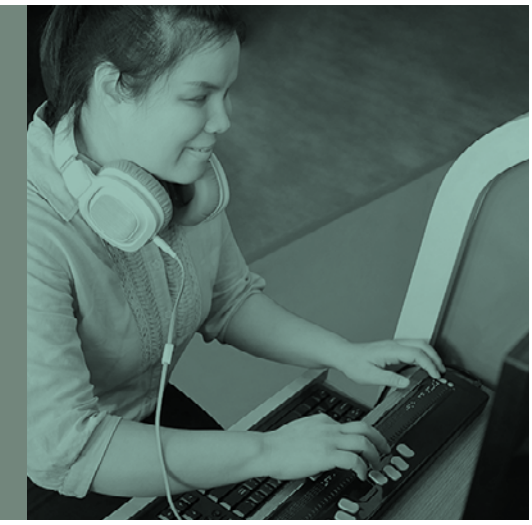








astho®

CSTE

Understanding Disability and Health through Data Linkage

A Toolkit for Public Health Practice

Executive Summary

This toolkit offers a high-level overview of data linkage as a strategy to address gaps in disability and health data by improving the completeness, consistency, and usefulness of data across populations, highlighting its value for improving population estimates, enhancing service coordination, and deepening understanding of health inequities. While disability data are often collected across healthcare, public health, education, social service, and administrative systems, these data frequently remain disconnected and underutilized. Data linkage provides an opportunity to leverage existing data resources to understand the relationship between disability and health better and to inform public health action. While the toolkit supports a wide range of public health applications, it does not provide detailed guidance for specific subgroups. Instead, it emphasizes areas where subgroup-specific issues in definition, measurement, and interpretation require additional expertise. This framework was developed with input from the ASTHO-CSTE Disability Health Data Workgroup, whose composition and contributions informed its design. This toolkit outlines data linkage rationale, describes its phases in detail, and introduces tools to support a systematic approach. The toolkit is structured as a step-by-step process, guiding users from defining purpose through implementation, evaluation, and use of linked data. Consistent with the principles described throughout this toolkit, accessibility and inclusion were considered during development, and jurisdictions should ensure that adapted materials and dissemination products are accessible to the communities they are intended to serve.

How Should This Toolkit Be Used

This toolkit is designed to support public health practitioners and their partners in planning, conducting, and interpreting data linkage efforts for understanding disability and health. This toolkit is intended to be used as:

- **A roadmap.** Follow the process outlined in the toolkit to guide planning and implementation.
- **A reference.** Use specific sections to address challenges, such as understanding how disability is measured across data sources, navigating governance considerations, or determining whether data can be effectively combined.
- **A decision-support tool.** Use guiding questions, visuals, and examples to assess whether linkage is suitable, practical, and responsible in a given context.

Users should review the toolkit early in the planning process to understand the scope of work, including the required expertise, time, and resources, as well as potential challenges. This toolkit includes examples, considerations, and resources to support decision-making. However, not all approaches or tools will be applicable in every context. Users should adapt the guidance based on their specific data, partnerships, and goals.

Who This Toolkit Is For

This toolkit is intended for state, tribal, local, and territorial public health practitioners and their partners who are interested in better understanding disability and health by linking data to inform public health practice, policy, and research. Because data linkage is inherently collaborative, this toolkit may also be useful for a range of partners involved in or supporting this work, including:

- Epidemiologists, data analysts, and informatics staff.
- Program managers and policy staff.
- Data governance and legal personnel.
- Academic partners and researchers.
- Community-based organizations and disability partners.

This toolkit will be most useful for individuals or teams who:

- Have a foundational understanding of public health data and are comfortable working with data from multiple sources.
- Are interested in using data linkage to answer specific questions or inform decision-making.
- Understand their organizational context well enough to assess feasibility (e.g., data access, governance requirements, and partnerships).
- Can identify and engage relevant partners, including subject matter experts, technical experts, and community partners with lived experience.
- Are prepared to navigate the practical and ethical considerations involved in data linkage, including data limitations, representation, and responsible interpretation.

This toolkit is designed to help both technical and non-technical partners understand the process, ask the right questions, and contribute meaningfully to data linking efforts. Data linkage is not only a technical process but also a collaborative and interpretive one. This toolkit is designed to support both the technical and decision-making aspects of this work.

Funding Acknowledgement

This toolkit was supported by The Centers for Disease Control and Prevention of the U.S. Department of Health and Human Services (HHS) as a part or component of a financial assistance award totaling $112,000 with 100 percent funded by CDC/HHS. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement, by CDC/HHS, or the U.S. Government.

Contents

Executive Summary **2**

How Should This Toolkit Be Used 2

Who This Toolkit Is For 3

Funding Acknowledgement 3

1. The Why **6**

1.1 | Motivation for Linking Data to Understand Disability and Health 6

1.2 | Policy Context and Drivers 7

1.3 | Ethics and Responsible Use 8

2. Understanding Existing Data **10**

2.1 | Disability Definitions and Classification Systems 10

2.2 | How Disability Appears in Different Datasets 12

3. Deciding Whether to Link **19**

4. Planning for Data Linkage **21**

4.1 | Governance, Representation, and Partner Engagement 21

4.2 | Privacy, Ethical, and Legal Considerations 26

4.3 | Defining Purpose and Scope 28

4.4 | Defining the Question 29

4.5 | Identifying and Assessing Data Sources 31

4.6 | Preparing Data for Linkage 33

5. Conducting Data Linkage **36**

5.1 | Selection of Linkage Variables 36

5.2 | Choosing a Linkage Method 37

6. Evaluating and Documenting Linkage Quality 39

7. Interpreting Results and Applying Findings 43

8. Addressing Common Challenges 47

9. Designing a Sustainable, Long-Term Linkage System 49

9.1 | Infrastructure Needs 49

9.2 | Maintenance and Updating Procedures 51

9.3 | Long-Term Partner Engagement 51

Final Thoughts 56

Appendix A
Data Linkage Readiness Assessment 57

Appendix B
Example Data Use Agreement 61

Appendix C
Example Data Linkage Project Documentation 64

Appendix D
Common Agreements and Documentation 65

Appendix E
Available Data Linkage Tools 66

Appendix F
Selecting a Linkage Tool 67

Bibliography 68

1 The Why

1.1 | Motivation for Linking Data to Understand Disability and Health

More than one in four Americans has a disability, and people with disabilities experience persistent disparities in health outcomes and access to care. Improving disability data is fundamental for identifying disparities, informing public health action, and improving outcomes across populations. Although health systems routinely collect large amounts of information about health outcomes, healthcare utilization, and service delivery, disability is not consistently identified across those systems. As a result, people with disabilities often remain invisible within otherwise robust health datasets. Disability is complex and multidimensional, reflecting the interaction between health conditions, individual functioning, and environmental, social, and structural factors. Together, these challenges contribute to gaps in understanding that data linkage can help address.

Importantly, people with disabilities are not a homogeneous population. Health outcomes, access barriers, and service needs vary substantially across disability subgroups, including differences by disability type, severity, onset, communication needs, and support requirements. Disability populations are also diverse with respect to age, race and ethnicity, gender, geography, socioeconomic status, and when disability was acquired, all of which may influence health experiences and outcomes. Research has demonstrated that aggregating disability populations can obscure meaningful differences in health experiences and reduce the effectiveness of policy and program responses. Such data are crucial for understanding the impacts of public health issues on people with disabilities and for developing, evaluating, and improving public health programs and policies.

Disaggregated data by disability status and type are essential for identifying health disparities and the structural factors behind them. However, today's siloed data landscape, spanning claims, health records, surveys, and public health databases, often leaves valuable disability data disconnected, inconsistently structured, or underused. Further, varying definitions, lack of standards, and inconsistent applications create challenges for public health agencies seeking to monitor disability, compare findings, and make data-informed decisions. Although public health and healthcare systems collect substantial amounts of data, disability information is not consistently included across these systems. As a result, jurisdictions may be able to measure health outcomes, service utilization, or program participation. Still, they may not be able to determine how those outcomes differ for people with disabilities.

Data linkage is the process of connecting records or information across two or more datasets to create a more complete understanding of individuals, populations, services, or outcomes.

Data linkage provides one strategy for addressing this gap by connecting disability information from one source with health information from another. In many cases, the most significant public health questions about disability cannot be answered using a single data source. In public health, linkage is often used to combine complementary information from different systems when no single dataset contains all the information needed to answer a question. Importantly, the challenge is not solely a lack of disability-related data. Many jurisdictions already collect information relevant to disability through surveys, administrative systems, healthcare records, education systems, and public health programs. Data linkage creates an opportunity to bring these existing sources together to answer questions that cannot be addressed through any single dataset alone. By bringing together complementary data, such as service use, health outcomes, and population characteristics, data linkage fills gaps in individual data sources. It provides a more complete understanding of disparities and system performance.

1.2 | Policy Context and Drivers

Policy initiatives increasingly emphasize improving population health, accountability, and the quality and use of data across populations. Healthy People 2030 includes disability-specific objectives and highlights persistent disparities in health outcomes, access to care, and social determinants of health among people with disabilities. At the same time, federal efforts such as the Federal Evidence Agenda on Disability and broader public health data modernization initiatives call for improved data integration, standardization, and cross-system use of data to better inform policy and program decisions. Disability-focused frameworks further emphasize the importance of representation, inclusion, and data for identifying and addressing structural barriers affecting health and well-being.

Accreditation and performance improvement efforts, including evolving standards from the Public Health Accreditation Board, increasingly emphasize data-informed decision-making, health equity, community engagement, and the ability to assess and respond to population health needs across groups. These policy drivers reflect a growing recognition that existing data systems are often fragmented and do not fully capture the experiences of people with disabilities. As a result, jurisdictions may lack the information needed to identify disparities, evaluate programs, or allocate resources effectively.

Data linkage efforts help jurisdictions strengthen these capacities by improving the completeness, relevance, and usability of data for assessment, planning, and evaluation. Data linkage is one approach jurisdictions use to meet these policy priorities. By connecting data across systems, such as health care, social services, and community-level data, jurisdictions can better understand population needs, improve program coordination, and generate evidence to inform policy and program decisions.

1.3 | Ethics and Responsible Use

A foundational principle of this toolkit is understanding that disability is not inherently a health deficit. People with disabilities can and do experience good health, but they often face barriers related to access, environment, and systems that contribute to disparities. Studies consistently report worse health outcomes for people with disabilities, reflecting substantial health disparities. Disability is shaped by the interaction between individual health conditions, functional status, and environmental or structural factors such as accessibility, policy, and social support. Because disability is experienced differently across populations and contexts, decisions about how disability is measured, linked, and interpreted can influence who is represented in the resulting data and whose experiences remain unseen.

Applying a thoughtful and inclusive approach to data linkage means examining not only what data is available, but also who is represented, who is missing, and how structural factors shape both. Without careful attention, linkage efforts may unintentionally reinforce existing disparities by underrepresenting marginalized populations or mischaracterizing their experiences. Although there are numerous benefits to data linking, there are also limitations, potential risks, and costs. Considering limitations and risks is part of responsible data stewardship and helps program staff be mindful of issues and plan mitigation strategies. Meaningful engagement with people with disabilities and disability-serving organizations can help identify potential harm, improve interpretation, strengthen trust, and ensure that linkage efforts reflect community priorities and lived experience.

Potential risks of data linkage include misinterpretation of findings, particularly when disability is inconsistently defined; unauthorized or unintended disclosure of sensitive information; and the potential for harm if data are used in ways that reinforce stigma, exclude certain populations, or inform inequitable policies. The legacy of unmitigated risks continues to contribute to disability stigma and to persistent physical and procedural barriers to accessing care. These impacts may be further compounded for people with disabilities who are also members of other historically marginalized groups, including communities of color, LGBTQI+ individuals, and older adults, who may experience overlapping forms of discrimination,

Disability plays multiple roles in data analysis and interpretation:

- Determinant: Disability status influences access to care, exposure to risk, or health outcomes.
- Outcome: Disability results from injury, chronic disease, or environmental conditions.
- Analytic construct: Disability is used to define populations for surveillance, program evaluation, or analysis.

exclusion, and barriers to health and healthcare. These risks highlight the importance of careful planning, clear documentation, and ongoing partner engagement, including people with lived experience, throughout the linkage process.

It is also necessary to understand that disability is not a single, fixed concept in data analysis. Depending on the purpose of the analysis and the data sources used, disability is defined and applied differently. These differences affect how populations are identified, how results are interpreted, and how findings are used to inform public health action.

In the context of data linkage, be explicit about how disability is being used. Linking data across systems often brings together datasets that define or measure disability differently. Without clearly defining the role disability plays in data linkage activities, results may be misinterpreted, or comparisons across datasets may be misleading.



2 Understanding Existing Data

Prior landscape work found that data are highly dispersed across systems and vary widely in how disability is defined and captured. These gaps limit programs' ability to fully understand health needs and disparities, underscoring the value of thoughtful data linkage. This section introduces common frameworks for measuring disability and describes how different data sources capture disability in distinct ways.

Definition and measurement of disability are central to public health surveillance and data linkage efforts, yet disability is not a single fixed or universally defined concept. Depending on the purpose of the data collection effort and intended use of the data, disability may be identified through functional limitations, participation restrictions, service eligibility, clinical diagnoses, or the interaction between individuals and their environments. Because disability is shaped by environmental support, policies, social attitudes, and context, different measurement approaches identify different populations and produce varying estimates. Recognizing these differences is essential for interpreting disability data practically and determining whether datasets can be usefully linked.

Disability measurement continues to evolve as public health, clinical, legal, and community perspectives change over time. As a result, disability data systems should be viewed as dynamic rather than fixed, and linkage efforts should remain adaptable to emerging standards, definitions, and community priorities.

2.1 | Disability Definitions and Classification Systems

The methods used to identify disability vary across datasets depending on the purpose of data collection, which influences who is represented in the data and how linked findings can be interpreted and compared. Clearly understanding how disability is identified within each dataset is a key step in planning and conducting linkage activities.

Many frameworks take a functional approach to defining disability, focusing on how policies, programs, and procedures address individual needs. In the United States, there are dozens of disability definitions in federal statute, each tailored to specific policy areas such as employment, education, or health care. For example, someone who qualifies as a person with a disability under the Americans with Disabilities Act may not necessarily meet the Social Security Administration's criteria for disability benefits, even if they experience substantial impacts on their daily life. Further, some federal definitions also reflect the historical context in which they were developed and use outdated terminology or conceptualizations of disability that differ from current public health or disability rights frameworks. These reflect different purposes and result in the identification of different populations. Table 1 reviews commonly used disability definitions and measurement frameworks relevant to data linkage.

Table 1. Examples of Disability Definitions and Measures

Developmental Disabilities Act Categories of functional limitations (e.g., three or more significant functional limitations in major life activities).	**American Community Survey Six Question Set** Six yes/no questions used in the American Community Survey to assess disability, focusing on functional limitations in hearing, vision, cognition, walking, self-care, and independent living.	**Washington Group Short Set** Six survey questions on functioning for use on national censuses and surveys. Unlike the ACS-6 yes/no format, the set uses graded response categories ranging from 'No difficulty' to 'Cannot do at all.' Although responses are often dichotomized for reporting purposes, the scale structure allows for a more nuanced assessment of functional difficulty.	**International Classification of Functioning** Uses a biopsychosocial model, considering disability as an interaction between health conditions and contextual factors (both environmental and personal).
Clinical Diagnostic Codes While diagnostic coding systems such as ICD and SNOMED do not directly define disability, they are often used as proxies in administrative and clinical data. Emerging approaches, such as syndromic surveillance definitions recently published by CSTE, have promising potential to measure disability using clinical diagnostic codes.	**Equal Employment Opportunity** Defines disability in employment contexts as a physical or mental impairment that substantially limits one or more major life activities, typically captured through employee self-identification or workplace reporting systems.	**Americans with Disabilities Act** Defines disability as a physical or mental impairment that substantially limits one or more major life activities, a history of such an impairment, or being regarded as having such an impairment; this broad, rights-based definition is used to support civil rights protections and reflects accommodation or service-related data.	**Social Security Disability Insurance** Defines disability as the inability to engage in substantial gainful activity due to a medically determinable impairment expected to last at least 12 months or result in death; this definition is used for program eligibility and is typically captured in administrative data, representing a subset of individuals with more severe or work-limiting disabilities.

Looking Ahead

Researchers, federal agencies, and disability advocates continue to explore ways to improve disability measurement. Emerging work emphasizes the importance of developing measures that are more inclusive, equitable, and responsive to the diversity of disability experiences while maintaining comparability across data systems.

Disability measurement continues to evolve, and no single approach fully captures the diversity and complexity of disability experiences. These frameworks reflect different purposes and approaches to identifying disability, which result in different populations being captured across systems. Measurement approaches translate these concepts into specific questions, eligibility criteria, diagnostic codes, or functional assessments. As a result, disability estimates and classifications vary across datasets, even when they appear to measure similar concepts.

2.2 | How Disability Appears in Different Datasets

Disability data are crucial for understanding experiences, identifying barriers, and informing policy. Each data source provides a partial view of disability and health and introduces its own types of bias, making it essential to understand these differences for interpretation. It is also important not to overinterpret limitations, as these data remain integral for capturing lived experience and are often the only source of population-level disability information. In addition, some commonly used public health datasets exclude or underrepresent subpopulations, including institutionalized populations, people living in long-term care facilities, or other congregate settings. As a result, linked datasets may not fully reflect the experiences or health outcomes of all people with disabilities. Table 2 summarizes common data sources used in analyses, along with their strengths, limitations, and typical uses. Although this toolkit focuses primarily on public health surveillance and health-related applications of data linkage, many additional data sources can provide important information about the interaction between disability and health. Education, employment, housing, transportation, social services, and other administrative datasets may offer valuable context for understanding health outcomes, access to services, and the broader factors that influence well-being among people with disabilities.

Table 2. Common Sources of Disability Data

Data Source	Disability Measurement	Strengths	Limitations	Typical Use
Surveys	Self-reported functional limitation	Captures lived experience, functional status	Proxy reporting, survey context, limited clinical detail	Prevalence, disparities
Administrative/ program	Eligibility categories	Large populations, service use	Limited disability definition, eligibility bias	Access, utilization
Clinical/claims	Diagnosis codes, service use	Detailed diagnoses, longitudinal	Proxy for disability, access bias	Health outcomes, costs
Linked data	Dependent on data source	More complete picture	Complex, resource-intensive	Policy, system-level insights

While Table 2 provides a high-level overview of common data sources, each type has nuances that affect how disability is measured and interpreted in practice. Some datasets rely on self-reported functional measures, while others identify disability through program eligibility, diagnostic codes, service use, or administrative criteria. The following sections provide additional detail on these data sources.

No disability measure captures the entire disability population.

Every disability measure reflects the purpose for which it was created. Survey measures, eligibility definitions, diagnostic codes, and administrative indicators identify different populations and may miss important disability experiences. As a result, estimates derived from different measures should not be assumed to be directly comparable, and linked analyses should carefully document how disability was identified and the implications for interpretation.

Survey Data

Survey data offers direct insight into individuals' experiences. These surveys, conducted at the national or local level, focus narrowly on disability status or include broader questions about health and social participation. Survey data are valuable for capturing lived experience and functional status, including among populations who may not be engaged in formal service systems. These data are often used to understand prevalence and disparities at the population level. For this data, interpretation should consider who completed the survey or form, how questions were asked, and whether responses reflect self-report or proxy reporting. While survey data relies on self-reports, many instruments are standardized and widely validated, making them an integral source for understanding population-level disability and lived experience.

An example is the ACS, which assesses six domains of functional difficulty: hearing, vision, cognition, mobility, self-care, and independent living. Such data quantify prevalence and reveal disparities across demographic groups. The ACS-6 is among the first successful efforts to standardize disability measurement across multiple federal surveys and has been widely adopted in public health surveillance.

American Community Survey Six-Question Set (Yes/No)

- Are you deaf or do you have serious difficulty hearing?
- Are you blind or do you have serious difficulty seeing, even when wearing glasses?
- Because of a physical, mental, or emotional condition, do you have serious difficulty concentrating, remembering, or making decisions? (5 years or older)
- Do you have serious difficulty walking or climbing stairs? (5 years or older)
- Do you have difficulty dressing or bathing? (5 years or older)
- Because of a physical, mental, or emotional condition, do you have difficulty doing errands, such as visiting a doctor's office or shopping (15 years or older)

Administrative Data

Administrative data are information collected primarily for operational, programmatic, or administrative purposes, such as billing, eligibility determination, service delivery, or record keeping, rather than for research or surveillance. These data can often be repurposed to support public health assessment, evaluation, and data linkage activities. Examples include Medicaid or Medicare claims, hospital discharge records, enrollment systems, and other operational databases.

Administrative data provides broad population coverage and detailed information on healthcare utilization, service patterns, costs, and interactions with public systems over time. Because they are collected continuously as part of routine operations, they are often valuable for longitudinal analyses and large-scale population monitoring.

However, administrative data are generally not designed specifically to measure disability. Disability is identified indirectly through diagnostic codes, service use patterns, or eligibility indicators, resulting in incomplete or inconsistent identification. These datasets also reflect patterns of healthcare access and system interaction rather than underlying disability prevalence or lived experience.

Case Study

Identifying Disability in Administrative Data

A Maine project demonstrated that disability may not be captured through a single dedicated disability variable. Instead, researchers developed an intellectual and developmental disability identifier using diagnosis codes across multiple administrative datasets. The project highlights how disability information may be embedded within healthcare records and how different identification approaches can influence who is included in analyses.

Program Data

Program data are information collected through the implementation and operation of specific programs or services to support service delivery, monitoring, reporting, quality improvement, or evaluation. These data often include participant characteristics, services received, program outcomes, and measures of engagement over time. These data are particularly useful for understanding how people interact with service systems, including patterns of enrollment, service delivery, participation duration, and pathways through systems of care.

Because program data are shaped by program rules, eligibility thresholds, and access to services, they do not reflect a uniform definition of disability. Instead, they capture the subset of individuals who qualify for or can access specific programs. As a result, they may exclude people with disabilities who do not meet the eligibility criteria, are not enrolled in services, or face barriers to accessing services.

Program data are especially valuable for evaluating service delivery systems, identifying gaps in access or utilization, and examining the impact of policies or programs over time. However, differences in eligibility rules and program structures across jurisdictions limit comparability and should be considered during interpretation.

Clinical or Diagnostic Data

Clinical data often identify disability indirectly through diagnostic codes, procedure codes, utilization patterns, or documented functional needs. Clinical, diagnostic, and claims/billing data provide detailed information on diagnoses, healthcare encounters, procedures, treatments, and service utilization over time. As a result, these data are particularly useful for examining health outcomes, healthcare utilization, costs, care coordination, and longitudinal patterns of health and service use. Because clinical data are collected routinely and often cover large populations, they can provide valuable insights into population health trends and support linkage across healthcare systems and administrative datasets.

However, disability is often inferred from diagnostic or utilization data rather than directly measured, and these datasets reflect patterns of healthcare access and interaction with healthcare systems rather than underlying need. They typically provide limited information about functional status, participation, environmental barriers, or lived experience. Coding practices also vary across providers, health systems, and billing environments, which can affect how disability is documented and identified.

Emerging Approaches

Recent work in syndromic surveillance has developed and validated disability-related indicators with input from disability community members and subject matter experts. These efforts reflect growing recognition that disability identification in healthcare data can extend beyond traditional approaches and may support more inclusive disability surveillance and public health monitoring.

Some disabilities may not appear in clinical or claims data if they are unrelated to the reason for care or not needed for billing, treatment, or reimbursement. As a result, the absence of a disability-related code should not be interpreted as evidence that a person does not have a disability. In some cases, disability-related diagnoses may only be documented when relevant to reimbursement or care delivery, resulting in under-identification or incomplete representation. In addition, disability may appear in clinical data as a diagnosis, a demographic characteristic, a functional need, or a health outcome, depending on the context of care.

Disability Reflects both Individual and Environmental Factors

Disability is shaped not only by health conditions or functional limitations, but also by environmental and structural factors such as accessibility, policies, and social conditions. Most existing data systems capture individual-level information but do not fully reflect environmental context, including physical accessibility, policy barriers, social support, and attitudes. As a result, many disability measures identify people at risk of participation restrictions without fully capturing whether environmental accommodations or barriers shape their lived experience. When possible, linking individual-level data with community- or system-level data may provide a more complete understanding of disability and health.

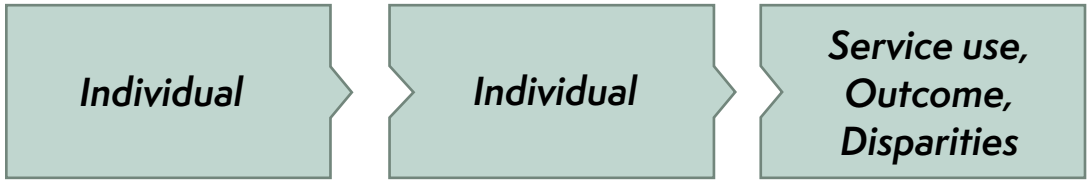
Linked data

Linked data combines information across multiple sources to provide a more complete and multidimensional view of disability and health. By connecting data across systems, linkage supports policy development, identifies service gaps, and generates cross-system insights. However, linkage efforts can be complex and resource-intensive, and results may be affected by linkage errors or underlying biases in the contributing datasets.

> *The challenge is often not a lack of health data, but a lack of disability information within health data systems. Linking datasets can help provide a more complete understanding of disability and health.*

Because no single data system fully captures disability and health, combining multiple sources provides a more complete and nuanced understanding of population needs, disparities, and outcomes. In many cases, linking these data sources addresses individual limitations and provides a more comprehensive understanding of disability and health.

There are two common approaches to data linkage in public health, each with its own strengths and weaknesses. The level at which data are linked affects how findings can be interpreted and generalized. Individual-level linkages support person-specific analysis, while community-level linkages provide broader contextual insights but do not reflect individual experiences. In individual-to-individual linkage, records from different datasets are linked at the person level to examine outcomes such as service use, health outcomes, and disparities across populations.



In contrast, individual-to-community linkage connects individual-level data with area- or community-level information (e.g., neighborhood characteristics or social determinants of health) to provide context for environmental and structural factors that influence health.



In some cases, linkage also occurs across the community level, where data from different systems are combined at a shared geographic level (e.g., county or ZIP code). Community-to-community linkage can be useful for examining patterns across populations, identifying geographic variation, and understanding how system-level factors, such as service availability or policy environments, relate to health outcomes.

Together, these approaches demonstrate how data linkage supports different types of public health questions by connecting individual experiences with broader system and environmental context.



3 Deciding Whether to Link

Not all public health questions require data linkage. Before initiating a linkage effort, programs should assess whether linkage is justified for the intended question, whether suitable data sources are available, and whether the effort is manageable given available resources, data access, timing, and governance considerations. Conducting this assessment early ensures linkage efforts are purposeful, achievable, and likely to provide actionable public health value.

When is linkage needed?

Consider using a **single dataset** when:

- The dataset contains all key variables needed.
- The population is well-represented.
- The question can be answered within a single program or data system.

Consider **linkage** when:

- No single dataset contains all required variables.
- You need to connect outcomes with exposures or services.
- You are examining disparities across systems.
- You need longitudinal or cross-sector insights.

Neglecting these considerations may lead to unanticipated barriers, an unstructured approach, and low-quality or inaccurate analyses. Figure 1 presents a workflow for assessing whether linkage is necessary, achievable, and likely to add value.

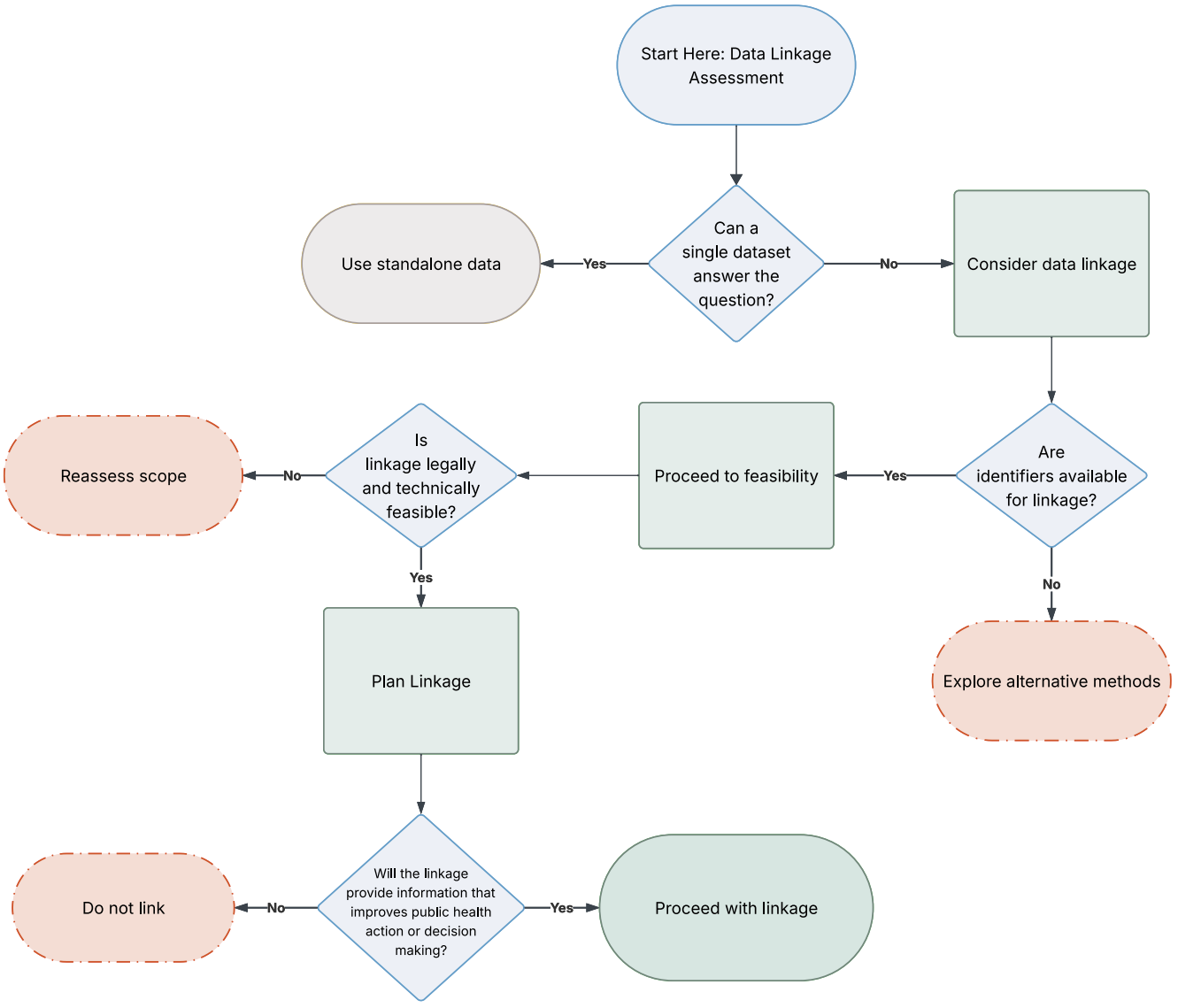
Additional considerations for deciding whether to link should include:

- How quickly is the information needed to support decision-making or public health action?
- Are resources (e.g., staff time, funding, technical capacity) available to support linkage?
- Are there simpler approaches that could produce actionable results more quickly?
- What are the anticipated costs of data access, linkage, and maintenance?

Cost considerations should be evaluated early as data linkage entails expenses such as data access, software, staff, legal agreements, and maintenance, which affect project feasibility and scope.

This toolkit includes a Data Linkage Readiness Assessment (Appendix A), based on the ASTHO Framework for Linking PRAMS with Administrative Data, to support this preparation. This tool consists of questions designed to help teams evaluate their readiness for a linkage project. The questions focus on existing infrastructure, capacity, leadership support, expertise, and data access. Users assess overall readiness and identify areas for development. The assessment underscores the key components of successful linkage projects and provides a way to evaluate them, identify potential gaps, and address areas needing attention before initiating data linkage. Once readiness and feasibility are established, the next step is to define the linkage's purpose and scope.

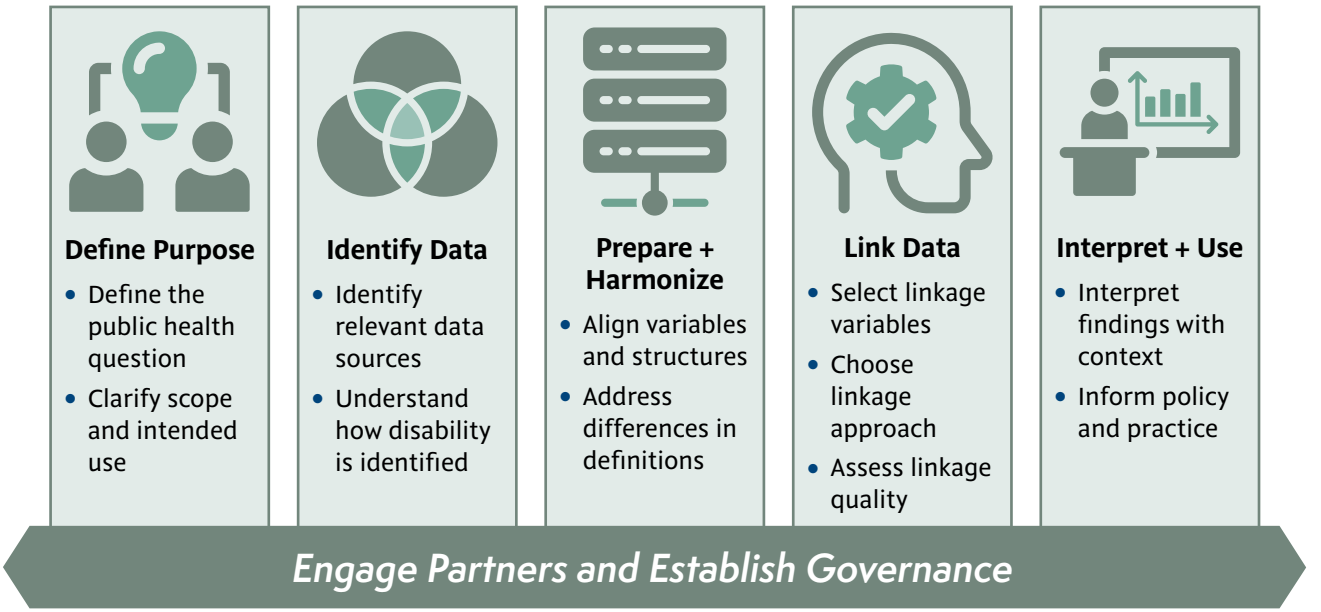
Figure 1. Data Linkage Assessment Workflow



4 Planning for Data Linkage

This toolkit does not prescribe a single type of data linkage. Instead, it encourages users to begin with the public health question they are trying to answer and the information needed to support that work, then determine which linkage approach is most suitable. Although the steps below are presented sequentially, data linkage is rarely linear. Teams may revisit earlier steps as new information emerges, such as identifying additional data needs during harmonization or refining questions based on data availability. Successful linkage efforts depend not only on technical infrastructure but also on ongoing engagement with leadership, community partners, data stewards, and subject matter experts throughout the process. Figure 2 illustrates the key steps in the data linkage process that are explored throughout this toolkit.

Figure 2. Planning and Conducting Disability and Health Data Linkage



4.1 | Governance, Representation, and Partner Engagement

Governance is a key component of disability data linkage projects. In the context of data linkage, governance refers to the structures, processes, agreements, and relationships that guide how data are accessed, shared, protected, interpreted, and used. Inclusive governance incorporates both technical expertise and lived experience to support responsible and substantive use of linked data.

Governance considerations span all stages of the data linkage process and are shaped through collaboration among partners. At each step, different decisions must be made regarding priorities, data use, interpretation, and communication.

Key Components of Governance

- **Establish purpose and oversight**
  Define project goals, scope, decision-making structures, and accountability mechanisms.
- **Define roles and responsibilities**
  Clarify who is responsible for data stewardship, linkage activities, analysis, review, and approval processes.
- **Guide data access and sharing**
  Develop data use agreements (DUAs), data sharing agreements, and procedures for managing access to sensitive information.
- **Protect privacy and confidentiality**
  Ensure compliance with legal, ethical, and organizational requirements related to data security and privacy.
- **Support data harmonization and documentation**
  Establish shared definitions, assumptions, and documentation practices to promote consistency and transparency.
- **Promote representative and responsible data use**
  Consider who is included in the data, potential sources of bias, and limitations that may affect interpretation.
- **Guide interpretation and dissemination**
  Establish processes for reviewing findings, incorporating disability expertise, and communicating results responsibly.

Effective governance depends on representation throughout the linkage process to define roles, establish shared expectations, guide decision-making, and address issues related to privacy, data use, and interpretation. Rather than functioning as a separate activity, representation, including people with disabilities, disability-led organizations, and disability advocacy groups, is integrated throughout the project lifecycle and serves as the foundation for developing and sustaining governance structures. Meaningful engagement extends beyond consultation and should create opportunities for people with disabilities to help shape project goals, influence decision-making, identify potential risks and unintended consequences, and interpret findings. Including disability expertise throughout the linkage process can strengthen the relevance, validity, and responsible use of linked data. Whenever possible, people with disabilities and disability-led organizations should have meaningful roles in governance structures, advisory groups, and project leadership rather than serving solely as reviewers or external stakeholders.

When this expertise is not available internally, jurisdictions should actively seek disability expertise throughout the project lifecycle, including from disabled researchers, disability-led organizations, and individuals with lived experience whose perspectives can strengthen study design, interpretation, and dissemination. As illustrated in Figure 2, governance and partner engagement operate alongside each phase of the linkage process rather than occurring at a single point in time.

Disability Data Justice

Disability data should not only be collected and interpreted about people with disabilities but also shaped by people with disabilities. Disability Data Justice emphasizes equitable participation in decisions about what data are collected, how disability is defined, how data are linked and interpreted, and how findings are used. Applying a Disability Data Justice lens helps ensure that data systems support community priorities, avoid reinforcing stigma or exclusion, and promote accountability to the populations represented in the data.

Throughout the project, it is critical to include internal and external partners such as tribal entities, epidemiology centers, content-area experts, individuals with lived experience, medical providers, and non-governmental organizations. Engaging them during development ensures their perspectives and needs are considered, increasing the likelihood that the data used will be relevant. Engaging partners at these points ensures that governance processes are inclusive, transparent, and aligned with both technical requirements and real-world context.

Partner engagement should not be viewed as a single step in the process, but rather as a continuous activity that informs and shapes each stage of the linkage effort, from defining the question through interpretation and use. Building partnerships involves identifying data stewards and establishing open communication. Early conversations should focus on understanding goals and constraints and on demonstrating the benefits of collaboration.

Case Study

Building Partnerships Across Sectors

Montana partners linked survey and Medicaid claims data through collaboration among researchers, state agencies, healthcare partners, economists, and Centers for Independent Living. Bringing together partners with expertise in disability, health services, and policy helped shape the project approach, interpret findings, and ensure results could inform decision-making. The project demonstrates how sustained partnerships can support both data linkage activities and the translation of findings into practice.

Partner engagement should also include identifying and addressing confidentiality concerns specific to disability communities, such as data sovereignty, privacy expectations, and potential risks associated with data use.

Gathering perspectives directly from the disability community is crucial for tailored planning. Engaging disability community representatives and partners in decision-making strengthens interpretation, enhances relevance, and promotes ethical data use. Advancing disability inclusion in public health data requires centering lived experience, fostering collaboration, and investing in inclusive, ethical, and representative practices that support the health and well-being of people with disabilities. Because individuals with disabilities make up over a quarter of the U.S. population and are affected by all public health issues, their inclusion in planning is essential. Excluding them may create barriers and worsen health disparities.

Table 3. Strategies for Partner Engagement

Strategies for Community Engagement
Engage early
Use accessible communication
Compensate participants
Involve the community in interpretation
Share findings back
Identify trusted intermediaries

Continued on next page

Approaches to Practical Engagement:
Disability advisory workgroups or steering committees.
Community listening sessions or town halls.
Public comment periods on data use or dissemination plans.
Surveys or questionnaires to gather community priorities and concerns.
Focus groups with people with disabilities or caregivers.
Partnerships with Centers for Independent Living or advocacy organizations.
Community review of findings, dashboards, or communication materials.
Accessible webinars or virtual feedback sessions.
Participatory interpretation sessions to contextualize findings.
Inclusion of disability organizations in governance or planning meetings.

Examples of disability partners include Centers for Independent Living, protection and advocacy organizations, disability-led advocacy groups, and community-based organizations representing specific disability populations. Disability-serving organizations also play a critical role in translating findings into real-world interventions, ensuring that analyses are actionable. Jurisdictions should also consider compensating community partners and individuals with lived experience for their time and expertise, recognizing that disability expertise and lived experience should not be treated as unpaid consultation.

Additional Resources

- National Disability and Public Health Resource Center for People with Disabilities
- ASTHO: Designing With, Not For: Stakeholder-Centered Approaches to Disability Health Data
- Oregon: Race, Ethnicity, Language, & Disability (READL) Implementation Guide

To support inclusive and ethical data practices, teams may consider applying a structured inclusive lens such as the IDEAS framework. This approach, developed within disability-inclusive public health research, emphasizes the importance of engaging people with disabilities, ensuring accessibility throughout the process, and accounting for intersecting identities and structural inequities in the design, linkage, and interpretation of data.

Reflection Questions for Inclusive and Responsible Data Practices (IDEAS)

- **Inclusion**: Are people with disabilities meaningfully included in this process?
- **Dignity:** Does this approach respect autonomy, consent, and lived experience?
- **Equality:** Are disparities being examined and addressed?
- **Accessibility:** Are processes, tools, and outputs accessible?
- **Intersectionality:** Are overlapping identities (e.g., race, disability, geography) considered?

Once governance structures and processes are established, teams can move forward with the technical and analytic steps of data linkage with a clear framework for decision-making and accountability.

4.2 | Privacy, Ethical, and Legal Considerations

Privacy, legal, and ethical considerations are central to all stages of the data linkage process, not just data access. Data linkage requires careful attention to privacy, confidentiality, and ethical considerations. Programs should ensure that data use complies with consent procedures, governance policies, and applicable legal requirements while maintaining transparency and safeguarding sensitive information. Data linkage efforts should also consider accessibility and inclusion in both process and outputs, including ensuring that findings, tools, and dissemination strategies are accessible to people with disabilities.

Projects should establish clear protocols for secure data storage, access controls, role-based permissions, data minimization, suppression practices, and procedures for handling potentially identifiable information. Privacy protections should be revisited throughout the project lifecycle as datasets, technologies, and intended uses evolve.

Additional considerations may be needed when working with rare conditions or small populations, where individuals may be more easily identifiable even in de-identified datasets. In these cases, extra safeguards, data suppression techniques, or aggregation are necessary to protect privacy. This consideration is especially critical in linked datasets, where combining sources increases the risk of re-identification.

Rural Data Considerations

- Small population sizes in rural areas increase suppression concerns and may reduce analytic flexibility.
- Some datasets do not support stable rural estimates annually or at smaller geographic levels.
- Rural disability populations may be geographically dispersed or concentrated near service centers.
- Limited sample sizes increase the risk of misinterpretation or exclusion from analyses.
- Partnerships with rural providers, community organizations, and local agencies can help contextualize findings and improve interpretation.

Additional Resources

- Small Numbers Policies and Data Suppression: What Health Department Attorneys Need to Know
- CSTE Data Release Reference Document
- Less than five is less than ideal: replacing the "less than 5 cell size" rule with a risk-based data disclosure protocol in a public health setting

Evaluation frameworks from public health, informatics, and inclusive disciplines emphasize that ethical data linkage requires more than technical rigor; it depends on strong data stewardship, accessibility, documentation, and accountability. Central to this is ensuring that people with disabilities are meaningfully included in data privacy and ethics efforts. Historically, people with disabilities have often been excluded from data collection or included without informed consent, contributing to gaps in understanding and inequities in services and policy. These considerations also include issues related to data sovereignty, where data governance expectations extend beyond standard privacy and legal requirements. As a result, participation must be approached thoughtfully, recognizing the potential for historical trauma and ensuring that individuals are not coerced or included without safeguards in place.

Transparency is core to building trust and accountability. Clearly communicating how data is collected, used, and protected ensures that individuals and communities understand and have confidence in data practices. Collaborative engagement goes beyond participation alone; it requires building trust, ensuring accessibility, and involving partners in decision-making processes, particularly around how data are interpreted and used. Early engagement with partners and clear agreements regarding data sharing, privacy, and roles help to build trust and facilitate collaboration. Regular updates and transparency promote trust and ensure compliance. Together, these practices ensure that data linkage efforts are not only technically sound but also ethically grounded and responsive to the needs and perspectives of the communities they serve.

Cross-agency data linkage often requires navigating different legal requirements. Formal agreements, such as DUAs or memoranda of understanding, are foundational for defining roles, responsibilities, and data protection expectations. These agreements address privacy concerns, clarify governance, and support long-term collaboration across agencies. A formalized agreement institutionalizes the project, describes expectations and responsibilities, provides legal protection for data sharing, and, among other things, paves the way for future linkage projects. Appendix D provides an overview of common types of formalized agreements used in data linkage projects. When establishing these agreements, an often-overlooked part is future use, reporting, and data sharing of the linked data resources created. These components should be clearly defined and documented, even when an agreement is not required. The process of establishing agreements should not be underestimated and can take multiple months to complete. Depending on the source, different agreements and approvals may be required. Each jurisdiction has unique processes that determine which types of agreements must be signed and approved. Typically, at a minimum, a DUA is required. Appendix B provides a DUA template that jurisdictions can use. Although most jurisdictions have standard language and formats for agreements, this template helps ensure that all elements are included.

Together, these considerations highlight that data linkage is not only a technical process but also an ethical and legal responsibility. Ensuring privacy, transparency, and authentic engagement throughout the process builds trust and supports ethical, equitable use of data. These practices should be integrated into all stages of data linkage, informing decisions about data access, preparation, analysis, and interpretation. These considerations align with broader disability-inclusive research frameworks, which emphasize inclusion, dignity, accessibility, and attention to intersectionality as core components of ethical data practice.

4.3 | Defining Purpose and Scope

A key success indicator is a well-defined linkage purpose. Clearly defining this purpose helps teams choose appropriate activities to address their questions. Teams with a specific purpose and scope face fewer barriers to approaching, executing, assessing, and planning for the sustainability of linkages.



Regardless of whether a comprehensive linked data infrastructure already exists, starting with a clearly defined question provides several key benefits:

- It ensures that data linkage is driven by a specific purpose, helping teams focus on answering impactful questions rather than building linkage for its own sake.
- It narrows the scope of the project by identifying the most relevant data sources, variables, and populations, reducing unnecessary complexity and technical burden.
- It supports more efficient planning and decision-making by clarifying what information is needed, how it will be used, and what level of linkage is reasonable.
- It provides a clear framework for engaging partners, aligning expectations, and communicating results in a way that is relevant and actionable.

Key questions to define purpose:

- What problem are you trying to address?
- What decision will the data inform?
- Why is data linkage necessary?
- What would success look like?

Ultimately, the clarity of the scope and purpose shape the design and direction of a data linkage effort. Starting with a focused purpose allows teams to define a manageable scope, identify data sources, and generate informed results. Because disability can be defined differently depending on the purpose of the analysis, clearly defining the intended use of the linkage effort is vital for selecting relevant measures, populations, and data sources. As capacity, partnerships, and infrastructure develop, linkage efforts can be expanded to address additional questions or broader populations. In this way, defining a clear purpose not only guides initial activities but also supports the long-term growth and sustainability of data linkage efforts.

4.4 | Defining the Question

Defining a clear, actionable question is a necessary first step in any data linkage effort. The question should guide all subsequent decisions, including which data sources to use, whether linkage is necessary, and how results will be interpreted and applied. A well-defined question ensures that data linkage is purposeful and realistic rather than exploration without a clear outcome. It also supports communication with partners, alignment among partners, and efficient use of resources.

When defining a question for data linkage, consider the following:

- **Which population are you interested in?** Define the population clearly, including how disability will be identified or operationalized.
- **What outcomes or exposures are you trying to understand?** Determine which variables are needed and whether they are available in a single dataset or across multiple sources.
- **What decision or action will this analysis inform?** Identify how the results will be used (e.g., for program improvement, policy development, or resource allocation).

Table 4: Common Public Health Questions for Data Linkage

Common use cases for disability data linkage include:
Identifying differences in health outcomes, access, and service utilization across disability populations and between people with and without disabilities.
Examining variation across disability subgroups, including by disability type, severity, or support needs.
Understanding service utilization patterns across systems.
Validating health outcomes using multiple data sources.
Supporting emergency preparedness and response.
Monitoring population health trends over time.
Common types of questions include:
Descriptive questions: What are the health outcomes among people with disabilities compared to those without disabilities?
Comparative questions: Are individuals with disabilities more or less likely to receive specific services or interventions?
Associative questions: How is participation in a program associated with health outcomes among people with disabilities?
System-level questions: How do individuals interact with multiple systems (e.g., healthcare, social services), and where are gaps or barriers?

Public health questions for data linkage often involve connecting information across data. Breaking the question into elements such as population, exposure, outcome, and data helps clarify whether linkage is necessary and which data sources are required.

Common pitfalls include asking overly broad or vague questions, starting with available data rather than the question at hand, failing to clearly define how disability is measured, overlooking feasibility, or trying to answer multiple unrelated questions in a single linkage effort. Once the question has been clearly defined, the next step is determining whether data sources exist to support the planned analysis.

Case Study

Examining Disability and COVID-19 Mortality

Researchers at the University of New Hampshire sought to understand whether counties with higher disability prevalence experienced higher rates of COVID-19 mortality. Because disability status was not available in mortality records, answering the question required combining multiple datasets, including American Community Survey disability estimates and county-level mortality data. Clearly defining the question at the outset helped determine which data sources were needed and guided the overall analytic approach. This example illustrates how a well-defined question can shape every subsequent decision in a linkage project.

4.5 | Identifying and Assessing Data Sources

Identifying and assessing data sources is a core step in determining whether linkage is practical and relevant. This process involves understanding what data are available, how disability is captured, and how well the data align with the question.

Guiding Questions for Data Source Selection

- What disparities are you trying to identify or measure?
- What populations are included or excluded in each dataset?
- How is disability defined and operationalized?
- What outcomes or services are captured?
- What time period is available?
- What geographic coverage is available?
- Who is missing or underrepresented in this dataset?
- What biases might this data introduce into the analysis?
- How complete, timely, and reliable are the data?
- Are there shared identifiers or variables that enable linkage across datasets?
- What access, governance, or privacy constraints affect use of this dataset?
- How well does the dataset support the intended analysis based on factors such as completeness, consistency, timeliness, population coverage, and availability of key variables?

Considering the original purpose of data collection and how it aligns with its intended use is crucial. Data gathered for billing, compliance, or operations may lack variables necessary for research or policy analysis. Secondary uses of data can provide valuable insights. Still, users should consider how the data were originally collected, document any limitations, and engage stakeholders to ensure the data are used and interpreted appropriately.

Case Study

Leveraging Existing Administrative Data Sources

Human Services Research Institute and the Maine Health Data Organization explored whether existing administrative health datasets could be used to identify people with intellectual and developmental disabilities. Rather than creating a new data collection system, the project assessed the strengths and limitations of Maine's All-Payer Claims Database and Hospital Encounter Database. It demonstrated how existing data assets could support analyses of disability and health. This approach highlights the value of evaluating available datasets before developing new data collection efforts.

Differences in data collection methods, eligibility criteria, access to services, and population coverage result in some groups being underrepresented or excluded. Clearly documenting who is included in the linked dataset, how disability is identified, and any known limitations in coverage supports more accurate interpretation and responsible use of findings. In some cases, linking more than two datasets provides additional context and improves the validity of findings. When disability is defined differently across systems, incorporating a third dataset may clarify inconsistencies, strengthen interpretation, and provide a more complete understanding of the population.

Evaluating data systems with structured criteria supports transparent, consistent, and inclusive decision-making. Clear documentation of a dataset's strengths, limitations, and contextual factors improves interpretation and determines whether a dataset is suitable for a specific purpose.

4.6 | Preparing Data for Linkage

Following selection, each data source should be assessed to understand its structure further before it is integrated. The following questions may help jurisdictions understand a source's data structure to inform their data linkage strategy:

- Are the source records event- and/or individual-based?
- Does the source contain a unique identifier for individuals or entities?
- Does it contain multiple records (rows) per individual or entity?
- Does it have identifiers that can be used to link to other data?
- Is the data longitudinal?
- Does the source have legal restrictions that prohibit or limit linkages?

Data quality affects the success of data linkage efforts, making harmonization a critical step in the process. Harmonization involves standardizing, cleaning, and aligning data from different sources to ensure consistency, compatibility, and interoperability. This includes assessing missing identifiers, removing invalid data, and aligning elements such as string types, date formats, and special characters. It also involves standardizing variables across datasets to ensure that units of measurement and coding schemes are consistent.

One dataset records sex as "M/F" while another uses "Male/Female." Harmonization standardizes these values to a common format to ensure consistency across datasets.

Inconsistencies in data, such as missing values, outliers, or conflicting information, need to be resolved through data cleaning. By standardizing the data, it can be transformed to a common scale or distribution, facilitating comparison and analysis across. In addition to standardizing formats and cleaning data, harmonization also involves aligning how key variables are defined and categorized across datasets. In some cases, variables can be directly aligned when definitions and categories are consistent across datasets. When differences exist, recoding may be used to standardize values into a common format, or mapping techniques may be applied to translate categories from one system to another using crosswalks. In other situations, creating broader groupings may facilitate comparability, particularly when one dataset contains more detailed categories.

Case Study

Preparing Administrative Data for Linkage

In Maine, claims and hospital encounter records were linked using a de-identified person-level identifier maintained by the Maine Health Data Organization. Project partners standardized records across datasets, developed a person-level disability flag, and documented inclusion criteria and assumptions. This work demonstrates the importance of data preparation, harmonization, and transparent documentation before conducting analyses with linked data.

Harmonization should also account for changes in disability measures over time. A single dataset may include both legacy and current disability questions, definitions, or response categories. Teams should document when measures changed, assess how those changes affect comparability, and determine whether categories need to be mapped, recoded, or aggregated for analysis.

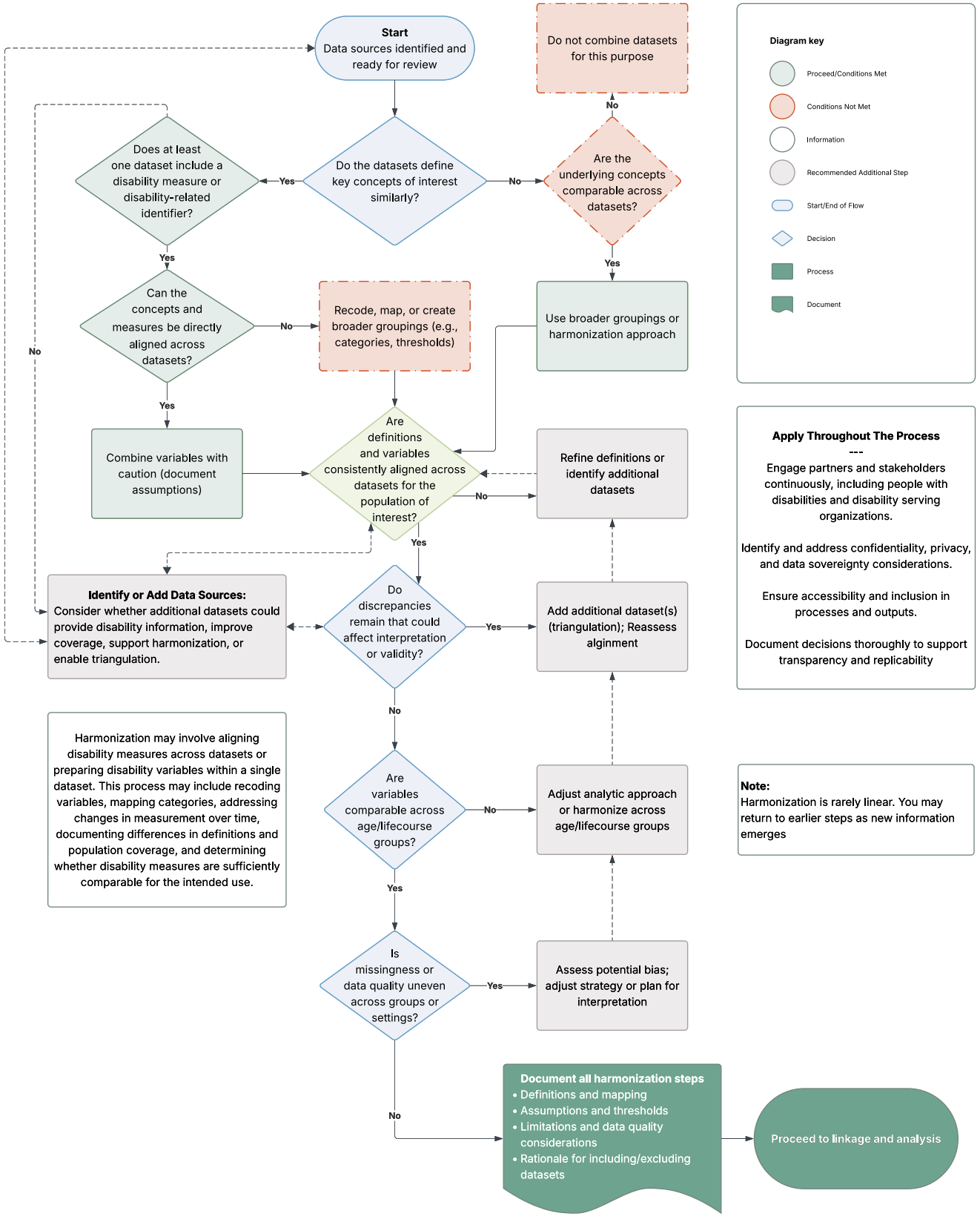
However, not all differences can or should be resolved. In some cases, definitions may be too conceptually distinct to combine without introducing misinterpretation or loss of meaning. In these instances, it may be more suitable to analyze variables separately or reconsider the scope of the linkage. Figure 3 outlines a structured approach to harmonization, guiding users through key decision points for determining how variables can be aligned across datasets.

In some cases, related data elements may provide context for missing or incomplete disability information, such as accommodation requests, assistive technology claims, interpreter requests, service utilization patterns, or open-text responses. These indicators should be used cautiously and documented clearly, as they may support interpretation but do not necessarily provide a complete or definitive measure of disability.

Harmonization is often iterative. As data sets are prepared, teams may identify inconsistencies, missing variables, or definitional differences that require revisiting earlier steps or incorporating additional data sources. It is common to discover that initial data sources are insufficient and that additional data are needed to support impactful analysis.

Documenting the harmonization process supports transparency and reproducibility, and it should include details of data transformations, standardization procedures, and any assumptions or decisions made during harmonization. Additionally, metadata describing the data's characteristics and lineage are important for understanding and interpreting the harmonized dataset.

Figure 3. Data Linkage and Harmonization Workflow



5 Conducting Data Linkage

Once planning and preparation are complete, the next step is conducting the data linkage. In practice, many projects use hybrid approaches combining several of the described methods to balance accuracy and feasibility.

5.1 | Selection of Linkage Variables

Linkage variables determine how records from different datasets are matched and whether they refer to the same individual, household, or entity. The choice and quality of these variables directly affect the accuracy of the linkage process. Poor or inconsistent variables may lead to missed or incorrect matches, introducing bias and affecting how results are interpreted. In some cases, linkage may be conducted using general identifiers (e.g., name, date of birth) rather than disability-specific variables. Disability status may then be identified in one or more linked datasets, enabling broader linkage across datasets that do not explicitly capture disability.

The types of variables used for data linkage depend on the level at which records are being linked. For individual-level linkage, direct identifiers, such as full name, date of birth, Social Security Number, or address, uniquely identify individuals and often support more accurate matching. However, access to these variables is often restricted by privacy, legal, or governance constraints. As a result, many linkage efforts rely on indirect identifiers (quasi-identifiers), such as partial names, birth year, ZIP code, or sex, which do not uniquely identify individuals on their own but can be used in combination to establish likely matches.

For community- or area-level linkage, identifiers are typically based on geographic or organizational units rather than individuals. These include variables such as ZIP code, census tract, county, or facility identifiers. While these linkages do not require personal identifiers, they depend on consistent geographic definitions and alignment across datasets, as with individual-level linkage, differences in how these variables are defined or recorded affect the accuracy and interpretability of results.

Selecting linkage variables requires careful consideration of both the data and the broader context in which the linkage will occur. At a minimum, variables must be available across all linked datasets. Even when variables are present, differences in naming conventions, formats, or coding affect their effective use. Data quality is also a key factor. Variables with high levels of missingness, inconsistency, or error reduce the reliability of the linkage.

Another consideration is how well a variable distinguishes among individuals. Variables with greater discriminatory power, such as the full date of birth rather than the year of birth alone, improve match accuracy, particularly when combined with other variables. At the same time, some variables, such as address or last name, change over time, complicating linkage, especially with longitudinal data.

In practice, selecting linkage variables often requires balancing accuracy and feasibility. Using more or higher-quality identifiers improves matching accuracy, but these variables may be inaccessible due to governance or legal restrictions. Conversely, using more readily available variables requires more complex linkage methods and additional validation to ensure reliable results.

Several common pitfalls undermine linkage efforts. These include selecting variables that are not present in all datasets, relying on variables with poor data quality, failing to account for formatting differences, or using too few variables to distinguish individuals confidently. Even when overall match quality appears high, assess whether linkage errors or missing data are systematically distributed across populations, as this can introduce bias into the results. Engaging governance and legal partners early, particularly when working with identifiable data, ensures that variable selection aligns with DUAs and privacy requirements.

As a practical starting point, users can identify all potential linkage variables across datasets and then assess each variable for availability, quality, and usefulness. In most cases, matching on multiple variables yields more reliable matches than relying on a single identifier. Thoughtful selection and preparation of linkage variables substantially improve both the quality of the linkage and the credibility of the resulting analyses.

5.2 | Choosing a Linkage Method

Data linkage can be performed using deterministic, probabilistic, or hybrid approaches, depending on the availability of identifiers, data quality, and project goals. Choosing the data linkage method depends on several factors, including the characteristics of the data to be linked, the linkage task at hand, available resources, and the desired level of accuracy. Identifying and selecting the best linkage method requires a clear understanding of the data structure and quality, the types of identifiers available, and the dataset sizes needed to estimate the computational resources required.

Deterministic linkage relies on predefined criteria to match records using exact or highly similar identifiers, such as names and dates of birth. Probabilistic linkage, by contrast, assigns weights to variables and calculates the likelihood of a match based on their similarity across records. Machine learning methods use algorithms to learn patterns and relationships in the data, enabling more flexible and adaptive linkage strategies.

Table 4. Comparison of Deterministic and Probabilistic Linkage Methods

Method	Example
Determinist	Matching records using exact SSN and DOB
Probabilistic	Matching based on name similarity + DOB
Machine Learning	Algorithm identifies likely matches across datasets.

Pilot testing is often conducted to evaluate these methods and refine them iteratively based on the results. Which linkage method to employ depends on a variety of factors, some scientific and others more subjective.

Deterministic record linkage links records across datasets using exact matches or predefined rules, without considering the probability of a match. It is often used when exact matches or specific criteria reliably identify true matches, such as merging datasets with unique identifiers or linking records within a single database. While deterministic linkage can be faster and simpler than probabilistic methods, it may overlook potential matches that do not meet the predefined criteria, leading to high specificity but lower sensitivity and potentially missing key relationships between records.

Probabilistic record linkage assesses the probability that two records belong to the same individual or entity. Similarities between attributes are used to calculate match probabilities. Probabilistic linkage methods are particularly useful when dealing with datasets that contain errors, inconsistencies, or missing information and have only partial identifiers. By considering match probabilities rather than relying solely on exact matches, probabilistic linkage can handle variations in data quality and account for potential discrepancies. Additionally, probabilistic linkage allows flexibility in the matching process, as weights can be adjusted and thresholds tailored to specific requirements.

Machine Learning Record Linkage (MLRL) is a modern, evolving approach to linking records across datasets using machine learning techniques. MLRL offers several advantages over traditional linkage methods, including automated linkage, scalability to large datasets, and the ability to handle complex data structures and relationships. However, MLRL also requires careful consideration of feature selection, model training, and evaluation to ensure accurate and reliable linkage results. It also may require increased staff training to ensure the process's long-term sustainability.

Multiple tools (Appendix E) are available to support data linkages. Choosing the best approach and tool (Appendix F) depends on factors such as purpose and specific requirements, budget, technical expertise, scalability needs, and the nature of the data being worked with. Linkage software should make it easy to obtain both matching and non-matching data and provide a mechanism to review and reconcile potential matches. Spend time understanding the underlying population demographics, distributions, and how different tools and algorithms perform with different naming structures.

6 Evaluating and Documenting Linkage Quality

Once the linkage is complete, evaluation ensures the accuracy and reliability of the chosen method, while thorough documentation and reporting provide transparency and accountability throughout the process. Continuous monitoring facilitates ongoing improvement and optimization of the linkage process, ensuring the integrity of linked data for research, analysis, and decision-making.

Evaluating linkage quality is not only a technical exercise but also an analytic and inclusive consideration. Linkage errors may not be evenly distributed across populations, potentially introducing bias into results. For example, individuals with disabilities may be less likely to be accurately linked if disability is inconsistently defined across datasets, if identifiers differ across service systems, or if they receive services from multiple providers that record information differently. Therefore, evaluating linkage quality should include both overall performance metrics and subgroup-specific assessments. This approach to evaluation promotes the accuracy and usefulness of the resulting dataset.

Linkage errors occur when records are incorrectly matched or missed. Measuring and validating linkage quality typically requires multiple approaches, such as developing training datasets and conducting validation across multiple data sources.

Table 5. Standard Error Reporting Measures

Area	What to Assess	Key Actions
Error Documentation	Identification of missed and incorrect links	Document all errors, including missed matches and false matches
Error Correction	Processes for resolving linkage errors	Establish procedures to correct errors and rerun linkage as needed
Error Rates	Frequency and types of linkage errors	Measure and report error rates, including missed and incorrect links
Data Completeness	Retention of records and key variables after linkage	Assess whether linked data maintains expected coverage and completeness
Data Consistency	Alignment of linked data with expectations	Check for logical consistency and alignment across datasets

Commonly used metrics for assessing the quality of data linkages include:

- **Precision:** The proportion of correctly linked records out of all linked records. High precision means fewer false positives (incorrect links).
- **Recall:** The proportion of correctly linked records out of all possible linked records. High recall means fewer false negatives (missed links).
- **F1 Score:** The F1 score is the harmonic mean of precision and recall. It provides a single metric indicating better data linkage quality.
- **Confusion matrix:** A confusion matrix provides a detailed breakdown of true positives, true negatives, false positives, and false negatives.
- **Record-level comparison:** This method involves comparing linked records against a gold standard or a subset of manually reviewed records.
- **Sample-based review:** Identifies issues that automated methods miss by manually reviewing a randomly selected subset of linked records to assess data quality.
- **Visualizations:** Visualizations include histograms, scatter plots, or linkage graphs and can assist in identifying patterns or anomalies in the linkage process.

All evaluation metrics have limitations. It may be important for users to consider the project's intent and linkage methods used when selecting an approach to evaluation. Regardless, all methods should retain information on linkage-match probabilities for each match, enabling researchers to refine and interpret their analyses. Additional measures, such as sensitivity, specificity, positive predictive value, and negative predictive value, may also be used, depending on the linkage context and available validation data.



Determining what level of linkage quality is acceptable depends on the purpose of the analysis and the potential impact of errors. To determine what level of quality is "good enough," teams must consider the project's specific context, including the data's purpose and the potential consequences of errors. Balancing accuracy of a linkage with acceptability is important and can save time and effort. Consider whether small adjustments to improve linkage will result in meaningful gains in accuracy. The acceptability of data linkage quality varies by several factors, and there is no one-size-fits-all answer. Consider the following factors when determining what level of quality is acceptable:

- **Context and Purpose:** The intended use of the linked data is a crucial factor. In applications such as healthcare, high-quality linkage is critical because of potential legal, ethical, or financial consequences. Each linkage project should also assess whether results are "good enough" for the intended purpose, which can be supported through sensitivity or extreme-case analyses.
- **Tolerance for Errors:** Consider the impact of false positives and false negatives on your analysis or decision-making. Some applications have a low tolerance for errors, while others tolerate a higher error rate. Tools such as the Pareto distribution may be used to establish linkage thresholds at which successful matches are statistically unlikely across all possible combinations.
- **Resource Constraints:** Assess the resources available for data linkage, including time, budget, and personnel. In some cases, starting with a simpler or lower-quality linkage approach is justified if resources are limited and there is a plan to refine the process over time.
- **Risk Assessment:** Conduct a risk assessment to understand the potential consequences of data linkage errors, including their impact on both the data and the inferences drawn. This assessment should consider both conceptual and analytical implications. Evaluating linkage quality across subgroups helps identify potential misclassification.
- **Continuous Improvement:** Consider the feasibility of ongoing monitoring and improvement of data linkage quality. Starting with a lower-quality linkage process is acceptable if there is a plan for continuous improvement. Regularly reviewing, reassessing, and refining linkage processes ensures that quality continues to meet evolving needs and expectations.

A hallmark of successful linkage projects is high-quality documentation. Thorough documentation is essential not only for transparency but also for reproducibility, allowing future analyses to replicate or build upon linkage methods. Documenting the linkage process, decisions made, and version control (Appendix C) ensures that the methods and approaches used can be critiqued, communicated, and replicated. Documentation of the linkage process should include the following:

- **Brief project overview:** Describe the project's purpose and goals and explain why data linkage is necessary.
- **List and description of data sources:** List all data sources used and considered. Describe the format, structure, size, included identifiers, years available, who owns or stewards the data, limitations on use, and population coverage. Include the paths where the original data are stored locally (if applicable).
- **Overview of data preparation:** Describe the steps taken to prepare and harmonize the data for linkages, and what cleaning, standardization, and transformations were completed. Include any changes to the original data sources or elements required for the linkages.
- **Linkage methods deployed:** Explain the linkage algorithms or methods used to match records, the parameters or thresholds applied, and how they were determined. Describe how potential errors, manual review, or uncertainty in the linkage process were resolved.
- **Matching quality assessment results:** Describe how the quality of the linkage results was evaluated.

Together, evaluation and documentation provide the foundation for transparent, reliable, and reproducible data linkage. By clearly documenting decisions and systematically assessing linkage quality, jurisdictions may build confidence in their data and ensure that results are reasonable for their intended use. No linkage process is perfect. The goal is not to eliminate all uncertainty, but to achieve a level of quality that is appropriate for the intended use while transparently documenting limitations and potential sources of bias.

7 Interpreting Results and Applying Findings

Once the data has been linked and validated, analysis to support the project's original purpose can proceed. In this phase, data elements are selected, coded, and organized to capture the health outcomes of interest. At this point, a range of analytical procedures can be applied to extract actionable insights and metrics. The synthesized results may then be compiled into detailed reports or information products, integrating visualizations to convey complex data in a digestible format. Ultimately, the value of linked data lies not only in generating new insights but also in informing decisions, such as identifying service gaps, targeting interventions, and evaluating program impact.

Case Study

Using Linked Data to Inform Policy and Practice

Researchers in Montana linked self-reported disability information with Medicaid claims data to better understand healthcare utilization, secondary conditions, and associated costs among people with disabilities. The linkage combined information that was not available in either dataset alone, providing a more complete picture of disability-related health outcomes. Findings helped quantify the impact of secondary conditions and informed discussions related to disability health promotion and policy. This example demonstrates how linked data can support decision-making and translate findings into action.



Interpreting linked data requires careful consideration of both data limitations and the broader analytic context. Without context, findings may be misinterpreted. For example, patterns observed in data may reflect underlying population characteristics (such as age distribution) rather than true differences in outcomes. Engage the community and subject-matter experts to identify these issues and improve interpretation. When interpreting linked data, consider the following:

- **Representation and coverage:** Linked datasets may not fully represent all populations. Differences in data collection methods, access to services, and linkage success rates can result in underrepresentation of certain groups.
- **Disability measurement:** Variations in how disability is defined and identified across datasets affect comparability and influence how findings should be interpreted.
- **Linkage quality:** Linkage errors, including false matches and missed matches, can affect results and should be considered when evaluating findings.
- **Within-group differences:** Whenever sample size and data quality permit, analyses should examine variation within disability populations rather than treating disability as a single category. Differences may exist across disability types and intersecting characteristics such as age, race and ethnicity, gender, geography, socioeconomic status, and disability acquisition.
- **Privacy and representation:** Suppression and aggregation decisions should balance privacy protection with the risk of obscuring small or underrepresented disability populations. When possible, teams should engage community partners to determine whether aggregation, broader categories, notation of small counts, or other disclosure-limitation approaches best support both confidentiality and representation.

These considerations underscore the importance of interpreting linked data with caution and transparency. Findings should be presented alongside clear explanations of uncertainty and limitations and should not be generalized beyond the populations represented in the data. At the same time, interpretation should remain focused on generating actionable insights that support programmatic and policy decision-making while acknowledging data constraints.

Community perspectives can strengthen the interpretation of linked data by helping teams understand how findings align with lived experiences, service systems, and local contexts. Engaging people with disabilities and disability-led organizations during interpretation can help identify assumptions, avoid misinterpretation, and ensure that conclusions accurately reflect the experiences of the populations represented in the data.

Case Study

Incorporating Community Perspectives into Data Interpretation

In 2024, the Health and Disability Program at the Massachusetts Department of Public Health analyzed communication preferences among people with disabilities using data from the Community Health Equity Survey. The analysis found that communication preferences varied by disability status and disability type. One finding stood out: among all disability subgroups, deaf and hard-of-hearing respondents reported the highest preference for receiving information through radio.

Rather than drawing conclusions based solely on the data, the team engaged members of the Deaf community to help interpret the findings. A community member involved in communications outreach asked whether the data included information on how or when individuals became deaf, noting that people with age-related hearing loss may continue to rely on communication methods they used earlier in life. Although the survey did not include that level of detail, the team returned to the data and found that deaf and hard-of-hearing respondents tended to be older than hearing respondents.

While the analysis could not determine whether age fully explained the result, the discussion provided important context for interpreting the finding. As a result, the team now includes age as a potential contributing factor when sharing and discussing these data. This example demonstrates how community perspectives can strengthen interpretation by identifying contextual factors that may not be immediately apparent from the data alone.

While scientific papers serve as one avenue for disseminating results, their impact within government and policy spheres may be limited without collaboration between researchers and policymakers. Alternative communication channels such as memos, presentations, or fact sheets may be another helpful way to present information. These formats allow for a concise and accessible presentation of findings tailored to the specific needs and interests of policymakers, public health officials, or other government officials. Visual aids and the use of plain language also play a crucial role in facilitating informed decision-making and program evaluation, empowering partners to use evidence-based insights to inform policy formulation and programmatic interventions.

How results are communicated should be tailored to the intended audience and format. For example, detailed descriptions of linkage methods and their limitations can be provided in supporting materials or online resources, while more concise products, such as infographics, may reference these materials rather than include full explanations.

In addition to communicating findings to policymakers and technical audiences, it is important to share results with the communities represented in the data, including people with disabilities. When sharing findings with the disability community, prioritizing accessibility and inclusivity in both format and delivery is paramount. Design communication products to meet diverse access needs, such as using plain language, providing alternative text for visual materials, ensuring compatibility with screen readers, and offering materials in multiple formats or languages. Engaging community partners in the development and dissemination of results further enhances relevance, clarity, and trust. Sharing findings in accessible and useful ways ensures that data linkage efforts not only generate insights but also support transparency, accountability, and community-informed action.

Linked disability data may also support monitoring and evaluation efforts related to accessibility, inclusion, nondiscrimination, and proportional participation requirements under frameworks such as the ADA, Section 504, and public health accreditation activities.

Regardless of format, clearly communicate the data sources, populations included, and key limitations so that audiences appropriately interpret the findings and build on them in future work.



8 Addressing Common Challenges

Data linkage efforts may face a range of technical, governance, and conceptual challenges. These challenges may affect not only whether linkage is achievable but also the quality and interpretability of the results. Recognizing potential barriers early enables teams to plan, engage the right partners, and set realistic expectations for what linked data can and cannot support.

Technical challenges are among the most visible barriers to data linkage and often stem from differences in how data are collected, stored, and maintained across systems. Data systems vary widely in maturity, integration, and capacity. Differences in technological infrastructure, data-sharing practices, and system interoperability create additional barriers to linking and using data effectively. Common technical challenges include:

- **Inconsistent or incomplete identifiers** across datasets, which make it difficult to match records accurately.
- **Data quality issues**, such as missing values, errors, or inconsistent formatting.
- **Lack of interoperability between systems**, including differences in data structures, coding schemes, or platforms.
- Datasets that are **difficult to interpret or use** because of limited documentation, lack of benchmarking resources, or insufficient contextual information.

These challenges affect both the feasibility of linkage and the accuracy of results. Addressing them requires data cleaning and standardization, and, in some cases, more advanced linkage methods.

Data linkage also often involves sharing or combining data across programs, agencies, or sectors, which introduces governance and legal challenges. Common governance and legal challenges include:

- **Data sharing restrictions and approval processes**, which may be time-consuming or complex.
- **DUAs** that require negotiation across multiple entities.
- **Privacy and confidentiality concerns**, particularly when working with identifiable or sensitive data.

These challenges may affect timelines and limit which data elements are eligible for linkage. Early engagement with legal, privacy, and data governance partners is essential to navigating these requirements effectively.

In addition to technical and governance barriers, data linkage efforts must address conceptual challenges regarding how data are defined, interpreted, and used.

Common conceptual challenges include:

- **Varying definitions of disability** across datasets, which complicate comparison or harmonization.
- **Differences in population coverage**, as datasets include different groups or exclude certain populations.
- **Misinterpretation of linked data results**, particularly when limitations or biases are not fully considered.
- Data linkage projects **may take longer than initially anticipated**, with a substantial portion of effort dedicated to data cleaning, harmonization, and governance processes.

These challenges are particularly important in disability and public health contexts, where definitions, measurement approaches, and lived experiences vary across systems. Careful attention to these issues is crucial to responsible analysis and interpretation.

Case Study

Revisiting Data Sources During Analysis

During a University of New Hampshire analysis of county-level disability prevalence and COVID-19 mortality, researchers identified limitations in the mortality dataset originally selected for the project. Specifically, the counties with missing mortality data had a higher rate of disability on average, which could lead to biased conclusions. The team revisited the analysis using a more appropriate data source and reassessed the results. Although this required additional work, the revised approach strengthened confidence in the findings. This experience highlights the importance of remaining flexible and reassessing assumptions as new information becomes available.

While these challenges are common, they are not insurmountable. Many can be addressed by:

- **Early planning and feasibility assessment.**
- **Clear documentation of decisions and assumptions.**
- **Engagement with partners,** including those with lived experience.
- **Iterative testing and validation of linkage approaches.**

A structured, collaborative approach helps teams navigate challenges, improve linkage quality, and ensure that results are relevant and actionable.

9 Designing a Sustainable, Long-Term Linkage System

While many data linkage efforts begin with a focused project or initial use case, their value often extends beyond a single analysis. Over time, linkage can evolve into a sustained capability that supports ongoing monitoring, program evaluation, and policy decision-making. Investing in long-term linkage efforts may enable jurisdictions to develop a more complete and responsive understanding of population health, particularly for people with disabilities whose experiences may not be fully captured by any single data source.

As partnerships, governance structures, infrastructure, and analytic capacity develop, linkage efforts may expand beyond initial projects to address additional public health questions, broader populations, and ongoing needs for surveillance or program evaluation. Establishing a clear purpose and question at the outset guides early linkage activities and creates a foundation for long-term scalability and sustainability.

Sustaining these efforts requires more than technical infrastructure. It involves maintaining strong partnerships, establishing clear governance processes, and ensuring that data are regularly updated, documented, and securely managed. Embedding disability representation and inclusion into these systems promotes relevant, accurate data in the long term. When data systems are connected and representative, agencies are better positioned to allocate resources, tailor interventions, and assess impact over time.

Planning for sustainability also requires early and ongoing leadership support. Aligning linkage efforts with organizational priorities helps ensure that the necessary resources, approvals, and partnerships are in place to support both initial work and long-term expansion. Clearly communicating the purpose and value of linkage promotes continued momentum and supports continued investment.

9.1 | Infrastructure Needs

Sustaining data linkage efforts requires establishing and maintaining reliable infrastructure to support data collection, storage, processing, and analysis over time. This includes not only technical systems but also the policies, workforce capacity, and governance structures needed to support ongoing linkage activities.

Key infrastructure components include secure data environments that support storage and access to sensitive data; standardized data pipelines that enable repeatable, efficient data linkage processes; and interoperable systems that facilitate data sharing and integration across programs and agencies. As linkage efforts grow, automation and scalable computing resources may be needed to support larger datasets and more complex analyses.

Data governance and security infrastructure are also essential. This includes systems and processes for managing data access, tracking data use, ensuring compliance with privacy requirements, and maintaining audit trails. Clear documentation of data flows, storage locations, security measures, and linkage processes supports transparency, reproducibility, and long-term sustainability.

Workforce capacity is another key component of infrastructure. Sustained linkage efforts require staff with expertise in data management, linkage methods, analysis, and governance. Cross-training, documentation, and partnerships can mitigate the impact of staff turnover and ensure continuity of work.

Collaborating with academic institutions or external experts may provide additional technical expertise and support, particularly for specialized linkage methods or system development. These partnerships have the potential to strengthen infrastructure capacity and support long-term sustainability.

Case Study

Leveraging Existing Data Systems

Researchers in Montana leveraged existing survey and Medicaid claims data to address questions about disability and health rather than creating a new data collection effort. By building on established data systems, the project generated valuable insights while minimizing burden on participants and organizations. The effort demonstrates how existing infrastructure can support meaningful disability and health analyses when data are linked appropriately. Investments in data access, linkage capacity, and analytic expertise can expand the value of existing data assets.

The infrastructure describe here align closely with broader public health data modernization initiatives, which aim to improve data interoperability, timeliness, and usability across systems. Investing in modern, flexible data systems supports not only individual linkage projects but also the development of more integrated and responsive public health data ecosystems.

Emerging technologies, including artificial intelligence and machine learning tools, may create new opportunities to support data linkage activities, such as record matching, data quality assessment, harmonization, and analysis. However, these tools should be used thoughtfully and in conjunction with established governance, privacy, and validation processes to ensure transparency, accuracy, and equitable outcomes.

9.2 | Maintenance and Updating Procedures

Developing standardized operating procedures and robust documentation is vital to continuity efforts and enables seamless onboarding of new team members. Adequate documentation ensures the project can be repeated, replicated, critiqued, and communicated effectively. Appendix C provides recommended documentation. At a minimum, documentation should include:

- All research notes/documentation for completed and ongoing projects.
- Specification of challenges and resolutions.
- Location of all raw, processed, and relevant databases or research datasets.
- Updated standard operating procedures (i.e., the linkage methods and routines established).
- Software used, cost, renewals, description of special waivers needed by IT, and experts who can be consulted.
- Any institutional knowledge related to policies, ethics, regulations, deadlines, or other information to mitigate loss of productivity or replication of prior efforts.

Any code developed should be well documented, with paths to the code and data used included in the protocol document. If non-code-based systems are used (e.g., Microsoft Excel), an explanation of how these tools were used, what was performed, and any other information that will facilitate replication should be provided.

9.3 | Long-Term Partner Engagement

Sustaining and scaling data linkage efforts depends on strong, ongoing partnerships. While initial linkage projects may begin within a single program or agency, expanding and maintaining these efforts over time requires coordination across partners who contribute data, expertise, infrastructure, and governance. Engaging partners early and consistently lays the foundation for moving from one-time analyses to more integrated, scalable systems.

As linkage efforts mature, partnerships may enable more advanced and sustained uses of data. For example, jurisdictions may expand from a single-linkage project to longitudinal linkage across multiple datasets and time periods, enabling analysis of trends, long-term outcomes, and the impact of programs or policies over time. These types of scalable linkage efforts require coordinated data sharing, consistent dataset updates, and a shared commitment among partners. Demonstrating the value of these efforts builds leadership support and secures sustained investment.

Case Study

Sustaining Partnerships Across Sectors

A Montana disability and health linkage project brought together university researchers, Medicaid agencies, Centers for Independent Living, CDC partners, and health economists. Each partner contributed unique expertise related to disability, healthcare utilization, policy, and data analysis. These relationships strengthened the project and supported the interpretation and application of findings. They also helped establish shared approaches to data privacy, confidentiality, security, and responsible data use, building trust among partners and supporting long-term collaboration. This example illustrates how sustained partnerships can enhance the quality, relevance, and long-term impact of linkage efforts.

Maintaining strong partnerships also supports the adaptability of data linkage efforts. Regular communication and sharing of findings with partners can help identify emerging priorities, address gaps, and refine approaches over time. Developing strategic partnerships address needs in technology, workforce capacity, data exchange, and legal or policy expertise, particularly as projects grow in scope and complexity.



To support long-term engagement and scalability, jurisdictions may consider:

- **Assessing needs and identifying gaps**.
  - Review existing data modernization initiatives, assessments, and priorities to identify areas where partner support is needed.
  - Engage partners to identify gaps in skills, infrastructure, or resources and co-develop strategies to address them.
  - Focus on partnerships that have the greatest impact on expanding or sustaining linkage efforts.
- **Aligning partnerships with agency goals**.
  - Identify how partnerships support both immediate project needs and long-term public health priorities.
  - Consider key areas such as data sharing, technology infrastructure, workforce capacity, analytics, and governance.
  - Prioritize partnerships that strengthen the ability to scale efforts over time.
- **Establishing and maintaining effective partnerships.**
  - Define roles, responsibilities, and points of contact to support coordination and accountability.
  - Maintain regular communication to ensure alignment and address challenges as they arise.
  - Identify shared goals and clearly articulate the value of the partnership.
  - Develop and maintain a partnership plan to track engagement, activities, and progress over time.

Long-term sustainability and scalability depend on maintaining strong partnerships across agencies, community organizations, and other stakeholders involved in linkage efforts. By investing in ongoing collaboration, shared governance, and coordinated infrastructure, jurisdictions can build more connected and resilient data systems that continue to support public health action over time.

Case Study

Building a Sustainable Public Health Data Warehouse

Massachusetts Department of Public Health: Developing Long-Term Data Linkage Infrastructure

The Massachusetts Department of Public Health (MDPH) developed the Public Health Data Warehouse (PHD), a long-term cross-sector data linkage system designed to support public health assessment, policy development, and program evaluation. Established through legislative authority in 2019 and sustained through interagency partnerships, the warehouse links data from more than 30 person- and community-level data sources, including vital records, healthcare claims, hospital data, behavioral health services, corrections, housing, social services, and public health surveillance systems.

The warehouse did not emerge from a single project. Instead, Massachusetts began by addressing a high-priority public health issue, the opioid overdose epidemic, in 2015 and used early analyses to demonstrate the value of linking data across systems. Early findings identified dramatically elevated overdose mortality among people leaving incarceration, which informed the development and evaluation of medications for opioid use disorder programs in county jails. Demonstrating tangible public health impact helped build support among agency leaders, policymakers, and data partners, creating momentum for continued investment and expansion.

Over time, Massachusetts expanded the warehouse by adding new data partners, refining governance structures, evolving data use agreements, strengthening privacy protections, and establishing mechanisms for ongoing collaboration. The state also developed a Community Advisory Board in 2024 to help guide priorities and dissemination efforts. Sustainability has been supported through braided funding approaches, partnerships with academic institutions and other organizations, and a commitment to ensuring that analyses address real-world policy and program needs.

Today, the Public Health Data Warehouse supports a wide range of public health priorities, including maternal and child health, substance use, health equity, disability and health, COVID-19, and other emerging issues.

Continued on next page

Continued from previous page

| Case Study

Analyses conducted through the warehouse have informed legislative reports, peer-reviewed publications, public-facing data briefs, and programmatic decision-making. Examples include assessments of severe maternal morbidity among people with disabilities, opioid-related overdose mortality among injured workers, and racial and ethnic inequities in maternal health outcomes.

Key Takeaway: Sustainable data linkage systems are built over time through demonstrated value, strong governance, trusted partnerships, flexible infrastructure, and ongoing investment. While not every jurisdiction has the resources to develop a large-scale data warehouse, Massachusetts illustrates how linkage efforts can evolve from a single high-priority project into a long-term asset that supports public health decision-making across programs and populations.



Final Thoughts

Data linkage is a valuable method for enhancing understanding of disability and health, but it must be conducted carefully and ethically, with input from all partners and the community. Emphasizing disability inclusion throughout the process is crucial to ensuring data are used in ways that reflect lived experiences, prevent harm, and enable meaningful action. This involves prioritizing transparency, safeguarding privacy, and involving partners, including people with disabilities, at every stage. Applying representative approaches, such as the IDEAS framework, ensures that data linkage efforts not only generate insights but also reflect the values of inclusion, dignity, and accessibility. Additionally, data linkage efforts need to be sustainable and adaptable, as jurisdictions differ in their systems, resources, and priorities. Many jurisdictions already collect the data needed to better understand disability and health. Data linkage provides an opportunity to connect those existing data sources and generate insights that would not be possible using any single dataset alone.

The strategies in this toolkit are designed to be flexible for local use. Ongoing learning, documentation, and collaboration are key to improving practices over time. Future efforts may include developing training materials, workshops, or implementation resources to support jurisdictions in applying the concepts presented in this toolkit. By adopting inclusive, responsible, and innovative approaches to data, public health systems can better use disability data to guide decisions, address disparities, and promote the health and well-being of people with disabilities.

Appendix A

Data Linkage Readiness Assessment

This readiness assessment was adapted from the ASTHO PRAMS Data Linkage Toolkit and revised to support broader data linkage efforts for understanding disability and health.

	Question	Score (0-5)
1	Does your program have a clear purpose for undertaking a data linkage project? • Are the key objectives well defined? • Will the linked data provide something that isn't already available? • What are the expected impacts of linking these data, particularly for understanding disability and health disparities? • How will the results inform programs, policies, or services for people with disabilities?	
2	Do you have a well-defined project plan? • Do you have a clear timeline? • Do you have specific milestones? • Will the project involve ongoing or single (static) linkages? • Do you have anticipated products or other deliverables? • Have you considered how disability-specific findings will be communicated to partners, including people with disabilities?	
3	Does your program have the necessary technical infrastructure and expertise for data linkage? • Will you require staff training on linkage methods or the software used to conduct linkages? • Do you have a data integration solution, and has it been calibrated to your population? • Is the data linkage solution affordable and sustainable? • Does the data linkage solution require additional IT support? • Will the linkage approach scale to include multiple data sources over time?	
4	Have you identified relevant data sources and addressed their quality and compatibility? • How is disability defined and measured in each dataset? • Do the data have identifiers, and can you access them? • Does the data have a cost? • Are agreed data standards and dictionaries in place, or is there a process to develop them? • Are the data comparable and combinable across iterations? • Do the years of data available coincide? • Do the data cover a jurisdiction-wide population or only a subset? • Do you have data cleaning and harmonization standards?	
5	Can you access and link the data sources? • Do policies and procedures exist for requesting and using the data? • Are there shared identifiers or variables that support linkage? • Are there limitations that may affect linking disability data across systems?	

Continued on next page

Appendix

A

	Question	Score (0-5)
6	Does your agency have an existing data governance framework? • Does the framework cover privacy, security, and confidentiality of data? • Are there legal and regulatory frameworks in place to govern data sharing and data protection? • Is the process documented and clearly defined? • Does the existing data governance framework include data linkages? • Is an IRB required for linkages? • Does governance cover all participating data sources? • Does the framework consider ethical use of disability data, including representation and potential harm?	
7	Do you have key partners engaged? • Have you identified and engaged all required partners? • Have you engaged disability partners? • Have you established roles and responsibilities? • Are people with disabilities meaningfully included in planning or interpretation?	
8	Have you conducted a comprehensive risk assessment for data linkage, including potential privacy and security risks? • Have you identified privacy and security risks? • Do you have a plan to mitigate risks related to data sharing, linkage, storage, and use? • Have you considered risks related to misinterpretation or misuse of disability data?	
9	Have you conducted a pilot study to test the data linkage process? • Do you have metrics to evaluate quality and linkage performance? • Do you have thresholds for acceptable linkage quality? • Do you have methods to assess bias or differential linkage across populations (including people with disabilities)?	
10	Have you established the rules related to analyzing the linked data? • Are there provisions for sharing with internal and/or external partners? • Do they include review and approval processes for products? • Do they address appropriate interpretation of disability data? • Do they include guidance for communicating findings in an accessible and responsible way?	
	Total	

For each question, assign a score from 0 to 5 based on your program's current level of readiness. Use the guidance below to ensure consistent scoring across all domains:

0 – Not Started: There is no evidence of planning, discussion, or activity in this area.

1 – Minimal: There has been initial awareness or informal discussions have occurred, but no formal plans or actions are in place.

2 – Early Development: Some elements are in progress, but efforts are incomplete, inconsistent, or not yet documented.

3 – Moderate: Key components are in place and functioning, but gaps remain in consistency, documentation, or implementation.

4 – Strong: Most elements are well established, documented, and consistently applied, with only minor gaps or areas for improvement.

5 – Fully Established: This area is fully developed, documented, and consistently implemented. Processes are sustainable, regularly reviewed, and aligned with best practices.

Users should consider both formal documentation and actual practice when assigning scores. Scoring should reflect how well disability-specific considerations (e.g., representation, accessibility, and ethical use) are incorporated into each area.

Overall readiness to link data: Total the sum of your scores from each question.

41-50: Advanced Capacity

Your organization demonstrates strong readiness across governance, partnerships, data infrastructure, and implementation planning and is well-positioned to sustain and expand linkage efforts.

Recommended sections:

- Section 6: Evaluating and Documenting Linkage Quality
- Section 7: Interpreting Results and Applying Findings
- Section 8: Addressing Common Challenges
- Section 9: Designing a Sustainable, Long-Term Linkage System

Consider:

- longitudinal linkage
- advanced analytics
- scaling efforts
- mentorship and peer learning

31-40: Ready for Implementation

Many foundational elements are already in place. Your organization appears prepared to begin or expand linkage activities while continuing to strengthen specific areas of capacity.

Recommended sections:

- Section 5: Conducting Data Linkage
- Section 6: Evaluating and Documenting Linkage Quality
- Section 7: Interpreting Results and Applying Findings
- Section 9: Designing a Sustainable, Long-Term Linkage System

Focus on:

- implementation
- linkage quality
- interpretation
- sustainability

Appendix

A

21-30: Early Planning

You have several areas to address and develop that will improve the likelihood of a successful data linkage project. Important building blocks are in place, but additional planning, partnership development, infrastructure, or governance activities may be needed before implementation.

Recommended sections:

- Section 3: Deciding Whether to Link
- Section 4: Planning for Data Linkage
- Section 5: Conducting Data Linkage
- Section 9: Designing a Sustainable, Long-Term Linkage System

Focus on:

- project planning
- governance and partnerships
- data access
- linkage design

<20: Building Foundation

Your organization may benefit from focusing on foundational readiness activities before pursuing linkage efforts. You may benefit from reviewing the entire toolkit, with particular attention to:

Recommended sections:

- Section 3: Deciding Whether to Link
- Section 4: Planning for Data Linkage
- Section 5: Conducting Data Linkage
- Section 9: Designing a Sustainable, Long-Term Linkage System

Focus on:

- identifying partners
- clarifying goals
- understanding available data
- assessing readiness

This assessment serves only as general guidance, as each state or other jurisdiction may have additional factors to consider. This assessment should not replace a comprehensive evaluation.

Domain-Specific Recommendations

Domain	Recommended sections:
Governance and Partner Engagement	Section 4.1 and Section 4.2
Data Sources	Section 2.2 and Section 4.5
Linkage Planning and Implementation	Section 4.6 and Section 5
Linkage Quality	Section 6
Interpretation and Use	Section 7
Sustainability	Section 9

Appendix

B Example Data Use Agreement

The contents below are intended as an illustrative example of common components included in data use agreements for linkage projects. They should be reviewed and adapted in consultation with legal, privacy, and data governance staff.

This Data Use Agreement ("Agreement") is entered into between [*Name of First Agency*], located at [*Address of First Agency*], hereinafter referred to as the "Data Provider," and [*Name of Second Agency*], located at [*Address of Second Agency*], hereinafter referred to as the "Data Recipient."

Effective Date: [*Date*]

Purpose and Scope: Describe the purpose and necessity of the data linkages. This section should typically explain the legal authority for collecting and using identifiable data and clarify that the agreement defines how the data is used, shared, and linked. It should also specify the datasets involved, give a general overview of each, and indicate which elements will be utilized in the data.

Data Access and Use: *Typically, this section describes the following:*

- The Data Provider agrees to grant the Data Recipient access to the specified datasets for the purpose of performing data linkage and analysis.
- Data Recipient agrees to use the data solely for the purpose stated and to comply with all applicable laws and regulations.
- Data Recipient shall not use the data for any unauthorized purpose.
- Either specifically or broadly, individuals who will have access to the data (define who is an "authorized user").

Data Security and Confidentiality: Outline the security measures to protect confidentiality, restrict access, and utilize only the necessary data components. This section may also describe how data will be secured during transmission and storage, specify the tools employed for sharing, and detail the sharing process. Generally, both parties implement responsible steps, use approved technology, and adhere to organizational safeguards. Most agreements explicitly state that "Data Recipient shall restrict access to the data to authorized personnel only," or a similar clause.

Privacy and Legal Compliance: Outline the responsibilities of both parties to comply with all applicable data protection and privacy laws, clearly listing the relevant laws and regulations. Additionally, specify the steps the Data Recipient will take in case of a data breach or unauthorized disclosure. This typically includes procedures for addressing the breach, the notification timeframe for the Data Provider, and the legal actions to be undertaken.

Appendix

B

Data Linkage and De-identification: Describe the methods used for linkage and de-identification, including any tools involved, particularly if they utilize external servers or web-based platforms. This section should generally outline the identifiers used for linking data and explain how the research data will be de-identified to align with organizational policies. If applicable, include language about the potential re-identification of de-identified datasets. If external vendors, cloud-based platforms, or third-party linkage tools are used, the Agreement should specify approved environments, applicable security requirements, and any restrictions related to storage, transmission, or processing of identifiable data.

Data Sharing and Reporting: Explain the process and timing for the Data Recipient to share de-identified and aggregated data analysis results with other parties for research or public health purposes. Typically, shared results or datasets will not include personally identifiable information. Still, any provisions for sharing should be detailed here, such as in an IRB-approved data use agreement signed by all data stewards. Data shared under this Agreement should be limited to the minimum necessary information required to accomplish the approved public health, research, or linkage activities. Agreements may also include requirements related to small-cell suppression, aggregation thresholds, or other disclosure-avoidance practices to reduce the risk of re-identification in shared results. Dissemination products developed from linked data should consider accessibility and usability for intended audiences, including the use of accessible formats.

Ownership and Intellectual Property: Describe the original raw data and the information products derived from the linked data. In state-based agreements, this generally includes provisions for the Data Provider to review these information products, specify the review duration, and establish a process for resolving disagreements. Additionally, this section should outline how the Data Recipient can share de-identified data with other internal or external researchers. Usually, the Data Provider maintains ownership of the original data, with the Data Recipient acknowledging this ownership. Conversely, the Data Recipient retains intellectual property rights in analyses and research conducted using the data. The Agreement should define processes for review and approval of publications, presentations, dashboards, or other dissemination products derived from the linked data, including timelines for partner review and procedures for resolving disputes.

Duration and Termination: Describe the length of the agreement, how it can be terminated, and the conditions for data use after termination. Generally, open-ended arrangements are preferred over fixed-term limits, and clauses for the ongoing use of linked data are incorporated, along with provisions for the deletion of raw data or identifiers used for linkages. The Agreement should specify timelines and procedures for the return, destruction, or continued authorized retention of identifiable data, linkage keys, and analytic datasets following project completion or termination.

Governing Law and Jurisdiction: This section identifies the legal authorities, regulations, and organizational policies governing data sharing, access, and linkage activities under the Agreement.

Amendments and Modifications: This section can specify how amendments will be handled. Any changes to this Agreement must be made in writing and signed by authorized representatives of both parties. Agreements should identify responsible data stewards or points of contact for oversight, approvals, incident response, and ongoing coordination.

Signature:
[*Signature of Authorized Representative of Data Provider*]
[*Date*]
[*Name of Authorized Representative of Data Provider*]
[*Title of Authorized Representative of Data Provider*]

[*Signature of Authorized Representative of Data Recipient*]
[*Date*]
[*Name of Authorized Representative of Data Recipient*]
[*Title of Authorized Representative of Data Recipient*]

Appendix

C

Example Data Linkage Project Documentation

Background [3-4 paragraphs that provide details on the motivation for the linkage project. This should provide enough background to understand the project's purpose.]

Purpose [One sentence statement on the overarching goal of the project.]

Objectives [Clearly stated objectives. This should be as specific as possible.]

Research Questions [Questions should be grouped by anticipated short-, mid-, and long-term analyses.]

Data infrastructure [Brief description (1-2 paragraphs) outlining any prior linkage work and existing infrastructure that will assist or limit the linkage activities.]

Population [Describe the population to be linked, including the specified years and the data that will be used]

Agreements and approvals [A list of all existing agreements and ones that are needed, the process for completing them, key contacts if available, expiration dates, and other relevant information.]

Linkage plan [A short description of how the actual linkages will be accomplished. To support the description, include a simple outline like the following.]

Tools [Short description of the linkage tool used, any source code needed, cost, methods for renewal, or any other relevant information.]

Methods [Briefly describe the expected linkage methods, specifying whether they are deterministic, rule-based, probabilistic, machine learning, or a combination. Since most successful linkages are iterative, this section should be adaptable, starting with your initial approach and then detailing the method used. Include information on manual reviews, threshold setting, data cleaning procedures, and error handling or assessment strategies.]

Validation [This section briefly explains how you will validate the linkages and estimates. It should describe the main tools you'll use to check linkage accuracy, identify false positives and negatives, and evaluate overall correctness. Any previous linkage work can provide valuable insights. After linking the data, you will create a research dataset containing all records and the outcomes of interest from the linked sources. Ensure that no duplicates are created or omitted during this process.]

Analysis and Reporting [Provide a brief overview (1-2 paragraphs) detailing the analyses and statistical tests to be performed, along with how the data and results will be shared. Include the types of information products that will be generated.]

Sustainability [A brief overview of factors for improving or maintaining sustainability should be included in 1-2 paragraphs. This might cover network paths to files and code, approvals, key contacts, and funding notes. Since this section is likely to evolve, it should remain adaptable. Teams need to keep key information for each linked source current and update it annually to support ongoing projects.]

Appendix

D

Common Agreements and Documentation

A brief description of common agreements and other documents relevant to the linkage project is provided below. Not all of the following documents and agreements will be relevant to all projects.

Data Use Agreements (DUAs): Legal contracts that define the terms for accessing, using, and sharing data. They usually cover data security protocols, confidentiality requirements, allowed data uses, and limits on data sharing.

Institutional Review Board (IRB) Approval: Many research projects involving human subjects, including data linkage studies, require IRB approval. Researchers need to submit their study protocols for review to verify that the project adheres to ethical standards and protects participants' rights and well-being.

Data Sharing Agreements: Document how data is shared with external partners or collaborators. They specify the terms of the data exchange, covering data security measures, usage restrictions, and liability clauses.

HIPAA Authorization: Data that includes identifiers might be classified as individually identifiable health information by certain health departments that are covered entities. To use and disclose this linked data, covered entities may need a HIPAA authorization or a valid exemption.

Confidentiality Agreements: These may be required when sharing data with external entities or individuals. Such agreements specify the responsibilities for safeguarding the confidentiality and security of the data and often include clauses for data destruction or return upon project completion.

Data Security Plans: Data security plans outline measures to protect data from unauthorized access, disclosure, or misuse. These plans should address physical, technical, and administrative safeguards, such as encryption, access controls, and employee training, to ensure data security throughout the project lifecycle.

Consent Forms: If data linkage between datasets is not a function already covered in current consent procedures, additional consent from the participants may be required to proceed with a data linkage project. The reviewing IRB or other scientific review committee can provide guidance about whether this is required. Typically, IRBs waive this requirement when obtaining consent is not feasible, and the data is used for research or public health surveillance purposes.

Appendix

E

Available Data Linkage Tools

The resources included in this appendix are intended as examples of commonly used tools for public health data linkage and are not a comprehensive list of available software or resources. Jurisdictions should select tools based on their specific project goals, data environment, technical capacity, and organizational requirements.

Tool	Description
Link King	Link King is a SAS tool that includes a data-importing and formatting wizard, artificial intelligence to ensure the use of appropriate linking protocols, an interface for manual review, the ability to generate random samples of links for validation, and "point-and-click" editing. Link King uses both probabilistic and deterministic algorithms.
Link Plus	Link Plus was created by the Centers for Disease Control and Prevention's Division of Cancer Prevention and Control to assist cancer registries. Link Plus can detect duplicate records and link registry files to external files. It computes probabilistic record linkage scores using value-specific, last name, first name, middle name, date, Social Security number, generic string, or ZIP code matching methods. Missing values in matching variables are automatically treated as null or empty. Link Plus facilitates blocking, comparing pairs with identical values on at least one variable, and has two phonetic coding systems.
Match*Pro	Conducts probabilistic record linkage based on the Fellegi-Sunter model, using tools to conduct data validation using pre-defined or custom validators. The user-friendly interface enables manual review with flexible configurations, including specifying blocking and matching methods, adjusting blocking sensitivity, defining unknown values, setting weights, and performing substitutions.
R: RecordLinkage	Provides functions for linking and de-duplicating datasets with supervised and unsupervised classification. Methods are based on a stochastic approach using an expectation-maximization algorithm. Employs machine learning methods, and thresholds can be determined by tools based on extreme value theory.
R: FastLink	FastLink uses the Fellegi-Sunter probabilistic linkage method with the expectation-maximization algorithm. This method allows for missing data and the inclusion of auxiliary information.
Python: recordlinkage	This toolkit provides most of the tools required for data linkage and deduplication of small and medium-sized files. It includes indexing methods and functions to compare records and classifiers.
Python: Splink	Splink uses probabilistic record linkage to deduplicate and link datasets without unique identifiers.
SQL: PIRL	Point-of-contact Interactive Record Linkage prospectively links data using a probabilistic linkage algorithm based on the Fellegi-Sunter method.

Appendix

F

Selecting a Linkage Tool

The following considerations can guide the selection of a linkage tool.

Define Requirements: Consider factors such as the volume of data, types of data sources, data quality, types of identifiers available in each source, security and compliance requirements by your IT department, and any specific matching algorithms or rules you need.

Flexibility of Matching Algorithms: Look for tools that offer a variety of matching algorithms. Some tools enable different algorithms to be applied to each element and weighted, or allow for other flexible refinements to be implemented. Depending on your data, you may need different algorithms to achieve accurate linkage.

Scalability: Consider the scalability of the tool and your need for the ability to handle increasing volumes of data.

Integration Capabilities: Determine whether the tool can integrate with your existing systems and data sources. It should support the data formats and protocols used in your environment.

Data Quality and Cleansing: Check if the tool provides features for data quality assessment, cleansing, and standardization. Data quality issues can significantly impact the accuracy of linkage results.

Security and Compliance: Ensure that the tool meets security and compliance requirements, especially when dealing with sensitive or regulated data. Look for features such as encryption, access controls, and compliance with standards.

Ease of Use and Maintenance: Consider the tool's usability and whether it requires extensive technical expertise to set up and maintain. A user-friendly interface and comprehensive documentation can simplify the implementation process.

Performance and Speed: Evaluate the performance and speed of the tool, particularly if you have large volumes of data. The tool should be able to process data efficiently without significant latency.

Cost and Licensing: Compare the pricing models of different tools, including upfront costs, licensing fees, and ongoing maintenance expenses. Consider both the initial investment and long-term operational and staff training costs.

Trial and Proof of Concept: Whenever possible, try out the tool through a trial or proof of concept to assess its suitability for your specific use case.

Threshold establishment and manual record review: Review whether the tool has features to support establishing acceptance and review regions. Additionally, review how manual record review is completed and integrated into the final matched results. Disparate approaches result in data degradation, and tools that provide interactive reviews of “possible matches” and reconciliation support help maintain data integrity.

Bibliography

1. Centers for Disease Control and Prevention. Disability Impacts All of Us Infographic. Disability and Health. Published December 10, 2024. Accessed April 27, 2026. https://www.cdc.gov/disability-and-health/articles-documents/disability-impacts-all-of-us-infographic.html

2. Krahn GL, Walker DK, Correa-De-Araujo R. Persons with Disabilities as an Unrecognized Health Disparity Population. *American Journal of Public Health*. 2015;105(S2):S198-S206. doi:https://doi.org/10.2105/ajph.2014.302182

3. Chimienti M, Morlino G, Ndreca D, et al. Barriers to Accessing Healthcare for People with Disabilities: a Systematic Review. *Frontiers in Public Health*. 2026;14. doi:https://doi.org/10.3389/fpubh.2026.1765145

4. National Institutes of Health. Strategic Plan for Disability Health Research. Nih.gov. Published March 26, 2026. https://dpcpsi.nih.gov/disabilityhealthresearch/strategicplan

5. Iezzoni LI. Eliminating Health and Health Care Disparities among the Growing Population of People with Disabilities. *Health Affairs*. 2011;30(10):1947-1954. doi:https://doi.org/10.1377/hlthaff.2011.0613

6. Lundberg DJ, Chen JA. Structural Ableism in Public Health and Healthcare: a Definition and Conceptual Framework. *The Lancet Regional Health - Americas*. 2024;30(100650):100650. doi:https://doi.org/10.1016/j.lana.2023.100650

7Jain A, Chughtai Z, Westermann H. How to Modernize Data Infrastructure. Astho.org. Published June 9, 2025. Accessed April 27, 2026. https://www.astho.org/topic/toolkit/how-to-modernize-data-infrastructure/

8. Office of Disease Prevention and Health Promotion. Healthy People 2030. Healthy People 2030. Published 2024. Accessed April 27, 2026. https://odphp.health.gov/healthypeople

9. Disability Data Interagency Working Group of the National Science and Technology Council. *Federal Evidence Agenda on Disability*. Disability Data Interagency Working Group of the National Science and Technology Council; 2025. Accessed April 27, 2026. https://bidenwhitehouse.archives.gov/wp-content/uploads/2025/01/Federal-Evidence-Agenda-on-Disability.pdf

10. National Council on Disability. *Health Equity Framework for People with Disabilities*.; 2022. Accessed April 29, 2026. https://www.ncd.gov/assets/uploads/reports/2022/ncd_health_equity_framework.pdf

11. Public Health Accreditation Board. Standards and Measures. Phaboard.org. Published 2026. Accessed May 20, 2026. https://phaboard.org/accreditation-recognition/standards-measures/

12. Umucu E, Vernon AA, Pan D, et al. Health Inequities among Persons with Disabilities: a Global Scoping Review. *Frontiers in Public Health*. 2025;13. doi:https://doi.org/10.3389/fpubh.2025.1538519

13. World Health Organization. Disability and Health. World Health Organization. Published March 7, 2023. https://www.who.int/news-room/fact-sheets/detail/disability-and-health

14. Yee S, Breslin M. *Risks and Rewards of Demographic Data Collection: How Effective Data Privacy Can Promote Health Equity This Data, Not That Data: Big Data, Privacy, and the Impact on People with Disabilities Disability Rights Education and Defense Fund.* Disability Rights Education and Defense Fund; 2023. Accessed April 27, 2026. https://healthlaw.org/wp-content/uploads/2023/03/This-Data-Not-That-Data_Disability-Rights-Education-and-Defense-Fund_FINAL.pdf

15. Transform Health. Health Data Governance Principles. Healthdatagovernance.org. Published 2026. Accessed June 12, 2026. https://healthdatagovernance.org/principles

16. Holliman BD, Stransky ML, Dieujuste N, Morris MA. Disability Doesn't Discriminate: Health Inequities at the Intersection of Race and Disability. *Frontiers in Rehabilitation Sciences*. 2023;4. doi:https://doi.org/10.3389/fresc.2023.1075775

17. Nilz M, Chughtai Z. Creating Inclusive Disability Data Systems. Astho.org. Published July 14, 2025. Accessed April 27, 2026. https://www.astho.org/communications/blog/2025/creating-inclusive-disability-data-systems/

18. Leonardi M, Lee H, Kostanjsek N, et al. 20 Years of ICF—International Classification of Functioning, Disability and Health: Uses and Applications around the World. *International Journal of Environmental Research and Public Health*. 2022;19(18):11321. doi:https://doi.org/10.3390/ijerph191811321

19. Lauer EA, Henly M, Coleman R. Comparing Estimates of Disability Prevalence Using Federal and International Disability Measures in National Surveillance. *Disability and Health Journal*. 2019;12(2):195-202. doi:https://doi.org/10.1016/j.dhjo.2018.08.008

20. National Council on Disability. *National Disability Policy: A Progress Report*.; 2019. https://www.ncd.gov/assets/uploads/reports/2019/ncd_progress_report_federal-enforcement-of-disability-rights-laws-part-2.pdf

21.U.S. Department of Justice Civil Rights Division. The Americans with Disabilities Act (ADA) . ada.gov. Published 2020. Accessed May 19, 2026. https://www.ada.gov/

22. Social Security Administration. Who Can Get Disability. https://www.ssa.gov/disability/eligibility. Published 2025. Accessed May 19, 2026. https://www.ssa.gov/

23. Administration for Community Living. The Developmental Disabilities Assistance and Bill of Rights Act of 2000. acl.gov. Published March 25, 2021. https://acl.gov/about-acl/authorizing-statutes/developmental-disabilities-assistance-and-bill-rights-act-2000

24. US Census Bureau. How Disability Data Are Collected from the American Community Survey. Census.gov. Published October 17, 2017. https://www.census.gov/topics/health/disability/guidance/data-collection-acs.html

25. US Census Bureau. American Community Survey (ACS). The United States Census Bureau. Published November 5, 2023. Accessed April 29, 2026. https://www.census.gov/programs-surveys/acs.html

26. Washington Group on Disability Statistics. WG Short Set on Functioning (WG-SS). The Washington Group. Published 2026. https://www.washingtongroup-disability.com/question-sets/wg-short-set-on-functioning-wg-ss/

27. World Health Organization. International Classification of functioning, Disability and Health. World Health Organisation. Published 2001. https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health

28. Centers for Medicare and Medicaid Services. ICD Code Lists. www.cms.gov. Published October 2025. https://www.cms.gov/medicare/coordination-benefits-recovery/overview/icd-code-lists

29. National Library of Medicine . SNOMED CT. www.nlm.nih.gov. Published September 18, 2019. https://www.nlm.nih.gov/healthit/snomedct/index.html

30. Massoudi BL, Chester KG. *Disability Definitions Implementation Toolkit: Enhancing Syndromic Surveillance for Emergency Response.* Council of State and Territorial Epidemiologists; 2026. Accessed April 27, 2026. https://cdn.ymaws.com/www.cste.org/resource/resmgr/pdfs/pdfs2/Implementing_Disability_Defi.pdf

31. U.S. Equal Employment Opportunity Commission. Job Applicants and the ADA. U.S. Equal Employment Opportunity Commission. Published October 7, 2003. Accessed April 27, 2026. https://www.eeoc.gov/laws/guidance/job-applicants-and-ada?

32. U.S. Department of Justice - Civil Rights Division. Guide to Disability Rights Laws. ADA.gov. Published May 5, 2023. Accessed April 27, 2026. https://www.ada.gov/resources/disability-rights-guide

33. Social Security Administration. Substantial Gainful Activity. www.ssa.gov. Published 2026. Accessed April 27, 2026. https://www.ssa.gov/oact/cola/sga.html

34. Landes SD, Swenor BK, Clark MA, et al. A Research Roadmap toward Improved Measures of Disability. *Health Affairs*. Published online July 9, 2024. doi:https://doi.org/10.1377/forefront.20240708.306851

35. Dusetzina SB, Tyree S, Meyer AM, Meyer A, Green L, Carpenter WR. *Linking Data for Health Services Research: A Framework and Instructional Guide*. Agency for Healthcare Research and Quality; 2014. Accessed April 27, 2026. https://www.ncbi.nlm.nih.gov/books/NBK253312/

36. Centers for Disease Control and Prevention. Disability Datasets. Disability and Health Data System (DHDS). Published December 10, 2024. Accessed May 19, 2026. https://www.cdc.gov/dhds/datasets/index.html

37. Okoro CA, Hollis ND, Cyrus AC, Griffin-Blake S. Prevalence of Disabilities and Health Care Access by Disability Status and Type among Adults — United States, 2016. *MMWR Morbidity and Mortality Weekly Report*. 2018;67(32):882-887. doi:https://doi.org/10.15585/mmwr.mm6732a3

38. Brault M, Stern S. *Evaluation Report Covering Disability*.; 2007. Accessed June 10, 2026. https://www.census.gov/content/dam/Census/library/working-papers/2007/acs/2007_Brault_01.pdf

39. Kline N. *Using Administrative Data in Social Policy Research*. Insight Policy Research; 2022. https://acf.gov/sites/default/files/documents/opre/administrative_data_brief_feb2023.pdf

40. Human Services Research Institute, Maine Health Data Organization. *Using Administrative Data to Identify Individuals with Intellectual and Developmental Disabilities: An Exploration in Maine*.; 2026. Accessed June 10, 2026. https://www.hsri.org/app/uploads/2026/04/AdminData_HSRI_Maine_March2026.pdf

41. Ya-Chen Tina Shih, Bradley C, K Robin Yabroff. Ecological and Individualistic Fallacies in Health Disparities Research. *Journal of the National Cancer Institute*. 2023;115(5). doi:https://doi.org/10.1093/jnci/djad047

42. Parrish J, Strahle S, Vance S. *A Framework for Linking PRAMS with Administrative Data Enhancing PRAMS through Data Linkages*. Association of State and Territorial Health Officials; 2024. Accessed April 27, 2026. https://www.astho.org/4a4b2a/globalassets/report/a-framework-for-linking-prams-with-administrative-data.pdf

43. Swenor BK. A Need for Disability Data Justice. *Health Affairs Forefront*. Published online August 22, 2022. doi:https://doi.org/10.1377/forefront.20220818.426231

44. Ravesloot C, Seekins T, Traci M, et al. Living Well with a Disability, a Self-Management Program. *MMWR Supplements*. 2016;65(01):61-67. doi:https://doi.org/10.15585/mmwr.su6501a10

45. Popkin SJ, Webley H. Stronger Data Privacy Protections Are Needed to Protect People with Disabilities. Urban Institute. Published October 6, 2025. Accessed May 19, 2026. https://www.urban.org/urban-wire/stronger-data-privacy-protections-are-needed-protect-people-disabilities

46. Centers for Disease Control and Prevention. Disability and Health Data Now. Disability and Health. Published December 13, 2024. Accessed May 19, 2026. https://www.cdc.gov/disability-and-health/articles-documents/disability-and-health-data-now.html

47. Centers for Disease Control and Prevention. Disability Barriers to Inclusion. Disability Inclusion. Published December 10, 2024. Accessed May 19, 2026. https://www.cdc.gov/disability-inclusion/barriers/index.html

48. Association of State and Territorial Health Officials. Improving Disability Inclusion. Astho.org. Published July 31, 2025. Accessed April 27, 2026. https://www.astho.org/topic/health-equity/disability-inclusion/improving-disability-inclusion/

49. Administration for Community Living. Centers for Independent Living | ACL Administration for Community Living. Acl.gov. Published 2014. Accessed May 19, 2026. https://acl.gov/programs/aging-and-disability-networks/centers-independent-living

50. American Association on Health and Disability. National Disability and Public Health Resource Center for People with Disabilities. AAHD. Published August 17, 2020. Accessed June 18, 2026. https://aahd.us/national-center-on-research/

51. Nilz M. Stakeholder-Centered Approaches to Disability Health Data. Astho.org. Published July 22, 2025. Accessed April 27, 2026. https://www.astho.org/communications/blog/2025/stakeholder-centered-approaches-to-disability-health-data/

52 McGee M. *Race, Ethnicity, Language and Disability (READ) Implementation Guide.* Oregon Health Authority, Equity and Inclusion Division; 2020.

53 Berghs M, Atkin K, Graham H, Hatton C, Thomas C. Implications for Public Health Research of Models and Theories of disability: a Scoping Study and Evidence Synthesis. *Public Health Research.* 2016;4(8):1-166. doi:https://doi.org/10.3310/phr04080

54North Carolina Department of Health and Human Services. *Data Suppression and Working with Small Numbers.* Division of Public Health, Injury and Violence Prevention Branch (IVPB), Epidemiology, Surveillance, and Informatics Unit; 2025. Accessed May 19, 2026. https://www.dph.ncdhhs.gov/media/5708/open

55. Wasserman C, Ossiander E. *Department of Health Agency Standards for Reporting Data with Small Numbers*. Washington Department of Health; 2018. Accessed May 19, 2026. https://doh.wa.gov/sites/default/files/legacy/Documents/1500//SmallNumbers.pdf

56. Supriya Lal Kapur, Sabatello M. People with Disability and Privacy in Precision Medicine research: What's at stake? *Trends in Genetics*. 2023;39(5):335-337. doi:https://doi.org/10.1016/j.tig.2023.01.001

57. Public Health Law Center at Mitchell Hamline School of Law. Small Numbers Policies and Data Suppression: What Health Department Attorneys Need to Know. Publichealthlawcenter.org. Published 2026. Accessed June 18, 2026. https://www.publichealthlawcenter.org/resources/small-numbers-policies-and-data-suppression-what-health-department-attorneys-need-know

58. Wilkinson K, Green C, Nowicki D, Von Schindler C. Less than Five Is Less than ideal: Replacing the "Less than 5 Cell Size" Rule with a risk-based Data Disclosure Protocol in a Public Health Setting. *Canadian Journal of Public Health.* Published online March 11, 2020. doi:https://doi.org/10.17269/s41997-020-00303-8

59. Council of State and Territorial Epidemiologists. *Data Release Reference Document Tabulated, Aggregated, and Macro Data Policy Development Recommendations for Public Health Agencies.* Accessed June 18, 2026. https://cstelearndev.wpengine.com/wp-content/uploads/2025/10/CSTE_Data-Release-Reference_FINAL.pdf

60. Carroll SR, Rodriguez-Lonebear D, Martinez A. Indigenous Data Governance: Strategies from United States Native Nations. *Data Science Journal.* 2019;18(1):31. doi:https://doi.org/10.5334/dsj-2019-031

61. Christen P. *Data Matching : Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection.* Springer; 2012. https://dl.acm.org/doi/10.5555/23441085

62. Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health Services Research and Data Linkages: Issues, Methods, and Directions for the Future. *Health Services Research*. 2010;45(5p2):1468-1488. doi:https://doi.org/10.1111/j.1475-6773.2010.01142.x

63. Harron KL, Doidge JC, Knight HE, et al. A Guide to Evaluating Linkage Quality for the Analysis of Linked Data. *International Journal of Epidemiology.* 2017;46(5):1699-1710. doi:https://doi.org/10.1093/ije/dyx177

64. Herzog TN, Scheuren FJ, Winkler WE. *Data Quality and Record Linkage Techniques.* Springer New York; 2007.

65. Harron K, Doidge JC, Goldstein H. Assessing Data Linkage Quality in Cohort Studies. *Annals of Human Biology.* 2020;47(2):218-226. doi:https://doi.org/10.1080/03014460.2020.1742379

66. Harron K, Goldstein H, Dibben C, eds. Methodological Developments in Data Linkage. *Wiley Series in Probability and Statistics.* Published online November 10, 2015. doi:https://doi.org/10.1002/9781119072454

67. Bohensky MA, Jolley D, Sundararajan V, et al. Data Linkage: a Powerful Research Tool with Potential Problems. *BMC Health Services Research*. 2010;10(1). doi:https://doi.org/10.1186/1472-6963-10-346

68. Horner-Johnson W, Senders A, Higgins Tejera C, McGee MG. Sexual Health Experiences among High School Students with Disabilities. *Journal of Adolescent Health.* 2021;69(2). doi:https://doi.org/10.1016/j.jadohealth.2021.03.001

69. Massachusetts Department of Public Health. Community Health Equity Initiative Data. Mass.gov. Published 2023. Accessed June 11, 2026. https://www.mass.gov/community-health-equity-initiative-data

70. Oliver K, Innvar S, Lorenc T, Woodman J, Thomas J. A Systematic Review of Barriers to and Facilitators of the Use of Evidence by Policymakers. *BMC Health Services Research*. 2014;14(1).

71. Government Services Administration. Section 508. Section508.gov. Published 2018. Accessed May 19, 2026. https://www.section508.gov/manage/publish-accessible-content/

72. Harron K, Dibben C, Boyd J, et al. Challenges in Administrative Data Linkage for Research. *Big Data & Society*. 2017;4(2):205395171774567. doi:https://doi.org/10.1177/2053951717745678

73. Association of State and Territorial Health Officials. Overcoming Common Barriers to Data Linkage. Astho.org. Published May 4, 2022. Accessed May 19, 2026. https://www.astho.org/topic/brief/overcoming-common-barriers-to-data-linkage/

74. Henly M. Community-to-community Linkage Analysis Examining Disability and COVID-19 Mortality in New Hampshire . In: *CSTE Annual Meeting*. Council of State and Territorial Epidemiologists; 2025.

75. Arnesen M, Leavitt I. Massachusetts Harnesses Data from Multiple Agencies to Improve Public Health. Pew.org. Published July 16, 2025. Accessed April 27, 2026. https://www.pew.org/en/research-and-analysis/issue-briefs/2025/07/massachusetts-harnesses-data-from-multiple-agencies-to-improve-public-health

76. Massachusetts Department of Public Health. Public Health Data Warehouse Data Sources. Mass.gov. Published 2026. Accessed June 12, 2026. https://www.mass.gov/info-details/public-health-data-warehouse-data-sources

77. Massachusetts Department of Public Health. Public Health Data Warehouse Community Advisory Board. Mass.gov. Published 2024. Accessed June 12, 2026. https://www.mass.gov/info-details/public-health-data-warehouse-community-advisory-board

78. Vance S, Kuruganti S, Strahle S, Kipling L, Cox S. Multi-State Network of Linked PRAMS and Clinical Outcomes Data for Patient-Centered Outcomes Research. Astho.org. Published July 1, 2024. Accessed April 27, 2026. https://www.astho.org/topic/report/multi-state-network-linked-prams-clinical-outcomes-data-patient-centered-outcomes-research/

79. The Center for IDEA Early Childhood Data Systems. Data Linking Toolkit. Dasycenter.org. Published July 2022. Accessed April 27, 2026. https://dasycenter.org/data-linking-toolkit/

80. Division of Analysis and Epidemiology, National Center for Health Statistics, Centers for Disease Control and Prevention . *The Linkage of the National Center for Health Statistics (NCHS) Survey Data to U.S. Department of Housing and Urban Development (HUD) Administrative Data: Linkage Methodology and Analytic Considerations.* Division of Analysis and Epidemiology, National Center for Health Statistics, Centers for Disease Control and Prevention ; 2014. Accessed April 27, 2026. https://www.cdc.gov/nchs/data/datalinkage/NCHS-HUD-Linked-Data-Methodology-and-Analytic-Considerations.pdf

81. Kearly A, Hluchan M, Brazeel C, et al. Health Service Utilization Patterns among Medicaid Enrollees with Intellectual and Developmental Disabilities before and during the COVID-19 Pandemic: Implications for Pandemic Response and Recovery Efforts. *Journal of Public Health Management & Practice.* 2024;30(6):857-868. doi:https://doi.org/10.1097/phh.0000000000001957

82. Mays VM, Ponce NA, Washington DL, Cochran SD. Classification of Race and Ethnicity: Implications for Public Health. *Annual Review of Public Health.* 2003;24(1):83-110. doi:https://doi.org/10.1146/annurev.publhealth.24.100901.140927

83. National Library of Medicine. SNOMED CT to ICD-10-CM Map. www.nlm.nih.gov. Published March 1, 2021. Accessed April 27, 2026. https://www.nlm.nih.gov/research/umls/mapping_projects/snomedct_to_icd10cm.html

84. Centers for Disease Control and Prevention. About NCHS Data Linkage. Data Linkage. Published April 22, 2026. Accessed April 27, 2026. https://www.cdc.gov/nchs/linked-data/about/index.html

85. Centers for Disease Control and Prevention. National Hospital Care Survey Linked Data. Data Linkage. Published April 22, 2026. Accessed April 27, 2026. https://www.cdc.gov/nchs/linked-data/nhcs/index.html

86. U.S. Equal Employment Opportunity Commission. Disability Discrimination. www.eeoc.gov. Accessed April 27, 2026. https://www.eeoc.gov/youth/disability-discrimination-0

87. Cook L, Zhu M. *Injury Data Linkage Toolkit.* Council of State and Territorial Epidemiologists; 2022. Accessed April 27, 2026. https://cdn.ymaws.com/www.cste.org/resource/resmgr/injury/CSTE_Injury_Data_Linkage_Too.pdf

88. Council of State and Territorial Epidemiologists. Tribal Epidemiology Toolkit. Cste.org. Published 2014. Accessed April 27, 2026. https://www.cste.org/page/TribalTookitLanding

89. National Association of Health Data Organizations. Data Enhancement and Linkage. Nahdo.org. Published 2026. Accessed April 27, 2026. https://www.nahdo.org/data_resources/data_enhancement_linkage

90. Agency for Healthcare Research and Quality. Race and Ethnicity Data Improvement Toolkit. hcup-us.ahrq.gov. Published 2014. Accessed April 27, 2026. https://hcup-us.ahrq.gov/datainnovations/raceethnicitytoolkit/home_race.jsp

91. UK Office of National Statistics. Developing Standard Tools for Data linkage: February 2021 - Office for National Statistics. www.ons.gov.uk. Published February 2021. Accessed April 27, 2026. https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/developingstandardtoolsfordatalinkagefebruary2021

92. State of Queensland (Queensland Health). *Queensland Data Linkage Framework*. State of Queensland (Queensland Health); 2020. Accessed April 27, 2026. https://www.health.qld.gov.au/__data/assets/pdf_file/0020/1211483/qlddatalinkageframework.pdf

93. U.S. Census Bureau Data Stewardship Program. *Census Bureau Data Stewardship Program DS014 -Data Linkage Policy.* US Census Bureau Data Stewardship Program; 2026. Accessed April 27, 2026. https://www2.census.gov/foia/ds_policies/ds014.pdf

94. Chugani V. Understanding the Pareto Distribution: a Comprehensive Guide. Datacamp.com. Published April 10, 2025. Accessed April 27, 2026. https://www.datacamp.com/tutorial/pareto-distribution

95. Karr AF, Taylor MT, West SL, et al. Comparing Record Linkage Software Programs and Algorithms Using real-world Data. Wang H, ed. *PLOS ONE*. 2019;14(9):e0221459. doi:https://doi.org/10.1371/journal.pone.0221459

96. International Population Data Linkage Network. The International Population Data Linkage Network Conference. IPDLN. Published March 11, 2026. Accessed April 27, 2026. https://ipdln.org/

97. Association of State and Territorial Health Officials. Data and Disability Inclusion. Astho.org. Published July 31, 2025. Accessed April 27, 2026. https://www.astho.org/topic/health-equity/disability-inclusion/data-and-disability-inclusion/

98. Hanson-DeFusco J. What Data Counts in Policymaking and Programming Evaluation – Relevant Data Sources for Triangulation According to Main Epistemologies and Philosophies within Social Science. *Evaluation and Program Planning.* 2023;97(102238):102238. doi:https://doi.org/10.1016/j.evalprogplan.2023.102238

99. Syracuse University. Disability Data Equity Research Working Group. Aging Studies Institute. Published 2025. Accessed June 11, 2026. https://asi.syr.edu/disability-health-equity-research-network-dhern/disability-data-equity-research-working-group/

100. Massachusetts Department of Public Health. About the Public Health Data Warehouse. Mass.gov. Published 2026. https://www.mass.gov/info-details/about-the-public-health-data-warehouse

101. Aitken Z, Walmsley S, M Bishop G, Badji S, Fortune N. Methods Used to Construct Disability Indicators in Linked Administrative datasets: a Systematic Scoping Review. *Population Health Metrics.* 2025;23(1). doi:https://doi.org/10.1186/s12963-025-00386-w

102. Administration for Community Living. Aging and Disability Partners and Providers. Acl.gov. Published 2026. Accessed July 1, 2026. https://acl.gov/help/aging-and-disability-partners-providers

103. American Association on Health and Disability. State Resources. American Association on Health and Disability. Published August 17, 2020. Accessed July 1, 2026. https://aahd.us/national-center-on-research/state-resources/

104. Harron KL, Doidge JC, Knight HE, et al. A Guide to Evaluating Linkage Quality for the Analysis of Linked Data. *International Journal of Epidemiology.* 2017;46(5):1699-1710. doi:https://doi.org/10.1093/ije/dyx177

105. National Cancer Institute. Match*Pro Software - SEER Registrars. Surveillance, Epidemiology and End Result Program (SEER). Published 2018. Accessed July 1, 2026. https://seer.cancer.gov/tools/matchpro/

106. R Project. Record Linkage Functions for Linking and Deduplicating Data Sets. *R-projectorg*. Published online January 25, 2026. doi:https://cran.r-project.org/package=RecordLinkage

107. Parrish J. *Record Linkage Using R*. Association of State and Territorial Health Officials; 2024. Accessed July 1, 2026. https://www.astho.org/globalassets/resources/record-linkage-using-r.pdf

108. King Memorial. The Link King. The Link King Memorial Archive. Published 2026. Accessed July 1, 2026. https://the-link-king.party/

109. Centers for Disease Control and Prevention. Link Plus. National Program of Cancer Registries. Published 2024. https://www.cdc.gov/national-program-cancer-registries/registry-plus/link-plus.html

110. De Bruin J. recordlinkage. PyPI. Published July 20, 2023. https://pypi.org/project/recordlinkage/

111. Lindsay S, Manassis T, Hepworth T, Bond A, Kennedy R. splink. PyPI. Published June 4, 2025. https://pypi.org/project/splink/

112. Rentsch CT, Whiteson Kabudula C, Catlett J, et al. Point-of-contact Interactive Record Linkage (PIRL): a Software Tool to Prospectively Link Demographic Surveillance and Health Facility Data. *Gates Open Research.* 2018;1(8):8-8. doi:https://doi.org/10.12688/gatesopenres.12751.2

113. Asher J, Resnick D, Brite J, Brackbill R, Cone J. An Introduction to Probabilistic Record Linkage with a Focus on Linkage Processing for WTC Registries. *International Journal of Environmental Research and Public Health.* 2020;17(18):6937. doi:https://doi.org/10.3390/ijerph17186937

114. Association of State and Territorial Health Officials. Succession Planning for Public Health Agencies: a Three-Part Webinar Series. Astho.org. Published October 6, 2025. Accessed July 1, 2026. https://www.astho.org/communications/videos/succession-planning-for-public-health-agencies/