CASE WESTERN RESERVE UNIVERSITY
Jack, Joseph and Morton Mandel School of Applied Social Sciences
Center on Poverty and Community Development

The Child & Household Integrated Longitudinal Data System

A guide to an Integrated Data System Resource for Cuyahoga County

Center on Poverty and Community Development
Case Western Reserve University

January 22, 2026

Acknowledgments

The CHILD System at the Center on Poverty and Community Development is partially made possible through the ongoing support of:

- The George Gund Foundation
- Cuyahoga County Office of Early Childhood
- Case Western Reserve University
- Jack, Joseph and Morton Mandel School of Applied Social Sciences

Our Center is a member of the Actionable Intelligence for Social Policy (AISP) Network and the National Neighborhood Indicators Partnership (NNIP), through which the CHILD System benefits from shared knowledge and best practices related to data governance and responsible data use. However, it is worth noting that AISP and NNIP are not responsible for CHILD's procedures, limitations, or governance.

This report uses the Sullivan Business Report template from Overleaf.

Authors

Francisca Richter
Joseph Andre
Tsui Chan
Michael Henderson
Stephen Steh

Center Leadership

Rob Fischer & Francisca Richter, Co-Directors

Copyright

Contact

Contact: Povertycenter@case.edu

Changelog

v1.0	2017-12	First Edition
v1.1	2026-01-27	Second Edition

Table of Contents

1 Introduction

Governmental and nonprofit agencies and institutions routinely generate electronic administrative records related to the population they serve. These records capture information about individual and family characteristics, eligibility, risk assessment, services received, and information about outcomes or results. But, individuals often traverse multiple systems as they move along in their development. The ongoing integration of administrative records across agencies and time has the potential to provide new types of information that can be used to evaluate outcomes, drive decision making, target resources and gain an understanding of how the collective work of agencies and systems are addressing the needs and concerns of individuals and communities.

The Child & Household Integrated Longitudinal Data (CHILD) System is composed of linked administrative records of individuals in Cuyahoga County, Ohio, born since 1989 and continuing up to the present. The linkage of records across time and systems is performed via a combined approach involving probabilistic and deterministic matching followed by a robust manual verification process to ensure the highest level of accuracy. Many individual-level records contain geographic information that enables linking with other data systems at various levels of geographies such as parcel, address, or census tracts. The end result is a longitudinal data system in which individuals are observed if and when they have interacted with at least one of the numerous agencies and systems that contribute electronic records to CHILD.

The CHILD System was born in the late 90's as part of the Cuyahoga County Invest in Children (IIC) initiative. IIC is a community-wide, public-private partnership of government leaders and agencies, non-profit organizations and local foundations whose goal is to assure that all the County's young children, their families, and communities receive the support they need to enter kindergarten and succeed. Because no one agency or system within the initiative can achieve this outcome on their own, the CHILD System was required to support joint planning and the evaluation of collective impact. Since its inception, the CHILD System has gone from comprising information from seven to 35 administrative systems and has been extended longitudinally to cover individuals' system interactions as they transition into adulthood (see

To learn about the Invest In Children Initiative, see Fischer, R. L., Lalich, N., & Coulton, C. (2008). Taking it to scale: Evaluating the scope and reach of a community-wide initiative on early childhood. Evaluation and Program Planning, 31, 199-208.

Figure 1).

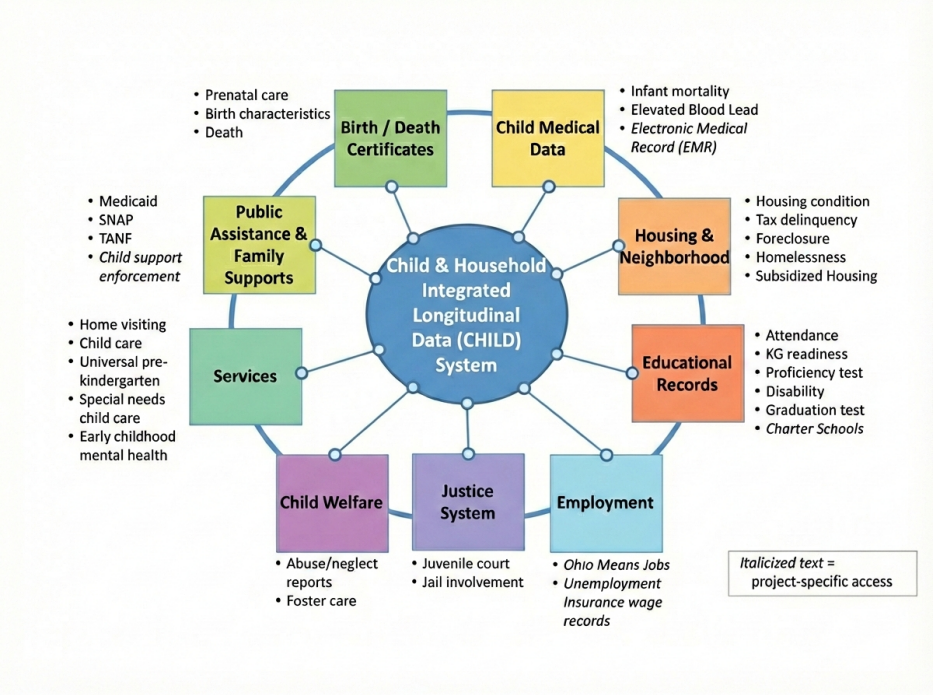

Figure 1: Data systems linked in CHILD.

CHILD integrates data from 35 administrative systems represented in 9 broad categories.

From the 35 administrative data systems, CHILD holds millions of records representing one or multiple interactions of over 850,000 individuals with agencies in Cuyahoga County (see Table 1). The CHILD System currently contains information related to the following observable events:

Birth, death, prenatal and ongoing home visiting, early childhood mental health services, subsidized child care, public assistance benefits (e.g., Cash Assistance, the Supplemental Nutrition Assistance Program -SNAP), housing subsidies and public housing, lead test results, special needs child care, public preschool and UPK attendance, kindergarten readiness assessments, public school attendance and test scores, child welfare involvement, juvenile court involvement, homelessness services and County jail spells.

In addition, the following data sources have been linked to CHILD occasionally for specific projects: Ohio Means Jobs, unemployment insurance wage records, and Breakthrough charter schools.

Individuals are assigned a unique CHILD identifier in the registry, a CHILD id. Along with their CHILD id, we keep all identifiers assigned to each individual within each administrative system (school ID, homeless services ID, etc.).

2 Structure, Matching and Updating Procedures

On a regular schedule, new data becomes available from the various administrative entities that provide data to the CHILD System. The workflow begins with data transfer, where most data requires a formal request and verification of Data Use Agreements (DUAs).

Various methods are employed for data transfer, including File Transfer Protocols (FTP), secure site login for Ohio Department of Health (ODH) data, Python scraping for open data, and direct uploads from partners to the Secure Research Environment (SRE). Once the data is transferred, it undergoes an import and pre-geocoding process. Some datasets require pre-processing to create SAS datasets, particularly those with significant import issues.

SAS macros are used to clean and standardize dates, addresses and other matching variables such as name, date of birth, sex, race, and guardian information. This step ensures the data is ready for geocoding. Mapmarker and MapInfo Pro are used for geocoding due to their speed and high accuracy when paired with the SAS Address cleaner macro. We use desktop versions of the software within a secure research environment that lacks internet access. After geocoding, the cleaned dataset is updated with geocoded addresses, and any final cleaning is performed.

While Mapmarker and MapInfo Pro are now available exclusively as a cloud-based service, we use previously licensed desktop versions within a secure, offline research environment, permitted under our agreement.

Table 1: CHILD Data Sources and Approximate Individual Count by Source (as of 12/2025)

#	CHILD Data Source	Data Years	Individuals	Notes
	Vital Records			
1	Births	1989-2024	624,500	Vital Statistics
2	Deaths	1992-2024	10,300	Vital Statistics
	Human Services			
3	DCFS Child Abuse and Neglect Reports	1989-2023	282,300	Demographics, placements, etc.
4	Public Assistance	1992-2024	474,300	SNAP/TANF (Medicaid to 2015)
5	MomsFirst Prenatal Home Visiting	2007-2023	14,600	Demographics, assessments
6	Help Me Grow Visiting	2000-2014	109,300	Demographics, assessments
7	Early Childhood Mental Health	2012-2022	3,900	Demographics, service use
8	County Newborn Home Visiting	2010-2022	12,200	Demographics, service use
9	Homeless Services	2009-2024	39,200	Entry/exit information
10	Cuyahoga Metro Housing	1989-2020	71,300	Public Housing application data
	Education / Child Care			
11	Universal Pre-K (UPK)	2007-2023	29,500	Attendance, assessments
12	Special Needs Child Care	2001-2022	12,600	Providers, service use
13	Cleveland Metro Schools	2005-2025	196,200	Enrollment, attendance
	Suburban Districts			*14 districts included below:*
14	Bedford	2009-2025	15,000	
15	Berea	2011-2025	18,400	
16	Brooklyn	2009-2025	5,500	
17	Cleveland Hts/Univ Hts	2011-2025	18,500	
18	East Cleveland	2009-2024	15,300	
19	Euclid	2017-2023	14,000	
20	Garfield Heights	2009-2025	18,000	
21	Lakewood	2009-2025	21,600	
22	Maple Heights	2009-2022	16,000	
23	Parma	2006-2019	41,900	
24	Richmond Heights	2009-2022	4,500	
25	Shaker Heights	2024-2025	5,700	
26	S. Euclid/Lyndhurst	2009-2025	17,200	
27	Warrensville Heights	2012-2025	9,200	
28	Subsidized Child Care	1997-2022	170,500	Demographics, service use
29	Starting Point Data	2008-2024	3,400	Provider type, slots, quality
	Corrections			
30	Juvenile Court Data	2000-2024	58,300	Cases and charges
31	Cuyahoga County Jail	2002-2021	33,500	Entry/exit dates
	Health			
32	Blood Lead Tests	2000-2024	382,700	Lead test results

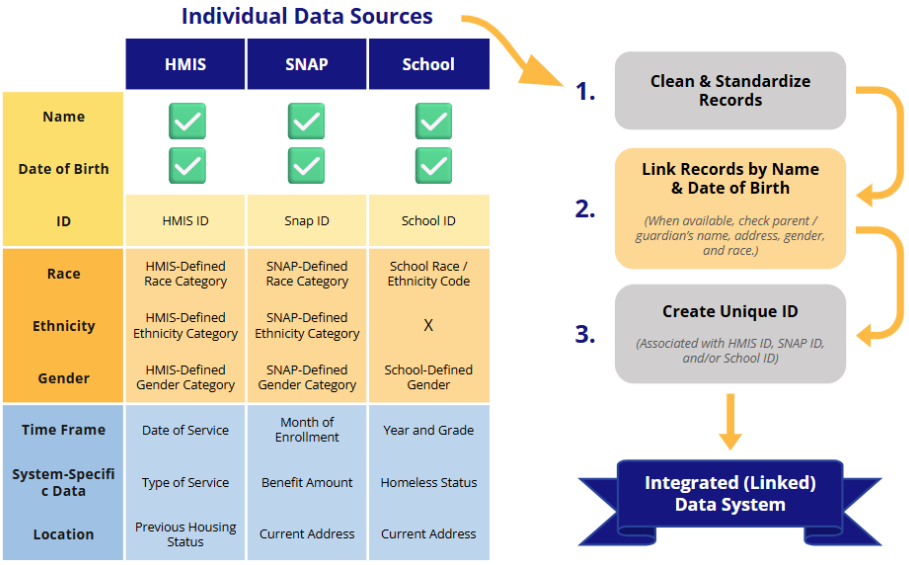

Figure 2: Simplified depiction of the linking process.

A SAS macro - Linkpro - is used for routine linkages of the CHILD System. LinkPro automatically calculates and applies probabilistic weights in order to estimate the likelihood that a pair of records from separate files corresponds to the same individual. The Linkpro package has also been buffered with in-house SAS macros to handle division of records that can be accepted as matches and those that need to be manual reviews. These additional algorithms look for typical errors that have been found over years of database curation such as hyphenated last names, month and day date of birth swaps, changes in last names used, and minor misspellings.

Gender and race identities may be captured via different classification schemes in various administrative data systems, underscoring the nature of these variables as social constructs.

The output generated includes a detailed report of the linking process, probabilistic weight calculations, and SAS data sets containing linked, unresolved and non-linked records. The output also includes Access databases with easily human reviewable predefined forms for manual review of records.

The linking of incoming partner data with existing individuals in the CHILD registry is carried out through a four step process.

Step 1: Exact blocking on core identifiers

We block records using First name, Last name, date of birth (DOB) and only a piece of guardian information (typically mother's last name or DOB) as additional matching criteria. Prior exploration of linking using this criteria in a multi-year,

state level birth dataset showed that these perfect block matches are highly specific and can successfully be used to match a large percent of new records.

However, matching rates vary by administrative system. As an example, 87% of children enrolled in Universal Pre-Kindergarten and 64% of people using homeless services in 2024 were identified and had their records linked to CHILD via this blocking step.

Step 2: Probabilistic matching within DOB blocks

The remaining records undergo probabilistic matching in Step 2. Here, DOB serves as the blocking variable and the algorithm compares all other available fields including First name, Last name, Middle name, Sex at birth, Guardian First name, Guardian Last name, Guardian DOB, Address (Number, Street, City, Zip code all individually compared), as well as SSN when present.

We use gender and racial identity data with caution as classification schemes for these social constructs vary across administrative data systems. An additional 6% of children enrolled in Universal Pre-Kindergarten and 9% of people using homeless services in 2024 had records linked to CHILD via Step 2.

Step 3: Phonetic matching and resolving possible DOB inaccuracies

The remaining records are processed in Step 3, which involves blocking on the Soundex of names, an algorithm that turns First names into text strings defined by phonetics. Within these phonetic blocks, all other fields (including DOB) are compared as described in Step 2. This step also resolves possible DOB inaccuracies (such as day/-month switches or Year+-1). An additional 3% of children enrolled in Universal Pre-Kindergarten and 6% of people using homeless services in 2024 were found to have already a CHILD id via Step 3.

For more information on Soundex see Knuth, D. E. 1998. The Art of Computer Programming, Volume 3. Sorting and Searching. Reading, USA: Addison-Wesley.

Step 4: Final duplicate check and assignment of new CHILD id

Records that do not match any existing individual in CHILD after Steps 1-3 are treated as new cases. Before creating a new CHILD id, a final duplicate review of the incoming data is performed to check whether the new individual has du-

plicate administrative IDs. Once an unduplicated list is obtained, new CHILD ids are generated and the associated administrative identifiers (such as school ID, homeless services ID) are assigned to the new CHILD id.

Via this step 5% of children enrolled in Universal Pre-Kindergarten and 21% of people using homeless services in 2024 were assigned a new CHILD id and added to the CHILD System.

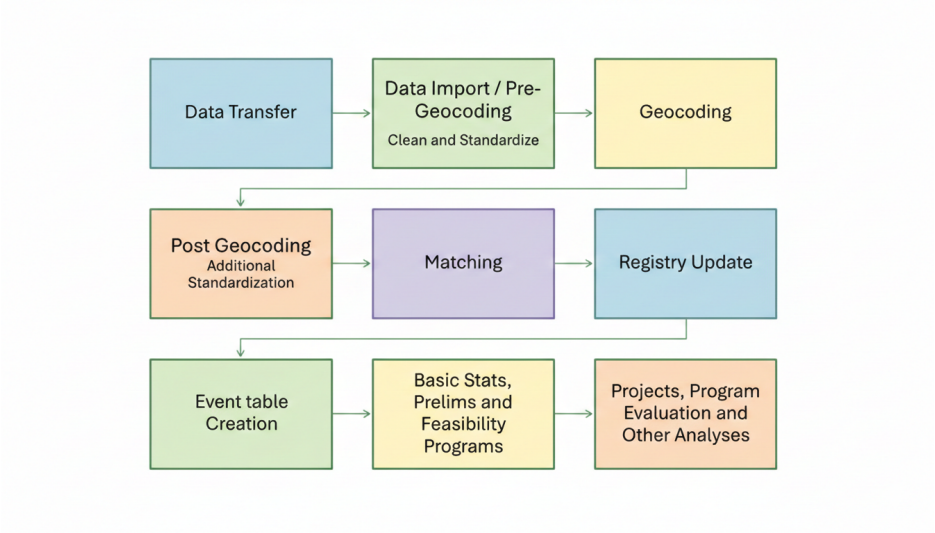

Figure 3: Schematic of the Data Intake and Linking Process in CHILD

3 Population Represented in the CHILD System

Representation in the CHILD System depends on whether a person interacted with the administrative agencies whose data make part of the IDS. Thus, all individuals born in Cuyahoga County, Ohio since 1989 will have at least one record in CHILD, their birth certificate. The administrative data can track one-time events such as births, or recurrent events such as home visits and school enrollment; however, the majority of events represented in the CHILD System occur more than once with specific dates attached. As the schedule of appearance in administrative records is highly variable across children, researchers routinely create count variables of service receipt over a specified period of time, such as year or quarter, or identify spells of service engagement and other metrics.

For some children, birth certificate data will be the only time they appear in the CHILD System, because they have not experienced any of the events captured in the administrative records. However, the majority of individuals in CHILD have multiple records in the system. 80 percent of individuals with a first recorded address in Cleveland have more than one record and 71 percent of individuals in CHILD have more than one record. This proportion will continue to increase as additional suburban school districts and charter schools enter into data sharing agreements with CHILD.

We are working on linking children to their parents or guardians for multigenerational analyses. This linkage will allow us to identify key events that are associated with generational poverty and more importantly, opportunities for interrupting this cycle over the life course. Two-generation data could also be used to better understand the ripple effect of a major life event throughout a family system. On a project basis, we have been able to integrate workforce data into our system allowing further tracking of individuals into adulthood.

For an example of a project that linked workforce data to CHILD, see Coulton, C., Richter, F. G.-C., Cho, Y., Park, J., Jeon, J., Fischer, R., 2023. Making the case for lead safe housing: Downstream effects of lead exposure on outcomes for children and youth. Health & Place, Volume 84, 103118, ISSN 1353-8292.

4 Linking the CHILD System to Neighborhood or Address Level Data

The residential addresses and geocodes in the CHILD System provide a means to link data from CHILD with data from other systems that contain information at the address or neighborhood level (using geocodes such as block group, census tract, or other geographical unit). For example, the Center on Urban Poverty and Community Development (Poverty Center) at Case Western Reserve University has a free and publicly accessible longitudinal database of social, economic, and property data called Northeast Ohio Community and Neighborhood Data for Organizing (NEOCANDO). Neighborhood level variables can be extracted from the NEOCANDO system and appended to individual records based on geocodes. Examples include rates of poverty, unemployment, foreclosures, violent and property crimes, or vacant housing.

Individual addresses can also be linked to property records compiled in another Poverty Center database, the Neighborhood Sta-

bilization Technology integrated parcel information system (NST). Historical data from NST include information at the address level on housing type, conditions, and values, land use, mortgage originations, deed transfers, foreclosure filings and completions, vacancies, code violations, demolitions, tax delinquencies, building permits, and community development investments.

NEOCANDO and NST are accessible via `http://neocando.case.edu`. For more on NST see Hirsh, A., Schramm, M., & Coulton, C. (August 2012). Neighborhood Stabilization Team Web Application. Briefly Stated No. 12-04. Center on Poverty and Community Development.

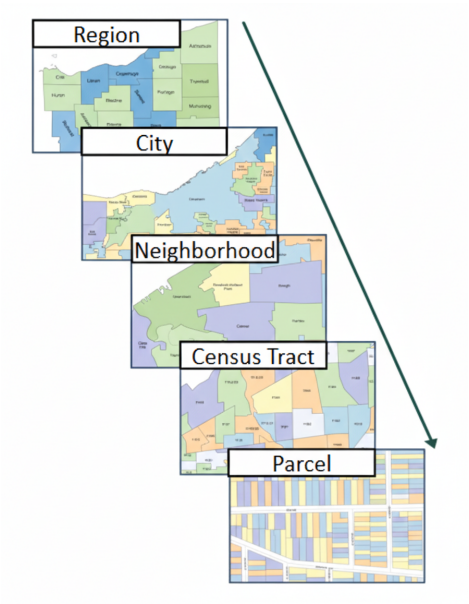

Figure 4: NEOCANDO and NST provide aggregated and parcel-level indicators, respectively

5 Confidentiality and Sharing Identifiable Information

Creating an integrated data system (IDS) remains a difficult endeavor despite their growing presence in the field of human services. In particular, the process of accessing data from agencies is governed by federal regulations such as the Privacy Act, HIPAA, and FERPA, as the release of PII is a risk to personal privacy and confidentiality.

Federal, state, and local regulations dictate strict limitations as to what identifiable information can be released. Due to university policies, university legal staff typically work with lawyers from data contributing agencies to enter into a process of data sharing bound by a legal document, often called a data use agreement

(DUA) or memorandum of understanding (MOU), governing acceptable uses of the data. Most IDS, including the CHILD System, also submit protocols for specific projects and receive approval from an Institutional Review Board (IRB) charged with the protection of human subjects.

The CHILD System is protected through a strict set of procedures. Only a small number of staff who are certified in human subjects protection and have signed oaths of confidentiality work with identifiable records. All work with these records is done in a highly secure research environment and personal identifiers are stored separately and linked through a random ID. Researchers and analysts, who are also certified by the IRB, work with de-identified data sets within the SRE to analyze the data

6 Governance

The CHILD System has a multi-level governance model designed to ensure excellent data stewardship. Strict DUAs are executed between the Poverty Center and all data partners that share data. These agreements explicitly state the terms of use and expectations of confidentiality and security and through them data providers govern the use of their data. The CHILD System is also governed by the IRB at Case Western Reserve University. The IRB assures that all research conducted using data from the CHILD System are in compliance with all federal protections of human subjects, including privacy and risk. In 2016, the CHILD Advisory Group was first assembled to provide guidance for the CHILD System. The purpose of the Advisory Group is to facilitate and maximize the application of CHILD in community planning, policy and evaluation, as well as provide advice on governance and sustainability initiatives. The Advisory Group can be particularly helpful in developing a framework for the effective use and development of the CHILD System as a community and governmental resource. This advisory body comprises representatives from several key constituencies including:

- Data partners: public agencies and nonprofit organizations that supply data to the Poverty Center for inclusion in CHILD.
- System partners: organizations that serve many of the same children and families, but do not currently provide data to

CHILD.

- Funders: governmental, federated and philanthropic funding entities that invest resources in serving the populations represented in CHILD.

The Poverty Center also routinely collaborates with other advisory groups for specific projects or grant applications to increase capacity and generate discussion among the agencies that supply data, the foundations that fund grants, and the officials who develop policies at local, state and federal levels.

In addition to local governance structures, the CHILD System benefits from participation in national networks dedicated to strengthening responsible data use for the public interest. The Poverty Center is a member of the Actionable Intelligence for Social Policy (AISP) Network and the National Neighborhood Indicators Partnership (NNIP). Both networks provide ongoing guidance, best practices, and peer learning to support ethical and equitable integrated data work. AISP advances standards for responsible data governance and cross-sector data sharing, while NNIP promotes community-centered data use and capacity building. Through these memberships, the Poverty Center receives field-informed advice that complements local governance processes and helps ensure that the CHILD System aligns with national norms, equity principles, and innovations in IDS practice.

7 Limitations

The CHILD System emerged from an ad hoc network of distinct but thematically related program evaluations and research efforts. Over the years, the Poverty Center and its data partners have put in place written data-use agreements, ethical review board protocols, and technical tools that make it possible to draw meaningful insights from complicated administrative data. These agreements hinge on trusted relationships between data partners and the Poverty Center respecting each data partner's own administrative and programmatic needs. At its core, the system is passive in that it does not place requirements on data partners but instead operates within the constraints of each partner's existing administrative and programmatic structures and accepts

their built-in requirements as the breadth and bounds of what is possible.

Partner data providers are as unique and complex as their data. Some collect administrative data passively (e.g., lead blood testing), while others are more exhaustive and active in their collection (e.g., district-level school records). In this regard, the system is not a collection of opt-in data or a collection of comprehensive data but a hybrid model that itself is non-exhaustive. For example, it does not collect data from all public school districts in the county nor does it capture every shelter stay, or for that matter every incidence of homelessness in or out of shelters.

Regarding variations in the scope and depth of data collected, some systems collect data at a household level, others at a family level, or at individual levels with markers for household or family that are unclear. One district may provide highly detailed data points on academic outcome, but if all others only provide mandatory state testing and GPA then that granularity cannot be an advantage to the system as a whole.

The frequency and timing of partner data represents a limitation as well. The time lag between when events occur and when data becomes available for analysis varies across sources, impacting the immediacy of cross-system comparisons. A program evaluation of a summer break intervention may wish to examine current-year school attendance after the Fall term, but most districts provide annual attendance after the close of the academic year, creating a necessary lag. Because data collection and delivery depend on each partner's internal schedules and the Poverty Center's processing timelines, the system cannot support real-time or near-term decision-making. Maintaining real-time, secure data pipelines would impose extensive labor and technological demands on both partners and the Center, undermining the goal of minimizing administrative burden.

Data currency is also shaped by the relational foundations of the CHILD System and the lasting effects of the COVID-19 pandemic. Staff turnover, shifting priorities, and transitions in internal data systems disrupted institutional knowledge and partner relationships, and because CHILD integrates data across systems, delays or gaps at one partner can constrain what can be linked or contextualized for others.

The result is a unique and complex data landscape, far from a uniform, harmonious dataset. Understanding this nuanced infrastructure, common to integrated administrative data systems, is key to implementing careful and ethical data analytics. Harnessing the power of integrated data requires acknowledging the limitations that arise from using data not designed for CHILD research, but rather generated from social interactions embedded in systems shaped by inequity and marginalization. But acknowledging the limitations alone is not sufficient. We have developed an explicit framework, FAIR2, to address discrimination bias reflected in administrative data, enrich our understanding of the data, and guide data analytics with insights from people represented in it.

For more information on the FAIR2 framework, visit `https://cwru-dsci.org/`

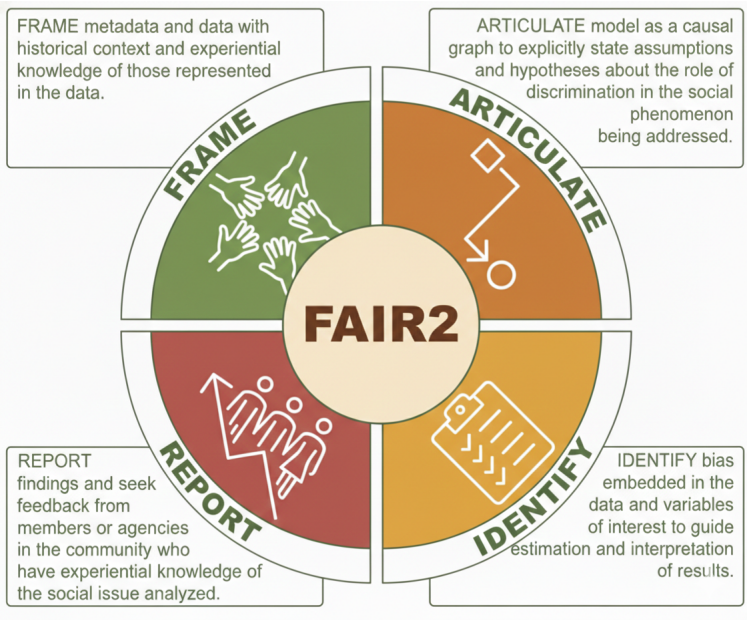

Figure 5: FAIR2 is a Public Interest Technology framework for integrating community knowledge in data analytics

8 Examples of the CHILD System in use

The following selected examples illustrate how the CHILD System can be used for program planning, improvement, evaluation and policy innovation.

Program Planning

- Community Need: In 2011, the Sisters of Charity Foundation of Cleveland engaged the Poverty Center in a planning process to create a Promise Neighborhood in Central, one of Cleveland's most economically distressed neighborhoods. Data from the CHILD System were used to describe the magnitude of need in the community. This descriptive information informed the identification of partners and service providers who could participate in revitalizing Central. With a collaborative team in place, data from the CHILD System have been used to monitor and describe demographic, educational, and social service involvement changes in the neighborhood over time.
- Cross Service Use: Recent analyses using data from CHILD explored the extent to which families experiencing homelessness relied upon public assistance in the years leading up to and after shelter entry. This study provided insights to a multi-agency partnership in Cuyahoga County that was working towards making sure that people experiencing and exiting homelessness have access to the public benefits and employment supports they need to achieve stability.

Program Improvement and Evaluation

- Medical Home Study: Ongoing monitoring of health insurance coverage showed nearly universal coverage of the child population under age 6 in Cuyahoga County. But, using CHILD to review data on receipt of well-child visits showed that only 50% of infants on Medicaid were receiving the recommended number of well-child visits in the first year of life. Thus, access to medical care did not translate to utilization of medical care.

  As a result, the County developed a medical home model in which patient advocates in health clinic settings conducted outreach to pregnant women. The patient advocates also served as liaisons between healthcare providers and families.

  Initially, patient advocates were embedded in two health care clinics via a pilot project. CHILD was used to evaluate the effectiveness of the medical home pilot. The completion of scheduled well-child visits in the pilot sites was about double the rate found in similar populations in Cuyahoga County (86% v. 40%).

Policy Innovation

- Partnering for Family Success: The CHILD System was used to inform the County's decision to explore performance-based contracting (Pay for Success or PFS) for financing human services. Through a partnership between government, nonprofit, philanthropic organizations, and private investors, PFS provides immediate capital to implement innovative social programs targeting entrenched social issues. If successful, governments see cashable savings and a return to investors. Using the CHILD System, the Poverty Center identified several cross-service utilizations representing high costs for the County and reflecting high levels of distress for families involved. Ultimately, after thorough exploration of service receipt patterns, the County launched Partnering for Family Success to more quickly reunite parents experiencing homelessness with their children who were placed in out-of-home foster care.

9 Future Directions

Since its founding, the Poverty Center has been committed to finding better solutions to problems associated with urban poverty by using data and research on a project-by-project basis to support targeted planning or evaluation efforts. This piecemeal approach has garnered impressive results, though the Poverty Center's leadership recognizes that a structured and strategic investment in its current IDS infrastructure is needed to enable further expansion of its innovative efforts to inform effective and efficient decision-making.

In the near term, the Center has identified two areas that require the investment of significant time and resources: 1. Improving the efficiency and functionality of the system; and 2. Enhancing the impact and value proposition of CHILD for the community.

System improvements

- As the CHILD System continues to expand, future enhancements to the matching process may include integrating more advanced entity-resolution tools, such as ChoiceMaker or machine-learning–based approaches, to improve

accuracy and reduce manual rule-tuning. Additional improvements, such as expanded fuzzy-matching techniques and routine auditing of match quality and bias, would further strengthen the reliability and equity of the linkage process.

- Strengthen processes that help researchers select study populations and create analysis files.
- Expand integration to include additional data sources and improve the ability to track individuals, families, and multiple generations over time.
- Enrich and standardize metadata resources including a more comprehensive data codebook.
- Develop a business model and governance structure that supports long-term stability and excellent stewardship.

Value Proposition

- Engage the institutions, agencies and programs that can benefit from analysis of CHILD System data.
- Support decision-making on the impact of social programs to ensure effective funding allocations by local, state and federal governments.
- Contribute to the understanding of best practices in government and human services programming.
- Enhance existing, and enable the creation of new community partnerships and solutions.
- Inform philanthropy in its efforts to make program-related and mission-driven investments.
- Encourage multi-sector collaborations and investments through social innovation financing models.