

International Journal of Population Data Science

Journal Website: www.ijpds.org



Swansea University
Prifysgol Abertawe

A landscape overview of integrated data systems funding and staffing models across the U.S.

Amy Hawn Nelson^{1,*}, Sharon Zanti², Rebecca S. Pepe³, Isabel Algrant¹, Renette Lee³, and Della Jenkins¹

Submission History

Submitted:	23/01/2026
Accepted:	17/03/2026
Published:	xx/xx/20xx

¹ Actionable Intelligence for Social Policy (AISP), University of Pennsylvania, School of Social Policy and Practice, 3701 Locust Walk, Philadelphia, PA, 19104, United States

² Iowa State University of Science and Technology, Ames, IA, 50011, United States

³ University of Pennsylvania, School of Social Policy and Practice, 3701 Locust Walk, Philadelphia, PA, 19104, United States

Abstract

Introduction

Funding and staffing structures for data sharing and integration vary widely across the U.S., with few recommendations for optimal resource allocation.

Objectives

We sought to deepen our understanding of how Integrated Data Systems (IDS) are funded and staffed over time. This work examines how IDS have developed to provide practical guidance for practitioners, policymakers, and academic audiences interested in establishing and expanding integrated data efforts.

Methods

Since 2008, Actionable Intelligence for Social Policy (AISP) has supported jurisdictions through the process of developing integrated data systems, offering guidance, technical assistance, and training opportunities. We regularly interact with and convene a network of 40+ sites across the U.S. and conduct a biennial network survey. This paper provides insights from survey results (2023, 2025) and qualitative interviews conducted with 24 sites (2024).

Results

Interviews revealed three broad categories of IDS. Small efforts, with budgets between \$250,000 - \$500,000, tend to serve smaller populations, focus on a specific domain, with 2-3 full-time staff members. Medium efforts, which have budgets between \$500,000 - \$1,000,000, tend to be county or state level, often based at universities, with 3-6 full-time staff. Large efforts, which have budgets over a million dollars, tend to be county or state level, operate within government, and have up to 20-40 full-time staff members. Funding comes from a range of sources—federal, local, state, philanthropic, and fee-for-service. Across all categories, sites spend the majority of their budgets on staffing and little, comparatively, on technology and infrastructure.

Conclusion

Key themes persist across the unique landscape of IDS efforts. Regardless of budget and funding mechanisms, sites spend more money on personnel than any other expense. Successful efforts tend to focus initial investments on partnership development and data governance, rather than technical infrastructure, and diversify funding sources over time.

Keywords

integrated administrative data systems; capacity building; sustainable funding

*Corresponding Author:

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

Introduction

This article provides an overview of budget and staffing models for U.S. based Integrated Data Systems (IDS), defined here as secure data infrastructure designed to share, integrate, and analyse administrative data across multiple agencies and drive policy and programme improvement. For this analysis, we are specifically referring to IDS with established data governance bodies and legal frameworks that facilitate routine cross-sector data integration at the person level to guide decision-making in state and local government. These systems support day-to-day analytics, research and evaluation, and service delivery and operations, all while ensuring privacy, legal authority, and increasingly, community engagement [1]. In the U.S. context, IDS are also commonly referred to as data collaboratives, data intermediaries, data hubs, or data trusts [2–4].

The development of IDS is critical to the work of population data science. These systems securely and ethically bring together person level data across domains to identify population level insights and create positive impact for those represented in the data. Although this article is based on state and local government IDS within the U.S., the insights and considerations on staffing models and budgets are relevant to IDS efforts worldwide.

IDS Management Models

Although every IDS is distinct, key differences emerge when comparing systems across three core organising principles: geography (statewide or local), management model (the type of organisation that hosts and governs the system), and purpose (how and why the IDS is used). To support understanding of these variations, AISP created a taxonomy and quality framework based on these dimensions [1, 5], which highlights commonalities and distinctions among IDS initiatives nationwide (see Figure 1).

Executive-Led

These sites are hosted by an executive-level office with close proximity to decision-makers (e.g., Office of a County Executive, Mayor, Governor, or Office of Policy or Budget and Management).

Agency-Led

Agency-led sites are hosted by a large umbrella agency (i.e., Department of Health and Human Services, Department of Education) that encompasses multiple programmes with proximity to both the people represented within the data and the practitioners implementing services. Within the Agency-led and Executive-led sites, there are several that identify as providing “shared services,” meaning that their core role is to conduct data integration and then provide linked data to credentialed users, rather than drive an inquiry cycle.

University-Public Partnership

Such sites are usually hosted by a university, often in collaboration with or in service of government agencies at either the state or local level. Staffing models of such sites typically include student workers or graduate interns.

Community Based Organisation (CBO)

CBO-led sites are hosted by a non-profit agency or backbone organisation (e.g., United Way), often with external partners (private, university, agency, or other) performing other key functions/activities.

Growth of IDS in the U.S.

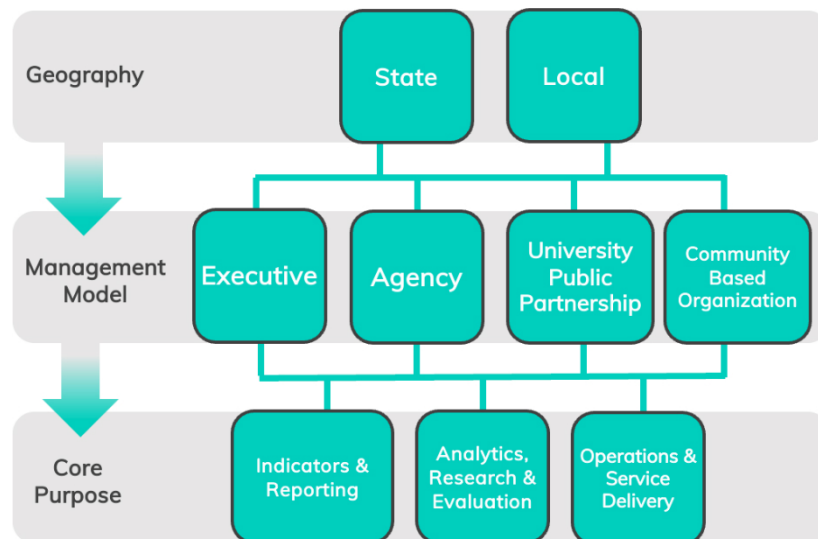
The field of IDS is not new [6] but has expanded rapidly in the past decade. An initial scan conducted by Actionable Intelligence for Social Policy (AISP) in 2008 identified just 8 U.S.-based IDS. Selection criteria specified that an IDS must have an established legal framework and governance structure and be routinely linking cross-sector data for programme and policy analysis[7]. Since this initial scan, AISP has supported state and local jurisdictions building shared data infrastructure through training, technical assistance, and best practice development. Additionally, AISP convenes a network of IDS across the U.S and conducts a biennial survey that captures the state of the field and growth over time. Despite the challenges of routinely linking administrative data [8–10], IDS efforts across the U.S. have been successful in creating shared data infrastructure that includes collaborative governance, clear legal authority, adaptable technology, and appropriate staff capacity [1].

Record linkage has many potential approaches and core purposes, so it is difficult to determine a valid count of data integration efforts in the U.S. However, the parameters of the AISP Network are explicitly defined. To be a member of the AISP Network, sites must have a clear legal framework and data governance to guide the access and use of cross-sector administrative data, with demonstrated linkage and use of data for impact. The AISP Network now includes over 45 sites [11]. The Network operates on a voluntary, no cost membership model, and Network Sites collectively represent 53% of the U.S. population.

Internationally, AISP’s network building work is comparable to organisations such as Administrative Data Research UK (ADRUk) and Health Data Research Network Canada (HDRN). All three organisations support field building and best practices for networks of data holders across jurisdictions. However, unlike ADRUK and HDRN, AISP is not a data holder, does not operate a trusted research environment (TRE), and does not receive government funding for network activities. AISP’s network activities are funded by philanthropic partners and AISP is university-based, fielding a small team (<10 employees). AISP’s primary work is studying best practices for administrative data reuse and supporting the broader field to implement these best practices, while comparable organisations in other countries focus more on research.

The concept of IDS is quite simple—by linking data across domains at the person level, these systems allow partners to better understand programmes and policies and make decisions that support public good. They are an essential resource for evidence-based policymaking. However, significant barriers remain to their development and expansion into new jurisdictions across the U.S. Currently, the growth of the IDS field is constrained by limited understanding of how these organisations are developed, structured, funded,

Figure 1: AISP Taxonomy for IDS



and staffed. Recent publications have begun to address these questions specific to state longitudinal data systems in the education domain [12, 13] but broader efforts across housing, health and human services, child welfare, and beyond remain under-studied. Drawing on results from two surveys of the AISP network and in-depth qualitative interviews, this article describes the typical development trajectory, budgetary, and staffing structures of IDS and explores how resourcing and the chosen management model affect capacity and IDS impact. In doing so, we aim to provide valuable insights for practitioners, policymakers, and researchers seeking to enhance the effectiveness and sustainability of their administrative data reuse efforts.

Methods

Since 2008, AISP has conducted a biennial survey of the network sites to capture the landscape of active integrated data systems across the U.S. This article builds upon surveys disseminated to local and state data integration efforts in the AISP Network in February 2023 (N=37) and June 2025 (N=42), as well as semi-structured qualitative interviews conducted with 24 Network members in 2024 [14]. Together, these sources were used to address the following research questions:

1. How do IDS develop? How are the structure, purpose, and management model determined?
2. How are IDS funded? What does it cost to operate an IDS on an annual basis?
3. How are IDS staffed? How does staff capacity support or constrain the impact of IDS activities?

Surveys

Both surveys included three main sections: (i) IDS purpose and management model; (ii) practices and approaches across the five core quality framework components (governance,

legal, technology, capacity, impact), [1]; and (iii) current data holdings and practices around data use and access. See Appendix A for an overview of the survey instrument. The first section asked respondents to share their management model (e.g., executive-led, agency-led, university-public partnership, community based organisation led) and how integrated data are used (e.g., conduct research, inform policy, evaluate programmes, support case management and service delivery, etc.). In the second section, respondents were asked about governance structure, forms of legal authority used to facilitate interagency data integration, technical approaches to data integration, information about staffing and budgets, and the observed impact of data use across policy domains. In the final section, respondents reported the types of data that were integrated over the previous two years (e.g., data on vital records, health, child welfare and/or adult protection, early childhood, education, legal and law enforcement engagement, economic security and employment, housing and homelessness, or other local, state, or federal data sources). Multiple-choice and free-response questions were included to enable qualitative explanations.

Survey data from 2023 and 2025 were analysed using descriptive statistics to identify patterns and shifts. When discrepancies emerged between years, 2025 data were used as the default point of comparison to reflect the most current trends. Survey findings were further contextualised with interview data to deepen understanding of participant perspectives.

Interviews

We conducted semi-structured interviews to contextualise and fill gaps in the information gleaned from the surveys. These interviews were conducted in person during our AISP National Meeting (June 2024) and virtually (June and July 2024). Interviewees represented a diverse group of practitioners and experts from local and state government agencies, universities, and non-profits. The title and duties of interviewees varied, but all led IDS efforts, including some sites that have

been AISP Network members since as early as 2009, as well as newer members. After participants provided consent, the interviews, which lasted approximately 30 minutes, were recorded, and notes were taken by the interviewer and/or a dedicated notetaker. Interviewees were not compensated for participation. See Appendix B for a copy of the interview protocol.

Deductive coding was used given an a-priori set of codes derived from findings of the 2023 survey data. The interviewer and one additional author reviewed each interview recording for pre-specified information related to the site's focus, partnerships, budget, funding, charge model, actual and ideal staff positions, effort, and structure, as well as lessons learned. The lead author (AHN) reviewed the pulled information and resolved any discrepancies. Social desirability was minimised by framing the interview as a means for new data integration efforts to learn from more established ones. The interviewers emphasised that there were no correct answers and ensured anonymity in result dissemination. The sample size was chosen, in part, to ensure sufficient interviews to reach thematic saturation. A condensed version of both the network survey and interview protocol can be found in Appendix A and B.

Our approach of collecting survey data over two years (2023 and 2025) provides a wealth of longitudinal information, but it also introduces analytical challenges. For one, sites changed during this time period, experiencing shifts in funding sources, budget sizes, and staffing structures. Further, the U.S. federal data landscape has changed significantly since January 2025, whereby a new presidential administration enacted dramatic cuts to research funding, national statistical infrastructure, and grantmaking to states and localities. As a result, almost all IDS—especially those operating within academic institutions—are now operating with less funding and staffing than at the time of data collection. At the same time, there has also been an increase in federal demand for and ingestion of state administrative data, leading to increased concerns about privacy, security, governance, and publicly available metadata. In response, AISP has stopped publishing detailed information on the data holdings of Network members, data that were previously publicly available from 2012-2025. Given this broader context, we prioritised reporting overarching trends and have not included site names to protect their anonymity.

Results

The survey respondents (2023 N=37, 2025 N=42), included IDS representing every major region of the continental U.S., with high representation of coastal states and cities. All sites surveyed have established data governance and legal authority procedures, but vary widely in maturity, scope, purpose, and approach. The 2023 survey results showed significant variation in site development, funding, and staffing, which raised questions about the resources needed to build and maintain state and local IDS. We relied on interview data to fill these gaps and provide the rich, contextual detail needed to answer our core research questions. The following sections report on the combined knowledge generated from surveys and interviews.

How Do IDS Develop?

Change is a defining characteristic of complex systems. IDS naturally grow and shift over time in ways unique to each context; however, the chosen management model is a helpful entry point to understand the typical paths these efforts take. Where these efforts are located (within or outside of government) influences how they are able to develop, the data that is accessible, the legal challenges they face, and the funding streams they can access. Figure 2 shows the distribution of AISP network sites across the four management models: Agency-led, Executive-led, University-public partnership, and Community based organisation led.

Survey and interview findings indicate that agency-led efforts are the most common approach in the AISP Network (See Figure 2). This is likely because agencies tend to have the capacity to address the complex web of U.S. federal and state privacy laws when building IDS. Management models are also influenced by the purpose they aim to serve (See Table 1). For example, a Health and Human Service Agency may develop an IDS to better understand cross-agency enrollment, programme participation, and downstream social service use. Executive-led models tend to start smaller, are often created for a specific initiative (e.g., improving maternal health outcomes, reducing homelessness), and evolve to support interagency, evidence-based policy making. University-public partnerships and CBO-led IDS tend to be more non-linear in their development, and are often situated outside of government, for reasons such as an interest in integrating data within a neutral institution to improve trust or in response to constrained staff capacity in government.

How are IDS Funded?

Funding Sources

As Table 1 demonstrates, IDS funding can come from local, state, federal, and philanthropic funds. Additionally, sites can operate in a fee-for-service model whereby they charge for distinct tasks such as preparation of an analytic dataset or completion of a research project. Each funding source has advantages and disadvantages. Philanthropic funding, which supported 11 sites in 2025, is often short-term (one to three years) and tied to a specific focus (e.g., early childhood) and/or geography (Table 1). Local (n=6), state (n=17), and federal (n=13) government funds offer more long-term stability, but come with strict criteria on how funds can be used (e.g., human capital vs. infrastructure). Given the ongoing demands of maintaining institutional knowledge and data system upkeep, tethering efforts to government funding can be precarious due to changing administrations and shifting priorities at the local, state, and federal level. Fee-for-service offers flexibility to support diverse activities, but the overhead costs of governance, maintenance, and integration of new data sources make it challenging to recoup all costs through this model alone. Therefore, all 12 sites that employ fee-for-service do so in conjunction with other funding sources. Given the disproportionate need for staff effort at the outset of IDS creation, a fee-for-service funding model is only possible once the site is past the development stage.

Figure 2: Count of IDS by Management Model in 2025 (N=42)

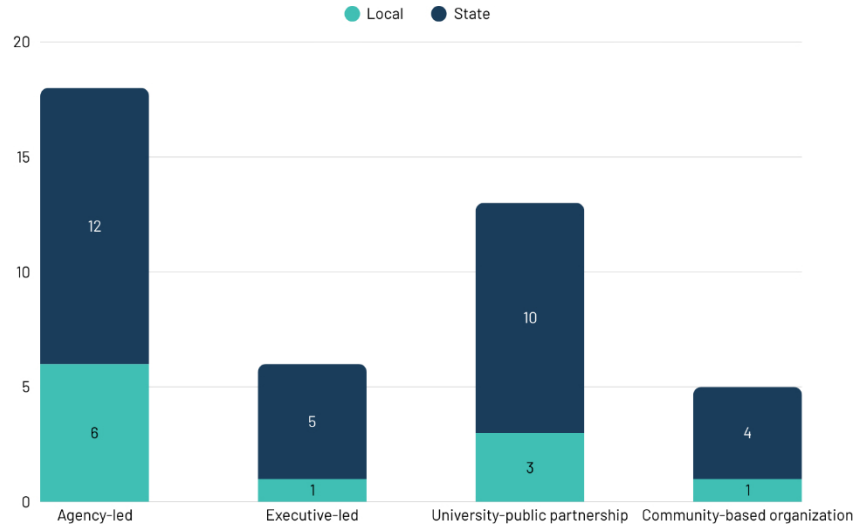


Table 1: Summary Table of Respondents to AISP Network Surveys (2023, 2025) and Qualitative Interviews (2024)

Geography/ Population	Model*	Purpose**				Date of Founding	Shared Service	Operating Cost	Funding Sources***					# FTE Staff
		ER	IR	ARE	CM				L	S	F	P	Fee	
City, 500K	CBO	X	X	X		2017	no	\$250K – \$499K				X		1.5
State, 11M	UPP	X	X	X		2017	no	\$250K – \$499K				X	X	4
City, 120K	CBO	X	X	X		2019	no	\$250K – \$499K			X	X	X	3.5
State, 3.5M	UPP	X	X	X		2016	no	\$250K – \$499K		X	X		X	4.5
State, 6M	UPP	X	X	X		2017	no	\$250K – \$499K		X	X	X	X	3.5
County, 3M	UPP	X	X	X		2014	no	\$250K – \$499K	X			X	X	2
County, 500K	CBO	X	X	X		2010	no	\$500K – \$999K		X		X	X	2
County, 2M	AL	X	X	X	X	2017	yes	\$500K – \$999K	X	X	X			4
State, 7M	AL	X	X	X		2017	yes	\$500K – \$999K		X	X			4
State, 11M	AL	X	X	X		2012	no	\$500K – \$999K		X	X			1.5
State, 4M	EL	X				2005	yes	\$500K – \$999K		X	X		X	4
State, 1M	EL	X	X	X		2009	no	\$500K - y\$999K		X	X		X	7
State, 9.5M	UPP	X				2020	no	\$500K – \$999K		X		X	X	2.5
State, 8M	AL	X	X	X	X	1995	yes	\$500K – \$999K		X	X	X	X	7
City, 8.5M,	EL	X	X	X	X	2012	no	\$1M – \$2.9M	X			X		6
County, 1M	UPP	X	X	X		2005	no	\$1M – \$2.9M	X	X		X	X	4
State, 3.5M	EL	X	X	X		2014	yes	\$1M – \$2.9M		X	X			6
City, 1.5M	EL	X	X	X	X	2002	no	\$1M – \$2.9M	X	X				11
State, 6M	UPP	X	X	X		1985	no	\$1M – \$2.9M		X	X	X	X	22
County, 1.3M	AL	X	X	X	X	2000	yes	\$5M+	X	X	X			40
State, 7M	EL	X	X	X	X	2014	yes	\$5M+		X				40
State, 1M	AL	X	X	X	X	2016	no	\$5M+		X	X			4.5

Table Notes: This Table provides an overview of 22 sites that participated in both surveys and the qualitative interview and provided permission to share this information.

*Management Model: AL = Agency-led; EL = Executive-led; CBO = Community based organisation; UPP = University public-partnership.

** Purpose: ER = entity resolution; IR = indicator and reporting; ARE = analytics, research, evaluation; CM = case management.

***Funding Sources: L = local; S = state; F = federal; Fee = fee-for-service.

Each site has a slightly different origin story, but philanthropic and federal funding are often involved in the initial development, with sites moving to more diverse funding sources as the effort matures. When describing their site's mix of funding sources, one participant reflected "So it is totally

blended, which I'll say is a challenge, managing that, but we also see it as kind of protective. . . regardless of what happens, we know that if any one piece falls out, if any one piece goes away, there is enough other stuff that we feel confident that the core will be sustained." Most importantly, successful

sites use their initial funding to develop relationships, establish data governance, and collaboratively create legal frameworks, rather than building technical infrastructure. This can be a challenge if funders do not see this as a priority, one participant recounts, “so our advice is like, don’t underestimate that time... The relationship building on all different levels, and that is also the hardest fund. Our [funder] realises the immense capacity it takes to build that trust, and then the sharing of data, but not all funders are patient.”

Operational Budgets

Across the types of funding sources, the surveys and interviews revealed three broad categories of budget size: Small efforts are defined as having annual budgets between \$250,000 and \$499,000 and included seven sites in 2025 (Figure 3). Five sites were medium efforts, defined as having budgets between \$500,000 and \$999,999. Large efforts are those with budgets over one million dollars and represented 16 sites in 2025. In general, small efforts tend to serve smaller populations, focus on a specific domain, and employ two to three full-time staff members. Medium efforts often operate at the county- or state level, tend to be based at universities, and employ three to six full-time staff, in addition to student and faculty affiliates. Large efforts tend to operate at the county- or state level, are usually situated within government, and have up to 20-40 full-time staff members.

Interestingly, the costs for development within these bands have remained fairly stable since we began collecting this information in 2008. At their outset, small-sized efforts often operate around \$350,000 and as they develop and grow into medium-sized efforts, expand to around one million dollars. Sites that serve larger geographies are associated with larger budgets (Figure 3) and are the sites most likely to integrate data for case management purposes (Table 1). The largest budget in our Network is about \$10 million dollars annually while six sites are in early development and currently unfunded and seven have budgets less than \$250,000 (Figure 3). As sites mature, the AISP Network has seen a 300% increase in sites that have budgets of more than \$5 million from 2023-2025, increasing from one to four sites. These increases were largely due to existing IDS expanding analytic and reporting capacity. We expect that budgets will continue to rise as local and state governments need more access to linked data.

How are IDS staffed?

Sites spend the vast majority of their budgets on staffing and require comparatively little on technology. The success of a data integration effort is driven by the people who do the work. No matter the size, scope, or purpose of an effort, it is first and foremost a relational endeavour that needs to be sustainably resourced and effectively staffed [15, 16]. Yet effectively staffing an IDS is a challenge.

Interviews indicate that staffing challenges are somewhat universal, regardless of budget and staff size. A small site lead described an ideal state of staffing saying, “If I was in charge, I would create an integrated data model... I think a project manager would help. It’d be nice to have somebody [de]voted solely to governance. [It would] Be nice to have

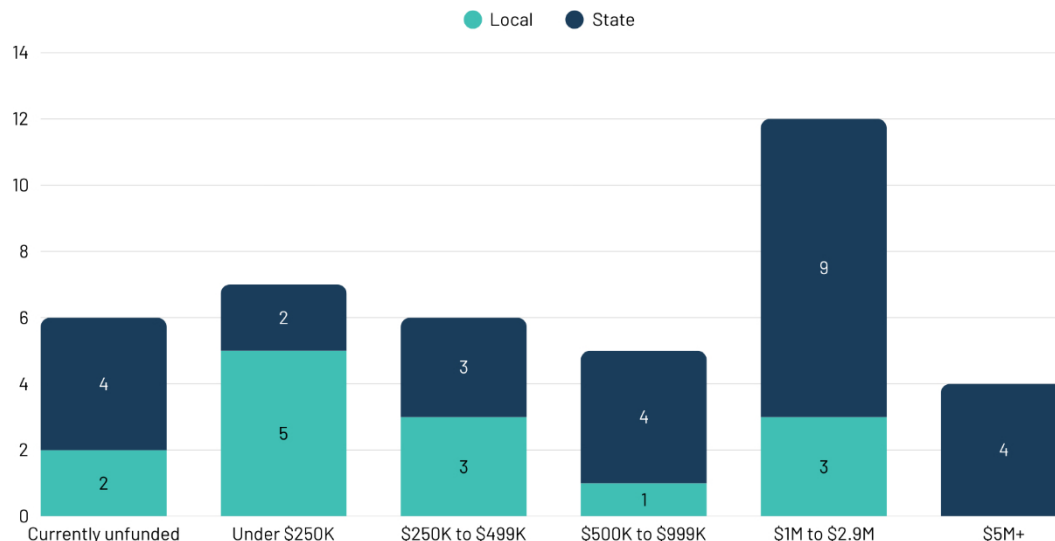
research analysts, some devoted to creating these linkage projects, some devoted to creating products that people can use. You know, that are housed within [the integrated data system]. And then just other people to maintain relationships with all of the data sources.” As the lead recounted this ideal state, they pointed out that currently these roles are: “my role as well.” A larger site described the challenges of hiring for specific roles saying “Oh, we are always hiring” and building their administrative capacity saying “we need someone full-time... and again a lot of us do some of it, but we just sort of feel like we could centralise that, so over time we’ve built up sort of an administrative core, that’s been essential... I don’t know how you can do it without having some administrative core, whether it is handling budgeting and contracting and hiring, there is just a lot of stuff like that.”

As Table 1 indicates, small efforts tend to have less than five full-time staff members (mean = 3.2), medium efforts tend to have three to six full-time staff members (mean = 4), and large efforts may operate with as few as four or as many as 20-40 full-time staff (mean = 16.7). Essential roles include leadership, project management, data governance and administrative support, data analysis, and data management. Depending on the size of the effort, the same person may hold multiple roles or there may be entire teams dedicated to one function. To note, almost all IDS in our network have a similar configuration of staffing during their development (see Appendix C: “Recommended Roles for IDS Staff in Early Development”). More mature efforts have more analytic capacity and robust dissemination practices, including publicly available dashboards and datasets, which expand the staffing needs of an effort. See Wiegand et al. for guidance in strengthening analytic capacity in government agencies [17].

Recruitment, hiring, and training are challenges for IDS, given the need for staff who are highly skilled, have received advanced training, and often are required to work across multiple domains requiring diverse technical expertise. A 2023 survey respondent shared that “we’ve struggled to hire specific high skilled positions (e.g. data engineer) in the past...we often have a hard time attracting good candidates for positions.” Another responded that it “would be nice to have dedicated staffing especially when it comes to the technical infrastructure development,” but many shared the compensation is often a barrier. One respondent explained that “In-house technology staff are underpaid in government. It is very challenging to hire and retain good talent here because of the salary structures imposed by the state.” Given restrictions of rigid salary bands in government positions, one large IDS noted in their interview that while it can be hard to recruit for technical and analytic positions because the agency cannot match the salaries offered by the private sector, their site has been able to remain attractive to hires by “making a case for the mission and working with people that are on the same page” as well as offering flexible schedules and “work-life balance, [where it otherwise] might not exist.” There is no clear training path for IDS staff and as a result, the AISP Network is a multi-disciplinary community representing a wide range of degrees, training, and previous work experience, with members representing expertise across a wide range of content areas and outsourcing others.

Across sites of varying management models, staff sizes, and budgets, most report being understaffed and continually

Figure 3: 2025 Survey, Annual Budget Ranges by Geography (N=42)



stretching limited resources. Staff frequently operate in cross-functional roles, particularly within university- and CBO-based IDS, where individuals may hold multiple titles (e.g., an IDS Director who also serves as university faculty). Balancing competing priorities with constrained capacity remains a consistent challenge, even as sites strive to stay innovative and agile within a rapidly evolving field.

Discussion

As the results show, there is no single way to design, develop, staff, and operate an IDS. Despite the persistent sense that an IDS archetype exists, IDS are unique and develop according to their purpose and context. This diversity, however, serves to highlight important commonalities among sites, including developmental milestones and challenges. The following sections discuss common themes, challenges, and opportunities presented by the collective knowledge base.

IDS Development

IDS development varies across sites, with almost every respondent describing their structure and approach as different from other AISP sites due to factors like age, staffing structures, or funding mechanisms. This diversity suggests that in practice, there is no single conventional model to IDS development and maintenance, even as many still implicitly reference a perceived archetype. Across contexts, personnel costs consistently represent the largest investment, reinforcing that IDS work is fundamentally a people-centred endeavour. Success relies on relationships, buy-in, and operational capacity, while the technical components are often seen as comparatively easier to resource. Because IDS work develops iteratively, systems must evolve as needs and capacities change. As integrated data efforts grow, some sites aspire to become the state's singular data hub, while others intentionally define a narrower niche to avoid serving as the sole statewide resource. This landscape can create competitive dynamics, where multiple IDS efforts coexist, underscoring

the continually shifting environment in which these systems operate.

Challenges and Opportunities

Management Model

While some challenges are universal, there are distinct challenges and opportunities based on the management model. Within the agency, executive, and university setting, there is a persistent need to educate the institution about how IDS work differs from standard processes, causing potential for friction when existing systems and expectations don't align with what integrated data work requires. For example, sites in these settings face issues of procurement (e.g., How can we craft a sole source contract for this particular approach?) and salary bands (e.g., How can we pay our data scientists close to what private industry offers?). Conversely, this is where CBO-led efforts shine, and the origin story of AISP Network Sites that are situated within CBOs often include anecdotes that the management model was selected to avoid complex bureaucracy and allow more nimble funding and staffing. However, being located in a CBO is often less stable, without economies of scale to facilitate niche staffing positions present in larger university or government institutions (e.g., database architect or ethics officer).

Universities tend to have more stable staffing than government and CBO's, with greater staff retention, faculty often serving in lifetime appointments, and access to graduate students that may support the work for 2-6 years. Universities also have access to staff with deep expertise that can be utilised for consultation and support through a dedicated portion of staff effort. Universities are used to funding a percentage of a person's time on a project, an approach that is atypical in other contexts. Universities play a particularly important role in the university-public partnership by offering continuity and stability that governmental partners often lack. Within these partnerships, the university often serves as the source of institutional memory, providing long-term

consistency and helping sustain efforts despite turnover, shifting priorities, or structural changes within government agencies. However, it is important to note that public partnerships require consistent engagement and support from government and CBO partners. Data partners must be invested to participate in governance activities, which are often outside of formal job duties.

Funding

Funding for IDS work is shaped by a core sustainability lesson: diversify revenue sources. Sites emphasise beginning with investments in partnership development, data governance, and relationship building—rather than starting with infrastructure—because trust and priority alignment are foundational to success. Philanthropic dollars often support early stages through planning grants and start-up operating costs, while federal funds typically finance technical infrastructure and major initial investments, with the goal of eventually shifting to ongoing appropriations. Approaches vary and some sites leverage project funding to strengthen infrastructure, whereas others intentionally pursue infrastructure-specific funds and avoid project-based support. Many sites note that the slow, relational nature of this work is difficult to fund. Even when systems are functioning well, securing new investment can be challenging because leadership may not see an urgent need. One respondent reflected *“when something works and people see projects in place. How do you convince them that you actually need to make investments... because they're like ‘It works. You guys are here’.”*

In the interviews, all sites distinguished hard versus soft funding. Hard funding is stable, ongoing support—often from government appropriations or permanent budgets—that an organisation can reliably expect year to year. Soft funding is temporary or uncertain support—like grants, philanthropy, or short-term contracts—that must be renewed or replaced regularly. Sites often weigh the benefits and drawbacks of these funding types. Avoiding constant grant-seeking can be beneficial, but reliance on government appropriations can be precarious amid administrative turnover. Fee-for-service models offer flexibility but reflect a specific perspective on data rights and access. Specifically, several respondents discussed the contradiction of viewing data as a public resource that should be made available regardless of ability to pay, but recognising that sites must cover data management and analytic costs to provide that information.

Staffing

Respondents frequently discussed the importance of consistent staffing and ensuring transfer of institutional knowledge. They emphasised the importance of documenting processes early and often, both to maintain continuity and to distribute knowledge across the team. Consultants or “pseudo employees” can help fill gaps, but systems should avoid relying on “unicorns”—individuals who hold all the institutional knowledge and who cannot be replaced. One respondent noted *“but you don't wanna have a unicorn, and that's also the kiss of death. You want a process that's replicable, and safety nets that people new can come in, learn and follow.”* Instead, as the quote suggests, key information should be codified

in procedures, shared among multiple staff members, and supported by redundancy and cross-training to ensure the work remains resilient even amid turnover or change. Respondents also discussed the challenge in funding capacity for analysis and dissemination. In small- and medium-sized IDS, staff are driving all parts of the data life cycle. Often this limits their capacity to devote time to important tasks that allow for meaning-making after initial data products are delivered, such as dissemination, strategic communication, and partner engagement. The major distinction between sites lies in their capacity to afford personnel with expertise across stages of the data life cycle, meaning separate staff to support public involvement, conduct analyses, design data products internally, and communications professionals that can use these products to communicate externally and drive action.

Supporting IDS Capacity

In both the surveys and qualitative interviews, respondents were asked about how AISP and its network of sites could help to address needs and challenges. Assistance with capacity constraints was a consistently identified need.

As noted above, sites indicated that they are understaffed across a range of areas—particularly in legal, technical, communications, engagement, and project management roles. Several sites referenced difficulty recruiting and supporting data scientists. Reliance on project-based or soft-money funding undermines long-term sustainability, while dissemination and grant writing remain underfunded. Many expressed interest in strategies such as student engagement, shared services, and multi-site collaborations, along with a need for clear job descriptions and role definitions to streamline hiring. Sites indicated interest in business planning support for fee-for-service or diversified funding models, shared staffing structures across legal, administrative, and technical functions, examples of sustainable funding approaches in universities and government agencies, and a resource bank of job descriptions and hiring strategies.

Conclusion

While no single model exists for developing, funding, and staffing an IDS, we found that several key themes persist across all efforts, regardless of budget and staffing structures. Based on these findings, we recommend that building integrated data capacity begin with upfront investment in partnership development and data governance, rather than focusing initial funds on technical infrastructure. This is foundational to demonstrating early value and building long-term success. Although diversifying revenue sources supports sustainability, realistically, sites may rely on one main source of funding to launch their IDS and then work on securing more durable and diversified funding sources over time. IDS sites also find success by drawing on the unique benefits of their managing organisation—whether it be a government entity, university, or CBO—and addressing the inherent challenges of their management model through strategic funding and staffing decisions. Despite its moniker, IDS work is inherently people-centric: it requires a large investment in skilled personnel to navigate governance and legal processes with

care, manage and analyse highly sensitive data, navigate increasingly complex analytic needs, and generate actionable insights to support improved programmes, policy, and practice.

It is important to acknowledge that the U.S. context has experienced significant changes at all levels of government since 2025, which has ripple effects on AISP Network sites and other administrative data efforts across the country. Federal funding for both research and state and local programmes has decreased. At the same time, the federal government has been attempting, sometimes successfully, to access and use state and local administrative data for enforcement actions against immigrants and other groups. Trust in public institutions is at an all-time low, and IDS are working with fewer resources, more public scrutiny, and greater concern regarding social licence. In the wake of these changes, investment in state and local data infrastructure that centres legal authority and participatory governance—all components of AISP Network sites—is critical. The field of IDS has demonstrated that taking the time to build relationships, protocols, and systems is critical for ethical data use, and in these unsettling times, IDS act as an important layer of privacy and security, especially for vulnerable people represented in the data. Our goal is for the work of the AISP Network to provide guidance to efforts in the U.S. and globally to focus investments in this critical infrastructure with data ethics at the forefront.

Acknowledgements

We would like to acknowledge all members of the AISP Network who are active participants in our learning community—attending our National meetings, completing our biennial survey, contributing to network events, and connecting with each other to build the field. We give special thanks to the 24 sites who participated in interviews for this study and provided invaluable insights.

AISP has been generously supported by The Robert Wood Johnson Foundation, The Annie E. Casey Foundation, The Gates Foundation, The Walton Family Foundation, and The Ford Foundation, among others. The work of AISP is collaborative and would not be possible without the contributions of many current and former team members, including: Emily Berkowitz, TC Burnett, Dennis Culhane, John Fantuzzo, Deja Kemp, Jessie Rios Benitez, Kristen Smith, and many others who contribute to and inspire us daily in their work.

Statement on Conflicts of Interest

None declared.

Ethics Statement

The AISP Network Survey research (2023, 2025) is covered by the approved research protocol titled IDS Landscape Analysis, University of Pennsylvania, IRB Protocol #: 834988. The qualitative interview research (2024) is covered by the approved research protocol titled Actionable Intelligence for Social Policy-Qualitative Research and Case Studies, University of Pennsylvania, IRB Protocol #: 850559. All

participants gave informed consent before participating in the surveys and qualitative interviews.

Data Availability Statement

Information on individual sites within the AISP Network are available at <https://aisp.upenn.edu/about-aisp-network/>. Findings from the 2023 Network Survey have been published in a series of briefs, available here: <https://aisp.upenn.edu/network-survey/>. We anticipate the Network Survey Briefs being updated with data from the 2025 survey by Fall 2026.

AI Disclosure Statement

The authors declare that no generative AI tools were used in the preparation of this manuscript.

References

1. Actionable Intelligence for Social Policy. Quality Framework for Integrated Data Systems [Internet]. Philadelphia (PA): Actionable Intelligence for Social Policy; 2025 [cited 2026 Jan 5]. Available from: <https://www.aisp.upenn.edu/quality-framework-for-integrated-data-systems>
2. Hawn Nelson A, Jenkins D, Zanti S, Katz M, Burnett T, Culhane D, et al. Introduction to Data Sharing and Integration [Internet]. Philadelphia (PA): Actionable Intelligence for Social Policy; 2020 [cited 2026 Jan 5]. Available from: <https://aisp.upenn.edu/introduction-to-data-sharing>.
3. Susa I, Marijn J, Verhulst S. Data Collaboratives as a New Frontier of Cross-Sector Partnerships in the Age of Open Data: Taxonomy Development. 50th Hawaii International Conference on System Sciences. 2017. <https://doi.org/10.24251/HICSS.2017.325>
4. Goerge RM. Barriers to Accessing State Data and Approaches to Addressing Them. The ANNALS of the American Academy of Political and Social Science. 2017; 675(1). <https://doi.org/10.1177/0002716217741257>
5. Hawn Nelson A, Kemp D, Jenkins D, Rios Benitez R, Berkowitz E, Burnett TC, et al. Finding a Way Forward: How to Create a Strong Legal Framework for Data Integration [Internet]. Philadelphia (PA): Actionable Intelligence for Social Policy; 2025 [cited 2026 Jan 5]. Available from: https://aisp.upenn.edu/wp-content/uploads/2025/12/AISP_FindingAWayForward_2025.pdf
6. Zanti S, Berkowitz E, Katz M, Hawn Nelson A, Burnett T, Culhane D, Zhou Y. Leveraging integrated data for program evaluation: Recommendations from the field. Evaluation and Program Planning. 2022; 95. <https://doi.org/10.1016/j.evalprogplan.2022.102093>

7. Culhane D, Fantuzzo J, Rouse H, Tam V, Lukens J. Connecting the Dots: The Promise of Integrated Data Systems for Policy Analysis and Systems Reform [Internet]. Philadelphia (PA): Actionable Intelligence for Social Policy; 2010 [cited 2026 Jan 5]. Available from: <https://aisp.upenn.edu/resource-article/connecting-the-dots-the-promise-of-integrated-data-systems-for-policy-analysis-and-systems-reform>
8. Coulton C, Goerge R, Putnam-Hornstein E, de Haan B. Harnessing Big Data for Social Good: A Grand Challenge for Social Work [Internet]. American Academy of Social Work and Social Welfare, Grand challenges for social work initiative working paper no.11, p.1-20; 2015 [cited 2026 Jan 5]. Available from: <https://grandchallengesforsocialwork.org/wp-content/uploads/2015/12/WP11-with-cover.pdf>
9. Harron K, Dibben C, Boyd J, Hjern A, Azimae M, Barreto ML, et al. Challenges in Administrative Data Linkage for Research. Big Data & Society. 2017; 4(2). <https://doi.org/10.1177/2053951717745678>
10. Johnson D, Massey C, O'Hara A. The Opportunities and Challenges of Using Administrative Data Linkages to Evaluate Mobility. The ANNALS of the American Academy of Political and Social Science. 2014; 657(1), 247-264. <https://doi.org/10.1177/0002716214552780>
11. Actionable Intelligence for Social Policy. Integrated Data Systems Map [Internet]. Philadelphia (PA): Actionable Intelligence for Social Policy; 2025 [cited 2026 Jan 5]. Available from: <https://aisp.upenn.edu/integrated-data-systems-map>.
12. Data Quality Campaign. Funding What Matters: What Building and Sustaining a Statewide Longitudinal Data System Costs [Internet]. Data Quality Campaign; 2025 [cited 2026 Jan 5]. Available from: <https://dataqualitycampaign.org/resource/funding-what-matters>.
13. Boer B. Developing and Costing a State Longitudinal Data System Organization [Internet]. EdSystems Center at Northern Illinois University; 2025 [cited 2026 Jan 5]. Available from: <https://edsystemsniu.org/developing-and-costing-a-state-longitudinal-data-system-organization>.
14. Actionable Intelligence for Social Policy. About the AISP network [Internet]. Philadelphia (PA): Actionable Intelligence for Social Policy; 2025 [cited 2026 Jan 5]. Available from: <https://aisp.upenn.edu/about-aisp-network>.
15. Gibbs L, Hawn Nelson A., Dalton E, Cantor J, Shipp S, Jenkins D. IDS Governance: Setting up for Ethical and Effective Use [Internet]. Philadelphia (PA): Actionable Intelligence for Social Policy; 2017 [cited 2026 Jan 5]. Available from: <https://aisp.upenn.edu/wpcontent/uploads/2016/07/Governance.pdf>.
16. Berkowitz E, Jenkins D, Hawn Nelson A. Network Survey Brief: Capacity [Internet]. Philadelphia (PA): Actionable Intelligence for Social Policy; 2025 [cited 2026 Jan 5]. Available from: <https://aisp.upenn.edu/capacity>.
17. Wiegand E, Monahan W, Geoghegan R, Wavelet M, Goerge R. Strengthening Analytics in Government Agencies: A Toolkit for Sustainable Data Use [Internet]. Washington (DC): U.S. Department of Health and Human Services, Office of Planning, Research, and Evaluation, Administration for Children and Families, OPRE Report 2023-148; 2023 [cited 2026 Jan 5]. Available from: <https://aisp.upenn.edu/resource-article/strengthening-analytics-in-government-agencies-a-toolkit-for-sustainable-data-use>.

Abbreviations

IDS:	Integrated Data Systems
CBO:	Community Based Organisation
UPP:	University Public-Partnership
AL:	Agency-led
EL:	Executive-led
ER:	Entity Resolution
IR:	Indicator and Reporting
ARE:	Analytics, Research, Evaluation
CM:	Case Management



Appendix A: AISP Network Survey

These are the elements collected from AISP's 2025 Landscape Survey of Integrated Data Systems in the U.S.

Section 1: Contact Information, Purpose, Management Model

- Name of data integration effort
- Contact information for main point of contact
- Other organisations who lead or co-lead the data integration effort
- Jurisdiction level (e.g., state, city, county, region)
- Data integration model (e.g. executive-led, agency-led, centralised IT, university-public partnership, community based organisation partnership)
- Purpose for data sharing (e.g., indicators and reporting; analytics, research, and evaluation; operations and service delivery)
- How integrated data are used (e.g. research, inform policy, programme evaluation, outcomes oriented/ performance-based contracting, case management, contract management, inform funding allocation/budget decisions, community needs assessment, grant reporting, planning, and development, predictive analytics/ forecasting, etc.)
- Primary characterisation of data integration effort (e.g. Integrated Data System, Policy Lab, Data Hub, Data Trust, Data Collaborative, Data Intermediary, Community Information Exchange, State Longitudinal Data System, Early Childhood Integrated Data System, etc.)
- Year that planning for data integration effort began

Section 2: Development and Approach across Core Components of the IDS

Governance

- Mission, vision, or principles to data sharing
- Components of governance structure (e.g. executive board or steering committee, data and/or research subcommittee, community/resident advisory committee, etc.)
- Representation in governance structure (e.g. government partners, university partners, philanthropic partners, advocates, community members, etc.)
- Number of government entities on governance board
- Content included in governance documents (e.g. references to legal authority, mission/vision/purpose statement, policies and procedures, commitment to equity, processes related to risks vs. benefits of data sharing, etc.)

- Whether IRB approval is required for research requests
- Data request process
- How projects are prioritised (e.g. funding, potential impact, government agenda/ priorities, order received, etc.)
- Method to support social licence and public trust (e.g. community review or engagement in projects, community representation on executive board, community advisory group, in development, etc.)
- Governance challenges, if any, which require support or resources

Legal

- Forms of legal authority used to facilitate interagency data sharing (e.g. legislation for cross-agency data sharing, legislation specific to priority use case or population, executive order, policy/rule, contracts such as MOU/DSA/DUA/DUL, etc.)
- Whether a tiered legal approach is used
- Whether consent is required for any current use of integrated data
- Number of partners with signed agreements
- Whether an electronic system is used for tracking/ monitoring updates to legal documents
- Language of legal agreements (i.e. whether in plain or legal language)
- Whether agreements are publicly available
- Approach for renegotiating legal agreements (e.g. standard term, varied terms, ad hoc, etc.)
- Source of legal support (e.g. in-house counsel, outside counsel, etc.)
- Legal challenges, if any, which require support or resources

Technology/Technical

- Technical approach for data integration (e.g. centralised, decentralised/federated/distributed/ hybrid-federated, etc.)
- Whether there is a master person/ patient/ client index
- Tool used to link data (e.g. Databricks, Fastlink, Link King, Python, R, SQL, Splink, Custom Logic, etc.)
- Privacy Preserving Technologies (PPTs) used (e.g. secure hashing, secure multiparty computation, homomorphic encryption, differential privacy, private query server, synthetic data, etc.)
- Whether technical approaches to sharing, linking, providing data access has changed in the past two years, and whether changes are being considered
- Data Access and Security Procedures

- Management of protected data (e.g. secure room, secure device, VPN, secure transfer method, etc.)
- Whether analytic data sets are provided to external researchers or users, and how access is granted to them

● Ensuring Data Quality

- Processes for assessing and documenting data quality (e.g. metadata standards, data quality dimensions framework, routine data quality checks, collecting information about quality from users, etc.)
- Tool used for data documentation and metadata management
- Identification of data owners for metadata
- Whether metadata is available publicly

- Technical challenges, if any, which require support or resources

Capacity

- Staffing source of governance activities
- Staffing source of data analytics
- Staffing source of technical infrastructure & integration
- Involvement of Chief Data Officer, if any
- Total FTE for personnel working directly for data sharing effort
- Types of personnel working directly for data sharing effort (e.g. full-time, part-time, contractors, students/ interns, etc.)
- Work nature of staff roles (e.g. research & analysis, technology, privacy and legal, project management/ coordination, communications, partner engagement, community engagement, etc.)
- Staffing challenges, if any
- Data integration effort's total annual operating budget
- Whether there has been significant changes to operating budget in the past 6 months due to shifts in state or federal government funding/priorities/policies
- Anticipated changes to funding, portfolio of work and/or partner relationships in the next 6 months
- Percentage breakdown of annual operating budget for data integration effort by source (e.g. fees for services, data licence fees, federal grants, state grants, local grants, foundation grants, university support, etc.)
- Structure of data licence fees
- Capacity (staffing or budget) challenges, if any, which require support or resources

Impact

- Top five policy areas that take highest priority (e.g. adult legal system & law enforcement, child welfare, early childhood, employment/workforce training, healthcare service use, homelessness, housing, juvenile legal system involvement, K-12 education, mental health, postsecondary education, public health, substance use, two-generation programmes, etc.)
- Observed impacts of integrated data (e.g. increased capacity for evidence-based policymaking, improved ability to evaluate what works, improved care coordination and case management, cost savings and efficiencies, etc.)
- How findings are communicated with the community and/or people represented in data
- How findings are communicated to government partners and elected officials
- Whether consulting or research services are offered for other integration efforts
- Products that have been produced or made possible with data integration (e.g. reports or briefings for policymakers, contributions to public testimony, research publications, conference presentations or webinars, dashboards or other data visualisations, public-facing communications, annual reports, grant proposals, etc.)
- Total number of products that have been produced or made possible with data integration
- Impact (staffing or budget) challenges, if any, which require support or resources

Section 3: Data Holdings

Categories of data you have used/ accessed in the last 2 years as part of data integration effort and nature of data use/access (i.e. ad hoc or routine)

- Vital records (e.g. birth or death records)
- Health (e.g. payer health claims, medicaid, substance use, disease registries, COVID-19 vaccinations, other vaccinations, community health centres, lead exposure, intellectual and developmental disabilities, behavioural and mental health, emergency medical services, nursing facility MDS, etc.)
- Child welfare and/or adult protection (e.g. out-of-home foster care, dependency, juvenile justice placement, delinquency, abuse and neglect, in-home services, adult protective services, etc.)
- Early childhood (home visiting, early intervention, child care subsidy, early head start, head start, other publicly funded pre-K programmes, early childhood workforce, etc.)
- Education (e.g. K-12 public education, special education, charter schools, postsecondary education, FAFSA, etc.)

- Legal system & law enforcement (e.g. family court, juvenile court, adult court, juvenile probation, local jail, state corrections, law enforcement, etc.)
- Economic security & jobs (e.g. TANF, SNAP, SNAP for WIC, workforce training programmes, unemployment insurance/ wage data, LIHEAP, SSI, child support, EITC, child tax credit, etc.)
- Housing & homelessness (e.g. HMIS, PHA, ed. Records indicating homelessness, state housing subsidy, eviction records, housing code violations, mortgage and/or rental assistance programmes, etc.)
- State or local data sources (e.g. DMV, voter records, etc.)
- Federal data sources (e.g. Census Bureau, IRS, Veteran affairs data, etc.)
- Non-profit sources
- Timeframe of data in datasets (e.g. oldest integrated data, most recent integrated data, frequency of refreshed data)

Appendix B: interview protocol for IDS budget and staffing models

Section 1: Budget/Funding sources

Tell me about how your IDS is funded:

- What are the funding sources?
 - prompt if needed, Federal, state, local, philanthropic, fee-fo- service, fee for data request. . .)
- What functions/pieces of the data infrastructure do these sources fund?
 - Prompt if needed, governance activities, legal activities, software / hardware, data management support, analytic support, data visualisation support, convenings. . .

- What is your high-level operating budget?
 - Prompt if needed, a back of the napkin for a fiscal year is fine.

Section 2: Staffing Structure

- What are the staff positions that support your IDS? (FTEs and PT)
 - Prompt if needed,
 - Director, Asst director
 - Coordinator / project manager / administrative rolls
 - Legal support
 - Analytics, data scientist, analyst
 - Data management, IT services
 - Engagement coordinator
- What are their titles/roles?
- What % of their effort is dedicated to the IDS?
- Are there any new staff roles expected in the future?
- What would your ideal staffing structure be, if you had the resources to support it?

Section 3: Development

- We think it's really valuable to understand how sites grow their data integration efforts over time. Could you tell us about where your funding and staffing started, and how they evolved to where you are now?
- What challenges did you face around funding and staffing when the IDS began? How have those challenges changed over time/what are they now?
- Looking back, is there anything you would have done differently in terms of building funding and staffing capacity?



Appendix C: Recommended Roles for IDS Staff in Early Development

Role	Skills/Competencies
Director	• Nuanced understanding of human subjects research, social science research, and a wide range of research methodologies • Familiarity with IDS stakeholder groups • Comfort in navigating bureaucracy • Ability to establish, cultivate, and maintain a wide range of relationships • Excellent communication skills; ability to communicate effectively among a range of competing interests • Ability to keep abreast of IDS field with dynamic legal parameters, best practices, and research methodologies
Data/Research Coordinator	• Advanced project management skills • Strong communication skills • Knowledge of research methodologies • Experience with data set/data holdings management, including managing codebook
Research Analyst	• Demonstrated ability to support database management, including data cleaning, data set management, coding • Demonstrated ability to support analysis and reporting • Nuanced understanding of statistics
Data Scientist	• Demonstrated experience with algorithm development and data integration • Strong technical documentation skills • Demonstrated understanding of best practices for data security, data transfer, data cleaning, and data import • Strong communication skills; ability to translate technical concepts into accessible language for a variety of stakeholders • Basic understanding of social science methodologies
Database Administrator*	• Experience with design and development of flexible database solutions • Strong technical documentation skills • Demonstrated understanding of a variety of operating systems, memory and storage requirements, firewall restrictions and ports, and data security best practices
Data Security Officer*	• Nuanced understanding of data security best practices across data transfer, storage, and export
Legal Counsel*	• Nuanced understanding of legal issues pertaining to data sharing

*This position could be staffed part-time or on a consultant basis.

