

# Establishing a Standard Data Model for Large-scale IDS Use

Actionable Intelligence for Social Policy,  
Expert Panel Report

Prepared by

Fred Wulczyn, Richard Clinch, Claudia Coulton, Sallie Keller,  
James Moore, Clara Muschkin, Andrew Nicklin,  
Whitney LeBoeuf, and Katie Barghaus

MARCH 2017



***ACTIONABLE* INTELLIGENCE**

FOR SOCIAL POLICY

- I. Introduction** ..... 5
- II. Principles Guiding the Data Selection Process for an IDS** ..... 6
  - A. Organize Around the Life Course** ..... 6
  - B. Reflect the Development of Human Capital** ..... 7
  - C. Include Contextual Factors** ..... 7
  - D. Ensure Validity and Reliability** ..... 7
  - E. Align with Mathematical Models** ..... 7
- III. IDS Data Sources** ..... 8
  - A. Vital Statistics** ..... 10
    - 1. Birth Records (High) ..... 10
    - 2. Death Records (Medium) ..... 10
  - B. Healthcare Utilization** ..... 10
    - 1. Medicaid (High) ..... 10
    - 2. Behavioral and Mental Health (High) ..... 10
    - 3. Alcohol and Substance Abuse (Medium) ..... 10
    - 4. Department of Public Health (Medium) ..... 10
    - 5. State Children’s Health Insurance Program (SCHIP) (Medium) ..... 11
    - 6. Nursing Facility Minimum Data Set (MDS) (Medium) ..... 11
    - 7. All Payer Claims Databases (Low) ..... 11
    - 8. Community Health Centers (Low) ..... 11
    - 9. Developmental Disabilities (Low) ..... 11
    - 10. Emergency Medical Services (EMS) (Low) ..... 11
  - C. Child Welfare** ..... 12
    - 1. Abuse and Neglect, Out-of-Home Care, In-Home Services (High) ..... 12
  - D. Early Childhood** ..... 12
    - 1. Child Care Development Fund (CCDF) (Low) ..... 12
    - 2. Early Intervention (Low) ..... 12
  - E. Education** ..... 12
    - 1. K-12 Education (Medium) ..... 12
    - 2. Postsecondary Education (Medium) ..... 12
    - 3. K-12 Special Education (Medium) ..... 13

**Actionable Intelligence for Social Policy**

University of Pennsylvania  
 3701 Locust Walk, Philadelphia, PA 19104  
 215.573.5827 | www.aisp.upenn.edu

- F. Juvenile Justice** .....13
  - 1. Juvenile Justice Services (High) .....13
  - 2. Juvenile Courts (Low) .....13
- G. Adult Justice** .....13
  - 1. City or County Jail (High) .....13
  - 2. State Corrections (Medium) .....13
  - 3. Law Enforcement (Medium) .....13
- H. Employment** .....14
  - 1. Workforce Training Programs (Medium) .....14
  - 2. Unemployment Insurance (UI) Wages (Medium) .....14
- I. Public Assistance** .....14
  - 1. Temporary Assistance for Needy Families (TANF) (High) .....14
  - 2. Supplemental Nutrition Assistance Program (SNAP) (High) .....14
  - 3. Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) (Low) .....15
- J. Homelessness/Housing** .....15
  - 1. Homeless Management Information System (HMIS) (High) .....15
  - 2. Public Housing Agency (PHA) (Medium) .....15
  - 3. Education Homeless Records (Medium) .....15
- IV. Standards for Data Sources** .....16
  - A. Person** .....16
  - B. Encounter** .....16
  - C. Place** .....16
  - D. Time** .....17
- V. Standard Data Repurposing Process for IDS** .....17
  - A. Prerequisites** .....17
  - B. Profiling Source Data** .....17
    - 1. Data Structure .....17
    - 2. Data Quality .....18
  - C. Data Transformation** .....19
    - 1. Merging Data .....19
    - 2. Restructuring .....20

- 3. Cleaning .....20
- 4. Transforming .....21
- 5. Using Existing IDS Data .....21
- VI. Conclusions** .....21
- References** .....22
- Appendix A:** Colorado’s Opportunity Framework. ....26
- Appendix B:** Data sources included in AISP network sites’ IDS by domain of life experience. ....27
- Appendix C:** Data elements by domain and data source. ....29

## I. Introduction

Integrated data systems (IDS) offer a means for an ever-deeper understanding of how human-built systems affect the well-being of people in intended and unintended ways. IDS have the potential to paint a more complete picture of multifaceted social problems—such as those of children in foster care who encounter juvenile justice, and of families who interact with multiple public assistance and housing programs—thereby supporting more efficient multisystem collaboration and responses. To realize that potential, administrative data captured during the course of normal interactions between people and public services must be organized in line with how scientists approach complex research questions.

In this paper, we provide both general and specific guidance to states and localities interested in building robust IDS that take full advantage of all that these systems have to offer. Our guidance is motivated by designs that assess the impact of policies and practices on the public, although this is not the limit of IDS potential. Governments at all levels go to great lengths (and expense) to administer programs that are designed to affect outcomes at a population level. If and when policies have their intended effect, we want to recognize and amplify their impact. When that is not the case, we want to understand how investments in well-being can be productively redirected. IDS can help policy-makers and researchers unpack questions such as:

- ❖ Does low-level lead exposure have an effect on children’s cognitive development? (link birth records, blood lead level data, and academic performance)
- ❖ What is the impact of public housing programs on children’s educational achievement and progress? (link housing data and academic achievement and graduation)
- ❖ How do economic dislocations (e.g., job loss) affect local health care utilization and expenditures? (link employment and health care utilization data)
- ❖ What types of workforce investments sustain a resilient labor force in the face of changing labor markets? (link workforce program, public assistance, and employment data)

We take a broad view in defining an IDS. We view an IDS as any well-organized collection of disparate data that coheres around a common purpose. Data integration combines diverse types of information to support a common unit of analysis. IDS are person-centered and involve knitting together individual-level data from disparate sources. This narrative between people and organizations is bi-directional—people are affected by the organizations that deliver services, just as the organizations are affected by the people they serve. Understanding these narratives requires an explicit IDS structure that connects units of analysis to theoretical models of human behavior in the context of complex social and administrative systems.

In this report, we tackle questions of scientific merit and practicality. Administrative data systems have the ability to record *all* transactions that take place within an agency. This volume of data is often too large and complex to discern meaning. To develop meaning, or reveal narratives, a data system must be built around structural models of what transpires between individuals and the systems that serve them. This focus takes IDS beyond a conceptualization as a data repository, to a resource that preserves and reveals the narratives captured by the data. Building consensus for the motivation behind the goal of bringing the data together is critical, and discussed at length in “IDS Governance: Setting Up for Ethical and Effective Use” (Gibbs et al., 2017). Below, we outline principles for how a state or locality might approach the task of building an IDS with scientific integrity and utility.

This framework rejects the notion that heterogeneity among state and local data systems prevents data from being connected in useful ways. To the contrary, within the collective experience of IDS nationally, there are numerous examples of how linked data have been used to manage social programs, design interventions, and evaluate public policy, all within a rigorous, scientifically motivated framework that yields practical insights across multiple agencies/programs, both within and between states. The

key to building and then using an IDS is a framework that preserves the human experience captured within each of the underlying data systems. Data standards, such as those described here, are a key component of this strategy, as they ensure a common data framework and targets for harmonization that allow for cross-agency and cross-site work.

Our recommended approach balances short- and long-term objectives. Over the long term, the goal is to build IDS that span geographic, programmatic, and agency-level boundaries. The short-term goals are locally oriented and predicated on a belief that nimble, opportunistic designs are the ones most likely to deliver demonstration projects that beget future investment. The link between the short- and long-term goals rests with an appreciation for the kind of problem-solving that drives evolution over time around common principles and shared purpose.

## II. Principles Guiding the Data Selection Process for an IDS

Public programs are designed to support and positively impact an individual's life course. Thus, the essential question for evaluating policies and programs becomes: what did the public system contribute to the well-being of an individual? The goal of an IDS is to bring together, in one place, the capacity to answer questions about the efficacy of the programs used to support people. This capacity relies upon quality data that are safely linked across data silos and accessible to analysts, who can then support policy-makers. As described above, governments have an inordinate amount of data that is captured within enterprise data systems. How can these data be prioritized to meet this goal? We recommend the following five principles to guide the data selection process for an IDS.

### A. Organize Around the Life Course

The life course is constructed from patterns in the timing, duration, and sequence of events that accumulate over a lifetime. Many of the most important life course narratives take their meaning from the interplay between the social institutions that shape development and the underlying bio-physiology of development. In a life course context, research and policy questions are often framed in terms of transitions into and between life events. Although there is no single underlying outline of the life course, social institutions that support development are organized in ways that promote transitions over the life course.<sup>1</sup> The life course can serve as a guide for securing and organizing data.

The developmental overlay covers age-graded transitions over the lifespan from birth to death. In between, the transition into school, out of school, and into the world of adulthood and the accompanying role-specific expectations define the scope of what an IDS can encompass. This includes:

- ❖ Birth and infancy
- ❖ Early childhood
- ❖ School-age children
- ❖ Transition to adulthood
- ❖ Adults and parents
- ❖ Elderly and death

<sup>1</sup> For example, the state of Colorado uses the life course as a framework for their cross-agency collaborative, the Opportunity Project, which aims to promote successful outcomes in every stage of life through an integrated system of health, social, and educational well-being (see Appendix A). The Children's Cabinet in New York City is also using the life course perspective to organize the City's investment in young people. See NYC Children's Cabinet (2016).

Within life course passages, programs are provided to promote healthy development: prenatal care addresses the risks associated with starting out in life; early care education supports the transition to school; post-secondary education prepares youth for workforce entry; and so on.

### B. Reflect the Development of Human Capital

Human capital refers to the relational skills, hard skills, experience, education, and know-how needed to transition seamlessly through the life course. Accumulated human capital is what differentiates children from adults. Successful adults *know how* to be successful adults. They know how because they have acquired the skills needed along the way. If we want to know how people are doing, we have to understand how human capital takes shape as childhood unfolds into adulthood and what, if anything, is getting in the way. When times are tough, human capital formation slows, and this is often seen as a risk factor for individuals (e.g., homelessness). When times are good, human capital formation quickens, and this reflects a protective factor (e.g., gainful employment). These changes in human capital need to be adequately captured in an IDS in order to fully understand individuals' well-being across the life course.

### C. Include Contextual Factors

Life is lived in context. In order to better understand positive life course trajectories, IDS need to contain information beyond individuals that captures the context of their development. IDS should be capable of appreciating the nesting of time within individuals, and individuals within organizations, and geographical locations (e.g., students nested within schools located in neighborhoods, counties, and states). This temporal and place-based information reflects the organization interfacing with the individual as well as the timing and physical location of those interactions (which could be home-based, office-based, school-based, etc.).

### D. Ensure Validity and Reliability

An effective IDS is built around measures from which inferences can be made about people and people-based attributes, e.g., the places they were living when they were receiving services, the services received, and the time in their lives when they received services. Validity and reliability have both absolute and relative meaning. For example, an event date is when something happens. The record in the data has to link reliably to when something really happened. What the event represents is a matter of policy and practice—meaning is attached to the date based on the meaning the event acquires in the policy and practice narratives of interest. The connection between an event and its meaning is more directly a matter of validity. For example, when evaluating the impact of a social intervention such as a youth development program, the research must establish that the measured outcomes are related to this program rather than other changes (e.g., improved funding for schools and other support systems).

### E. Align with Mathematical Models

The utility of an IDS hinges on how well the data are aligned with the mathematical models used to extract those (causal) narratives that have the most salience to scientists, policy-makers, and practitioners. Three families of mathematical models provide the necessary structure: event history, multilevel, and population dynamics models.

Event history models summarize the experiences of people served by public programs in terms of a historical sequence of events (i.e., the life course) that traces the various status changes people may undergo during their involvement in the system. In multilevel models, time is nested within people, people are nested within social or administrative structures, and administrative structures are nested within geographic locations and policy contexts. Multilevel models preserve the underlying

hierarchical structures of time, people, and organizations that are relevant to most policy and administrative questions.

Finally, dynamic models capture population-level changes over time. IDS should be developed with these models in mind to ensure that meaningful narratives can be extracted from the data.

### III. IDS Data Sources

As described above, the goal of an IDS is to create, in one place, the capacity to answer questions about what transpires between individuals and the systems created to support their well-being. This paper organizes the data sources typically included in an IDS by domains of life experience (e.g., health utilization, education). The data sources discussed below represent individual-level data that can feasibly be integrated into an IDS at this time.

We define data that can be feasibly integrated by proof of concept—that an AISP site has successfully integrated the data source into its IDS on a routine basis rather than for a single use, as this demonstrates long-term data-sharing partnerships and agreements. These data sources can serve as potential starting points for developers of IDS and can be employed by users of IDS to inform their conceptualization of questions that can be practically answered with these systems.

Figure 1 lists each of the data sources, their domain of life experience, and the current frequency of inclusion in AISP network sites.<sup>2</sup> Data sources noted as having a “high” frequency of inclusion are integrated into at least two-thirds (7 to 11) of network sites. Sources noted as having a “medium” frequency of inclusion are integrated into 4 to 6 network sites, and those with a “low” inclusion frequency are integrated into only 1 to 3 sites. Appendix B details the inclusion of data sources by each AISP site.

We do not intend to suggest that the data sources listed below represent either an exhaustive list of potential sources or a minimum for an IDS to be established. There is no limit on the type or kind of data that could be integrated into an IDS. For example, workforce, finance, non-profit, geospatial, and system-level information can all be included, and we should aspire to do so in order for the full value of integrated data to be realized. Though not exhaustive, the data sources discussed below still represent an aspirational list that we envision as being incorporated into an IDS over time. Those hoping to establish an IDS should start with the institutions where there is interest and political will to integrate data.

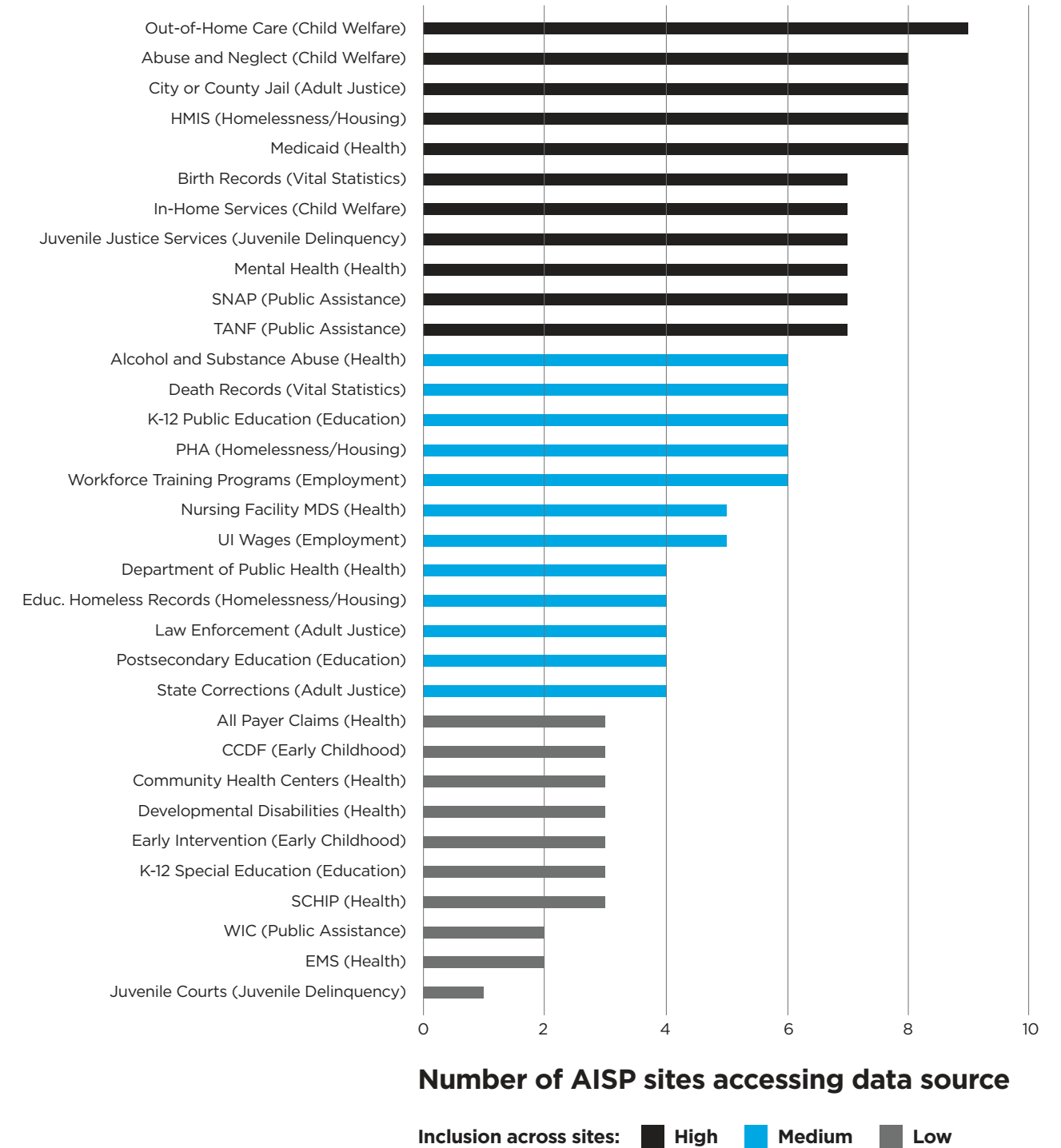


Figure 1: Data source inclusion across AISP sites.

<sup>2</sup> Eleven of thirteen established AISP sites were able to publicize their data holdings and are included here.

## A. Vital Statistics

### 1. Birth Records (High)

Birth records contain information related to maternal and child demographics and health. Each state is responsible for the collection of individual birth records. Data typically collected include (a) the birth parents, including age, marital status, race/ethnicity, and education level, (b) prenatal care, including number of visits and risk factors during pregnancy, and (c) birth of the child, including birth date, characteristics of labor and delivery, and infant health at time of birth (CDC, 2014).

### 2. Death Records (Medium)

Death records contain information related to demographics, health, and causes of death. States are required to maintain individual records on all deaths that occur within the jurisdiction. Data commonly collected include (a) death information, such as manner of death, place and date of death, whether an autopsy was performed, and cause of death, and (b) injury information, including whether the individual was injured prior to death (CDC, 2014).

## B. Healthcare Utilization

### 1. Medicaid (High)

Medicaid is a national health insurance program for low-income individuals funded by the federal and state governments. States collect Medicaid data through two systems: the Medicaid Statistical Information System and the Medicaid Budget and Expenditure System. See *All Payers Health Claims* below for more information on commonly collected health data.

### 2. Behavioral and Mental Health (High)

Behavioral and mental health refer to an individual's emotional and psychological well-being. Behavioral and mental health services are provided to individuals to promote this aspect of well-being. States provide mental health services with funding from the Federal Mental Health Block Grants, Medicaid, and State Children's Health Insurance Program (SCHIP) (Mental Health America, n.d.). See *All Payers Health Claims* below for more information on commonly collected health data.

### 3. Alcohol and Substance Abuse (Medium)

Alcohol and substance abuse services are provided to individuals suffering from substance use disorders characterized by the use of alcohol and/or drugs to significant impairment (Substance Abuse Mental Health Services Administration, n.d.). States provide alcohol and substance abuse services with funding from the Federal Substance Abuse Prevention and Treatment Block Grant (Substance Abuse Mental Health Services Administration, n.d.). See *All Payers Health Claims* below for more information on commonly collected health data.

### 4. Department of Public Health (Medium)

The Department of Public Health at the state or county level works to improve quality of life by providing access to health services, encouraging healthy living, and ensuring healthy environments. This includes monitoring the health status of the community, diagnosing and investigating health problems and health hazards, informing and educating people about health issues, and developing policies and plans for supporting individual health improvement. See *All Payers Health Claims* below for more information on commonly collected health data.

### 5. State Children's Health Insurance Program (SCHIP) (Medium)

SCHIP is a national insurance program for uninsured children from low-income families that do not qualify for Medicaid. The program is run in partnership by the federal and state governments. States collect SCHIP data through two systems: the Medicaid Statistical Information System and the Medicaid Budget and Expenditure System. See *All Payers Health Claims* below for more information on commonly collected health data.

### 6. Nursing Facility Minimum Data Set (MDS) (Medium)

The MDS is a federally mandated comprehensive assessment of the functional capabilities of residents in Medicare- and Medicaid-certified nursing homes. All certified nursing facilities are required to complete the MDS for each resident, regardless of source of payment for the resident's care, on admission, during the stay, and on discharge (Centers for Medicare and Medicaid, 2016). Data collected include a resident's (a) active diagnoses, (b) health condition, (c) treatment/procedures, (d) medication, (e) hearing, speech, and vision assessments, (f) cognitive patterns, (g) mood, (h) behavior, (i) preference for customary routines and activities, (j) functional status, (k) functional abilities and goals, (l) bladder and bowel condition, (m) swallowing/nutritional status, (n) oral/dental status, and (o) skin conditions (Centers for Medicare and Medicaid, 2016).

### 7. All Payer Claims Databases (Low)

All Payer Claims Databases are state-run systems that consolidate information from other data sources (e.g., Departments of Public Health, community health centers, Medicaid, SCHIP, and alcohol and substance abuse, mental health, developmental disabilities service providers), regardless of the health care provider type. Data typically collected include (a) health and sometimes dental claims, which include diagnosis codes, types of care received, insurance product type, facility type, cost, and provider information, and (b) unique identifiers, geographic, and demographic information of covered individuals (All-Payer Claims Database Council, 2011).

### 8. Community Health Centers (Low)

Community health centers are private, non-profit organizations that provide primary health and related services to residents of a particular jurisdiction who are medically underserved. Community health centers receive funding from the federal government and are reimbursed by Medicaid. They are also supported by other federal, state, and local grants or contracts. See *All Payers Health Claims* above for more information on commonly collected health data.

### 9. Developmental Disabilities (Low)

Developmental disabilities support services ensure an individual's health and safety, encourage participation in the community, increase opportunities for meaningful employment, and provide residential services and support from early childhood through adulthood. The services an individual receives are based on his or her needs, and are documented in an Individual Service Plan. Public funding may be provided at the state or county level and also reimbursed by Medicaid. See *All Payers Health Claims* above for more information on commonly collected health data.

### 10. Emergency Medical Services (EMS) (Low)

EMS are out-of-hospital acute medical care, transport to definitive care, and other medical transport to patients with illnesses and injuries that prevent them from transporting themselves. EMS data are collected by local EMS agencies. Data commonly collected include (a) information about agencies, (b) the unit/call information, (c) dates/times of the call, response, and incident, (d) patient characteristics, and (e) characteristics of the medical situation and response (National EMS Information System, n.d.).

## C. Child Welfare

### 1. Abuse and Neglect, Out-of-Home Care, In-Home Services (High)

Public child protective service agencies are charged with serving children who have allegedly been abused or neglected. States capture data on child welfare in terms of individual children's experiences of abuse and neglect, out-of-home care, and in-home services. Data collected include (a) person identifiers that uniquely identify individuals related to a case (e.g., child, caregiver, perpetrator), such as name, date of birth, and Social Security number, (b) person descriptors that describe the individuals related to the case (e.g., race, ethnicity, gender), (c) event information, including the type of event (e.g., report, investigation, disposition, service, out-of-home placement) and the location, and (d) timing information to establish when events occurred (Center for State Child Welfare Data/Chapin Hall, 2016).

## D. Early Childhood

### 1. Child Care Development Fund (CCDF) (Low)

The CCDF is a source of funding for states, territories, and tribes to provide child care for low-income family members so they can work or attend school or job training, and to provide child protective services. Data collected include (a) families' demographics, (b) types of care, (c) reasons for receiving care, (d) time spent in care, (e) amount of subsidies, and (f) family reported income and other public support (Research Connections, Child Care & Early Education, 2009).

### 2. Early Intervention (Low)

Early intervention (EI) refers to federally funded services provided to help children from birth through two years of age with mental or physical disabilities and their families. Data collected include (a) early intervention service(s) provided, (b) reason for service(s) ending, (c) eligibility and use of preschool services, and (d) child demographics, including race/ethnicity, limited English proficiency status, gender, disability category, and risk of having substantial developmental delays (Individuals with Disabilities Education Act, 2004).

## E. Education

### 1. K-12 Education (Medium)

Education aims to support the development of individuals' human capital from kindergarten through high school. Each state and local education authority collects information about their students. The National Center for Education Statistics' Common Education Data Standards (CEDS) provide "a set of commonly agreed upon names, definitions, option sets, and technical specifications for a given selection of data elements" (CEDS, 2015). Data commonly collected include (a) student demographics, (b) enrollment information, (c) academic assessment information, (d) disciplinary action information, and (e) exit information.

### 2. Postsecondary Education (Medium)

Postsecondary education aims to support the development of individuals' human capital following compulsory education. Data commonly collected include (a) student demographics, (b) admission information, (c) enrollment information, and (d) exit information.

### 3. K-12 Special Education (Medium)

Special education is purposefully designed instruction to meet the unique needs of a child with a disability, provided at no cost to families. In order to receive special education funding through the Individuals with Disabilities Education (IDEA) Act, states must collect data on children served. The CEDS provide information about common data elements recorded in the area of special education. Data commonly collected include (a) student demographics, (b) disability category, (c) timing of disability diagnosis, (d) special education participation, and (e) timing of special education entry and exit (CEDS, 2015).

## F. Juvenile Justice

### 1. Juvenile Justice Services (High)

Juvenile justice services include intervention activities to support youth involved with the justice system (e.g., prevention, rehabilitation, detention). Local and state agencies are responsible for providing juvenile justice services. Data commonly collected include (a) demographics of the juvenile, (b) dates of involvement in a service, and (c) service type (National Center for Juvenile Justice, n.d.).

### 2. Juvenile Courts (Low)

Juvenile courts aim to divert young offenders from the criminal courts and encourage rehabilitation based on the individual needs. County juvenile court systems are responsible for maintaining client-tracking or case-reporting information systems. Data collected commonly include (a) demographics of the referred youth, (b) the date and source of referral, (c) the offenses charged, (d) detention, (e) petitioning, and (f) the date and type of disposition (National Juvenile Court Data Archive, 2014).

## G. Adult Justice

### 1. City or County Jail (High)

City or county jails are correctional facilities that confine adult offenders and juveniles under certain circumstances who are awaiting trial or sentenced to one year (12 months) or less. These facilities are run by a local law enforcement agency, such as a sheriff's office or local corrections department, which maintains data on the jail population. Data typically collected include (a) demographics, (b) dates of entry and release, and (c) reason for release.

### 2. State Corrections (Medium)

State corrections refers to the supervision of individuals arrested for, convicted of, or sentenced for criminal offenses. States collect data on those under such supervision. The National Corrections Reporting Program for the Bureau of Justice Statistics collects information from states annually to create standardized national data. Data commonly collected include (a) prison admissions and releases, (b) parole entries and discharges, (c) demographic information, (d) conviction offenses, (e) sentence length, (f) minimum time to be served, (g) credited jail time, (h) type of admission, (i) type of release, and (j) time served (National Archive of Criminal Justice Data, n.d.).

### 3. Law Enforcement (Medium)

Law enforcement agencies are responsible for the prevention, detection, and investigation of crime, and the apprehension and detention of individuals suspected of law violation. Local agencies collect data related to these activities. The National Incidence-Based Reporting System and the National Crime Statistics Exchange are federal efforts to support data standardization.



These systems call for data collection to include (a) demographics of the victim, offender, and arrestee, (b) types of victims and offenders, (c) characteristics of the incident and arrest, and (d) dates and location information (Bureau of Justice Statistics, 2014).

## H. Employment

### 1. Workforce Training Programs (Medium)

Workforce training programs, such as Job Corps, are designed to support individuals who are looking for employment but do not have the financial resources for job search, training, and placement services. Typically, these programs are operated by state or local agencies. Recently, the U.S. Departments of Education and Labor have provided Statewide Longitudinal Data System and Workforce Data Quality Initiative grants to build data systems linking education, workforce training, and employment records. The CEDS (described in the K-12 education section) provide data standards that can be applied to workforce training program data. CEDS workforce data include (a) program participant demographics, (b) program enrollment information and credentials earned, and (c) post-participation employment (CEDS, 2015).

### 2. Unemployment Insurance (UI) Wages (Medium)

The Federal-State Unemployment Insurance Program provides unemployment benefits to eligible workers who are unemployed through no fault of their own, and meet state eligibility requirements. Each state collects information on those receiving unemployment insurance. This information has two components: UI benefits data, and UI wage record data (i.e., linked earnings data). UI benefits data include (a) financial information, including benefits paid, initial claims, first payments, and weeks compensated, and (b) recipient demographics, including gender, ethnicity, race, and age (U.S. Department of Labor, Employment & Training Administration, 2016). UI Wage Record data include quarterly data on individual employment and earnings. Subject to state-level data use and confidentiality restrictions, this UI Wage Record data can be linked to other administrative data to assess the employment and earnings outcomes of various policy interventions (U.S. Department of Labor, Employment & Training Administration, 1997).

## I. Public Assistance

### 1. Temporary Assistance for Needy Families (TANF) (High)

The TANF program is designed to help needy families achieve self-sufficiency (Administration for Children and Families [ACF], 2016). States receive block grants to design and operate TANF programs. Each state reports data to the ACF, Office of Family Assistance. Data collected include (a) TANF recipients' employment and earnings, (b) characteristics and financial circumstances of TANF recipients, (c) program expenditures and finances, (d) program performance measures, and (e) interactions between TANF and child support (Office of Management and Budget, Office of Information and Regulatory Affairs, 2008).

### 2. Supplemental Nutrition Assistance Program (SNAP) (High)

SNAP is the largest nutrition assistance program for low-income individuals and families (U.S. Department of Agriculture [USDA], 2016). Each state operates the program in its area and reports monthly and annual data to the USDA Food and Nutrition Service. Data collected typically include (a) services provided, (b) individual and household demographics, (c) participation characteristics, and (d) costs (USDA, 2016).

### 3. Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) (Low)

WIC provides supplemental foods, screening and referrals to health care and other social services, and nutrition education to low-income pregnant, breastfeeding, and non-breastfeeding mothers, and to children up to age five who are at risk nutritionally (USDA, Food and Nutrition Service, 2008). States administer the program and report monthly and annual data to the USDA Food and Nutrition Service. Data collected include (a) number of pregnant women participating, (b) number of women fully breastfeeding and partially breastfeeding, (c) total number of breastfeeding women, (d) number of postpartum women, (e) total number of women, (f) number of infants who are fully and partially breastfed, (g) number of infants who are fully formula-fed, (h) total number of infants, (i) total number of children, (j) total number of participants, (k) average food cost per person, (l) food costs, (m) total amount in rebates, and (n) cumulative cost of nutrition services and administration (USDA, Food and Nutrition Service, 2008).

## J. Homelessness/Housing

### 1. Homeless Management Information System (HMIS) (High)

HMIS databases collect information on the provision of housing and services to homeless individuals and families and persons at risk of homelessness as well as data on the clients served (U.S. Department of Housing and Urban Development [HUD], 2016a). The Homeless Emergency Assistance and Rapid Transition to Housing Act of 2009 requires all communities to have an HMIS. Thus, HMIS are locally run databases. In 2014, HUD, Health and Human Services, and Veterans Affairs released an HMIS Data Dictionary and Data Manual outlining data element requirements. Required data include (a) data elements that allow for the ability to record unique, unduplicated individual records (e.g., name, Social Security number, date of birth), (b) participation in a homelessness service, (c) individuals present for each homeless episode, and (d) length of stay (i.e., shelter entry and exit dates) (HUD, 2016a).

### 2. Public Housing Agency (PHA) (Medium)

Funded through HUD, PHAs provide rental assistance to low-income families in the private rental market. HUD requires local PHAs to collect and provide annual data on the "Picture of Subsidized Households." Data collected include (a) public housing occupants, including race, gender, and age, (b) family characteristics, including average household income, two-parent households, and single-parent households, (c) assistance characteristics, such as how long they spent on the waiting list and how long since they had moved in, and (d) housing characteristics, such as the number of bedrooms in the unit (HUD, 2016b).

### 3. Education Homeless Records (Medium)

The McKinney-Vento Education of Homeless Children and Youth Assistance Act ensures that homeless students have equal access to education opportunities as their more affluent peers, including the revision of compulsory residency laws and the provision of education and health services that are necessary for student achievement. Individual schools and local educational agencies collect data relevant to the services provided under the Act. Data collected include (a) student demographics, such as migratory, IDEA, and limited English proficient status, (b) homelessness status, (c) primary nighttime residence, (d) services received from the state's McKinney-Vento programs, (e) whether the student is unaccompanied by a parent or legal guardian, and (f) start and end dates of the homelessness episode (CEDS, 2015).

## IV. Standards for Data Sources

Data standardization is critical for IDS, because it allows for comparison of similar data across sources *within* an IDS, as well as uniformity in the definition of variables *across* IDS when interested in cross-site comparisons. This section of the report provides a detailed listing of standard data elements that can be accessed in the data sources identified above. A data element is defined as information that has been recorded on individuals who have had an encounter at a particular data-sharing agency, and standard data elements are those that can be expected to exist across jurisdictions. It is common to find differences across jurisdictions in the data-recording process (e.g., person recording the information, formatting and naming of data elements), but the meaning behind the data elements must be consistent across jurisdictions.

The standard data elements from the data sources are presented in Appendix C. The data elements listed in this table are not meant to be exhaustive of what information is available. Nor does it reflect the variation that exists among counties and states in terms of the level of detail of the information they collect. Rather, the goal is to surface a minimal set of common data elements that are widely available and for which standards have been articulated. In cases where national data standards have not been specified (e.g., SNAP), commonly used data elements across AISP sites are provided. This effort will help cultivate a universal set of minimum data elements for an IDS that includes any particular data source. The standard data elements are categorized by the following units of analysis that are typically available in these data sources:

### A. Person

For an IDS to be used effectively, the person has to be well defined. How one sees a person in the context of narrative depends on a number of factors—age (child vs. adult); the program; membership in a family, a household, or both—but the person at the center of those questions is unique. A unique identity can be maintained in a number of ways, but it must be unique if the other parts of the data are going to tie back to a given individual. Along with identifying the person, data sources contain recorded characteristics of the individuals identified in the data source (e.g., demographics, income, marital status).

### B. Encounter

People served by public systems encounter those systems in myriad ways, both formal and informal. Encounters are a primary unit of analysis—who was involved, why, how long, for what reason, and toward what end. Most critically, for encounters to fit the human narrative, each encounter has a timestamp that nests that encounter within the life course perspective. Going to school, attending class in a community college, being investigated for child abuse and neglect, seeing a doctor, and applying for TANF are all types of encounters that an IDS could integrate. As with the person, the encounter has descriptors that are recorded characteristics of the event, such as results from a child abuse investigation or screening results from a doctor visit.

### C. Place

Context is usually in reference to place, but it need not be. Place is geo-social in that the interplay between physical geography and social life is a large part of culture. In a social environment, political and administrative boundaries are important constructs when aggregating people in the way needed to understand the interplay of policy, practice, and context. Organizations are akin to places in that they represent an aggregation of people and infrastructure: people receive services from organizations, people work for organizations, organizations approach the same work differently, organizations are more or less effective. Organizations have attributes just as people do; an IDS cannot be limited to knowing more about people without also knowing more about the organizations that serve them.

Organizations can be independent business units such as non-profits or government agencies. Within an organization, there may be subunits made up of offices, units, or workers connected to each other through a common supervisor.

### D. Time

From a public health perspective, a well-designed IDS provides robust estimates of both incidence and prevalence from the same data set. For this to happen, time requires special attention. Structurally, “through time” and “in time” are the essential design requirements: one has to see in the data how a person changes through time; one also has to know something about everybody present at a moment in time. Because what is true at a moment in time is inextricably linked to who was in the system at that time, every unit within the IDS has to be connected to the other units in time.

## V. Standard Data Repurposing Process for IDS

Repurposing data for an IDS requires obtaining data, preparing data for inclusion, importing data, and setting up automation to repeat these steps in perpetuity. Maintaining good documentation for each step of this process is critical to the long-term health of an IDS. This section primarily focuses on preparing a new source of data for import, as this step can be time-consuming and complex. Tasks described here may also be useful when preparing existing data in an IDS for additional research objectives. Preparing a new source of data for import involves profiling the data source and planning how the data will be transformed (i.e., cleaned, restructured, and merged; performing the data transformations and documenting the entire process). This section details key considerations during this process and concludes with a brief statement on the use of data already in an IDS. The material in this section has been derived from key research on IDS and data quality (Fantuzzo and Culhane, 2015; Hellerstein, 2008; Keller et al., 2016; Wickham, 2014).

### A. Prerequisites

Before beginning the process of preparing a new source of data for import into an IDS, several pieces of information are needed. First, it is necessary to have documentation and a clear understanding of the goal(s) for adding the new data source. This will help to clarify how these new data may need to be adapted before bringing them into the IDS. Second, access to, or a copy of, the source data is required. Lastly, it is important to have documentation about the source data, if it exists, and access to people who are familiar with them. This will help resolve questions about how the data have been collected, what they mean, and how they are used.

### B. Profiling Source Data

Data profiling captures the data structure and quality of new data sources for the IDS. During this step, it is important to identify and describe issues with the data, and equally important not to resolve these issues until later in this process. Many extract-transform-load software tools include at least some ability to automatically profile data, but human review is essential.

#### 1. Data Structure

Provenance and metadata are of vital importance to understanding data structure. *Provenance* is the history and process of data collection and maintenance. It describes where the data came from and what the data are, including inception, history of access, transmission, or modification in terms of both what operations were performed and by whom. It provides a context for better understanding, interpretation, and inference. *Metadata* are a way of tracking whether data sources

(data sets and tables), their attributes (fields/columns), and their observation units (records/rows) are consistently named, sufficiently described, and appropriately formatted for analysis and for combination with other data.<sup>3</sup> At a minimum, metadata should include the following:

- ❖ **Data sources:** a name, high-level description of the included data, contact information for the owners/maintainers, the date the data were last updated, expected duration between updates, and the data's provenance. If the data have restrictions imposed by copyright, contracts, or other agreements, such as retention period or terms of use, these should also be noted.
- ❖ **Attributes:** A list of fields/columns, including names, description of what they capture, types of data (for example, "date," "integer," "true/false"), expected values, and provenance.

Provenance and metadata are often missing in early data requests if codebooks and other documentation are not made available. In that case, the metadata will need to be developed and documented through the profiling process. Provenance, if not provided, will not be easily recreated. In addition, if the data have been acquired from commercial data aggregators, the provenance is frequently considered proprietary and will not be made available.

## 2. Data Quality

Once the provenance and metadata have been gathered, they should be reviewed to help identify systemic challenges with the new data. A review of the documented data structure (without yet looking at the actual data) will reveal the following insights:

- ❖ **Relevance:** The source data may include columns or tables beyond what is needed to meet the goal of their inclusion in the IDS. Identifying irrelevant data at this step reduces the volume of data quality assessment.
- ❖ **Missing field names or descriptions:** Missing metadata may result in valuable information being excluded or unused.
- ❖ **Combined fields:** A column may represent more than one type of data, particularly if the data contain codes or abbreviations, or were open to free text entry.
- ❖ **Multiple structural directions:** The structure of data is defined through both columns and rows. This is very common when data are extracted from spreadsheets or fixed-width outputs designed for printing or viewing on older terminal screens.
- ❖ **Divided/duplicated values:** The same data element may exist in multiple tables from different data sources. This introduces a challenge to data quality of reconciling potentially different values stored in those duplicative fields.

Following a review of the data structure, the data are reviewed. Generally, data are evaluated for completeness, value validity, default values, consistency, uniqueness, and duplication, though deeper analyses may be needed.

*Completeness* is a characterization of missing data, and is application-specific. A set of data is complete with respect to a given purpose if the set contains all the relevant data for that purpose. Data that are missing can be categorized as record fields not containing data, records not containing necessary fields, or data sets not containing the requisite records. A common measure of completeness is the proportion of the data that has values to the proportion of data that should have values.

Data elements with proper values should also have *value validity*. The percentage of data records that possess values within the range expected for a legitimate entry is a measure of value validity. Checking for value validity generally comes in the form of straightforward domain constraint rules. Invalid values

<sup>3</sup> See metadata examples from Chapin Hall ([http://www.chapinhall.org/news/spotlight/chicago\\_data\\_dictionary](http://www.chapinhall.org/news/spotlight/chicago_data_dictionary)) and Washington (<https://data.wa.gov/browse>).

occur most frequently in free text entry columns, but are not limited to this scenario. For example, a record may have invalid data if a yes/no field is left blank or contains a value other than "yes" or "no"; a date of birth is set in the future; or an age is more than 120 years. Invalid data may still contain usable information, depending on the intended use. For example, if the question at hand simply requires a count of how many properties are "residential," it may be possible to transform existing entries with incorrect apartment numbers to adequately represent whether or not properties are "residential." Some data systems will automatically assign *default values* in cases where data have not explicitly been entered. For example, a new case record might automatically be given a status of "open" if a different value isn't specified. Default values should be well documented so they may be factored into any analyses. In extreme cases, they can make portions of records meaningless.

The degree of logical agreement between data values in either a single data set or between two or more data sets is *consistency*. Consistency is generally checked through two mechanisms. The first is the use of a master list. A common source of inconsistency comes from situations in which locally derived information (such as a list of clients) is provided with no associated master list or file. This is extremely common in human and health services, where multiple data systems are capturing information about the same people. A second mechanism for checking consistency is the creation of dependency constraints that specify logical relationships between different types of values. A simple example of a dependency constraint violation would be a location disagreement like a zip code that does not agree with a state code. Another might be the identification of a male who is also pregnant. Causes of inconsistency are varied.

The number of unique valid values that have been entered in a record field, or as a combination of record field values within a data set, is *uniqueness*. Uniqueness is not generally associated with data quality, but for answering research questions, the variety and richness of the data are of paramount importance. If a data set column has very little value uniqueness (for example, entries in the field "state" for an analysis of housing within a single county), then its utility is quite low and it can be considered of low relevance or quality in terms of the goal(s) in mind. In contrast, *duplication* refers to the degree of replication of distinct observations per observation unit type. For example, in state-level secondary-education registration records, greater than 1 registration per student per official reporting period would represent duplication. While duplication can occur as a result of the accidental entering of the same information multiple times, duplication can occur many times as a direct result of the choice of level of aggregation, e.g., aggregating to a single student registration per academic year when registration information is actually collected multiple times per academic year.

Once the data source has been profiled, the next step is to plan how it will be merged with other data in the IDS and prepared for import.

## C. Data Transformation

Planning how the new data will be changed before incorporating them into the IDS allows all the issues identified during data profiling to be addressed. Perhaps more importantly, it also explicitly defines how the incoming data will be linked with data that are already in the IDS or coming from other data sources.

### 1. Merging Data

Merging data is the combining of information across multiple sources. This can involve reconciling definitions and descriptions between the sources, as well as directly linking entities within the data sources. Merging data from new data sources into an IDS involves two activities: ontology mapping and record linking.

*Ontology mapping* matches consistent record values from different data sources. For example, one system may use "resolved" to indicate that a case is no longer open, whereas another may use "completed," but the IDS may use "closed." Ontology mapping may also be helpful in merging

records of different types. For example, one data set may represent condominiums by the quantity of units within a single building, as a single record. Another data set may represent condominiums as a collection of records of single owner-occupied units. To link these two data sets, the addresses in both data sources are used to aggregate into one record all the condominiums associated with a physical location. As an IDS matures, it will have its own ontology, which is different from the original data sources— which is why documenting the data as they exist in the IDS is very important.

*Record linking* identifies records in a data set that refer to the same entity in other data sets that have already been incorporated into the IDS. Record linkage is crucial, if not foundational, to the entire data science process. In social services, the most common link between records is a person or client, but there are others, such as geographic area, caseworker, and service(s) provided. While identifying opportunities for record linkage is relatively easy, accomplishing the linking is far more challenging, which is why it must be planned in advance. Often, multiple steps are required to achieve success, including multiple matching and validation steps such as deterministic geocode matching, probabilistic name matching, and human review (Kumar, 2015).

## 2. Restructuring

To address issues of structure discovered during data profiling, it may be necessary to restructure the data into multiple new data sets that are more facile, or align to existing data structures already in the IDS. This activity is akin to database normalization—the process of organizing the columns (attributes) and tables (relations) of a relational database to minimize data redundancy. It also includes rescaling to better align the data to the proposed goals for use of the data.

*Normalization* is used to bring a data field or variable to a common scale. This can include simple standardization or a more complicated shifting of scales to facilitate comparisons across other data sources in the IDS. *Feature extraction and construction* can be used to create new and useful variables.

## 3. Cleaning

Data cleaning is the process of resolving previously identified quality issues. This step may involve planning how to fix or remove data that are incorrect, incomplete, improperly formatted, or duplicated. Data profiling identifies what data need to be cleaned; this planning step defines *how* it will be addressed. Developing the cleaning process after planning the linking strategy ensures that necessary data aren't removed. A plan for cleaning data typically includes handling the following:

- ❖ **Missing values:** These can be inferred from other data, reconstructed during human review, or automatically set to a default. If the missing values are important to the goal, this step cannot be overlooked.
- ❖ **Date and time formatting:** This ensures that the source data are adjusted to IDS requirements. This may require splitting date values into separate year, month, and day fields, or it may require combining dates and times into one value.
- ❖ **De-duplication:** This detects and then merges or deletes all but one unique data record, according to the application of some algorithm for determining whether data contain duplicates.
- ❖ **Outlier reconciliation:** This resolves data that are beyond the expected range of values as identified in the data profiling value validity step. Corrective actions may include removing the value, adjusting it by hand, or coercing it to be a valid value through the use of an algorithm.

## 4. Transforming

Once the new data source is profiled and plans for restructuring and cleaning are in place, it is time to carry out the transformation of the data. Most transformation is performed using automated tools (e.g., various Microsoft, Oracle, and IBM products) so that the processes can be scripted and repeated as updated information becomes available from data sources. However, while modern machine-learning-based tools are becoming quite sophisticated, they cannot yet replace human review. Processing errors during the data transformation phase may identify new challenges that were not previously recognized during the data profiling and planning steps. The documentation developed during those previous steps should be updated as new issues are observed and corrected.

## 5. Using Existing IDS Data

Just as new source data must be profiled before they are included in the IDS, the data within the IDS must be profiled as well. This helps IDS users understand the information that has been incorporated, identify pitfalls and information gaps, and trace data back to original sources. Furthermore, data that have already been brought into the IDS are often repurposed as new goals and research opportunities are identified.

# VI. Conclusions

Within policy circles and academic research centers, there is a growing appreciation for the utility embedded in administrative data systems. Yet the full utility can only be realized when data holdings are brought together in a deliberate fashion with a clear vision in mind. One significant hurdle to overcome is resolving the inherent differences that exist between and within agencies at the local, state, and national levels in the design of data capture systems. Historically, these differences have been a barrier to the productive use of data for research and other types of operational support.

The work of IDS nationally demonstrates that these differences can be overcome. Legal, governance, and technical challenges are addressed in other papers in this series. Here, our focus has been on the processes that define access to and processing of administrative records. Our organizing principles are pragmatic. IDS promise considerable utility provided the issues of harmonization over people, place, and time can be resolved. Our recommendations are centered on a clear understanding of the issues at hand. First and foremost, an IDS is meant to inform how well systems serve people. If we want to know how people are doing and whether public investments help or hinder development over the life course, then people and their development become the core rationale for resolving differences in data systems. Fortunately, human experience is a powerful organizing framework. Human capital formation over the life course provides the conceptual structure needed to understand how place, time, and systems are represented in what happens to people and why.

Harmonization of disparate data structures is about decision making. When differences exist, the resolution of those differences has to follow a rigorous process that is replicable. The framework we provide is not meant to be rigid or formulaic. We see building an IDS as an evolutionary process prompted in many cases by opportunities that emerge in local, issue-specific contexts. Nevertheless, the lives people live are interconnected. They touch and are touched by diverse interactions with diverse systems. Policy-makers and scientists cannot expect to maximize the utility of programs without seeing those interconnections. For that reason, short-term opportunities have to be guided by what the long term is likely to demand: a holistic view of investments meant to improve the human condition. For that goal to be realized, it is best to start with a set of standards that shape how each locality approaches its work. As a starting point, the standards laid out here are meant to guide the work done today so that the promise of integrated data systems that eclipse the boundaries of person, place, system, and time can be realized.

## References

Fantuzzo, John, and Dennis P. Culhane (Eds.). (2015). *Actionable Intelligence: Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government*. New York: Palgrave Macmillan US.

Gibbs, Linda, Amy Hawn Nelson, Erin Dalton, Joel Cantor, Stephanie Shipp, and Della Jenkins. (2017). *IDS Governance: Setting Up for Ethical and Effective Use*. Actionable Intelligence for Social Policy, Expert Panel Report, University of Pennsylvania.

Hellerstein, Joseph M. (2008). Quantitative data cleaning for large databases. United Nations Economic Commission for Europe. <http://db.cs.berkeley.edu/jmh/papers/cleaning-unece.pdf>

Keller, Sallie, Stephanie Shipp, Mark Orr, Dave Higdon, Gizem Korkmaz, Aaron Schroeder, Emily Molfino, Bianica Pires, Kathryn Ziemer, and Daniel Weinberg. (2016). *Leveraging External Data Sources to Enhance Official Statistics and Products*. Report prepared for the U.S. Census Bureau. Arlington, VA: Social and Decision Analytics Laboratory (SDAL), Biocomplexity Institute of Virginia Tech. <http://cdn.vbi.vt.edu/mc/SDAL/leveraging-external-data-sdal-2016.pdf>

Kumar, Prashant. (2015). An overview of architectures and techniques for Integrated Data Systems Implementation. In John Fantuzzo and Dennis P. Culhane, Eds., *Actionable Intelligence: Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government*, pp. 105-124. New York: Palgrave Macmillan US.

NYC Children's Cabinet. (2016). *Growing Up NYC: A Policy Framework*. City of New York, NYC Children's Cabinet. [http://s-media.nyc.gov/agencies/childrenscabinet/NYCDOH\\_GrowingUP\\_Policy\\_Brochure\\_For\\_WEB.pdf](http://s-media.nyc.gov/agencies/childrenscabinet/NYCDOH_GrowingUP_Policy_Brochure_For_WEB.pdf)

Shank, Nancy C. (2009). *Understanding Human Services Utilization: Opportunities for Data Sharing between Federally Funded Programs*. University of Nebraska-Lincoln Public Policy Center, Nancy Shank Publications. Paper 6. <http://digitalcommons.unl.edu/publicpolicys Shank/6>

Wickham, Hadley. (2014). Tidy data. *Journal of Statistical Software*, 59(10). <http://www.jstatsoft.org/article/view/v059i10>

### Vital Statistics

#### Birth Records:

Centers for Disease Control and Prevention. (2014). *User Guide to the 2014 Natality Public Use File*. [http://www.cdc.gov/nchs/data\\_access/vitalstatsonline.htm](http://www.cdc.gov/nchs/data_access/vitalstatsonline.htm)

#### Death Records:

Centers for Disease Control and Prevention. (2014). *User Guide to the 2014 Mortality Multiple Cause-of-Death Public Use Record*. [http://www.cdc.gov/nchs/data\\_access/vitalstatsonline.htm](http://www.cdc.gov/nchs/data_access/vitalstatsonline.htm)

### Healthcare Utilization

#### Medicaid:

Medicaid and CHIP Data Collection Systems. (n.d.). <https://www.medicaid.gov/medicaid/data-and-systems/collection-systems/index.html>

#### Mental Health:

Mental Health America. (n.d.). *The Federal and State Role in Mental Health*. <http://www.mentalhealthamerica.net/issues/federal-and-state-role-mental-health>

#### Alcohol and Substance Abuse:

Substance Abuse and Mental Health Services Administration. (n.d.). <http://www.samhsa.gov/>

#### State Children's Health Insurance Program (SCHIP):

Medicaid and CHIP Data Collection Systems. (n.d.). <https://www.medicaid.gov/medicaid/data-and-systems/collection-systems/index.html>

#### Nursing Facility Minimum Data Set:

Centers for Medicare & Medicaid Services. (2016, October). *Long-Term Care Facility Resident Assessment Instrument 3.0 User's Manual, Version 1.14*. [https://downloads.cms.gov/files/draft\\_mds\\_30\\_rai\\_manual\\_v114\\_may\\_2016.pdf](https://downloads.cms.gov/files/draft_mds_30_rai_manual_v114_may_2016.pdf)

#### All Payer Claims:

All-Payer Claims Database Council. (2011). *All-Payer Claims Database Council Proposed CORE Set of Data Elements*. <http://www.apcdouncil.org/standards>

#### Emergency Medical Services (EMS):

National EMS Information System [NEMESIS]. (n.d.). *NEMESIS Data Dictionary v2.2.1*. <http://nemsis.org/v2/downloads/datasetDictionaries.html>

### Child Welfare

Center for State Child Welfare Data/Chapin Hall. (2016). *The Multistate Child Welfare Database*.

### Early Childhood Services

#### Child Care and Development Fund:

Research Connections, Child Care & Early Education. (2009). *Child Care and Development Fund Policies Database*. <http://www.researchconnections.org/childcare/studies/32261> UED

#### Early Intervention:

Individuals with Disabilities Education Act, Part C, Sec. 631, as amended; 20 U.S.C. 1431 et seq. 2004.

### Education

#### K-20 Education:

Common Education Data Standards [CEDS]. (2015). *Common Education Data Elements Version 6*. <https://ceds.ed.gov/elementsCEDS.aspx>

#### K-12 Special Education (IDEA):

Common Education Data Standards [CEDS]. (2015). *Common Education Data Elements Version 6*. <https://ceds.ed.gov/elementsCEDS.aspx>

### Juvenile Justice

#### Juvenile Courts:

National Juvenile Court Data Archive. (2014). *Data Set User's Guides*. <http://www.ojjdp.gov/ojstatbb/njcda/asp/guide.asp>

*Juvenile Justice Services:*

National Center for Juvenile Justice. (n.d.). *National Projects: Juvenile Justice Model Data Project*. [http://www.ncjj.org/Projects/National\\_Projects.aspx](http://www.ncjj.org/Projects/National_Projects.aspx)

**Adult Justice/Incarceration***Law Enforcement:*

Bureau of Justice Statistics. (2014). *Data Collection: National Incident-Based Reporting System (NIBRS)*. <http://www.bjs.gov/index.cfm?ty=dcdetail&iid=301#Documentation>

*State Corrections:*

National Archive of Criminal Justice Data. (n.d.). *National Corrections Reporting Program Variable List*. <http://www.icpsr.umich.edu/icpsrweb/NACJD/ssvd/series/38/variables>

**Employment***Unemployment Insurance (UI) Wages:*

U.S. Department of Labor, Employment & Training Administration. (2016). *Unemployment Insurance Data Summary*. <http://oui.doleta.gov/unemploy/content/data.asp>

U.S. Department of Labor, Employment & Training Administration. (1997). *State Unemployment Insurance Program Wage Records: Access and Use Issues*. <https://wdr.doleta.gov/opr/fulltext/document.cfm?docn=5809>

*Workforce training programs:*

Common Education Data Standards [CEDs]. (2015). *Common Education Data Elements Version 6*. <https://ceds.ed.gov/elementsCEDs.aspx>

**Public Assistance***Temporary Assistance for Needy Families (TANF):*

Office of Management and Budget, Office of Information and Regulatory Affairs. (2008). *TANF Data Report for Families Receiving Assistance under the TANF Program: Instructions and Definitions*. <https://www.reginfo.gov/public/do/DownloadDocument?objectID=24411801>

Administration for Children and Families [ACF]. (2016). *Temporary Assistance for Needy Families (TANF)*. <https://www.acf.hhs.gov/ofa/programs/tanf/about>

*Supplemental Nutrition Assistance Program (SNAP):*

U.S. Department of Agriculture [USDA]. (2016). *Supplemental Nutrition Assistance Program (SNAP)*. <http://www.fns.usda.gov/snap/supplemental-nutrition-assistance-program-snap>

*Women, Infants, and Children (WIC):*

U.S. Department of Agriculture [USDA], Food and Nutrition Service. (2008). *Functional Requirements Document for a Model WIC Information System*. [https://www.fns.usda.gov/sites/default/files/4.2\\_Data\\_Code\\_Tables.pdf](https://www.fns.usda.gov/sites/default/files/4.2_Data_Code_Tables.pdf)

**Homelessness and Public Housing***Homeless Management Information System (HMIS):*

U.S. Department of Housing and Urban Development [HUD]. (2016a). *HMIS Data Standards Data Manual*. <https://www.hudexchange.info/resource/3826/hmis-data-standards-manual/>

*Education Homeless Records (McKinney-Vento):*

Common Education Data Standards [CEDs]. (2015). *Common Education Data Elements Version 6*. <https://ceds.ed.gov/elementsCEDs.aspx>

*Public Housing Agency (HUD):*

U.S. Department of Housing and Urban Development [HUD]. (2016b). *Family Report, Form 50058 and Owner's Certification of Compliance with HUD's Tenant Eligibility and Rent Procedures, Form 50059*. [https://portal.hud.gov/hudportal/HUD?src=/program\\_offices/administration/hudclips/forms/hud5](https://portal.hud.gov/hudportal/HUD?src=/program_offices/administration/hudclips/forms/hud5)

THE OPPORTUNITY FRAMEWORK		
LIFE STAGES	INDICATORS <sup>1</sup>	
<p>Planned pregnancy, born at healthy birth weight to a dual parent household without maternal depression</p> <p><b>FAMILY FORMATION</b></p> <p>Conception to birth</p>	<ul style="list-style-type: none"> <li>Rate of low birth weight</li> <li>Family income</li> <li>Maternal depression</li> <li>Single- or dual-parent household</li> <li>Intendedness of pregnancy</li> <li>Early screening and intervention - maternal</li> </ul>	<p><b>INDIVIDUALS LIVING WITH DISABILITIES<sup>3</sup></b></p> <ul style="list-style-type: none"> <li>Family-directed planning</li> <li>Living in the community</li> <li>Access to assistive technology</li> <li>Receiving needed services</li> <li>Person-directed planning</li> <li>Living in the community</li> <li>Access to assistive technology</li> <li>Receiving needed services</li> </ul>
<p>School readiness, healthy social and emotional skills &amp; family access to affordable, nutritious food</p> <p><b>EARLY CHILDHOOD</b></p> <p>Ages 0 - 5</p>	<ul style="list-style-type: none"> <li>% of parents concerned about child's emotions, concentration, behavior or ability to get along with others</li> <li>% of families relying on low cost food</li> <li>School readiness</li> <li>Early screening and intervention - child</li> </ul>	
<p>Math/reading skills &amp; healthy social emotional skills</p> <p><b>MIDDLE CHILDHOOD</b></p> <p>Ages 6 - 11</p>	<ul style="list-style-type: none"> <li>Standardized test: math scores</li> <li>Standardized test: reading scores</li> <li>% of parents concerned about child's emotions, concentration, behavior or ability to get along with others</li> </ul>	
<p>Graduates from high school on time, develops healthy social emotional skills, not convicted of a crime, not a teen parent</p> <p><b>ADOLESCENCE</b></p> <p>Ages 12 - 17</p>	<ul style="list-style-type: none"> <li>High school graduation status</li> <li>Violent arrest &amp; property arrest rates</li> <li>Teen parent status</li> <li>% of students who report ever feeling sad/hopeless or have considered suicide</li> <li>% of young adults who are currently depressed</li> </ul>	
<p>Sustainably employed, attended post-secondary, education &amp; good physical/mental health</p> <p><b>TRANSITION TO ADULTHOOD</b></p> <p>Ages 18 - 29</p>	<ul style="list-style-type: none"> <li>Employed status of population</li> <li>% FPL/family income</li> <li>Attending post-secondary training/education</li> <li>Avg. # of days poor physical/mental health prevented usual activities, like self-care, work or recreation</li> </ul>	
<p>Employment status, good physical/mental health &amp; middle class household (300% FPL)</p> <p><b>EARLY ADULTHOOD</b></p> <p>Ages 30 - 40</p>	<ul style="list-style-type: none"> <li>Avg. # of days poor physical/mental health prevented usual activities, like self-care, work or recreation</li> <li>% of FPL/family income</li> <li>Employment status of population</li> </ul>	
<p>Independence at home, employment, social connectedness, preventive services</p> <p><b>ADULTHOOD Independent</b></p> <p>Ages 41 and older</p>	<ul style="list-style-type: none"> <li>Employment status</li> <li>Affordable, accessible housing</li> <li>Social connections</li> <li>Primary care provider</li> </ul>	
<p>At risk due to medical complexity: independence at home with support, avoidance of excessive healthcare risks, and caregiver support</p> <p><b>ADULTHOOD With Support<sup>2</sup></b></p> <p>Ages 60 and older</p>	<ul style="list-style-type: none"> <li>Employment status</li> <li>Affordable, accessible housing</li> <li>Social connections</li> <li>Primary care provider</li> <li>Sufficient financial resources</li> <li>Access to a formal or informal caregiver</li> <li>Falls</li> <li>Preventable hospitalizations</li> </ul>	
<p>Quality care that is aligned with personal goals, effective support</p> <p><b>DECLINE &amp; END-OF-LIFE</b></p> <p>Last 12 months</p>	<ul style="list-style-type: none"> <li>Advance care plan</li> <li>Early enrollment and hospice use</li> <li>Falls</li> <li>Dying in preferred place</li> </ul>	

<sup>1</sup> Indicators in the Colorado Opportunity Project Framework have been vetted and finalized by the Departments of Public Health and Environment, Human Services and Health Care Policy & Financing with the support of the Brookings Institution (indicators of Family Formation through Early Adulthood are based on available statewide data and quality measures).  
<sup>2</sup> Eligibility for Older Americans Act programs begins at age 60.  
<sup>3</sup> Living well with disabilities, including effective support and services.

Last Update Nov, 2016

Source: <https://www.colorado.gov/pacific/hcpf/colorado-opportunity-project>

Data Sources	Washington	South Carolina	Rhode Island	Philadelphia, PA	New York, NY	Milwaukee, WI	Mecklenburg, NC	Los Angeles, CA	Florida	Cuyahoga, OH	Allegheny, PA	Access Point
Vital Statistics												
Birth records		X	X								X	County or state
Death records								X		X		County or state
Healthcare Utilization												County or state
All Payer health claims								X				County or state
Alcohol and substance abuse								X				County or state
Community health centers												Local or county
Developmental disabilities									X			County or state
Medicaid								X	X			County or state
Mental health								X	X			County or state
Public health										X		County or state
SCHIP										X		County or state
Nursing facility MDS									X			County
EMS									X			County
Child Welfare												County or state
Out-of-home care									X			County or state
Abuse and neglect									X	X		County or state
In-Home services									X	X		County or state
Early Childhood												County or state
CCDF										X		County
Early intervention										X		County
Education												District or state
K-12 public education												District or state

This table continues on the following page.  
<sup>1</sup> Available only for certain sub-populations.

Data Sources	Access Point	Allegheny, PA	Cuyahoga, OH	Florida	Los Angeles, CA	Mecklenburg, NC	Milwaukee, WI	New York, NY	Philadelphia, PA	Rhode Island <sup>1</sup>	South Carolina	Washington
Education (cont.)												
K-12 special education	County or state		X			X					X	
Postsecondary education	State or federal	X	X						X		X	
Juvenile Delinquency												
Juvenile justice services	County or state	X	X		X			X	X		X	X <sup>1</sup>
Juvenile courts	County											X
Adult Justice												
City or county jail	Local or county	X	X		X	X	X	X	X			X
State corrections	State			X						X	X	X
Law enforcement	Local, county, or state			X			X				X	X
Employment												
Workforce training programs	County or state		X		X			X		X	X	X <sup>1</sup>
UI Wages	State		X		X					X	X	X
Public Assistance												
TANF	County or state	X <sup>1</sup>	X		X	X		X		X	X	X
SNAP	County or state	X <sup>1</sup>	X		X	X		X		X	X	X
WIC	County or state				X					X	X	X
Homelessness/Housing												
HMIS	County or state	X	X		X	X		X	X		X	X
PHA	Local, county or federal	X	X		X	X					X	X
Education homeless records	District, county or state	X			X	X					X	

<sup>1</sup> Available only for certain sub-populations.

Domain and Data Source	Person	Person Descriptor	Encounter	Encounter Descriptor	Place	Time
<b>Vital Statistics</b>						
Birth records	<ul style="list-style-type: none"> <li>• Expectant mother</li> <li>• Expectant father</li> <li>• Child</li> </ul>	<ul style="list-style-type: none"> <li>• Expectant mother                             <ul style="list-style-type: none"> <li>- Age</li> <li>- Marital status</li> <li>- Education</li> <li>- Race/Ethnicity</li> <li>- Sex</li> <li>- Height</li> <li>- Weight—pre, during pregnancy</li> <li>- Did mother get WIC</li> </ul> </li> <li>• Expectant father                             <ul style="list-style-type: none"> <li>- Age</li> <li>- Education</li> <li>- Race/Ethnicity</li> <li>- Sex</li> </ul> </li> <li>• Child                             <ul style="list-style-type: none"> <li>- Sex</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Prenatal care</li> <li>• Birth of a child</li> </ul>	<ul style="list-style-type: none"> <li>• Prenatal care                             <ul style="list-style-type: none"> <li>- Total number of prenatal care visits</li> <li>- Previous live births</li> <li>- Cigarette smoking history</li> <li>- Risk factors in pregnancy during pregnancy</li> <li>- Infections present &amp; treated during pregnancy</li> <li>• Birth of child                                     <ul style="list-style-type: none"> <li>- Obstetric procedures</li> <li>- Onset of labor</li> <li>- Characteristics of labor &amp; delivery</li> <li>- Method of delivery</li> <li>- Maternal morbidity</li> <li>- Abnormal conditions of newborn</li> <li>- Congenital anomalies of newborn</li> <li>- Infant living at time of delivery</li> <li>- Infant birth weight</li> <li>- Infant APGAR scores</li> </ul> </li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Mother's residential address</li> <li>• Facility name code</li> <li>• Type of place of birth code</li> </ul>	<ul style="list-style-type: none"> <li>• Prenatal care                             <ul style="list-style-type: none"> <li>- Date of first visit</li> <li>- Date of last visit</li> <li>• Birth of child                                     <ul style="list-style-type: none"> <li>- Date of birth</li> </ul> </li> </ul> </li> </ul>
Death records	<ul style="list-style-type: none"> <li>• Decedent</li> </ul>	<ul style="list-style-type: none"> <li>• Decedent                             <ul style="list-style-type: none"> <li>- Age</li> <li>- Sex</li> <li>- Race/Ethnicity</li> <li>- Occupation code</li> <li>- Veteran status</li> <li>- Education</li> <li>- Marital status</li> <li>- Pregnancy status</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Injury</li> <li>• Death</li> </ul>	<ul style="list-style-type: none"> <li>• Injury                             <ul style="list-style-type: none"> <li>- Injury at work</li> <li>- Place of injury</li> </ul> </li> <li>• Death                             <ul style="list-style-type: none"> <li>- Referred to coroner/medical examiner</li> <li>- Autopsy performed</li> <li>- Manner of death</li> <li>- Did tobacco contribute to death</li> <li>- Transport accident code</li> <li>- Underlying cause of death</li> <li>- Contributing cause of death</li> <li>- Multiple conditions</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Decedent's address</li> <li>• Facility code</li> <li>• Place of death</li> </ul>	<ul style="list-style-type: none"> <li>• Date of injury</li> <li>• Date of death</li> </ul>



Healthcare Utilization	Person	Person Descriptor	Encounter	Encounter Descriptor	Place	Time
Medicaid Mental Health Alcohol and Substance Abuse Public Health SCHIP All Payer Claims Databases Community Health Centers Developmental Disabilities	<ul style="list-style-type: none"> <li>• Subscriber</li> <li>• Member/ Patient</li> <li>• Service Provider</li> </ul>	<ul style="list-style-type: none"> <li>• Member/Patient</li> <li>- Gender</li> <li>- Age</li> <li>- Individual relationship Code</li> </ul>	<ul style="list-style-type: none"> <li>• Admission</li> <li>• Diagnosis</li> <li>• Procedure</li> <li>• Discharge</li> </ul>	<ul style="list-style-type: none"> <li>• Admission</li> <li>- Admission Type</li> <li>- Admitting Diagnosis</li> <li>- E-Code (Principal Diagnosis)</li> <li>• Diagnosis</li> <li>- Principal Diagnosis</li> <li>- Other Diagnoses (1 to 12)</li> <li>• Procedure</li> <li>- Procedure Code</li> <li>- Procedure Modifier – 1</li> <li>- Procedure Modifier – 2</li> <li>- ICD-9-CM Procedure Code</li> <li>• Discharge</li> <li>- Discharge Status</li> <li>- Charge Amount</li> <li>- Paid Amount</li> <li>- Prepaid Amount</li> <li>- Co-pay Amount</li> <li>- Coinsurance Amount</li> <li>- Deductible Amount</li> <li>- Claim Status</li> <li>- Type of Bill – Institutional</li> <li>- Billing Provider Number</li> <li>- National Billing Provider ID – “Biller”</li> <li>- Biller Name</li> </ul>	<ul style="list-style-type: none"> <li>• Member/Patient</li> <li>- City Name</li> <li>- State or Province</li> <li>- ZIP Code</li> <li>• Service Provider</li> <li>- City Name</li> <li>- State or Province</li> <li>- ZIP Code</li> <li>- Country Name</li> <li>- Number</li> <li>- Tax ID Number</li> <li>- National Service Provider ID</li> <li>- Entity Type Qualifier</li> <li>- First Name</li> <li>- Middle Name</li> <li>- Last Name or Organization Name</li> <li>- Suffix</li> <li>- Speciality</li> </ul>	<ul style="list-style-type: none"> <li>• Admission</li> <li>- Date</li> <li>- Hour</li> <li>• Service</li> <li>- Date of Service</li> <li>- From</li> <li>- Date of Service</li> <li>- Thru</li> <li>- Date Service</li> <li>- Approved/ Accounts Payable Date/Actual Paid Date</li> <li>• Discharge</li> <li>- Date</li> <li>- Hour</li> </ul>
Nursing Facility (MDS)	<ul style="list-style-type: none"> <li>• Resident in Medicare or Medicaid certified nursing facility</li> </ul>	<ul style="list-style-type: none"> <li>• Resident</li> <li>- Age</li> <li>- Race/Ethnicity</li> <li>- Gender</li> <li>- Marital status</li> <li>- Language</li> </ul>	<ul style="list-style-type: none"> <li>• Admission</li> <li>• Assessment</li> <li>• Discharge</li> </ul>	<ul style="list-style-type: none"> <li>• Assessment of resident's...</li> <li>- Hearing, speech, and vision</li> <li>- Cognitive patterns</li> <li>- Mood</li> <li>- Behavior</li> <li>- Functional status</li> <li>- Functional abilities and goals</li> <li>- Bladder and bowel</li> <li>- Health condition</li> <li>- Swallowing/nutritional status</li> <li>- Oral/dental status</li> <li>- Skin conditions</li> <li>- Medications received</li> <li>- Special treatment, procedures, and programs</li> <li>- Restraints used with resident</li> <li>- Participation in assessment and goal setting</li> <li>• Diagnoses</li> <li>- Active diagnoses</li> </ul>	<ul style="list-style-type: none"> <li>• Facility provider</li> </ul>	<ul style="list-style-type: none"> <li>• Admission Date</li> <li>• Assessment Date</li> <li>• Prior MDS assessment Date</li> <li>• Speech-Language, Occupational, Physical Therapy start date</li> <li>• Speech-Language, Occupational, Physical Therapy end date</li> <li>• Discharge Date</li> </ul>

Healthcare Utilization	Person	Person Descriptor	Encounter	Encounter Descriptor	Place	Time
EMS	<ul style="list-style-type: none"> <li>• Patient(s)</li> </ul>	<ul style="list-style-type: none"> <li>• Patient(s)</li> <li>- Gender</li> <li>- Race/Ethnicity</li> <li>- Age</li> </ul>	<ul style="list-style-type: none"> <li>• Emergency call/ Dispatch</li> <li>• Scene/Situation response</li> </ul>	<ul style="list-style-type: none"> <li>• Emergency call/Dispatch</li> <li>- Type of service requested</li> <li>- Type of delay</li> <li>- Dispatch unit number</li> <li>- Complaint reported by dispatch to responding unit</li> <li>• Scene/Situation response</li> <li>- Response mode to scene</li> <li>- Number of patients</li> <li>- Mass casualty incident</li> <li>- Prior aid</li> <li>- Prior aid performed by</li> <li>- Outcome of prior aid</li> <li>- Possible injury</li> <li>- Chief complaint</li> <li>- Primary and associated symptoms</li> <li>- Provider's impression</li> <li>- Cause of injury</li> <li>- Cardiac arrest and etiology</li> <li>- Resuscitation attempted</li> <li>- Barriers to patient care</li> <li>- Alcohol/drug use indicators</li> <li>- Medication given</li> <li>- Medication complication</li> <li>- Procedure</li> <li>- Number of procedure attempts</li> <li>- Procedure successful/ complication</li> <li>- Incident/Patient disposition</li> </ul>	<ul style="list-style-type: none"> <li>• Emergency call/ Dispatch agency</li> <li>• Scene/situation response</li> <li>- Incident address</li> <li>- Incident location type</li> <li>- Destination</li> <li>- Reason for choosing destination &gt; Type of destination</li> </ul>	<ul style="list-style-type: none"> <li>• Emergency call/ Dispatch</li> <li>- Call date/time</li> <li>- Unit notified by dispatch date/ time</li> <li>- Unit en route date/time</li> <li>- Scene/Situation response</li> <li>- Unit arrived on scene date/time</li> <li>- Unit arrived at patient date/time</li> <li>- Unit left scene date/time</li> <li>- Patient arrived at destination date/ time</li> </ul>

Child Welfare (data sources may exist in one or more data systems)	Person	Person Descriptor	Encounter	Encounter Descriptor	Place	Time
Abuse and neglect	<ul style="list-style-type: none"> <li>Child on the case</li> <li>Caregiver</li> <li>Perpetrator</li> <li>Siblings</li> <li>Relatives</li> </ul>	<ul style="list-style-type: none"> <li>All individuals                             <ul style="list-style-type: none"> <li>Gender</li> <li>Race/Ethnicity</li> <li>Relationship to individuals on case</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Child abuse or neglect</li> </ul>	<ul style="list-style-type: none"> <li>Report</li> <li>Report unique identifier</li> <li>Report source</li> <li>Allegation</li> <li>Allegation report unique identifier</li> <li>Maltreatment type</li> <li>Investigation</li> <li>Investigation report type</li> <li>Disposition                             <ul style="list-style-type: none"> <li>Dual track referral</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Child residence (parent or guardian)</li> </ul>	<ul style="list-style-type: none"> <li>Report date</li> <li>Investigation date</li> <li>Disposition date</li> </ul>
Out-of-home care	<ul style="list-style-type: none"> <li>Child on the case</li> <li>Caregiver</li> <li>Perpetrator</li> <li>Siblings</li> <li>Relatives</li> <li>Foster care</li> </ul>	<ul style="list-style-type: none"> <li>All individuals                             <ul style="list-style-type: none"> <li>Gender</li> <li>Race/Ethnicity</li> <li>Relationship to individuals on case</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Out of home placement</li> </ul>	<ul style="list-style-type: none"> <li>Report</li> <li>Investigation</li> <li>Disposition</li> <li>Placement type</li> <li>Exit destination</li> </ul>	<ul style="list-style-type: none"> <li>Facility type</li> <li>County of residence at first placement</li> <li>Responsible administrative or regional unit at first placement</li> </ul>	<ul style="list-style-type: none"> <li>Date of entry</li> <li>Date of exit</li> <li>Date of placement change</li> </ul>
In-home services	<ul style="list-style-type: none"> <li>Child on the case</li> <li>Caregiver</li> <li>Perpetrator</li> <li>Siblings</li> <li>Relatives</li> </ul>	<ul style="list-style-type: none"> <li>All individuals                             <ul style="list-style-type: none"> <li>Gender</li> <li>Race/Ethnicity</li> <li>Relationship to individuals on case</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Service encounter</li> </ul>	<ul style="list-style-type: none"> <li>Service name</li> </ul>	<ul style="list-style-type: none"> <li>Service location</li> </ul>	<ul style="list-style-type: none"> <li>Start date</li> <li>End date</li> </ul>
<b>Early Childhood</b>						
CCDF	<ul style="list-style-type: none"> <li>Child</li> <li>Parent/Caregiver</li> </ul>	<ul style="list-style-type: none"> <li>Child</li> <li>Parent/Caregiver                             <ul style="list-style-type: none"> <li>Single parent</li> <li>Total monthly income</li> <li>Sources of income</li> <li>Family size</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Subsidy receipt</li> </ul>	<ul style="list-style-type: none"> <li>Reason for receiving subsidy care                             <ul style="list-style-type: none"> <li>Total monthly copay</li> <li>Care type</li> <li>Hours</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Parent/guardian's residential address</li> <li>Provider location</li> </ul>	<ul style="list-style-type: none"> <li>Date assistance started</li> </ul>
Early Intervention	<ul style="list-style-type: none"> <li>Child</li> </ul>	<ul style="list-style-type: none"> <li>Child                             <ul style="list-style-type: none"> <li>Age</li> <li>Gender</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Early Intervention assessment</li> <li>Early intervention service delivery</li> </ul>	<ul style="list-style-type: none"> <li>Early intervention assessment                             <ul style="list-style-type: none"> <li>Referral source</li> <li>Early intervention service delivery                                     <ul style="list-style-type: none"> <li>Service type</li> <li>Discharge code</li> </ul> </li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Parent/guardian's residential address</li> <li>Provider location</li> </ul>	<ul style="list-style-type: none"> <li>Early intervention assessment                             <ul style="list-style-type: none"> <li>Date of referral</li> <li>Date of eligibility assessment</li> </ul> </li> </ul>

Education	Person	Person Descriptor	Encounter	Encounter Descriptor	Place	Time
K-12 public education	<ul style="list-style-type: none"> <li>Student</li> </ul>	<ul style="list-style-type: none"> <li>Gender</li> <li>Race/Ethnicity</li> <li>Age</li> <li>Gifted and talented</li> <li>Limited English Proficiency status and level</li> <li>Eligibility for free or reduced lunch program</li> <li>Migrant student status</li> <li>Title I indicator</li> <li>Medical condition</li> <li>Immunization flag</li> </ul>	<ul style="list-style-type: none"> <li>School enrollment</li> <li>Assessment</li> <li>Disciplinary action</li> </ul>	<ul style="list-style-type: none"> <li>School enrollment                             <ul style="list-style-type: none"> <li>Enrollment status</li> <li>Grade level</li> <li>Enrolled days</li> <li>Excused and unexcused absences</li> <li>Exit or withdrawal status</li> <li>Exit or withdrawal type (transfer, graduation, etc.)</li> <li>District identifier</li> <li>School identifier</li> <li>School type (regular, charter, magnet, etc.)</li> <li>School grades offered</li> <li>School improvement status</li> <li>Title I school status</li> <li>Teacher identifier</li> </ul> </li> <li>Assessment                             <ul style="list-style-type: none"> <li>Course grades</li> <li>Grade Point Average</li> <li>Standardized achievement test name</li> <li>Standardized achievement test score</li> <li>Standardized achievement test proficiency</li> </ul> </li> <li>Disciplinary action taken                             <ul style="list-style-type: none"> <li>Discipline reason</li> <li>Suspension indicator/days</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Student address</li> <li>School address</li> </ul>	<ul style="list-style-type: none"> <li>School enrollment entry/exit date</li> <li>Assessment date</li> <li>Disciplinary Action date</li> </ul>
K-12 Special Education	<ul style="list-style-type: none"> <li>Student</li> </ul>	<ul style="list-style-type: none"> <li>Gender</li> <li>Race/Ethnicity</li> <li>Age</li> <li>Gifted and talented</li> <li>Limited English Proficiency status and level</li> <li>Eligibility for free or reduced lunch program</li> <li>Migrant student status</li> <li>Title I indicator</li> <li>Medical condition</li> <li>Immunization flag</li> </ul>	<ul style="list-style-type: none"> <li>Disability screening</li> <li>Special education participation</li> </ul>	<ul style="list-style-type: none"> <li>Disability screening                             <ul style="list-style-type: none"> <li>Disability status</li> <li>Disability primary type</li> <li>Disability secondary type</li> </ul> </li> <li>Special education participation                             <ul style="list-style-type: none"> <li>Special education full-time equivalency</li> <li>Program participation</li> <li>Program exit reason</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Student address</li> <li>School address</li> </ul>	<ul style="list-style-type: none"> <li>Disability screening date</li> <li>Special education start date</li> <li>Special education end date</li> </ul>

Education	Person	Person Descriptor	Encounter	Encounter Descriptor	Place	Time
Postsecondary Education	<ul style="list-style-type: none"> <li>Higher Education Student</li> </ul>	<ul style="list-style-type: none"> <li>Gender</li> <li>Race/Ethnicity</li> <li>Age</li> <li>Disability status</li> <li>Disability condition type</li> <li>Dependency status (Title IV Federal aid)</li> <li>Financial aid income level</li> <li>Financial aid award status, type, and amount</li> <li>Citizen status</li> <li>Limited English Proficiency</li> <li>status</li> </ul>	<ul style="list-style-type: none"> <li>School enrollment</li> <li>Assessment</li> </ul>	<ul style="list-style-type: none"> <li>School enrollment</li> <li>School identifier</li> <li>Level of institution (2-year, 4-year, etc.)</li> <li>Public or private institution</li> <li>Tuition</li> <li>Program of study</li> <li>Exit or withdrawal type</li> <li>Diploma or credential awarded</li> <li>Assessment</li> <li>Standardized admission test type</li> <li>Standardized admission test score</li> <li>Grade Point Average</li> </ul>	<ul style="list-style-type: none"> <li>Student home address</li> <li>Student residence address</li> <li>School address</li> </ul>	<ul style="list-style-type: none"> <li>School enrollment</li> <li>Entry date</li> <li>Exit date</li> <li>Diploma or credential award date</li> </ul>
Juvenile Delinquency	<ul style="list-style-type: none"> <li>Youth on the case</li> <li>Mother, siblings</li> </ul>	<ul style="list-style-type: none"> <li>Youth, mother, siblings</li> <li>Age</li> <li>Gender</li> <li>Race/Ethnicity</li> <li>Relationship to youth on the case</li> </ul>	<ul style="list-style-type: none"> <li>Juvenile detention</li> <li>Court-mandated juvenile delinquency services</li> </ul>	<ul style="list-style-type: none"> <li>Service name</li> <li>Service category</li> <li>Admission code</li> </ul>	<ul style="list-style-type: none"> <li>Residential address</li> </ul>	<ul style="list-style-type: none"> <li>Service begin date</li> <li>Service end date</li> </ul>
Juvenile Courts	<ul style="list-style-type: none"> <li>Youth on the case</li> </ul>	<ul style="list-style-type: none"> <li>Youth</li> <li>Age</li> <li>Gender</li> <li>Race/Ethnicity</li> </ul>	<ul style="list-style-type: none"> <li>Juvenile court referral</li> <li>Juvenile court charge</li> </ul>	<ul style="list-style-type: none"> <li>Referral</li> <li>Referral type</li> <li>Referral reason</li> <li>Charge</li> <li>Charge code</li> <li>Condition code</li> <li>Disposition code</li> <li>Resolution/Finding</li> <li>Case resolution/Finding code</li> <li>Case outcome</li> </ul>	<ul style="list-style-type: none"> <li>Referral jurisdiction</li> <li>Referral agency</li> <li>Court code</li> </ul>	<ul style="list-style-type: none"> <li>Date of referral</li> <li>Date of case resolution/finding</li> <li>Date of disposition</li> <li>Date of case outcome</li> </ul>

Juvenile Delinquency	Person	Person Descriptor	Encounter	Encounter Descriptor	Place	Time
Adult Justice						
City or County Jail	<ul style="list-style-type: none"> <li>Incarcerated Individual (pretrial/sentenced)</li> </ul>	<ul style="list-style-type: none"> <li>Incarcerated Individual (pretrial/sentenced)</li> <li>Age</li> <li>Gender</li> <li>Race</li> </ul>	<ul style="list-style-type: none"> <li>Detainment in the county jail for 0 to 24 months</li> </ul>	<ul style="list-style-type: none"> <li>Reason for inmate release</li> </ul>	<ul style="list-style-type: none"> <li>Residential address prior to incarceration</li> </ul>	<ul style="list-style-type: none"> <li>Incarceration begin date</li> <li>Incarceration exit date</li> </ul>
State Corrections	<ul style="list-style-type: none"> <li>Incarcerated Individual</li> </ul>		<ul style="list-style-type: none"> <li>Incarceration and parole</li> </ul>	<ul style="list-style-type: none"> <li>Current offense (#1-3)</li> <li>Counts for offense (#1-3)</li> <li>Sentencing length</li> <li>Minimum prison term to be served</li> <li>Total time served</li> <li>Reason for release</li> <li>Inmate escape</li> <li>On community release prior to prison release</li> <li>Length of community release</li> </ul>	<ul style="list-style-type: none"> <li>County in which sentence was imposed</li> <li>Jurisdiction on date of admission</li> <li>Location where inmate serves sentence</li> <li>Agency that assumed custody at release</li> </ul>	<ul style="list-style-type: none"> <li>Incarceration</li> <li>Entry date</li> <li>Exit date</li> <li>Parole</li> <li>Entry date</li> <li>Exit date</li> </ul>
Law Enforcement	<ul style="list-style-type: none"> <li>Victim</li> <li>Offender</li> <li>Arrestee</li> </ul>		<ul style="list-style-type: none"> <li>Incident</li> <li>Arrest</li> </ul>	<ul style="list-style-type: none"> <li>Incident (where applicable)</li> <li>Offense attempted/completed</li> <li>Cleared exceptionally</li> <li>Offender suspected of using</li> <li>Offender use of bias motivation</li> <li>Incident number</li> <li>Group A or Group B Offense</li> <li>Type of incident</li> <li>Circumstance</li> <li>Property loss</li> <li>Type of property</li> <li>Property description</li> <li>Value of property</li> <li>Relationship of victim to offender</li> <li>Type of weapon/force</li> <li>Type of injury</li> <li>Number of premises entered</li> <li>Method of premise entry</li> <li>Criminal activity/gang involvement</li> <li>Suspected drug type</li> <li>Drug quantity</li> <li>Arrest</li> <li>Arrest number</li> <li>Type of arrest</li> <li>Multiple arrestee indicator</li> <li>Arrest offense code</li> <li>Arrestee was armed</li> <li>Disposition of arrestee under 18</li> </ul>	<ul style="list-style-type: none"> <li>Incident location type</li> </ul>	<ul style="list-style-type: none"> <li>Incident</li> <li>Incident Date/ Hour</li> <li>Exceptional clearance date</li> <li>Date property recovered</li> <li>Arrest date</li> </ul>

Employment	Person	Person Descriptor	Encounter	Encounter Descriptor	Place	Time
Workforce Training Programs	<ul style="list-style-type: none"> <li>Program participant</li> </ul>	<ul style="list-style-type: none"> <li>Claimant                             <ul style="list-style-type: none"> <li>Gender</li> <li>Age</li> <li>Race/Ethnicity</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Program participation</li> <li>Employment</li> </ul>	<ul style="list-style-type: none"> <li>Workforce program participation                             <ul style="list-style-type: none"> <li>Program type (Job Corps, TANF program, etc.)</li> <li>Number of credits earned</li> <li>Diploma or credential awarded</li> </ul> </li> <li>Employment                             <ul style="list-style-type: none"> <li>Employed while enrolled</li> <li>Employed after exit</li> <li>Employment type</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Participant address</li> </ul>	<ul style="list-style-type: none"> <li>Program participation                             <ul style="list-style-type: none"> <li>Start date</li> <li>End date</li> </ul> </li> <li>Employment                             <ul style="list-style-type: none"> <li>Employment start date</li> </ul> </li> </ul>
		<ul style="list-style-type: none"> <li>Claimant</li> </ul>	<ul style="list-style-type: none"> <li>Claimant                             <ul style="list-style-type: none"> <li>Gender</li> <li>Age</li> <li>Education</li> <li>Race/Ethnicity</li> <li>Language</li> <li>Veteran status</li> <li>Disability</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>UI Initial claim</li> <li>Continued claim</li> <li>Combined wage claim</li> </ul>	<ul style="list-style-type: none"> <li>Claimant occupation</li> <li>Industry of the employer</li> <li>Temporary layoff</li> <li>Union position</li> <li>Total wages</li> <li>Taxable wages</li> <li>Benefits paid</li> <li>First payment/Final payment</li> <li>Weeks claimed</li> <li>Weeks compensated</li> <li>Exhaustions (Final payments)</li> <li>Eligibility determination</li> </ul>	<ul style="list-style-type: none"> <li>Employer address</li> </ul>
Unemployment Insurance (UI)		<ul style="list-style-type: none"> <li>Employed Covered Worker SSN</li> </ul>	<ul style="list-style-type: none"> <li>Employment status by quarter</li> </ul>	<ul style="list-style-type: none"> <li>Total Wages</li> </ul>	<ul style="list-style-type: none"> <li>Employer address (may not be actual place of employment)</li> </ul>	<ul style="list-style-type: none"> <li>Employment status by quarter</li> </ul>
	Wage Record	<ul style="list-style-type: none"> <li>Covered Worker</li> </ul>				

Public Assistance	Person	Person Descriptor	Encounter	Encounter Descriptor	Place	Time
TANF	<ul style="list-style-type: none"> <li>Primary TANF recipient</li> <li>Minor child receiving assistance</li> <li>Minor siblings of child receiving assistance</li> <li>Persons with income counted toward eligibility</li> </ul>	<ul style="list-style-type: none"> <li>TANF Household members                             <ul style="list-style-type: none"> <li>Type of family for work participation</li> <li>Receives subsidized housing</li> <li>Receives medical assistance</li> <li>Receives SNAP</li> <li>Amount of SNAP assistance</li> <li>Receives subsidized child care</li> <li>Amount of subsidized child care</li> <li>Amount of child support</li> <li>Amount of family's cash resources</li> </ul> </li> <li>TANF assisted individuals                             <ul style="list-style-type: none"> <li>Family affiliation</li> <li>Parent with minor child</li> <li>Noncustodial parent indicator</li> <li>Race/ethnicity</li> <li>Gender</li> <li>Receives disability benefits</li> <li>Marital status</li> <li>Relationship to head of household</li> <li>Education level</li> <li>Citizenship</li> <li>Employment status</li> <li>Work-eligible individual</li> <li>Work participation status</li> <li>Cooperation with child support</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>TANF assistance</li> </ul>	<ul style="list-style-type: none"> <li>TANF Assistance - Family (Amount &amp; Months)                             <ul style="list-style-type: none"> <li>Cash and cash equivalent</li> <li>TANF child care</li> <li>Transportation</li> </ul> </li> <li>TANF Assistance - Individuals (Amount &amp; Months)                             <ul style="list-style-type: none"> <li>Transitional services</li> <li>Transportation</li> </ul> </li> <li>Needs of pregnant woman</li> <li>New child only family?</li> <li>Exemption from federal time-limit provisions</li> <li>Months toward federal time limit</li> <li>Countable months remaining under State limit</li> <li>Reason for amount of reductions</li> <li>Reason for closure</li> <li>TANF Assistance - Individuals (Amount &amp; Months)                             <ul style="list-style-type: none"> <li>Unsubsidized employment</li> <li>Subsidized private-sector employment</li> <li>Subsidized public-sector employment</li> <li>Work experience (hours, absences)</li> <li>On-the-job training</li> <li>Job search assistance (hours, absences)</li> <li>Community service (hours, absences)</li> <li>Vocational training (hours, absences)</li> <li>Job skills training (hours, absences)</li> <li>Providing child care (hours, absences)</li> <li>Number of deemed core work hours</li> <li>Amount of earned income</li> <li>Amount of unearned income</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Family address</li> </ul>	<ul style="list-style-type: none"> <li>Start date</li> <li>End date</li> <li>Reporting date</li> </ul>

Public Assistance	Person	Person Descriptor	Encounter	Encounter Descriptor	Place	Time
SNAP	<ul style="list-style-type: none"> <li>• SNAP participant</li> </ul>	<ul style="list-style-type: none"> <li>• SNAP participant</li> <li>- Social Security number</li> <li>- Date of birth</li> <li>- Race/ethnicity</li> <li>- Gender</li> <li>- Receives disability benefits</li> <li>- Marital status</li> <li>- Veteran status</li> <li>- Education level</li> <li>- Citizenship</li> <li>- Employment status</li> <li>- SNAP case number</li> <li>- Number of family members</li> <li>- Relationship to head of household</li> </ul>	<ul style="list-style-type: none"> <li>• SNAP participation</li> </ul>	<ul style="list-style-type: none"> <li>• SNAP status</li> <li>• Amount of SNAP assistance</li> </ul>	<ul style="list-style-type: none"> <li>• SNAP participant address</li> </ul>	<ul style="list-style-type: none"> <li>• Start date</li> <li>• End date</li> </ul>
WIC	<ul style="list-style-type: none"> <li>• WIC participant</li> </ul>	<ul style="list-style-type: none"> <li>• WIC participant</li> <li>- Race/Ethnicity</li> <li>- Age</li> <li>- Gender</li> <li>- Language</li> <li>- Marital status</li> <li>- Residential status (e.g., migrant, homeless)</li> <li>- Annual income</li> <li>- Education level</li> <li>- Employment status</li> <li>- Health change (e.g., weight)</li> <li>- Weight status (e.g., normal, obese)</li> <li>- Pregnancy outcome</li> <li>- Immunization status</li> <li>- Other services (e.g., Medicaid, homeless shelters)</li> </ul>	<ul style="list-style-type: none"> <li>• WIC participation</li> <li>• WIC complaint</li> </ul>	<ul style="list-style-type: none"> <li>• WIC assistance</li> <li>- Certification status</li> <li>- Priority level</li> <li>- Participant status (e.g., inactive, transferred)</li> <li>- Participant type (e.g., pregnant woman, infant)</li> <li>- Income documentation</li> <li>- Income eligibility programs</li> <li>- Ineligibility reason</li> <li>- Appointment type (certification, nutrition education, etc.)</li> <li>- Nutrition goal outcome</li> <li>- Nutrition risk code (e.g., maternal smoking)</li> <li>- Bloodwork type</li> <li>- Breastfeeding status</li> <li>- Breastfeeding discontinued reason</li> <li>- Dual participation match program</li> <li>- Classes offered (e.g., breastfeeding)</li> <li>- Termination/suspension reason</li> <li>• WIC complaint</li> <li>- Complaint type</li> <li>- Complaint status</li> </ul>	<ul style="list-style-type: none"> <li>• WIC participant address</li> </ul>	<ul style="list-style-type: none"> <li>• WIC assistance</li> <li>- Start date</li> <li>- End date</li> <li>• WIC complaint date</li> </ul>

Homelessness/Housing	Person	Person Descriptor	Encounter	Encounter Descriptor	Place	Time
HMIS	<ul style="list-style-type: none"> <li>• Head of household</li> <li>• Family relationship</li> </ul>	<ul style="list-style-type: none"> <li>• All individuals in family</li> <li>- Age</li> <li>- Race/Ethnicity</li> <li>- Gender</li> <li>- Relationship to head of household</li> </ul>	<ul style="list-style-type: none"> <li>• Singles and families entering the homeless shelter system</li> </ul>	<ul style="list-style-type: none"> <li>• Reason for shelter stay termination</li> </ul>	<ul style="list-style-type: none"> <li>• Client location</li> </ul>	<ul style="list-style-type: none"> <li>• Entry date</li> <li>• Exit date</li> </ul>
PHA	<ul style="list-style-type: none"> <li>• Head of household</li> <li>• Household member</li> </ul>	<ul style="list-style-type: none"> <li>• All tenants</li> <li>- Head of household designation</li> <li>- Relationship to head of household</li> <li>- Age</li> <li>- Race/Ethnicity</li> <li>- Sex</li> <li>- Citizenship</li> <li>- Income</li> <li>- Total number in family</li> <li>- Student status</li> <li>- Head of household only</li> <li>- Employment status</li> <li>- Date employment began</li> <li>- Employment benefits completed</li> <li>- Years of education received</li> </ul>	<ul style="list-style-type: none"> <li>• Individuals and families served by the U.S. Department of Housing and Urban Development</li> </ul>	<ul style="list-style-type: none"> <li>• Type of action</li> <li>• Total number in household</li> <li>• Expected family addition (pregnancy, adoption)</li> <li>• Program</li> <li>• Moving to Work action</li> <li>• Homeless at admission</li> <li>• Number of bedrooms in unit</li> <li>• Date unit last passed inspection</li> <li>• Year unit built</li> <li>• Structure type</li> <li>• Unit rent</li> <li>• Tenant rent</li> <li>• Family maximum subsidy</li> </ul>	<ul style="list-style-type: none"> <li>• Unit address</li> <li>• Agency name and PHA code</li> <li>• Zip code before admission</li> </ul>	<ul style="list-style-type: none"> <li>• Effective date</li> <li>• Date entered waitlist</li> </ul>
Education Homeless Records	<ul style="list-style-type: none"> <li>• Student</li> </ul>	<ul style="list-style-type: none"> <li>• Gender</li> <li>• Race/Ethnicity</li> <li>• Age</li> <li>• Gifted and talented</li> <li>• Limited English Proficiency status and level</li> <li>• Eligibility for free or reduced lunch program</li> <li>• Migrant student status</li> <li>• Title I indicator</li> <li>• Medical condition</li> <li>• Immunization flag</li> </ul>	<ul style="list-style-type: none"> <li>• Homelessness</li> </ul>	<ul style="list-style-type: none"> <li>• Homelessness status</li> <li>• Homeless primary nighttime residence</li> <li>• Homeless serviced indicator</li> <li>• Homeless unaccompanied youth status</li> </ul>	<ul style="list-style-type: none"> <li>• Student address</li> <li>• School address</li> </ul>	<ul style="list-style-type: none"> <li>• Start date</li> <li>• End date</li> </ul>



**Actionable Intelligence for Social Policy**

University of Pennsylvania

3701 Locust Walk, Philadelphia, PA 19104

215.573.5827 | [www.aisp.upenn.edu](http://www.aisp.upenn.edu)