CHAPIN HALL
AT THE UNIVERSITY OF CHICAGO

Chapin Hall Illinois DCFS Database (CHILD) Description and Documentation

DRAFT – May 2018

Emily Wiegand, David McQuown, Lia Amelia, and Molly Rossow

Table of Contents

About the CHILD

The Chapin Hall Illinois DCFS Database (CHILD) is a relational database that organizes information extracted from two Illinois Department of Children and Families (DCFS) information systems:

1. SACWIS (the Statewide Automated Child Welfare Information System) contains information about child abuse and neglect reports and investigations as well as some information on child welfare services. SACWIS replaced CANTS (Child Abuse and Neglect Tracking System), Chapin Hall's previous source for child abuse and neglect investigation information.[1] The CANTS data was previously tracked in DCFS-L, Chapin Hall's longitudinal DCFS database. Appendix 1 discusses similarities and differences between the CANTS data in DCFS-L and what we have extracted from SACWIS for the CHILD.

2. Legacy Golden Copy (LGC) contains information about child welfare services, particularly placement data. This system is often referred to as CYCIS (Child and Youth Centered Information System). It is accessed by DCFS workers through a portal called IMSA, although some LGC data is also fed to SACWIS and visible to workers through the SACWIS frontend. The raw LGC data is the same as what Chapin Hall has historically received from DCFS as "CYCIS", which was used to create the DCFS quarterly (and sometimes monthly) databases.

Users should remember that while Chapin Hall curates and stewards these data to facilitate our research, DCFS retains ownership of the data. In accordance with Chapin Hall's data governance policies, only staff members working on projects that have approval from DCFS to use the data are granted access to the database. For questions about this process or to request permission for a project to use DCFS data, Chapin Hall staff should contact the Data and Research Technology Liaison (helpdesk@chapinhall.org).

This documentation is intended to be comprehensive, but there may be specific questions that are not addressed here. We suggest that a researcher with questions about the DCFS data that go beyond what is described here contact the DCFS data user community at Chapin Hall for feedback and assistance (via the DCFS Data Users email list). For questions about data processing and to report suspected errors in the database or documentation, contact the Chapin Hall Integrated Database processing team (idb@chapinhall.org).

Users of DCFS data external to Chapin Hall should send questions about the data or processing to the Operations Database Administrator and the Chapin Hall Integrated Database processing team (idb@chapinhall.org).

Entity Relationship Diagram

Figure 1 presents the entities, attributes, and relationships in the CHILD.

[1] Although CANTS was sunset in about 2006, Chapin Hall continued to receive data in the CANTS format until the CHILD replaced DCFS-L. This data was created by DCFS from a SACWIS extract using NOMAD software.

Figure 1: Chapin Hall Illinois DCFS Database (CHILD) Entity Relationship Diagram

Last updated 5/25/2018

Legacy Golden Copy

SACWIS (investigations)

SACWIS (child services)

SACWIS (general)

Combined data

Assignment

case_id

assignment date

cmgr_id

Goal

case_id

date established

permanency goal

Case_Seq

case_id

openepcnt

open and close dates

open and close reason

Legal

case_id

legal date

legal hearing type

legal characteristics

Case_Mgr

cmgr_id

case manager RSF

casemanager characteristics

Case_Mgr_RSF

cmgr_id

source_month

case manager RSF

Case

case_id

child_cli_id (if child case)

fam_clus_id (if family case)

seq_case_nsf

case type

summary of case sequence information

Service Plan

service plan id

case_id

date created

Service Plan Member

service plan id

person id

role (adult, kid, etc)

Service Plan Outcome

outcome id

service plan id

person id

outcome dates

strengths and desired result (text)

Service Plan Action

action id

outcome id

action dates

task and responsible party (text)

Case_RSF

case_id

RSF start date

case RSF

Cluster

clusid

cli_id

date and status

HOH indicator

Incident

incident_id

incident type

date

case_id

person id

Meeting Participant

meeting id

person id

Meeting

meeting id

meeting type (CFTM, visitation)

date

supervision

Team

RSF

field office address ID

agency prv_id

Living Arrangement

case_id

livar_seq

provider id

openepcnt

other living arrangement characteristics

Client

cli_id (CYCIS registration id)

sacwis_person_id

static person characteristics

Person

person id

cycis_cli_id

static person characteristics

Allegation

allegation id

investigation id

victim's person id

perpetrator's person id

allegation code

finding and date

relationship between victim and perpetrator

Provider

provider id

provider RSF

provider characteristics

CH-generated surrogate FEIN (ch_prv_fein_num)

CH-generated surrogate SSN (ch_prv_ssn_num)

Client_Address

cli_id

source_month

CH address id

Investigation_Person_Address

investigation_id

person_id

CH address id

Protective Custody

person id (child)

investigation id

protective custody seq

PC information

Provider_Address

provider id

source_month

CH address id

Provider_Licensing

provider id

license sequence code

license supervisor RSF

provider license characteristics

Address

CH address id

address detail

Investigation_Address

investigation_id

CH address id

Investigation

investigation id

investigation attributes

Update Process for the CHILD

Preparing DCFS data for inclusion in the CHILD is a three step process:

1. Extract the raw data from SACWIS and LGC via Navicat and load into a PostgreSQL database. Scripts used to extract and load files can be found on the child-etl GitLab repository. A new script is created each month to load new data.

2. Transform the data to match the database master schema. Transformations include standardizing dates, joining tables, casting values to new types, and creating binary values of categorical data. All transformations and assumptions are outlined in detail in this document. This step also includes error corrections such as removing nonsensical characters. The script used to transform each file is saved in the child-etl repository on the Chapin Hall GitLab server. The same script is used each month to transform new data. Command line arguments specify the year and month.

   Transformed data is validated with scripts running in Jenkins. The validation script is run every night after working hours and can also be run manually at any time. Email alerts are sent to the database maintainers to report the results of the validation. If the validation scripts find errors in the transformed data the transformation scripts are corrected or issues in the raw data may need to be resolved.

3. Replace or append transformed files to the database tables. Scripts available in the child-etl repository on the Chapin Hall GitLab server replace most database tables with the updated version, since the data extracted from DCFS is cumulative. However, four historical tables (Client Address, Provider Address, Address, and Case Manager RSF) are appended over time, since these tables track changes over time in these data points. The same script is used each month to update the CHILD tables. Command line arguments specify the year and month.

   As in step 2, validation scripts are run through Jenkins to check that the data is free from errors after this work is completed.

At each step of the process, Python scripts and tool's from the ChapinPy data manipulation library are used to run and log the data processing. The actual ETL steps are written in SQL for clarity wherever possible.

Schemas Containing DCFS Data

The main schema for database users is called child. It contains the tables and fields outlined in this document, with names, SSNs, and free text fields redacted for research purposes. Researchers who have agency approval to access the full data, including identifiers, will find that data in the schema childfull.

Two additional schemas, childbackup and childbackupfull, provide additional resources. These schemas contains old copies of CHILD tables. Most tables in the CHILD are replaced each month using data pulled from DCFS. The DCFS datasets cover the entire time range included in the CHILD and, in theory, each new data receipt should only add data to previous ones but in reality,

this will not always be the case. Corrections or changes may be made to earlier data in the DCFS systems as well. To account for this and to improve the reproducibility of results using the CHILD, backup copies of CHILD tables are stored in the childbackup schema. Note that each month's copy of the tables are appended to the tables in the childbackup schema, so to view the data from a given point-in-time using this schema requires filtering on source_month.

The childstage and childscratch schemas exist only on the ETL server and are used during data development.

DCFS Data Not Included in the CHILD

Some DCFS data contained in LGC and SACWIS has not been incorporated in the CHILD to date. As available, documentation of the full DCFS databases is available on the DCFS Data Users Sharepoint site (or by request, for external users). Any additional tables can be pulled from Navicat on an ad hoc basis. Depending on project needs and funding availability, other tables may be added to the database at a later date. To inquire about expanding the CHILD with additional DCFS data, contact idb@chapinhall.org.

Table Descriptions

This section includes notes about the data sources for each table and any recoding or restructuring of the data that occurs in the process of database development. Variable names and definitions are documented in the data dictionary that follows this section. Coded database fields frequently have a corresponding descriptive field in the database, particularly when the source system is SACWIS, since code values are databased in SACWIS. Some data elements that originate from LGC/CYCIS have only coded values in CHILD. However, the code values for these variables are the same as in the old DCFS quarterly database.

General Usage Notes

Working with Longitudinal Data

Conducting research on administrative data usually requires that these data have a longitudinal component – i.e. the data describes not the current situation but also history. There are three common ways to store and access longitudinal data, all of which are represented in the CHILD.

In **event or transactional data**, each record represents a discrete event or transaction that has a time component. For example, an investigation is a discrete event that occurs at a specific time point. A child may have multiple investigations over time. This data is the most easily aggregated over time, since records are only added. Many of the tables in the CHILD are of this format. These files are replaced every month with a new, cumulative transactional file from DCFS' system. The source_month field on each of these tables gives the month of the extract[2] that last replaced the cumulative data.

[2] Chapin Hall's Operations Database Administrator pulls new data from DCFS on the first business day of the month. The source_month value for a given set of data is the month that is represented in the new pull, not the month when the data was pulled. For example, data pulled on April 1, 2018 would have a source_month value of 201803.

It is important to note that while the events captured in transactional data are discrete and do not change over time, sometimes the information reported about the events does. For example, when an investigation first appears in the data it may have a status of "pending", and at some point that status is updated to "indicated." For this reason, looking at the same data at different times can yield different results. Monthly versions of the CHILD are archived in the childbackup schema, and the most recent prior months' versions of the data are maintained on the database for easy access, so that analysts needing to look at a consistent version of the data have access to the version of the table from that point in time.

A **cumulative point-in-time file** includes a "snapshot" of a record at a series of time points. There are four history tables (three of them capturing address histories) in the CHILD that store data in this format. For example, the Case Manager RSF table includes one record for each case manager for each month, with the team to which the case manager was assigned during that month. Records in this table are uniquely identified by the combination of case manager's ID and a field that indicates the source month.[3] The most common use of a cumulative point-in-time file is to extract the data as it stood at a given point-in-time – for example, identifying the team a case manager was on at a given date. Similar to event or transactional data, cumulative point-in-time files are also ideally designed to underlie longitudinal data visualizations (for example, a plot of the number of providers in a given county over time), since these files already include a time variable. The cumulative point-in-time tables in the CHILD are the only tables that are not entirely replaced from DCFS systems each month. The data in these tables is not saved by DCFS – so when an address is changed in DCFS' databases, the prior address may be lost.[4] Chapin Hall finds value in locating addresses and team assignments historically to facilitate retrospective analysis, so the CHILD preserves this historical data and appends the snapshot records for each new month. As a result of this processing, the cumulative point-in-time files can be quite large, but they are indexed for fast querying.

A **spells file** includes data and a start and end date for when that data was applicable. Spells are ideally suited to assessing questions of duration and repetition (for example, identifying frequent placement changes). The Master Event and Master Spell derived tables present data in a spells format.

Core SACWIS Tables

Person

The Person table is intended to represent static person characteristics. This table is updated monthly. Every time the table is updated, it overwrites the existing data from previous month. It contains one record for each person. The Person table includes both DCFS workers and non-workers.

Each person is uniquely identified by person_id. This table also includes the ID for the person in LGC/CYCIS and his or her recipient number in DHS/HFS data. The DHS IDs are primarily used at Chapin Hall for record linkage purposes, and we suggest using IDB link files to identify the same

[3] In this case, not every record has the same source month. The source month value depends on what extract added a given record to the table.

[4] In practice, SACWIS seems to preserve some address history but address start and end dates are inconsistently populated, making this history challenging to use.

individual across databases. To understand when to use IDs in the database and when to use link files to connect individuals between SACWIS and LGC data, see Appendix 2.

There are several fields in the Person table that indicate the status of the record – some records are marked as merged or deleted by DCFS. However, these records are still include in the CHILD because we still see instances of these person IDs appearing in other tables. Use caution around merged or deleted person records. For more detail about these records, see Appendices 1 and 2.

Tables of Abuse and Neglect Investigation Information from SACWIS

Investigation

The Investigation table is intended to represent the administrative records of investigation. It contains one record per investigation id. This table is updated monthly. Each time this table is updated, it overwrites the existing data from the previous month. The fields of this table are derived from table Investigation in SACWIS, with a few descriptive values for codes added from the SACWIS Code Description table.

The primary key of this table is investigation ID (invst_id), but the table can usually also be uniquely identified (and can be compared to DCFS-L investigations) using the combination of scrnum and scrseq.[5] Scrnum can otherwise be used to identify the same household across multiple investigations. The investigation table also includes information about investigation dates and status.

This table also includes a field that indicates investigation records that are supposedly "merged" (merge code) as well as records that have been expunged or purged (status code). Investigations with these statuses have not been excluded from the CHILD because we found allegation fields still connected with those investigations. Analysts should exercise caution using "merged" or "purged" investigation records. For more discussions of these records, see Appendix 1.

Allegation

The Allegation table includes one record for every allegation id. Every time this table is updated monthly, it overwrites the existing data from previous months. The fields in this table are derived from joining multiple SACWIS tables, including Investigation Allegation, Investigation, Investigation Subject, and Relationship. Some additional fields to describe the codes are added from the Code Description and Code Description CH tables. The transformation code used to complete these joins and create table Allegation is available on the child_etl GitLab repository.

This table includes fields for victim id (person_id_vic) and perpetrator id (person_id_perp). These fields both contain person ID values and can be joined with the Person table to obtain further victim or perpetrator characteristics.

One investigation can have more than one allegation, with different combinations of victims, perpetrators, and/or allegation codes. A victim can have more than one household id (scrnum), which indicates that the child has changed households over time.

[5] There are a small number of cases (28) where the same scrnum/scrseq combination maps to more than one investigation. In all of these cases, there is no more than one investigation that is not marked expunged, purged, closed, or merged, using the status and merge codes.

Note that there are allegations that are linked to people or investigations that have been purged, deleted, or merged, according to the SACWIS codes included in the Person and Investigation tables. There are also allegations that include investigation IDs and person IDs that are not found in the Person or Investigation tables. These database integrity problems are found in the source data. We decided not to remove the allegations in order to more closely mirror the source, but we suggest caution in using these allegations. For more detail about this problem, see Appendices 1 and 2.

There is currently no coding in the database to indicate older allegations with code 60 ("Risk of Harm") which were mandated unfounded by court order (Julie Q. v. DCFS) in July 2012. Depending on the research question, some projects have treated these allegations differently from other unfounded allegations. If you are beginning a project where it is important that you be able to identify these allegations, contact idb@chapinhall.org.

Protective Custody

The Protective Custody table contains records of children placed in protective custody. Its contents are derived from the SACWIS Protective Custody and Investigation Subject tables, with some descriptions added from the Code Description table. The exact process used to create this table is documented in the transformation code available on GitLab.

Only person IDs associated with children from investigations are included in this table.

The same child may be taken into protective custody multiple times during the same investigation. The field pc_seq distinguishes these multiple protective custody incidents; pc_seq was assigned to unique combinations of investigation ID and person ID, ordered by the combination of the record creation timestamp and the timestamp of the record's last update. Higher sequence value indicate more recent protective custody records associated with a child in an investigation.

This table contains one record per the combination of investigation id, the child's person id, and the sequence value for protective custody incidents.

Investigation Address

This table includes the incident address for each SACWIS investigation. It is sourced from the SACWIS table address, which is filtered to only include incident addresses (CD_BASE_CATG_TYPE = 9). This table consists of investigation id (invst_id) and the Chapin Hall-assigned unique identifier for a DCFS address (addr_id). Not all investigations have a corresponding incident address; the CHILD includes addresses for investigations where they exist in the source data. This table is replaced every month, since these addresses exist at a single point in time (the time of the incident), so this is not an address history table.

Investigation Person Address

This table relates a person/investigation combination (where the person is either the perpetrator or the victim on the investigation) to that person's residential address (CD_BASE_CATG_TYPE = 1) in SACWIS as of the initial date of the investigation. Addresses associated with investigations in calendar year 2017 and before were sourced from address history information in Chapin Hall's old DCFS-L Database (and originally came from CANTS), because address start and end dates in SACWIS are inconsistent. Addresses are sourced from SACWIS table address for investigations starting in calendar year 2018. The address is the person's current address as of the month the

investigation first appears in Chapin Hall's SACWIS extract.[6] This table is not replaced month to month; instead, new investigation/person/address combinations are added.

This table consists of investigation id (invst_id), person id (person_id), and Chapin Hall-assigned unique identifier for a DCFS address (addr_id).

Tables of Child Welfare Service Information from SACWIS

Meeting

Note that the definitions of the meetings to be included in this table and the Meeting Participant table were derived from project-based definitions developed by Chapin Hall researchers Brian Chor and Zhidi Luo. Those definitions are continuing to evolve, and we are exploring the most appropriate definitions of a broader universe of "meetings" that will be of interest for various analyses. If you are interested in using meeting data for a project, check in with Brian and Zhidi (or contact idb@chapinhall.org) to learn more about how definitions are developed and potential limitations.[7]

The Meeting table includes records for all supervised and unsupervised visits and Child and Family Team Meetings (CFTMs). These records are a subset of records in SACWIS' Note table, though some fields are derived through a join with the Note Subcategory Type table in SACWIS. The logic to define supervised visits, unsupervised visits, and CFTMs was provided by DCFS and is documented on the DCFS Data Users Sharepoint site; its execution can be seen in the SQL code used to export table Note from Navicat (available in the child-etl GitLab repo).

The unique identifier for this table is called meeting_id. This is a field created at Chapin Hall; it is a concatenation of the SACWIS field id_note and cd_subcatg_type.

The field meeting_type was also created a Chapin Hall, and indicates whether the meeting record meets the criteria for a visit or a CFTM.

The field supervision_cd (or its descriptive complement, supervision_desc) distinguishes supervised and unsupervised visits.

Meeting Participant

The Meeting Participant table provides information regarding participants in visits and CFTMs. It consists of two fields: meeting_id and person_id, the ID of the person who participated in the meeting. Fields in this table are derived and extracted from multiple SACWIS tables: Note, Note Participant, and Case Type Member. This table is unique by the concatenation of meeting_id and person_id and is updated every month.

Incident

The Incident table includes records for youth incidents (significant events). The fields in this table are derived from joining multiple SACWIS tables: Incident Youth, Incident Circumstance, Case Type, Case Type Member, and Code Description. When joining table Incident Youth and Incident

[6] The combination of investigation/person may have multiple addresses. To get only one address per investigation/person, we extract only the residence type address (CD_ADDR_TYPE = 1) or the primary address (FL_PRIM = 'Y') or the most recent address based on the date/time when the address record was updated (TS_UPDT).

[7] Added by E. Wiegand, based on conversation with B. Chor and Z. Luo on 3/8/2018.

Circumstance, the incident category code from table Incident Circumstance was restricted to child/youth incident.

This table consists of the identifier of youth incident (incident_id), date of incident (incident_date), incident type code (circumstance_cd), the description of incident type code (circumstance_desc), CYCIS case ID associated with incident (case_id), and person associated with incident (person_id). It contains one record per the combination of incident id and incident type code.

This table is updated monthly. Every time it is updated, the new data overwrites the previous version of the table.

Service Plan

The Service Plan table is the basic record for the service plans associated with children in care. The fields in this table are derived from Service Plan table and Case Type table in SACWIS. This table is unique by service plan id. It consists of service plan id (serv_plan_id), date when the service plan was created (serv_plan_date), and the CYCIS case id associated with service plan (case_id).

Service Plan Outcome

This table includes records for outcomes (goals) associated with each service plan. All of the fields in this table are derived from Service Plan Outcome table in SACWIS. This table contains one record per service plan outcome id (outcome_id). A single service plan may have multiple (or no) outcomes. This table includes text descriptions for the outcome and strengths supporting the achievement of the outcome; however, these fields are only available in the fully identified copy of the database, since information such as names is sometimes included in the free text fields.

Service Plan Action

The Service Plan Action table has records for actions taken by the family, the caseworker, the caregiver, and other party to meet the goals for the child. Fields from this table are derived from Service Plan Action table in SACWIS. Service plan actions, uniquely identified by action_id, are affiliated with service plan outcomes. There may be multiple actions affiliated with a single outcome. There is no direct connection between service plan actions and service plans, either in the CHILD or in SACWIS, except through outcomes. However, there are some actions that are affiliated with outcome_id values that are not represented in the database, because these outcome records were also missing from SACWIS. These actions cannot be linked to service plans or cases.

Text fields for the task and party responsible are available for service plan actions but only in the fully identified version of the database, since information such as names is sometimes included in the free text fields.

Service Plan Member

The Service Plan Member table links service plans to children and adult participants. All fields in this table are derived from Service Plan Member table in SACWIS. This table is unique by the combination of service plan id (serv_plan_id) and person id (person_id). Each individual record in this table is identified as adult/caregiver or child. This table is updated monthly.

Tables from Legacy Golden Copy

Assignment

The assignment table relates a case to a case manager by assignment date and is sourced from the Legacy Golden Copy table CFTVCM9800 (IMSA screen CM06 Case Manager). Though this table contains the history of case manager assignments, it is truncated and entirely reloaded each month. This is because the history is maintained in the source DCFS system, not due to any additional processing on Chapin Hall's end.

Case

The case table stores summary information about the status of a case including case type, whether it is currently open, and dates and reasons for opening and closing. It is sourced from Legacy Golden Copy tables CFTVCM9000 (IMSA screen CM02 Case Open/Reopen), which includes case opening/reopening and limited related information; and CFTVCM9100 (IMSA screen CM05 Open/Transfer/Closing), which contains the case's open, (team) transfer, and closing dates, and is used here to obtain the date when the case was opened for a case service spell. This is done because CFTVCM9000 only contains information about the most recent opening of a case; for cases that have been closed and reopened, first_open_date and seq_open_date will differ. This table is truncated and entirely reloaded each month.

Case Manager

This table contains current case manager attributes, including RSF (region – site – field, also known as team) assignment and date, worker type and position, and whether the worker is currently active. It is sourced from the Legacy Golden Copy table CFTVMG9000 (IMSA screen MG02 Caseworker Information).

Case Manager RSF

This table contains a history of the RSF code of each case manager, also known as team. Each month, the latest version of the table is appended to the existing table in its entirety. The table as of a specific load may be accessed by filtering where source_month equals the month of interest. Data from December 2017 and before was transferred from Chapin Hall's older records; data beginning in January 2018 was loaded from Legacy Golden Copy (the same RSF data appended to Case Manager each month).

Case RSF

This table relates a case to an RSF code as of a particular start date and is sourced from the Legacy Golden Copy table CFTVCM9100 (IMSA screen CM05 Open/Transfer/Closing). Though this table contains the history of case RSF assignments, it is truncated and entirely reloaded each month. This is because the history is maintained in the source DCFS system, not due to any additional processing on Chapin Hall's end.

Case Sequence

This table stores information about the history of openings and closings on a case and is sourced from Legacy Golden Copy table CFTVCM9100 (IMSA screen CM05 Open/Transfer/Closing). Since each case may be opened and closed multiple times, this table is at the grain of case plus opening sequence number and includes current status and dates and reasons for each opening and

closing. It differs from the case table in that the case table only contains information about the most recent opening on a case. This table is truncated and entirely reloaded each month: the case history is maintained in the source DCFS system, not through additional Chapin Hall processing.

Client

This table stores demographic information on DCFS clients (from the LGC/CYCIS system) and is sourced from Legacy Golden Copy tables CFTVCR9000 (IMSA screen CR03 Client Basic Registration and IMSA Screen CM14 Case Eligibility) and CFTVCR9400 (IMSA screen CR06 Client Education). This includes information such as names, birthdates, social security number, gender, citizenship, and educational history. However, many of these fields are redacted from the version of the table stored on resdb. The race fields on this table are the result of transforming a series of categorical variables including race descriptions into a series of boolean columns indicating whether the specified race appeared in any of the original categorical fields. This table is truncated and reloaded each month.

Client Address

The client address table relates a CYCIS client to an address as of a given month. It is sourced from the Legacy Golden Copy table CFTVCR9000 (IMSA screen CR03 Client Basic Registration). This table is cumulative: each month, the latest version of the table is appended to the existing table in its entirety. The table as of a specific load may be accessed by filtering where source_month equals the month of interest. This table was seeded with the address history table that existed in the previous CYCIS database. All data from December 2017 and earlier was imported from Chapin Hall's previous database; records beginning in January 2018 were added from monthly LGC extracts.

Note that for children in care (i.e. open child cases), the "Client Address" value is not where the child lives but the address of the child when the case was opened. To determine the child's address while in care, use the Living Arrangement table joined with Provider Address on provider id.

Cluster

This table relates individual clients to clusters (roughly, households) and includes information about when a client was first added to a cluster, whether a given client is the cluster head of household, and the client status. This table is sourced from Legacy Golden Copy tables CFTVCR9100 (IMSA screen CR07 Family ID History), CFTVCR9500 (IMSA screen CR08), and CFTVCR9600 (IMSA screen CR08 Family Composition). It is truncated and reloaded entirely each month. Though it contains a history of additions to a cluster, this history is maintained in the source DCFS system, not due to additional Chapin Hall processing.

Goal

This table stores the history of permanency goals on a case and is sourced from the Legacy Golden Copy table CFTVCM9600 (IMSA screen CM11 Permanency Goal History). It is truncated and reloaded entirely each month. Though it contains a history of permanency goals, this history is maintained in the source DCFS system, not due to additional Chapin Hall processing.

Legal

This table stores the history of legal hearings attached to a case with information about each hearing including date, county of jurisdiction, and whether the hearing led to a change in legal status. This

table is sourced from Legacy Golden Copy table CFTVCM9400 (IMSA screen CM13 Court Hearing Results – Legal History). It is truncated and reloaded entirely each month. Though it contains a history of hearings, this history is maintained in the source DCFS system, not due to additional Chapin Hall processing.

Living Arrangement

This table stores the history of living arrangements on child cases including date of change, provider, reason for placement, and type of service. It is sourced from Legacy Golden Copy table CFTVCM9200 (IMSA screen CM07 Living Arrangement History). It is truncated and reloaded entirely each month. Though it contains a history of living arrangements, this history is maintained in the source DCFS system, not due to additional Chapin Hall processing.

Provider Address

The provider address table relates a CYCIS provider to an address as of a given month. It is sourced from the Legacy Golden Copy table CFTVPR1000 (IMSA screen PR02 Provider Registration). This table is cumulative: each month, the latest version of the table is appended to the existing table in its entirety. The table as of a specific load may be accessed by filtering where source_month equals the month of interest. This table was seeded with the address history table that existed in the previous Chapin Hall CYCIS database. All data from December 2017 and earlier was imported from Chapin Hall's previous database; records beginning in January 2018 were added from monthly LGC extracts.

Provider

This table stores information about providers, including RSF codes and names of individuals or facilities and is sourced from the Legacy Golden Copy table CFTVPR9000 (IMSA screen PR02 Provider Registration). While original social security numbers and federal employer identification numbers are redacted from resdb01, Chapin Hall-generated surrogate values are provided that may be used to group individuals or organizations with like SSNs for FEINs (in practice, this is used to identify multiple sites that are parts of the same agency). Each integer value in ch_prv_ssn_num or ch_prv_fein_num has a 1-to-1 relationship with a single string of up to 9 digits stored in a separate table on chpgdb01.

Provider License

This table stores changes in licensing status for providers and includes such information as application date, hours, age limits, and capacity. It is sourced from the Legacy Golden Copy tables CFTVLC1000 (IMSA screen unknown) and CFTVLC9100 (IMSA screen LC10).

Combined Tables

Address

The address table is a record of all unique source addresses from all DCFS sources. Each month, it is updated with client and provider addresses sourced from the Legacy Golden Copy tables CFTVCR9000 and CFTVPR1000, respectively, and SACWIS addresses sourced from the Address table. Only addresses that do not already exist in the table are loaded each month: the field source_month indicates the month where the address first appeared. Addresses in this table are not parsed or cleaned during the load process. In addition to the addresses loaded each month, this table

was initially seeded with all addresses that existed in the old Chapin Hall CYCIS and CANTS databases as of December 2017. These legacy records have a source_month of 0.

Derived Tables

The following three tables do not contain any additional data; they are derived from the main database tables. They are intended to reduce time spent by users coding frequently used logic. For practical purposes, these are equivalent to database views.

If you would like to suggest additional derived tables to minimize duplication of efforts or streamline routine processes, email idb@chapinhall.org.

Case County

This table lists three "county" affiliations for each case (by caseid) to simplify reporting by geography. The county options include:

1. County of current placement: This is derived from the current address of the provider on the current living arrangement. It is null if the child is not currently in placement with a provider. For DCFS purposes, this is a popular way to map and report on cases, though this value can change frequently for the same case.

2. Current legal county: This is the county of the case's most recent legal status. Legal county is very stable. DCFS often uses this value for figuring out worker resource needs by county.

3. Head of household's county: This is the county of the head of household in the child's family (i.e. parent's county). This value is important for visitation and development of family resources. This county should frequently align with legal county, but occasionally these counties are far apart, especially if the parent's housing is unstable. This value is calculated based on the current address for the head of household in the cluster.

Master Event

This table combines records from Case, Provider, and Living Arrangement and contains one row for each case opening or closing or placement event that a child experiences. Each record contains an event type, the event starting and ending dates, and a code for the next event the child experiences.

In this table, a placement is defined as the time a child is with a certain provider in a certain type of care. Case and placement events are defined by the value of the Type of Service (type_serv) variable. If the Type of Service variable has no value as in the case of unpaid placements, the 'type_code' variable is used to define the event. Both provider SSN and prvid are used to identify providers – if one of these two values are the same for adjacent records, the provider is the same.

Master Spells

Each record in the master spells table captures the time a child spends in out-of-home placement during an open child case. Out-of-home placement spells end with a case closing, or placement in the home of parent, home of adoptive parent, or subsidized guardianship. This table is generated using master event based on the following logic: Certain events are considered case opening events ('HAP','HMP','SGH','GDN','DEC','000','ZZZ','ZCA') and certain others are considered case closing

events (everything else). When closing or opening events are adjacent to each other, they get combined into one spell. New spells are only opened when an opening event appears as the first record or after a closing event. The censor field reflects whether a spell is currently open and, if not, what caused it to close.

Data Dictionary

Beginning on the next page is a full list of elements in each table, including data type and a basic description. All of this metadata is also maintained in the database schema and can be viewed in psql using the command "\d+ child.*".

Table 1: CHILD Data Dictionary (last updated 5/25/2018)

Table Name	Column Name	Primary Key	Data Type	Description
address	addr_id	PK	integer	Chapin Hall-assigned unique identifier for a DCFS address
address	city		character varying	City
address	cnty_cd		character varying	DCFS county code (Not FIPS)
address	source_month		integer	Indicates the month of the extract that added this record to the table
address	state		character varying	State
address	street_addr		character varying	Street address
address	zip		character varying	ZIP Code
allegation	alleg_cd		character varying	Code for type of allegation alleged (these values are defined by AFCARS reporting requirements)
allegation	alleg_desc		character varying	Description of type of allegation alleged
allegation	alleg_finding_cd		integer	Decision of the investigative worker about the validity of the allegations (code)
allegation	alleg_finding_desc		character varying	Decision of the investigative worker about the validity of the allegations (description)
allegation	alleg_id	PK	integer	Unique identifier for an allegation
allegation	invst_id		integer	Unique identifier for an investigation
allegation	person_id_perp		integer	Person ID for the alleged perpetrator
allegation	person_id_vic		integer	Person ID for the alleged victim
allegation	relation_cd		character varying	Relationship between alleged perpetrator and alleged victim (code)
allegation	relation_desc		character varying	Relationship between alleged perpetrator and alleged victim (description)
allegation	scrnum		character varying	Unique identifier for a household
allegation	scrseq		character varying	Investigation sequence number within a household

allegation	source_month		integer	Indicates the month of the extract that added this record to the table
assignment	assgn_date	PK	date	Case manager assignment date
assignment	case_id	PK	character varying	Unique identifier for a child or family case (equivalent to cli_id for the child or clus_id for the family)
assignment	cmgr_id		character varying	Unique identifier for a case manager
assignment	source_month		integer	Indicates the month of the extract that added this record to the table
case	case_id	PK	character varying	Unique identifier for a child or family case (equivalent to cli_id for the child or clus_id for the family)
case	case_status		character varying	Indicates whether the case is closed (C) or open (O)
case	case_type		character varying	Case Type - F (Family), C (Child), or J (???)
case	child_cli_id		character varying	For child cases, this is the unique identifier of the child
case	fam_clus_id		character varying	For family cases, this is the unique identifier of the family
case	first_open_date		date	Date of first ever case opening
case	seq_case_rsf		character varying	Region/site/field of most recent case sequence
case	seq_close_date		date	If most recent case sequence is closed, date of case closing
case	seq_close_rsn		character varying	If most recent case sequence is closed, reason for case closing
case	seq_open_date		date	Date of most recent case sequence opening
case	seq_open_rsn		character varying	Reason for most recent case sequence opening
case	source_month		integer	Indicates the month of the extract that added this record to the table
case_manager	cmgr_id	PK	character varying	Unique identifier for a case manager
case_manager	cmgr_spvs_id		character varying	Case manager ID for the case manager's supervisor
case_manager	posn_code		character varying	Position code (maps to job title)

case_manager	posn_date		date	Position date
case_manager	rsf_begn_date		date	Begin date for worker's region/site/field affiliation
case_manager	rsf_code		character varying	Current region/site/field affiliation of case manager
case_manager	source_month		integer	Indicates the month of the extract that added this record to the table
case_manager	status		character varying	Case manager employment status - I (inactive) or A (active)
case_manager	status_date		date	Start date of status
case_manager	type_wkr		character varying	Type of function/work done
case_manager	type_wkr_date		date	Date of work type
case_manager_rsf	cmgr_id	PK	character varying	Unique identifier for a case manager
case_manager_rsf	rsf_code		character varying	Historical region/site/field affiliation of case manager
case_manager_rsf	source_month	PK	integer	Indicates the month of the extract that added this record to the table
case_rsf	case_id	PK	character varying	Unique identifier for a child or family case (equivalent to cli_id for the child or clus_id for the family)
case_rsf	rsf		character varying	Region/site/field code for this case
case_rsf	rsf_start_date	PK	date	Start date for the region/site/field code for this case
case_rsf	source_month		integer	Indicates the month of the extract that added this record to the table
case_sequence	case_id	PK	character varying	Unique identifier for a child or family case (equivalent to cli_id for the child or clus_id for the family)
case_sequence	opseqcnt	PK	character varying	Sequence field indicating each unique case opening; most recent opening is always '01'
case_sequence	seq_close_date		date	Date this case sequence closed
case_sequence	seq_close_rsn		character varying	Reason why this case sequence closed
case_sequence	seq_open_date		date	Date this case sequence opened
case_sequence	seq_open_rsn		character varying	Reason why this case sequence was opened

case_sequence	seq_status		character varying	Indicates whether the case sequence is closed (C) or open (O)
case_sequence	source_month		integer	Indicates the month of the extract that added this record to the table
client	bdate		date	Date of birth
client	citizenship_cd		integer	Citizenship status code
client	citizenship_desc		character varying	Citizenship status description
client	cli_id	PK	character varying	Unique identifier for an individual in CYCIS
client	edlevel		character varying	Current level of education
client	ethnicity		character varying	Ethnicity
client	first_name		character varying	First name of client
client	gender		character varying	Gender
client	last_name		character varying	Last name of client
client	marital		character varying	Marital status
client	middle_name		character varying	Middle name of client
client	race_asian		boolean	Indicates whether the client has a race of Asian
client	race_black		boolean	Indicates whether the client has a race of black
client	race_hpi		boolean	Indicates whether the client has a race of native Hawaiian or other Pacific Islander
client	race_nat_amer		boolean	Indicates whether the client has a race of Native American
client	race_white		boolean	Indicates whether the client has a race of white
client	religion		character varying	Religion
client	rin		character varying	Recipient number from DHS/DHFS for individuals who are served by multiple systems
client	sacwis_person_id		character varying	Person ID for this individual in SACWIS

client	schprog		character varying	Type of educational program a child is enrolled in (individualized, vocational, etc.)
client	schstat		character varying	Educational enrollment status
client	source_month		integer	Indicates the month of the extract that added this record to the table
client	ssn		character varying	SSN
client_address	addr_id		integer	Chapin Hall-assigned unique identifier for a DCFS address
client_address	cli_id	PK	character varying	Unique identifier for an individual in CYCIS
client_address	source_month	PK	integer	Indicates the month of the extract that added this record to the table
cluster	cli_id	PK	character varying	Unique identifier for an individual in CYCIS
cluster	clus_id	PK	character varying	Unique identifier for a household in CYCIS
cluster	hoh_ind		boolean	Indicates if the individual is the head of household for the cluster
cluster	mbr_date	PK	date	Date this client was added to the household
cluster	mbr_stat_code		character varying	Role code - member's role in family
cluster	source_month		integer	Indicates the month of the extract that added this record to the table
goal	case_id	PK	character varying	Unique identifier for a child or family case (equivalent to cli_id for the child or clus_id for the family)
goal	goal_cd		character varying	Permanency goal code
goal	goal_date	PK	date	Date goal was established
goal	source_month		integer	Indicates the month of the extract that added this record to the table
incident	case_id		character varying	CYCIS case ID associated with incident (as entered in SACWIS)
incident	circumstance_cd	PK	integer	Type of incident (code)
incident	circumstance_desc		character varying	Type of incident (description)
incident	incident_date		date	Date of incident

incident	incident_id	PK	integer	Unique identifier of a youth incident ("significant event")
incident	person_id		integer	Person associated with incident
incident	source_month		integer	Indicates the month of the extract that added this record to the table
investigation	approx_date		boolean	Indicates that incident date specified is approximate
investigation	finding_cd		smallint	Investigation finding (code)
investigation	finding_date		date	Date the finding was updated
investigation	finding_desc		character varying	Investigation finding (description)
investigation	incident_date		date	Date of the incident under investigation.
investigation	init_date		date	Date when the first contact with the last victim in the investigation has been initiated
investigation	invst_id	PK	integer	Unique identifier for an investigation
investigation	merge_cd		integer	Value from SACWIS - indicates whether the investigation was merged and retained or removed (code)
investigation	merge_desc		character varying	Value from SACWIS - indicates whether the investigation was merged and retained or removed (description)
investigation	report_date		date	Report date associated with investigation
investigation	scrnum		character varying	Unique identifier for a household
investigation	scrseq		character varying	Investigation sequence number within a household
investigation	source_month		integer	Indicates the month of the extract that added this record to the table
investigation	status_cd		smallint	Current state of the investigation (code)
investigation	status_desc		character varying	Current state of the investigation (description)
investigation	unk_date		boolean	Indicates that incident date specified is unknown
investigation_address	addr_id	PK	text	Unique identifier for a DCFS address
investigation_address	invst_id	PK	integer	Unique identifier for an investigation

investigation_address	source_month		integer	Indicates the month of the extract that added this record to the table
investigation_person_address	addr_id	PK	integer	Unique identifier for a DCFS address
investigation_person_address	invst_id	PK	integer	Unique identifier for an investigation
investigation_person_address	person_id	PK	integer	Unique identifier of an individual
investigation_person_address	source_month		integer	Indicates the month of the extract that added this record to the table
legal	case_id	PK	character varying	Unique identifier for a child or family case (equivalent to cli_id for the child or clus_id for the family)
legal	docket_id		character varying	Docket ID
legal	guardian		character varying	Legal guardian
legal	lgl_cnty		character varying	Legal county
legal	lgl_date	PK	date	Date of legal status change
legal	lgl_hear_type	PK	character varying	Legal hearing type
legal	lgl_status		character varying	Legal status
legal	source_month		integer	Indicates the month of the extract that added this record to the table
living_arrangement	case_id	PK	character varying	Unique identifier for a child or family case (equivalent to cli_id for the child or clus_id for the family)
living_arrangement	cntr_id		character varying	Contract ID
living_arrangement	livar_date		date	Living arrangement change date
living_arrangement	livar_seq	PK	bigint	Living arrangement sequence number - used to disambiguate multiple living arrangements in the same day, ordered by time field in the source
living_arrangement	opseqcnt		character varying	Sequence field indicating each unique case opening; most recent opening is always '01'
living_arrangement	prv_id		character varying	Unique identifier for a provider

living_arrangement	reason		character varying	Reason placed code
living_arrangement	source_month		integer	Indicates the month of the extract that added this record to the table
living_arrangement	term_reason		character varying	Living arrangement termination reason code
living_arrangement	type_code		character varying	Living arrangement type
living_arrangement	type_serv		character varying	Type of service code
master_event	case_id	PK	character varying	Unique identifier for a child or family case (equivalent to cli_id for the child or clus_id for the family)
master_event	ch_fein_ssn_num		integer	Chapin Hall-generated surrogate FEIN or SSN for provider
master_event	close_reason		character varying	Reason for case closing (or LL for a living arrangement instead of case closing)
master_event	event		character varying	Event type
master_event	event_from_dt	PK	date	Event date
master_event	event_to_date		date	Date of following event
master_event	fein_ssn_num		character varying	Provider FEIN or SSN number
master_event	opseqcnt		character varying	Sequence field indicating each unique case opening; most recent opening is always '01'
master_event	perf_flag		integer	Indicates whether event has performance contracting type of service code (0 - no, 1 - yes, 2 - has both)
master_event	pre_adopt		integer	Indicates whether placement was ever identified as an adoptive placement (FHA). Intended to identify preadoptive placements where foster parents plan to adopt the child.
master_event	prv_id		character varying	Unique identifier for a provider
master_event	source_month		integer	Indicates the month of the extract that added this record to the table
master_event	type_serv		character varying	Numeric code for type of service provided
master_spell	case_id	PK	character varying	Unique identifier for a child or family case (equivalent to cli_id for the child or clus_id for the family)

master_spell	censor		integer	Flag indicating whether the spell has been censored (i.e. whether the spell was open at last update and update date was assigned as end date)
master_spell	event_from_dt	PK	date	Spell start date
master_spell	next_event		character varying	The event that ended the spell -- used to calculate censor
master_spell	next_event_from_dt		date	Spell end date (start date of event that closes the spell)
master_spell	opseqcnt		character varying	Sequence field indicating each unique case opening; most recent opening is always '01'
master_spell	source_month		integer	Indicates the month of the extract that added this record to the table
master_spell	speller		bigint	Chapin Hall-assigned spell number (numbers are assigned sequentially based on event_from_dt)
meeting	meeting_date		date	Date of meeting
meeting	meeting_id	PK	character varying	Unique identifier for meeting (the combination of "note ID" and "cd_subcatg_type" in SACWIS)
meeting	meeting_type		character varying	Type of meeting (distinguishes Child and Family Team Meeting or Visitation)
meeting	source_month		integer	Indicates the month of the extract that added this record to the table
meeting	supervision_cd		smallint	Distinguishes supervised from unsupervised visitation (code)
meeting	supervision_desc		character varying	Distinguishes supervised from unsupervised visitation (description)
meeting_participant	meeting_id	PK	character varying	Unique identifier for meeting (the combination of "note ID" and "cd_subcatg_type" in SACWIS)
meeting_participant	person_id	PK	integer	Unique identifier of an individual participating in meeting
meeting_participant	source_month		integer	Indicates the month of the extract that added this record to the table
person	bdate		date	Person's date of birth
person	citizenship_cd		smallint	Indicates whether the person is a US Citizen or their current status within the process of becoming a US Citizen (code)

person	citizenship_desc		character varying	Indicates whether the person is a US Citizen or their current status within the process of becoming a US Citizen (description)
person	cycis_cli_id		character varying	Client/registration ID for this person in CYCIS/IMSA
person	ethn_cd		smallint	Ethnicity of a person with Hispanic origin (code)
person	ethn_desc		character varying	Ethnicity of a person with Hispanic origin (description)
person	first_name		character varying	First name of person
person	fl_wkr		character varying	Indicates if the person works for DCFS
person	last_name		character varying	Last name of person
person	marital_cd		smallint	Marital status of person (code)
person	marital_desc		character varying	Marital status of person (description)
person	middle_name		character varying	Middle name of person
person	person_id	PK	integer	Unique identifier of an individual in SACWIS
person	person_status_cd		character varying	Value from SACWIS - indicates the status of this record (active, deleted, etc.) (code)
person	person_status_desc		character varying	Value from SACWIS - indicates the status of this record (active, deleted, etc.) (description)
person	race_asian		boolean	Indicates whether the person has ever reported a race of Asian
person	race_black		boolean	Indicates whether the person has ever reported a race of black
person	race_hpi		boolean	Indicates whether the person has ever reported a race of native Hawaiian or other Pacific Islander
person	race_nat_amer		boolean	Indicates whether the person has ever reported a race of Native American
person	race_white		boolean	Indicates whether the person has ever reported a race of white
person	rin		integer	Recipient number from DHS/DHFS for individuals who are served by multiple systems
person	sex		character varying	Gender (male, female, or unknown)

person	source_month		integer	Indicates the month of the extract that added this record to the table
person	ssn		character varying	Federal social security number of person
person	suffix		character varying	Name suffix (i.e. Jr, Sr, etc)
person	unkwn_name		boolean	Indicates when the worker entering the person record does not have enough information to search for the person
protective_custody	invst_id	PK	integer	The investigation during which the child was taken into protective custody
protective_custody	lgl_outcome_cd		smallint	Type of legal action resulting from the protective custody (code)
protective_custody	lgl_outcome_date		date	Date of legal action resulting from the protective custody
protective_custody	lgl_outcome_desc		character varying	Type of legal action resulting from the protective custody (description)
protective_custody	pc_date		date	Date of protective custody
protective_custody	pc_place_rmv_cd		smallint	Where the child was removed from (code)
protective_custody	pc_place_rmv_desc		character varying	Where the child was removed from (description)
protective_custody	pc_plct_cd		smallint	Type of placement the child was placed in as part of the protective custody (code)
protective_custody	pc_plct_desc		character varying	Type of placement the child was placed in as part of the protective custody (description)
protective_custody	pc_seq	PK	integer	Sequence value for protective custody incidents within an investigation/child combination (protective custody incidents are sorted by date/time protective custody was taken and then by date/time when data was created and was updated)
protective_custody	pc_taker_cd		smallint	Person who took the child into protective custody (code)
protective_custody	pc_taker_desc		character varying	Person who took the child into protective custody (description)
protective_custody	person_id	PK	integer	The person ID of the child taken into protective custody
protective_custody	source_month		integer	Indicates the month of the extract that added this record to the table
provider	aka_first_name		character varying	Provider alternate first name (only populated for individual providers)

provider	aka_last_name		character varying	Provider alternate last name (only populated for individual providers)
provider	aka_mi_name		character varying	Provider alternate middle name (only populated for individual providers)
provider	ch_prv_fein_num		integer	Chapin Hall-generated surrogate FEIN for provider
provider	ch_prv_ssn_num		integer	Chapin Hall-generated surrogate SSN for provider
provider	fein_ssn_ind		character varying	Indicates type of SSN/FEIN provided
provider	lic_prv_id		character varying	Licensing provider ID
provider	prv_fac_name		character varying	Corporate name of agency (only populated for facility providers)
provider	prv_fein_num		character varying	Provider Federal Employer Identification Number
provider	prv_first_name		character varying	Provider first name (only populated for individual providers)
provider	prv_id	PK	character varying	Unique identifier for provider
provider	prv_last_name		character varying	Provider last name (only populated for individual providers)
provider	prv_mi_name		character varying	Provider middle name (only populated for individual providers)
provider	prv_name_ind		character varying	Indicates whether the provider name is populated as a facility name (F) or an individual name (I)
provider	prv_rsf_code		character varying	Region/site/field code of provider
provider	prv_ssn_num		character varying	Provider social security number
provider	source_month		integer	Indicates the month of the extract that added this record to the table
provider_address	addr_id		integer	Chapin Hall-assigned unique identifier for a DCFS address
provider_address	prv_id	PK	character varying	Unique identifier for provider
provider_address	source_month	PK	integer	Indicates the month of the extract that added this record to the table
provider_license	day_age_from_ind		character varying	Indicates whether lower age limit for daytime care is in years, months, or weeks

provider_license	day_age_from_rnge		integer	Lower age limit for daytime care
provider_license	day_age_to_ind		character varying	Indicates whether upper age limit for daytime care is in years, months, or weeks
provider_license	day_age_to_rnge		integer	Upper age limit for daytime care
provider_license	day_capacity		integer	Daytime total capacity (includes extended hours children)
provider_license	day_open_from_hour		integer	Hour daytime care begins
provider_license	day_open_to_hour		integer	Hour daytime care ends
provider_license	ext_age_from_rnge		integer	Lower age limit for extended hours care (before/after school)
provider_license	ext_age_to_rnge		integer	Upper age limit for extended hours care (before/after school)
provider_license	ext_capacity		integer	Capacity for extended hours care (before/after school)
provider_license	lic_appl_date		date	License application date
provider_license	lic_begn_date		date	License beginning date
provider_license	lic_end_date		date	License ending date
provider_license	lic_lang_code		character varying	Language code of primary provider
provider_license	lic_per_code		character varying	License period code
provider_license	lic_pub_rlse_ind		character varying	Public release indicator (home/center agrees to be listed on any release of available homes)
provider_license	lic_seq_cnt	PK	character varying	Sequence count of provider's license
provider_license	lic_stat_code		character varying	Licensing status code
provider_license	lic_stat_date		date	Licensing status date
provider_license	lic_supv_by_id		character varying	Licensing supervisor ID (RSF code for DCFS, provider ID for other agencies)
provider_license	lic_type_code		character varying	License type code
provider_license	nit_age_from_ind		character varying	Indicates whether lower age limit for nighttime care is in years, months, or weeks

provider_license	nit_age_from_rnge		integer	Lower age limit for nighttime care
provider_license	nit_age_to_ind		character varying	Indicates whether upper age limit for nighttime care is in years, months, or weeks
provider_license	nit_age_to_rnge		integer	Upper age limit for nighttime care
provider_license	nit_capacity		integer	Nighttime total capacity
provider_license	nit_open_from_hour		integer	Hour nighttime care begins
provider_license	nit_open_to_hour		integer	Hour nighttime care ends
provider_license	prv_id	PK	character varying	Unique identifier for provider
provider_license	source_month		integer	Indicates the month of the extract that added this record to the table
service_plan	case_id		character varying	CYCIS case ID associated with the service plan (as entered in SACWIS)
service_plan	serv_plan_date		date	Date service plan was created
service_plan	serv_plan_id	PK	integer	Unique identifier for service plan
service_plan	source_month		integer	Indicates the month of the extract that added this record to the table
service_plan_action	action_id	PK	integer	Unique identifier for service plan action
service_plan_action	cmplt_date		date	Date service plan action was completed
service_plan_action	outcome_id		integer	Unique identifier for the outcome addressed by this action
service_plan_action	resp_party		character varying	Description of the party responsible for accomplishing the action
service_plan_action	rmv_date		date	Date service plan action was removed
service_plan_action	source_month		integer	Indicates the month of the extract that added this record to the table
service_plan_action	start_date		date	Effective date of the service plan action
service_plan_action	task		character varying	Description of task
service_plan_member	person_id	PK	integer	Unique identifier for person

service_plan_member	role		character varying	Indicates the individual's role in the service plan ("A" represents adult or caregiver; "K" represents child)
service_plan_member	serv_plan_id	PK	integer	Unique identifier for service plan
service_plan_member	source_month		integer	Indicates the month of the extract that added this record to the table
service_plan_outcome	estab_date		date	Date outcome was established
service_plan_outcome	outcome_desc		character varying	Text description of outcome
service_plan_outcome	outcome_id	PK	integer	Unique identifier for service plan outcome (an outcome is a goal of the service plan)
service_plan_outcome	person_id		integer	Person associated with this service plan outcome (may be the child or another person associated with the case, like a birth parent)
service_plan_outcome	rmv_date		date	Date outcome was removed
service_plan_outcome	serv_plan_id		integer	Unique identifier for service plan associated with this outcome
service_plan_outcome	source_month		integer	Indicates the month of the extract that added this record to the table
service_plan_outcome	strengths		character varying	Text description of strengths supporting achievement of the outcome
team	agency_prv_id		character varying	Team licensing agency, if not DCFS; links to provider table
team	office_addr_id		integer	Field office Chapin Hall address ID
team	rsf_code	PK	character varying	Team region/site/field code
team	source_month		integer	Indicates the month of the extract that added this record to the table

Appendix 1: Comparing Historical Data in CANTS and SACWIS[8]

Based on comparison of populations between DCFS-L and SACWIS, we see no evidence of systemic missingness in SACWIS that would suggest incorporating missing records from DCFS-L into the CHILD would improve the overall quality and representativeness of DCFS data in that database. Where there is evidence to explain why records appear in DCFS-L and not in SACWIS, it largely points to data errors.

The Chapin Hall Illinois DCFS Database (CHILD) derives its data directly from DCFS' SACWIS system. Chapin Hall's older database of abuse and neglect investigations (DCFS-L) was originally built to reflect DCFS' CANTS database, and it is currently (fiscal year 2017-2018) populated using data that is extracted from SACWIS and transformed in NOMAD by DCFS before being sent to Chapin Hall for further processing.

For any data that is overlapping between DCFS-L and SACWIS, we believe SACWIS is a stronger source,[9] since SACWIS is the data DCFS uses to analyze these issues, and we know that there have been issues with the NOMAD transformation in the last few years.

We do not have details on the transition from CANTS to SACWIS and do not know if all the historical records in CANTS that are contained in DCFS-L were transferred to SACWIS. This document reports on our exploration of that issue. We look at two types of records – investigations and people – to identify gaps in SACWIS' coverage.

Investigations

To identify the differences in investigation records between CANTS and SACWIS, we compared table **dcfsl_cps_inv** in DCFS-L and table **dbo.investigation** in SACWIS. We compared the tables based on the combination of household id and household investigation sequence. Respectively, these columns are called *scrnum* and *scrseq* in table **dcfsl_cps_inv** and *id_invst_case* and *id_scr_seq* in table **dbo.investigation**.

To reduce differences as much as possible, we attempted to compare these data sources at the same period in time. We used SACWIS tables extracted from DCFS' system on April 2, 2018 and DCFS-L tables updated with files sent by DCFS in early April (the data from both systems was updated through March 2018). Because we cannot control exactly when our monthly extract from DCFS is pulled and sent to NOMAD, we are not able to be 100% precise.

As depicted in Figure A, we identified 19,357 investigation records that can be found only in table **dcfsl_cps_inv** (DCFS-L), 5,926 records can be found only in table **dbo.investigation** (SACWIS),

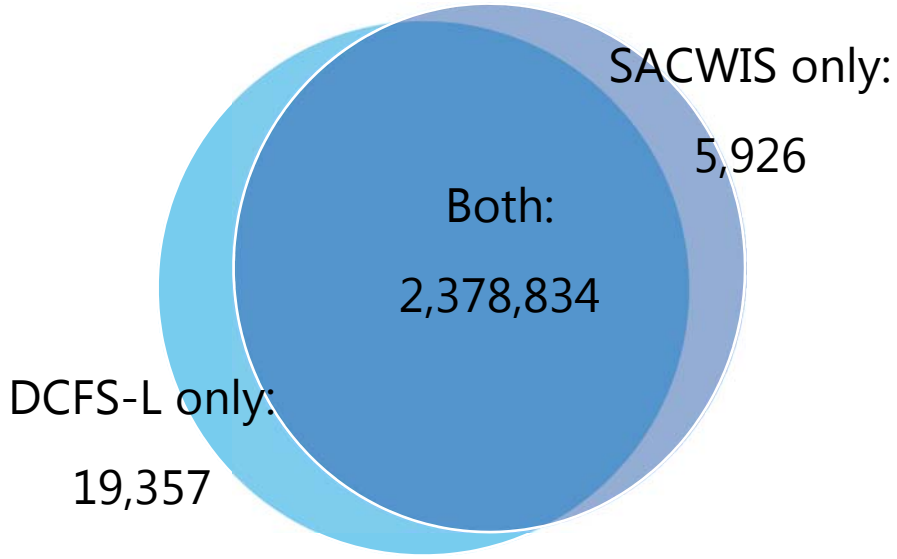
---

[8] This appendix summarizes work completed by Lia Amelia and Emily Wiegand, beginning in October 2017 and completed in May 2018.

[9] The only exceptions to this are instances where we suspect that DCFS may have removed or overwritten data that has value for research purposes. There are two possible instances of this: 1) individuals whose personal information has been purged or expunged from allegations – further research has shown that this data is retained in the SACWIS backend and can be queried, although it is masked from the frontend view; and 2) old allegation 60 records that were declared retroactively unfounded and may still represent examples of abuse and neglect for research purposes. See the discussion of allegation 60 above in the description of the CHILD's Allegation table.

and 2,378,834 records in common between the tables. In short, less than 1% of investigation records in DCFS-L are not found in SACWIS and vice versa.

Figure A. Overlap in Investigations between DCFS-L and SACWIS



We set aside the records that are found only in SACWIS – again, because SACWIS is DCFS' primary source, we trust these records are correct, and they will be included in CHILD by default. The records that are found only in DCFS-L would not be included in the CHILD without specific intervention, however.

In the interest of identifying possible reasons why some records are not represented in SACWIS, we looked at the distribution of number of investigation per year based on investigation initiation date (initdate). This should show whether there are records from a specific time that are left out of SACWIS – either older data that was not copied over or newer records that were added after our SACWIS data was pulled. Figure 2 shows the number of investigation only in DCFS-L per year.

Figure B. Number of investigations only in DCFS-L by year (initdate)

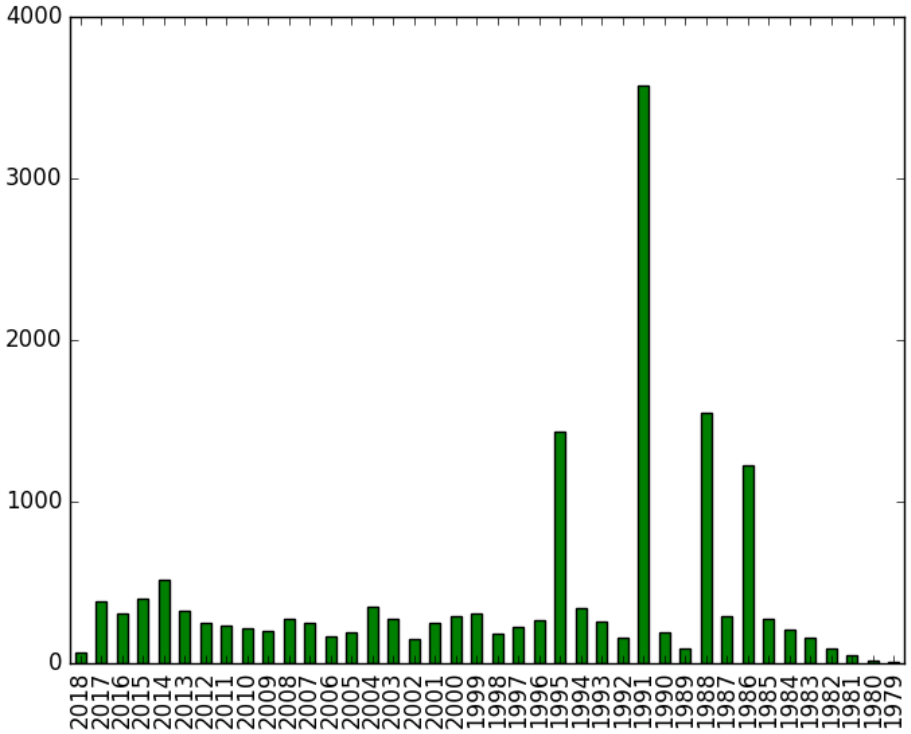


Figure B demonstrates that the number of investigation only in DCFS-L fluctuates from year to year.[10] Year 1991 has the highest number of investigations only in DCFS-L (3,574 investigations), followed by year 1988 (1,547 investigations), 1995 (1,430 investigations), and 1986 (1,228 investigations). Upon further investigation, in 1991, the majority of investigation are found on May (1,727 investigations) and June (1,061 investigations). An immediate cause for these chronological clusters is not clear. Certainly, there are more issues with older investigations (particularly those from 1995 and earlier), but the uneven distribution of missing records does not suggest systemic error—for example, a clear cutoff before which data was not transferred to SACWIS.

We then looked at the children associated with these missing investigations to see if any investigations associated with the children appeared in SACWIS – perhaps the missing investigations had been identified by DCFS as duplicate records. We joined two DCFS-L tables, **dcfsl_cps_inv** and table **dcfsl_cps_chinv** , to identify subjsids for children associated with each investigation. subjsid corresponds to the SACWIS value id_person. We found 37,653 children associated with the 19,357 investigation records that exclusively can be found in table **dcfsl_cps_inv**. Of these, 22,709 (60%) matched to records in table **dbo.person** in SACWIS.[11] Of the corresponding person records, 54% had a cd_pers_stat, or record status, of "merged", "deleted", or "merged-non-retained" in **dbo.person**.

[10] There are also a substantial number of investigations found only in DCFS-L, 3,279 investigations, that have an initdate in the year 1899. These investigations are generally suspect for obvious reasons.

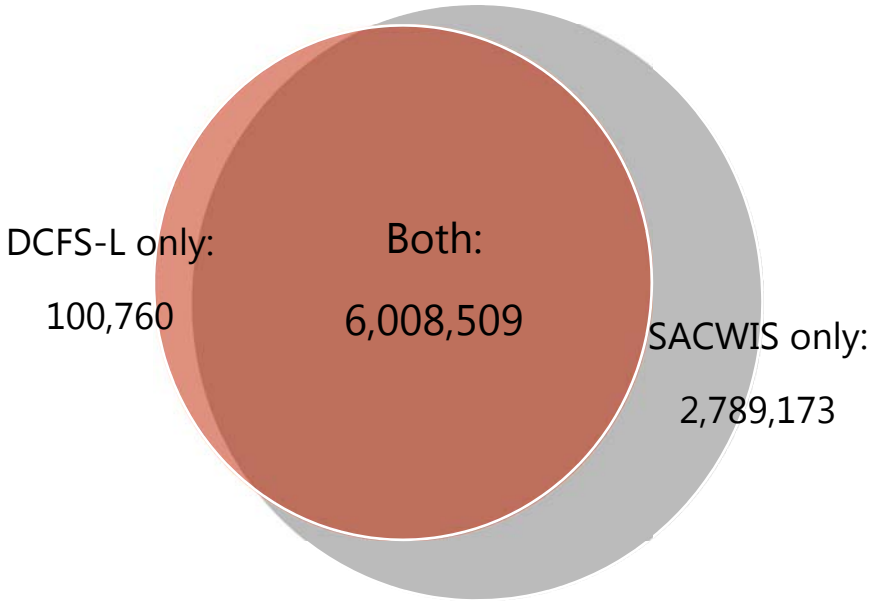
[11] Those that did not match are explored later in this appendix when we compare individuals between the two systems.

Overall, we found no evidence looking at investigations to believe that there are systemic gaps in the investigations in SACWIS. We believe the most likely hypothesis for the missing investigations is that they represent records that were cleaned up during or after the system conversion – the fact that over half of children associated with these investigations matched to person records in SACWIS that had been merged or deleted reinforces this hypothesis.[12]

Individuals

Next, we looked at individual records between the two systems. The basic comparison between the two systems is depicted in Figure C. The SACWIS database contains information not only on investigations but also on an array of child welfare services operated by DCFS, so the raw **dbo.person** in SACWIS is significantly larger than the corresponding table in DCFS-L, **dcfsl_individuals**.

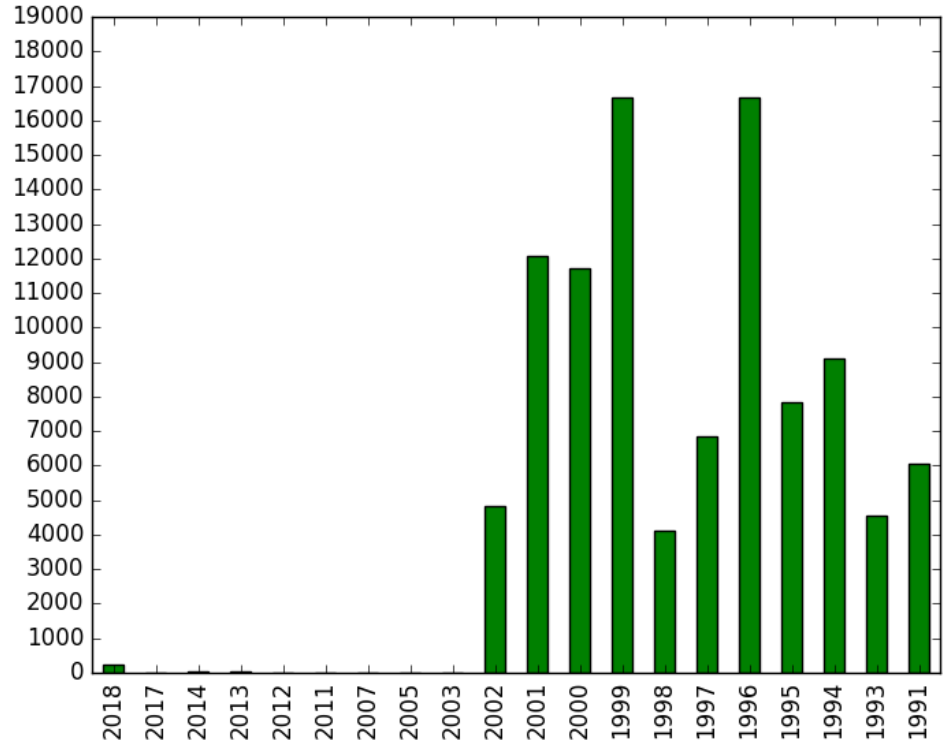
Figure C. Overlap in individuals between DCFS-L and SACWIS



There were 100,760 individuals found only in DCFS-L, representing about 1.7% of all individual records in DCFS-L. The distribution of these records over time, based on the pull date associated with their update id (the date of the extract that most recently changed the individual record in DCFS-L) is shown in Figure D.

[12] Subsequent investigation determined that DCFS may be merging or deleting records that have value for research. See Appendix 2 for more detail. We will continue to investigate this issue.

Figure D. Number of individuals only in DCFS-L by year (update id)



Note that the years 1992, 2004, 2006, 2008-2010, and 2015-2016 are omitted from Figure D because there are no individuals whose most recent update id was associated with a pull date in those years. Ignoring the small concentration of records with recent update ids (these probably represent problems matching the timing of the two sources precisely), Figure D shows that missing records are again primarily from older years, 1991-2002. Records are unevenly distributed over time, although there are reasons why the overall distribution of most recent update ids for individuals may also be irregular.[13]

We wanted to see if the person records exclusively in DCFS-L were perhaps duplicates of person records visible in SACWIS. We used a join between cycis_id in table **dcfsl_individuals** and id_cycis in table **dbo.person** as a quick method of seeking these duplicates. This check was only possible for 5,575 records (about 5.5%) because cycis_id was null in DCFS-L for the other 95,185 records. However, of the records with a cycis_id, 98.4% had a matching record in SACWIS. For those records that did match on cycis_id, similarity in other identifiers suggests that these records were duplicates. While this is only a small segment of the missing population, records with cycis_ids

[13] Records receive a new update id in DCFS-L when there is any change in the record from the previous month. Periodically, changes in the formatting of the extract or the raw data in certain fields may register as "updates" for large numbers of records, even though there has been no substantive change in the records. This means that, on the whole, some update ids are much more common than others.

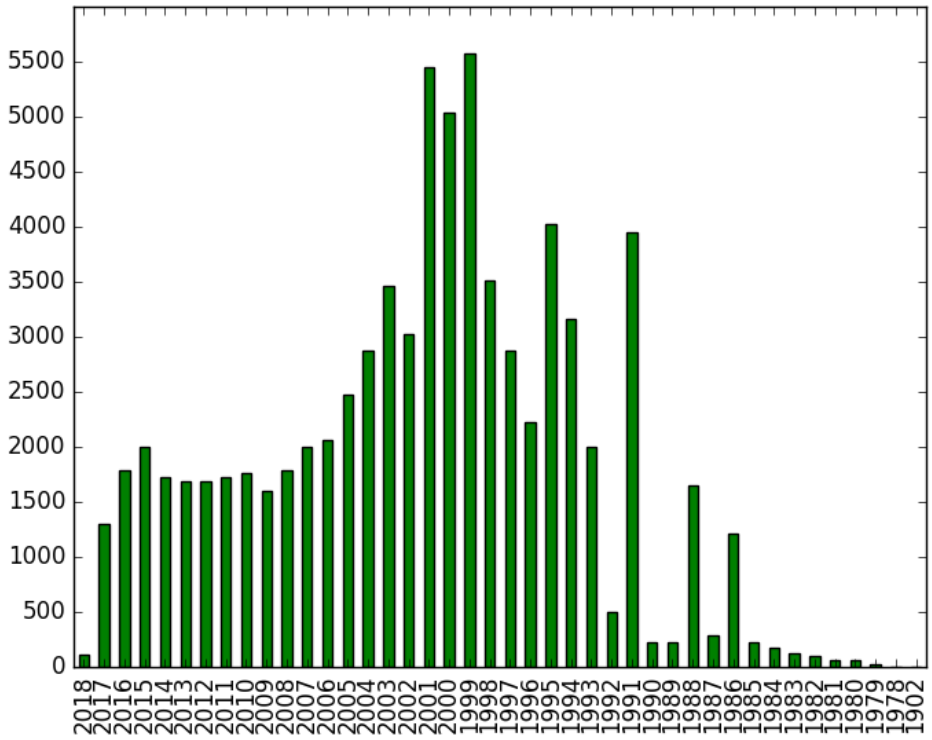
represent individuals who were most heavily engaged with DCFS; it is most important that we capture this population comprehensively in the CHILD.[14]

Next we concentrated our analysis on two subpopulations – children (victims) and caretakers (perpetrators). Missing individuals of these type would be more impactful for analyses than would missing records for other individuals associated with households.

Children

We compared childsid in table **dcfsl_cps_alleg** in DCFS-L and person_id_vic from table **dbo.invst_allegation** in SACWIS.[15] We identified 2,468,498 children in both tables, 13,331 only in SACWIS, and 68,138 children (2.6%) only in DCFS-L. These records peaked between 1999 and 2001 but were also spread across other years (based on investigation initiation date).

Figure E. Number of children/victims only in DCFS-L by year (initdate)



Among the 68,138 childsids that were associated with allegations in DCFS-L but do not match to SACWIS, just under half of them (n= 31,605) exist in SACWIS' table **dbo.person**. However, the vast majority of those (96%) have a status of "merged", "deleted", or "merged-non-retained" on the

[14] Note that this population overlaps with the population of individuals whose match using system identifiers could not be duplicated using a probabilistic match. As discussed in the next appendix, the SACWIS duplicate records are not linked to the old allegation histories. This represents an area that requires further investigation.

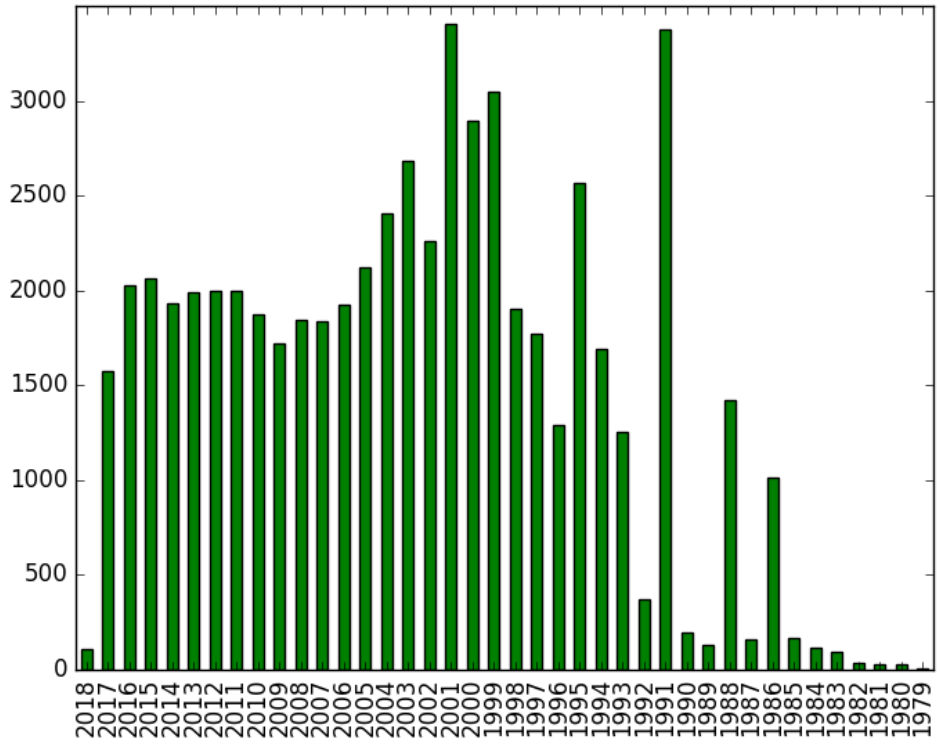
[15] Table **dbo.invst_allegation** needs to be joined with table **dbo.invst_subj** on id_invst_subj in order to get the id_pers associated with the victim or perpetrator.

person record. This suggests that at least a good portion of the victims that are identified in DCFS-L and not in SACWIS were removed from the latter system during data clean-ups.[16]

Caretakers

We compared ctkrsid in table **dcfsl_cps_alleg** in DCFS-L and person_id_perp from table **dbo.invst_allegation** in SACWIS.[17] We identified 2,006,326 perpetrators in both tables, 6,097 only in SACWIS, and 51,683 records (2.5 %) only in DCFS-L. As with children, more of the affected records are associated with investigations from 1999-2001 (though there is also a spike in 1991), but there are records spread across time (based on investigation initiation date).

Figure F. Number of caretakers/perpetrators only in DCFS-L by year (initdate)



Among the 51,683 caretaker ids that were associated with allegations in DCFS-L but do not match to SACWIS, more than half (n= 29,082) exist in SACWIS' table **dbo.person**. As with the missing child records, the vast majority (94.7%) have a status of "merged", "deleted", or "merged-non-retained" on the person record, again suggesting that at least a good portion of these records were removed from SACWIS during data clean-ups.[18]

[16] As noted above, there is reason to question whether these data should be discounted for research purposes. See Appendix 2.

[17] Table **dbo.invst_allegation** needs to be joined with table **dbo.invst_subj** on id_invst_subj in order to get the id_pers associated with the victim or perpetrator.

[18] As noted above, there is reason to question whether these data should be discounted for research purposes. See Appendix 2.

Appendix 2: Options for Linking Data between SACWIS and Legacy Golden Copy

Executive Summary

- For children who had open child cases or alleged incidents of abuse or neglect in the last 5-10 years), the links between SACWIS and Legacy Golden Copy (previously CANTS and CYCIS) provided by DCFS are generally quite accurate and may comfortably be used for routine reporting.
- The only exception to this rule is in the case of children with open cases who have an older (10+ years) history of abuse and neglect and whose original allegations do not seem to link with the open case. This affects about 10% of open cases. Research or reporting that draws on long term abuse and neglect history, even for currently open cases, should use the results of a probabilistic link.
- Any research working with children whose last allegation or last open case was more than 5-10 years ago would benefit from using a probabilistic link. This becomes increasingly true as the age of the data in question increases.
- Research using individual records that are not children in care or victims on allegations (for example, research looking to understand where DCFS-involved children reengage with the system as parents) should definitely use probabilistically linked data. We do not believe the data quality for other household members is as strong as it is for children, and it is particularly likely that individuals who engage the system in multiple contexts gather duplicate records.
- This analysis, in conjunction with our comparison of individuals and investigations between CANTS and SACWIS, has highlighted that there are odd patterns in the older investigation data regarding investigations and people that are marked as "merged" or "deleted" in SACWIS but that continue to be uniquely associated with allegations. We have identified this issue as a priority for Chapin Hall's DCFS Data Processing Team in fiscal year 2019. We will seek clarity about this historical data from DCFS and explore whether there is value in incorporating some of the oldest records or data from Chapin Hall's current DCFS longitudinal database into the CHILD.

We believe that regular (monthly/quarterly) links between datasets would only be required for research questions that fall under the first bullet above. Since these questions do not require a probabilistic match, we plan to reduce the frequency of these matches, conducting a routine annual link between datasets, while reserving capacity to conduct additional links during the year upon request.

Although it is our intention to update the CHILD database each month and provide updated data to external partners quarterly, the record linkage between SACWIS and Legacy Golden Copy (the new source system names for what were previously CANTS and CYCIS) is requires a constant compromise between rigor and automation. The best record linkage is never fully automated, so we

set out to determine if it was a good use of resources to update this match on a monthly or quarterly basis. In particular, we hypothesized that the use cases requiring frequent data updates (for example, routine reports to DCFS or monitoring of short term program outcomes) likely involve populations and data for which the cross-system identifiers recorded by DCFS (i.e. the SACWIS Person ID recorded in the client data from Legacy Golden Copy and the Client ID recorded in the person data from SACWIS) are entirely adequate. However, we believed that there are research questions, particularly those requiring historical data or thorough deduplication, for which these system values are no substitute for a probabilistic link.

To test these assumptions, we compared the results of the December 2017 probabilistic match between CANTS and CYCIS, to the links we would draw between these populations using the sacwis_person_id and cycis_cli_id values in CHILD.

We limited our tests to the universe of persons who were alleged victims of abuse and neglect on the CANTS/SACWIS side and the primary individuals in child cases on the CYCIS/Legacy Golden Copy side. For any research looking at the involvement of other clients and family members, it is necessary to use probabilistically linked data, as it is much more likely that there would be duplicate records or errors in data entry for these records at DCFS.

Comparing Probabilistic Match to DCFS IDs for Child Cases in CYCIS/Legacy Golden Copy

Table A summarizes the results of a comparison between the December 2017 probabilistic match and the value for subject ID (person ID) recorded in Legacy Golden Copy's client table (sacwis_person_id in CHILD).

Table A. Comparing probabilistic results with system values for CYCIS cases

	All children in care[19]	Children in care since 2010[20]	Children in care since 2014[21]	Children entering care since 2008[22]
Total Individuals (by caseid/caseno)	409,706	104,133	69,585	66,886
Total Individuals Matched Probabilistically	255,526	77,011	49,440	53,896
Matches Replicable Using subjid in LGC	115,427	50,820	34,461	39,766
Matches Partially Replicable Using subjid in LGC**	46,971	15,466	9,000	10,376
% of Matches Fully or Partially Replicable	63.6%	86.1%	87.9%	93.0%

**In these cases, multiple CANTS records were identified as matches for the CYCIS individual, and one of those matches is identified through the sacwis_person_id value.

It is clear from these results that the probabilistic match is especially important for older cases. For cases that have been open in the last 5-10 years, almost 85-90% of probabilistic matches could be duplicated using subject ID/person ID values from Legacy Golden Copy.

While these rates are fairly high, they still suggest that using the subject ID/person ID value in Legacy Golden Copy will miss around 1 in 10 potential matches to SACWIS. A series of spotchecks in this unmatched population found that most of the matches that could not be replicated were for older youth with long-term cases. The abuse and neglect records for these youth date back to the late 1990s and early 2000s. When the data in CANTS was compared to SACWIS, it appears that these individuals have been marked as "deleted" and new, duplicate person records created for the youth. However, the allegation histories are still linked to the deleted record. This investigation raises questions about how and why older records were merged or deleted, and what data may be lost in that process. We have opted to include all investigations and persons from SACWIS in the CHILD, regardless of the record's status, though we have also included those status fields so that users can identify records that are "merged" or "deleted" per DCFS. We did this in part because we found allegations linked to both deleted investigations and deleted persons.

Together with our findings from comparing the CANTS and SACWIS populations (Appendix 1), these results raise important questions about the quality of older investigation data and how those

[19] Includes all children from the 201712 Case table.

[20] Includes children with open cases or cases that most recently closed in 2010 or later.

[21] Includes children with open cases or cases that most recently closed in 2014 or later.

[22] Includes all children with cases that most recently opened in 2008 or later.

data should be used. We are targeting these questions as a priority for deeper investigation in fiscal year 2019 and will keep all CHILD users updated on the results of this analysis.

The final column of Table 1 focuses on children whose most recent case opening came in 2008 or later (though they may have had an older case). This should be a population with more recent allegations of abuse and neglect. As anticipated given the findings above, the subject ID/person ID values in Legacy Golden Copy better duplicate the probabilistic match rate for this population, with 93% agreement.

Comparing Probabilistic Match to DCFS IDs for Victims on Allegations of Abuse or Neglect in CANTS/SACWIS

Table B summarizes the results of a comparison between the December 2017 probabilistic match and the value for CYCIS ID (client ID) recorded in SACWIS's Person table (cycis_cli_id in CHILD). This is restricted to the universe of individuals found in both CANTS and SACWIS.[23]

Table B. Comparing probabilistic results with system values for CANTS/SACWIS cases

	All victims[24]	Victims (since 2010)[25]	Victims (since 2014)
Total Individuals (by subjid)	2,493,126	621,063	351,615
Total Individuals Matched Probabilistically	617,470	147,952	82,807
Matches Replicable Using cycisid in SACWIS	302,795	127,211	75,791
Matches Partially Replicable Using cycisid in SACWIS**	16,905	3,393	1,300
% of Matches Fully or Partially Replicable	51.8%	88.3%	93.1%

**In these cases, multiple CYCIS records were identified as matches for the CANTS individual, and one of those matches is identified through the cycis_cli_id value.

In the case of data on alleged victims of abuse or neglect, the importance of recency is particularly clear. For recent victims, the CYCIS client ID value captured in SACWIS largely mirrors the conclusions of the probabilistic match. For older victims on older allegations, however, it is important that analyses reference the probabilistic match results.

[23] See Appendix 1 for more detail about who is excluded from this overlapping population. Of particular note, excluding individuals found only in CYCIS impacted the overall number of victims for this analysis but had almost no effect on the victims with recent allegations (2010+).

[24] Includes all children indicated as victims on allegations in CANTS who also have records in the SACWIS Person table.

[25] Includes all children indicated as victims whose most recent investigation has an initial date in 2010 or later.