

FAQs about the CHILD Data System

What is an integrated data system?

The CHILD Data System is one of the oldest and most comprehensive integrated data systems in the country. An integrated data system holds administrative records from numerous agencies and organizations that are matched at the individual level. For example, data on early childhood lead exposure, public assistance receipt, child care enrollment and kindergarten readiness test scores are matched to inform the planning process for high quality pre-kindergarten programs in the region. By painting a comprehensive picture of need factors, the schools' programs can be better designed to s

powerful resource for planning, monitoring,
families.

What is administrative data?

Administrative dataD

researchers because it is not self-reported and provides a record of service uses and results across a wide spectrum of programs. Because this data is already being collected by agencies anyway, using it for the purposes of research is efficient, cost-effective and can provide statistically powerful results due to the breadth of coverage.

What administrative data does the CHILD Data System hold?

We are always in the process of negotiating new data sharing agreements with data providers and we currently have agreements with over 35 administrative systems including Cuyahoga County Health and Human Services, area school districts, home visiting service providers, the Ohio Department of Health, the Ohio Department of Job and Family Services and other local agencies.

Administrative entities that we already have established data sharing agreements with provide us with data on a regular schedule as new data become available, but with varying time intervals depending on the source.

The CHILD system currently captures administrative records for people born since 1989 living in Cuyahoga County, including more than 640,000 children, for a total of nearly 200 million records in the system.

What procedures are in place to protect this data?

The records imported into the CHILD Data System contain personally identifiable information so that they can be linked across agencies. The records are stored within a secure research environment that meets the highest standards of physical, administrative and technical controls.

No identifiable information is used outside this environment. Only summary statistics and tabulations can be removed, and these are carefully checked for the possibility of deductive re-identification before being released.

Ensuring that these data are used appropriately and ethically and that all data are secure are of utmost importance.

How is the CHILD Data System governed?

The CHILD Data System has a multi-dimensional governance model to ensure excellent data stewardship.

Strict data use agreements are executed between the Case Western Reserve University (Center on Urban Poverty and Community Development) and all data providers that explicitly state the expectations of confidentiality and security. Through the specifications in these agreements, the data providers govern the use of their data.

The CHILD Data system is also governed by the Institutional Review Board at Case Western Reserve University. The Institutional Review Board assures that all data with personally identifiable information (PII) are handled according to the highest standards for data security and protection. They also assure that research conducted using data from the CHILD Data System are in compliance with all federal protections of human subjects, including privacy and risk.

Additionally, the CHILD Data System has an Advisory Group that meets on a quarterly basis to represent the collective interests of public agencies and the community. The group works to ensure that the CHILD Data System adequately meets collective and community expectations of excellent data stewardship along the lines of governance, communication and sustainability.

How are the data matched?

When new data are submitted from administrative entities they are first geocoded and standardized. Standardizing the data entails using common abbreviations, ensuring everything is written in uppercase and removing punctuation. Then, the linkage of records across time and systems is performed via deterministic and probabilistic matching algorithms to determine whether the new records match those that are already in the system. After all new data have been matched and the system has been updated, data from the CHILD system are “self-matched” to eliminate possible duplicate records. The methods are continuously evaluated to ensure that the linkages are at acceptable levels of reliability and completeness.

