularly engaging with diverse and representative partners and perspectives. It is especially important to develop social license with Black, Indigenous, people of color, the economically disenfranchised, and other groups that have been disproportionately harmed by institutions. Additionally, to bring about social license, people represented “in” the administrative data and frontline program staff should be part of data governance processes and provided opportunities for authentic participation and decision-making [13].

A key component of developing social license is instituting a clear and thorough process to discern potential risks and benefits of data use with all relevant parties, especially data owners, data users, and those represented “in” the data. It’s important to draw upon a wide range of perspectives, as perceived risks and benefits will vary. Identity dimensions (e.g., race, ethnicity, sexual orientation, gender, age, citizenship status, etc.) often influence perspectives on data use, as does one’s role in an organization. For example, an executive leader, data analyst, and case worker within the same agency will likely hold different views on data access and use. It is important to consider perspectives of risk vs benefit across dimensions of identity, lived experience, role, and power or ability to influence decision-making. As shown in Figure 2, one way to do this is to have all partners carefully consider and categorize proposed data uses based on their perceived level of risk and benefit. This process often involves discernment and revision—which result from governance activities—to come to consensus as to which proposed uses are “red”, “yellow”, or “green.” Before moving forward with a project, partners should agree that the proposed data use is “in the green” by carrying relatively low risk and high benefit.

Is this a good idea?

In some instances, reusing administrative data may be both legal and ethical but still not feasible or a good idea in

Figure 2: Risk vs Benefit matrix for categorizing proposed data uses



the current moment. Data availability, resources, and action should also be carefully considered in the data governance process to ensure data sharing is a practical and worthwhile in a specific context.

Data availability

Given that administrative data are collected for operational rather than analytic purposes, the actual data and data quality may be inadequate for data sharing, integration, and/or analyses. For example, when race, ethnicity, and other demographic data have high levels of missingness, the data may not be of sufficient quality to examine disparities in service use by these characteristics. Similarly, partners may agree on the benefits of measuring household outcomes, but if the data source does not collect or link information on household members then it is not feasible to conduct this analysis.

Resources

Strategic data use requires ample resources, particularly to hire, train, and retain skilled staff as well as to procure technology. Though leveraging data to inform decision-making can yield cost-savings in the long-term as policies and programs are improved, in the beginning stages of data infrastructure development, this investment can reduce funds available for programmatic efforts. Even in a fully functioning IDS, undertaking specific data requests can limit available funds for other priorities. Tradeoffs in resource allotment can be a significant source of tension and can require careful discernment in the data governance process.

Action

Taking meaningful action based on findings is challenging work, particularly for agencies operating in complex political environments. Many analytic projects merely produce descriptions of already known problems—rather than insights that can lead to action that benefits the public good. For data use to clear the “good idea” bar, there must be intent, social license, resources, and a realistic plan to use findings to drive action that can improve the lives of those impacted by policies, programs, and services.

How do we know (and who decides)?

The three previous questions—Is this legal? Is this ethical? Is this a good idea?—are all answered through data governance. Strong and inclusive data governance practices are how we know if data sharing and integration is legal, ethical, and a good idea. Data governance involves the people, policies, and procedures that support how data are used and protected; and it guides decision-making to ensure that partners have carefully considered the risks and benefits. Cross-sector data sharing efforts may use a distinct governance process, rely on an agency’s existing policies and procedures, or involve a hybrid of the two.

Specific governance practices will vary widely based on the purpose, values, and guiding principles for data use

established by the data partners. For example, creating routine access to real-time integrated data for credentialed users to support case management will necessitate a different governance approach than an ad hoc data integration project to generate indicators and aggregated reporting metrics. We recommend that partners spend ample time up front—both internally and externally with partner organizations and community members—building social license for data sharing and integration, identifying shared goals, and establishing clear rules of engagement that best meet the needs of all partners. We define good governance practices as having these five qualities [2]:

- Purpose-, value-, and principle-driven
- Strategically located
- Collaborative
- Iterative
- Transparent

Purpose-, value-, and principle-driven

We encourage data integration efforts to start by identifying the purpose for data sharing, and then craft a vision, mission, and guiding principles. The mutual benefit for data partners and the broader community should be described with clear value statements during this process. Table 3 outlines three common purposes for sharing data and demonstrates how purpose informs the overall approach, governance requirements, and the most appropriate legal framework for integration.

Strategically located

After defining the purpose, values, and guiding principles, it is helpful to consider which partners will manage the core activities of data integration (e.g., hosting governance, managing technology, conducting analyses). In the U.S., data integration efforts are generally located within federal, state, or local contexts (e.g., city, county, region), with

Table 3: Core purposes and approaches for data sharing and integrationⁱⁱ

Purpose	Indicators and reporting	Analytics, Research, and Evaluation	Operations and service delivery
Approach	Data can be summarized and reported at the aggregate	Data must be curated, shared, linked, and then de-identified for statistical purposes, as required	Data must be identifiable and may include case notes to support client-level services
Governance	Often established by individual agencies	Clear parameters around access and use are required, shared processes involving all agencies	Clear parameters around access and use are required, shared processes involving all agencies
Legal Framework	Data may be publicly available (open) or may require a Data Use License to receive in de-identified format	Data access will generally require multiple agreements, including a Memorandum of Understanding, Data Sharing Agreement, and Data Use License to clearly outline permissible access, use, and outputs	Data access may require client consent and non-disclosure agreements. Data agreements must outline parameters for role-based, credentialed access
Data Frequency	Data may be updated based on reporting cycles, quarterly or annually	Archive of select data may be updated periodically depending on availability and analytic requirements	Daily or real-time updates of entire client records may be required
Privacy and Security	A lack of identifiers or small cell sizes in published results means minimal risk of redisclosure	Minimal access to identifiable data and a small group of approved users means that security requirements are essential and managed by a small group of expert users	Many users and identifiable data means that complex permissions, disclosure limitations, and an audit trail will be necessary
Example sites	Members of the National Neighborhood Indicators Partnership	Iowa’s Integrated Data System for Decision-Making (I2D2), Charlotte Regional Data Trust, UNC Charlotte	Allegheny County Data Warehouse, South Carolina Integrated Data System

ⁱⁱThis table was modified and used with permission from AISP (2022), *Finding a Way Forward: How to Create a Strong Legal Framework for Data Integration*.

the day-to-day activities supported by either an executive office (e.g., Mayor or Governor), agency (e.g., Department of Health & Human Services), university, or community-based organization. Determining which partner(s) are best positioned to carry out these activities depends on a variety of factors, such as which partner has legal authority to use the data as intended, staff capacity, sustainable funding, technical capacity, domain expertise, and perceived neutrality among data partners to support dispute resolution. Addressing these practical considerations early on to strategically locate the data integration effort can help prevent future obstacles in executing legal agreements and getting data to flow. Many data integration efforts, especially those early in development, divide the governance, technical, and analytic duties between multiple partners, whereas other efforts may rely on a single agency to carry out these activities.

Collaborative

Data governance is driven by people and should be developed collaboratively, with an emphasis on cultivating trust and building strong relationships across partners. In practice, this often means multiple layers of engagement between the executive leadership that supports strategic decision-making, a data subcommittee (including community partners) that reviews and oversees proposed projects, and the data integration staff that carry out daily operations. Staffing—particularly in terms of data integration staff and subcommittee members—is the essential component for building an effective data collaboration. It is critical to outline the duties of these two groups and provide sufficient resources to staff them, as they will be largely responsible for facilitating strong data governance.

The role of data integration staff is to carry out daily operations while informing and executing strategy. They manage all processes and procedures for data governance; facilitate stakeholder engagement; and often provide the initial review of incoming data requests, vetting for alignment with the effort's research agenda and values and appropriate risk mitigation strategies, before sending to a data subcommittee for further review. These staff should ideally represent a diversity of identities, competencies, and lived experiences to support both the relational and technical work of data sharing. Staffing the data integration effort with team members who bring diverse perspectives can also help in addressing obvious issues with the first three questions—*Is this legal? Ethical? And a good idea?*—before discussions with a broader group of partners.

A data subcommittee is typically responsible for making decisions about the data assets of the agencies represented. When thinking through data use that is legal, ethical, and a good idea, it is important to include data owners (signatory authority for use of data), data stewards (subject matter experts), and data custodians (charged with data security) in the discussion of proposed data uses, as each will articulate different perspectives on the risks, benefits, and limitations. For example, data owners often have a nuanced understanding of political implications regarding data use, while data stewards have a deep knowledge of potential bias and data quality concerns and data custodians know the details of security protocols. All of these roles are essential to engage

in the data governance process, though data stewards and data owners in particular should be involved in decision-making for cross-sector data efforts.

Iterative

Data governance should be an iterative process throughout the life of an IDS and each data project. It should also be revisited and honed regularly as the data sharing effort evolves. All processes and procedures should be thought of as living documents and continuously refined and improved.

Transparent

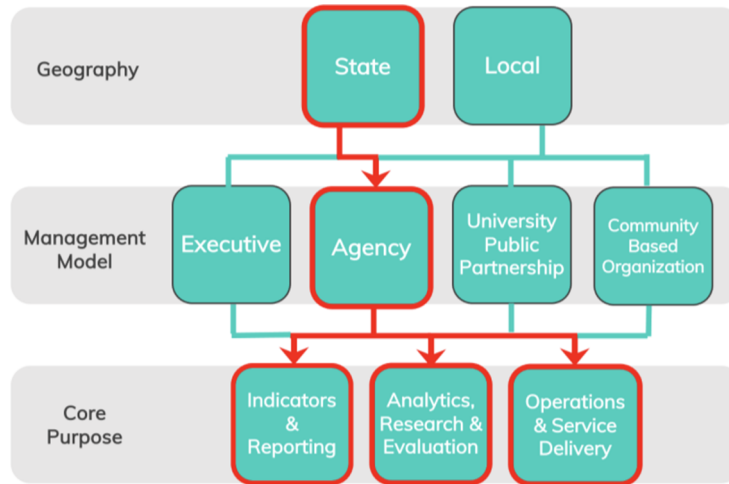
Most data integration efforts in the U.S. are largely funded with taxpayer dollars. Therefore, transparency around the purpose of data sharing, how decisions are made and who makes them, and what data are being shared is essential for accountability. Demonstrating and communicating the value of integrated data to diverse partner organizations and communities also builds social license. Policies, protocols, and documentation of the data integration effort—as well as any specific projects the effort is engaged in—should be readily available to the public in understandable and accessible formats. For example, multiple IDS in the U.S. make their data governance information publicly available, including Linked Information Network of Colorado; Hartford Data Collaborative; Iowa's Integrated Data System for Decision-Making (I2D2); and Connecticut Office of Policy & Management/P20 Win.

Discussion

Data governance for ongoing data sharing and integration should include clearly defined policies and processes to support decision-making, routine meeting structures, and well-documented proceedings—all fostering a culture of trust, collaboration, and openness that supports sustainability. While simple in concept, data governance is incredibly hard to operationalize and implement. The Four Questions can serve as an overarching governance framework for deciding if and how to establish an IDS or whether to approve a specific data use, yet these questions can also be applied within existing governance structures or decision-making frameworks. For example, the Four Questions can guide discernment of each of the Five Safes for data access—safe projects, safe people, safe data, safe settings, and safe outputs [3]. Alternatively, existing frameworks like the Five Safes can work well within the Four Questions, to ensure that IDS projects determined to be legal, ethical, and a good idea also meet the standards of safe data access. Importantly, data governance is highly context specific and should be structured in a way that makes sense for the unique constellation of partners involved.

Below we offer two very different examples of shared data infrastructure from the AISP network that incorporate a clear data governance framework that is purpose- and value-driven, strategically located, collaborative, iterative, and transparent. We then provide hypothetical scenarios that these sites could grapple with using the Four Question Framework to address a policy priority with integrated data.

Figure 3: North Carolina department of health & human services site overview



North Carolina Department of Health & Human Services (NCDHHS)

Lead Agency: North Carolina Department of Health & Human Services.

Data Partners: 33 Divisions & Offices within NCDHHS, Government Data Analytics Center (GDAC), Health Information Exchange, Department of Justice, North Carolina Department of Public Instruction, and other state departments.

Legal Authority: Authorizing legislation, Public Health Authority, Administrative Code, contracts and other agreements (e.g., Intradepartmental Memorandum of Understanding, Data Sharing Agreements, Data Use Agreements).

Funding: state, federal.

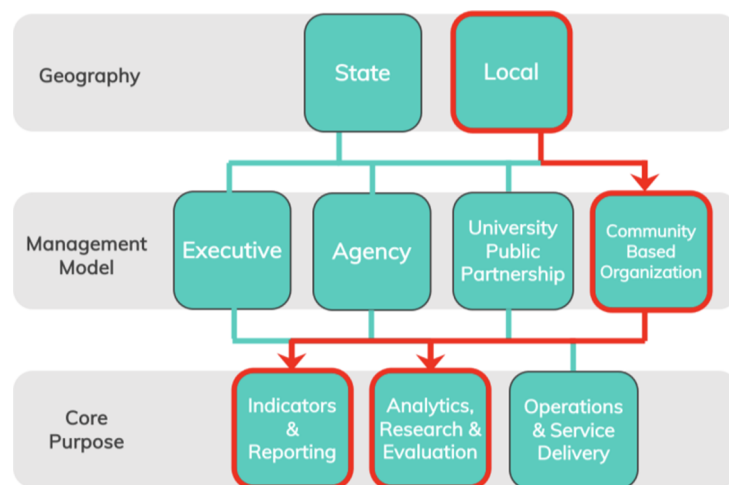
Data Governance Activities: Data Governance Executive Board, Data Governance Council, Data Office. See the NCDHHS Data Sharing Guidebook [14].

North Carolina Department of Health and Human Services

As shown in Figure 3, data integration within the North Carolina Department of Health & Human Services (NCDHHS) operates at the state level to support four core purposes—reporting, analytics, operations, and regulating for health and human services programs. NCDHHS is a large government

agency—\$26B budget, 33 Divisions and Offices, over 17,000 employees, serving the ninth most populated state in the U.S., and situated in a dynamic political context that includes a complex regulatory environment. Since 2019, NCDHHS has employed participatory practices to engage staff at every level of the department in developing and implementing a data governance approach, with steady commitment from leadership and staff to support this work [15].

Figure 4: Hartford data collaborative data integration site overview



Hartford Data Collaborative (HDC).

Lead Agency: CT Data Collaborative.

Data Partners: City of Hartford, Hartford Public Schools, Capital Region Education Council, Capital Workforce Partners, Our Piece of the Pie, The Village, CT Coalition to End Homelessness, and many other community-based organizations.

Legal Authority: contracts (Memorandum of Understanding, Data Sharing Agreement, Data Use License, Informed Consent).

Funding: Philanthropic partners, fee for service.

Data Governance Activities: Data Governance Committees, HDC Data Governance Manual, and Hartford Youth Data Fellows.

Hartford data collaborative

The Hartford Data Collaborative (HDC) is an initiative of the Connecticut Data Collaborative, a small community-based organization in Hartford, CT. Development for the HDC began in 2018 with philanthropic support acknowledging that data from community-based organizations were routinely being utilized for ad hoc data sharing, primarily for evaluation. This was a burden on small organizations and led to duplicative efforts without optimal outcomes. Initial development activities focused on where the data integration effort should be strategically located, and the CT Data Collaborative was chosen to lead the core activities of data integration due to their skilled staff, technical expertise, neutrality among data partners, and domain expertise. As shown in Figure 4, HDC operates at the local level to integrate data primarily for reporting and analytics.

Data access and use scenarios

The following scenarios pose realistic data use proposals and outline key considerations for the Four Question Framework. The question of “how do we know (and who decides)?” all comes down to data governance, which varies widely based on context as the examples of NCDHHS and HDC demonstrate.

These scenarios are hypothetical and not specific to either data integration effort described here. However, the established governance structures and legal framework of NCDHHS and HDC are well equipped to consider such scenarios using the Four Question Framework.

Hypothetical scenarios using the four question framework

The Four Questions are designed to be used within established data governance that includes clear decision-making processes. Each IDS context is different. In some systems, the data owner must approve each use of data (with power to veto use). In others, governance decisions mandate consensus or rely upon voting with majority rule. It is important to acknowledge that politics often influence decisions around IDS creation and use; however, a well-crafted governance process can redistribute power and provide a mechanism for all relevant partners to be engaged in decision-making in a way that makes sense for the context. The Four Questions are meant to guide discussion during such governance processes.

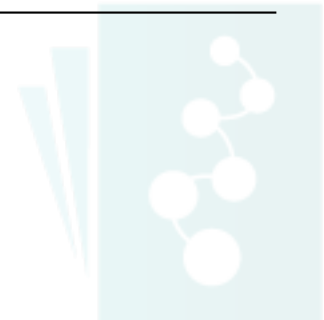
If you were using the Four Question Framework to determine whether to share and integrate administrative data for these proposed policy priorities, what would you recommend?

Scenario 1: Evaluating role of after-school enrichment programs in academic achievement

A large school system is interested in better understanding the connection between involvement in an after-school enrichment program (ASEP) that includes transportation and academic achievement as demonstrated by standardized test scores.

Context: To control for the mediating effect of attendance, the evaluation must include school and ASEP program attendance. The data system that includes school attendance and achievement data is not connected to the system that manages transportation. The transportation data is managed by a private firm, and extracting the data comes with significant cost. The ASEP program only collected age, rather than date of birth, so ASEP records cannot be linked by birthdate to school records.

Is this legal?	Unclear. Some ASEP programs are community-based organizations, and these data are private data, and individual-level consent is likely needed.
Is this ethical?	Unclear. Immediate benefit to students and families is not clear. Informed consent for use of ASEP data is not in place.
Is this a good idea?	Likely no. Data integration may not be possible at this time, particularly for community-based ASEP programs. To plan for this analysis in the future, registration forms could be amended to ask for date of birth (rather than age) and build optional consent into the ASEP program enrollment process for the next school year.
How do you know (and who decides)?	Governance across these partners is not clearly established. Data requests are made to individual data owners by the school system’s program evaluator.



Scenario 2: Dashboarding 2Gen public health indicators

A Department of Health is interested in conducting a community-wide engagement project to collaboratively design public health indicators, using integrated two-generation data on children and their families. These would be shared via an externally facing dashboard, with the goal to improve service interventions of community-based organizations.

Context: This department has experienced 40% turnover in the past 6 months. The governor's race is contentious, and the health commissioner is politically appointed.

Is this legal?	Yes. These data can be aggregated according to agreed-upon guidelines.
Is this ethical?	Yes. Based on involvement of community health clinics and community partners, these data will serve as a support to community-based organizations working in partnership with the Department of Health.
Is this a good idea?	Unclear. This project was a high priority for the previous commissioner and is named after them. It is unclear whether the administrative and technical processes are feasible because of staff turnover within the department, and it is likely that if the governor is not reelected, the project will be abandoned.
How do you know (and who decides)?	The Department of Health has a Data Release Committee with established bylaws yet has not met in 6 months due to staff turnover.

Scenario 3: Conducting an RCT on housing subsidy impacts

A philanthropic partner is interested in the causal impact of housing instability on educational outcomes for preschool age children. They have asked the local IDS to create a list of families who are housing unstable, as indicated by a shelter stay in the previous 18 months, identified as experiencing homelessness within their educational record (through Pre-Kindergarten student or sibling), and/or application for a housing voucher in the previous 24 months. They are requesting a list of individuals for a control and intervention group. The philanthropic partner wants to conduct a randomized controlled trial to evaluate the educational impacts of providing a large housing subsidy for 20 families.

Context: The local housing authority and family shelter are unwilling to partner with this philanthropic partner because of previous contractual issues. It is unclear who would administer the housing subsidy. Involved researchers have communicated to the philanthropic partner that a large body of research has already demonstrated that housing subsidies are effective for improving educational outcomes, but the philanthropic partner is resistant to eliminating the research component of the project. Informed consent is in place for use of identifiable data for service provision, but not research.

Is this legal?	Not currently. Appropriate legal authority is in place for operational use (identification of families for subsidy), but not for research purposes. The use of private data has not been evaluated by legal counsel.
Is this ethical?	Unclear. Families would receive a housing subsidy through random assignment, which is arguably an unethical way to distribute limited resources, particularly when there is strong evidence that the intervention is superior to the control condition. More subsidies could be provided by dropping the research component of this effort. The cost of the RCT research could fund 20 subsidies.
Is this a good idea?	No. Even if the housing subsidy is available and families are identified for receipt, there is no organization available to administer the subsidy program.
How do you know (and who decides)?	The IDS has a robust data governance structure and clear decision-making protocols. Data owners must approve all data use.

Scenario 4: Creating a task force report on racial disparities

The Governor convened a task force on violence prevention that wants to generate a report on the state of racial disparities in the state, with a focus on social determinants of health.

Context: There have been two reports created, using the same data sources, on related topics in the previous two years. All reports show the same general trends and disparities. Neither report included clear recommendations for action.

Is this legal?	Yes. Data use agreements are in place.
Is this ethical?	Yes. Unlike previous iterations, this task force will pay participants and will focus on generating policy and program recommendations.
Is this a good idea?	Unsure. The task force leader is committed to collaboratively generating recommendations for action, with concrete goals for a variety of sectors.
How do you know (and who decides)?	The taskforce includes a data subcommittee with data stewards from all data contributing agencies included in reports. Data stewards make recommendations regarding data access and use that are routed to data owners prior to release.

Conclusion

The Four Question Framework presented here provides a simple, accessible method for data partners to carefully consider the development of shared data infrastructure and proposed data integration projects. Not only must data use be legal to move forward, but it must also be ethical and a good idea, as determined through a robust governance process. Data sharing and integration carries risks that should be weighed alongside the benefits, which are context specific. This process is iterative and as relational as it is technical, which means authentic collaboration across partners should be priority throughout each stage of a data use project.

Acknowledgments

We would like to acknowledge the contributions of AISP team members: Della Jenkins, Dennis Culhane, TC Burnett, Emily Berkowitz, Jessie Rios Benitez, Kristen Smith, and especially Deja Kemp, a lead author of *Finding a Way Forward: How to Create a Strong Legal Framework for Data Integration*.

Thank you to Jessie Tenenbaum and Paul Hogle, North Carolina Department of Health & Human Services and Kate Eikel, CT Data Collaborative, for their on-going work in building this field, their review, and feedback.

We are also indebted to the extensive contributions of expert workgroup members who have shaped this guidance over the years, including: current (as of 2023) AISP Legal Advisory Workgroup members, Karen Barber, Richard Gold, Mark Humowiecki, Paul Hogle, Samuel Kohn, Elliot Regenstein, Joy Royes, Jennifer Cooper, Sean McDonald and Sarah Fulton Hutchins and Paul Stiles; the 2017 expert panel on legal issues, John Petrila, Barbara Cohn, Wendell Pritchett, Paul Stiles, Victoria Stodden, Jeffrey Vagle, and Mark Humowiecki; and the 2017 expert panel on data governance, Linda Gibbs, Amy Hawn Nelson, Erin Dalton, Joel Cantor, Stephanie Shipp, and Della Jenkins.

Statement on conflicts of interest

Amy Hawn Nelson served as a member of the Governance Expert Workgroup convened by AISP in 2016-2017. She also provided technical assistance to support the development of Hartford Data Collaborative from 2017-2021 and has served as a member of the North Carolina Department of Health & Human Services Data Office since 2019.

Ethics statement

Collaboratively generated group norms were created and agreed to by all participants during the first deliberative session. All members of the workgroups agreed to serve as subject matter experts, consented to participate in this collaborative process, and were financially compensated for their travel and received an honorarium for their participation where allowable.

References

1. Actionable Intelligence for Social Policy. About us [Internet]. Philadelphia (PA). Available from: <https://aisp.upenn.edu/>.
2. Actionable Intelligence for Social Policy. Quality Framework for Integrated Data Systems [Internet]. Philadelphia (PA). Available from: <https://aisp.upenn.edu/quality-framework-for-integrated-data-systems/>.
3. Ritchie F. The 'Five Safes': a framework for planning, designing and evaluating data access solutions. Data for Policy. 2017. https://uwe-repository.worktribe.com/index.php/preview/880718/99_Ritchie.pdf.
4. Abelson J. Using qualitative research methods to inform health policy: The case of public deliberation. The SAGE handbook of qualitative methods in health research. 2010:608–20. <https://doi.org/10.4135/9781446268247.n32>
5. Teng J, Bentley C, Burgess MM, O'Doherty KC, McGrail KM. Sharing linked data sets for research: results from a deliberative public engagement event in British Columbia, Canada. International Journal of Population Data Science. 2019;4(1). <https://doi.org/10.23889/ijpds.v4i1.1103>
6. Petrila J, Cohn B, Pritchett W, Stiles P, Stodden V, Vagle J, Humowiecki M, Rozario N. Legal Issues for IDS Use: Finding a Way Forward. Philadelphia (PA): Actionable Intelligence for Social Policy; 2017. <https://aisp.upenn.edu/wp-content/uploads/2016/07/Legal-Issues.pdf>
7. Gibbs L, Hawn Nelson A, Dalton E, Cantor J, Shipp S, Jenkins D. IDS Governance: Setting Up for Ethical and Effective Use. Philadelphia (PA): Actionable Intelligence for Social Policy; 2017. <https://aisp.upenn.edu/resource-article/ids-governance-setting-up-for-ethical-and-effective-use/>
8. Hawn Nelson A, Kemp D, Jenkins D, Rios Benitez R, Berkowitz E, Burnett TC, Smith K, Zanti S, Culhane D. Finding a Way Forward: How to Create a Strong Legal Framework for Data Integration. Philadelphia (PA): Actionable Intelligence for Social Policy; 2023. <https://aisp.upenn.edu/resource-article/finding-a-way-forward-how-to-create-a-strong-legal-framework-for-data-integration/>
9. Zanti S, Jenkins D, Berkowitz E, Hawn Nelson A, Burnett TC, Culhane D. Building and Sustaining State Data Integration Efforts: Legislation, Funding, and Strategies. Philadelphia (PA): Actionable Intelligence for Social Policy; 2021. <https://aisp.upenn.edu/resource-article/building-sustaining-state-data-integration-efforts-legislation-funding-and-strategies/>

10. Kemp D, Hawn Nelson A, Jenkins D. Yes, No, Maybe?: Legal & Ethical Considerations for Informed Consent in Data Sharing and Integration. Philadelphia (PA): Actionable Intelligence for Social Policy; 2023.
11. Feldstein S. The global expansion of AI surveillance. Washington, DC: Carnegie Endowment for International Peace; 2019 Sep 17.
12. U.S. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. United States; 1979 April 18. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html>.
13. Hawn Nelson A, Zanti S. A framework for centering racial equity throughout the administrative data life cycle. International Journal of Population Data Science. 2020;5(3). <https://doi.org/10.23889/ijpds.v5i3.1367>. <https://ijpds.org/article/view/1367>.
14. North Carolina Department of Health and Human Services. NCDHHS Data Sharing Guidebook. Raleigh (NC); 2022 May 13. 36 p. Available from: <https://www.ncdhhs.gov/about/administrative-offices/data-office/data-sharing-guidebook>.
15. Data integration in the time of covid: getting to yes with enterprise-wide data governance. Presented at: IPDLN 2022. Proceedings of the International Population Data Linkage Network Conference. 2022 Aug 25; Edinburgh, Scotland. <https://ijpds.org/article/view/1802>.

Abbreviations

IDS:	Integrated Data System
AISP:	Actionable Intelligence for Social Policy
NCDHHS:	North Carolina Department of Health & Human Services
HDC:	Hartford Data Collaborative



Appendix: Guiding questions to practice the four question framework

The following sets of questions are intended to be a starting point for applying the Four Questions to your own data sharing and integration work. We recommend considering these questions with all relevant partners and adapting them as needed to fit your context.

Question 1 Practice: defining access and use to determine legality

Use the following prompts and examples as a guide to clearly define your data access and use, which then allows you to determine legality.

WHY do you want to share and integrate data?

For example, to:

- Track indicators at the population level
- Identify a target population
- Describe cross-enrollment patterns
- Identify geographic areas of greatest impact
- Evaluate program outcomes
- Improve services at the point of intervention
- Conduct mandated reporting

WHAT type of data do you want to share and integrate?

Is it open, restricted, or unavailable?

For example:

- Information that does not identify individuals
- Information that does identify individuals
- Information that might identify a person
- Health information
- Educational records
- Housing status
- Demographics

WHO do you want to share it with, and who conducts the integration?

For example:

- Executive leadership
- Agency serving the same client
- Probation officers
- A community treatment provider
- A hospital emergency department
- A university-based researcher
- An agency-based analyst

HOW will you share the data?

For example, provide:

- Aggregate counts at the block group level
 - Credentialed access to source data
 - Access to public-facing dashboard
 - View-only access to data underlying a dashboard
 - Edit access to data underlying a dashboard
 - Row-level data with identifiers
 - Row-level data without identifiers
-

Question 2 Practice: Considering risks and benefits to determine ethical use

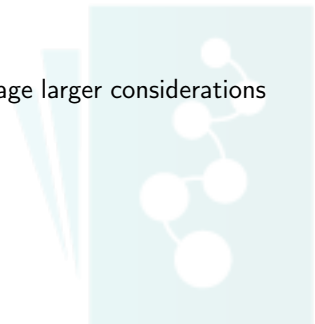
There is a lot to balance when deciding whether and how to use data. Use the following prompts and examples as a guide to consider risks and benefits to determine ethical use.

- Why is this data sharing and integration being conducted? Are data minimization principles considered? Can this question be addressed without the release of identifiable information?
- What are the risks of this data integration?
- What are the benefits of this data integration?
- Who will benefit from this data integration? In what ways?
- Who could be harmed from this data integration? In what ways?
- How are risks being mitigated?
- How will the data be shared to protect privacy and prevent redisclosure?

Question 3 Practice: This is legal and ethical. Is it a good idea?

Now that you've spent time determining legality and ethical use—an important first step—we also encourage larger considerations of the practicalities of data sharing and integration. Specifically:

- Are available data of sufficient quality to answer the question at hand?
- What action can be taken as a result of this analysis?
- How will programs/policies/lives be improved by this use of integrated data?
- What can reasonably be changed or improved based upon the findings? What cannot be changed?
- Has this question already been answered?



- Will the resources needed to conduct this integration yield more benefit than using these same resources for programmatic or direct funding?
- What is the sociopolitical context of this data integration? Is this building upon previous work? Is this work supplanting previous efforts? Is there a related effort that “went wrong” or needs to be acknowledged in some way?
- What are the political implications of this data use?
- Who is conducting this integration and analysis? Do they have sufficient understanding of the program/policy/population that is being studied?
- Who is “asking” the question? Is this topic of interest to the broader community? Do community members, including those “in” the data, know about and support this work?

