PSID Family Identification Mapping System User Manual

The PSID Family Identification Mapping System user manual was prepared by Noura Insolera and Mohammad Mushtaq.

This document should be cited as follows:
PSID Family Identification Mapping System User Manual. Institute for Social Research, University of Michigan, August, 2023.

Contents

1. Introduction

Due to its genealogic design, the Panel Study of Income Dynamics (PSID) is one of the few nationally representative US datasets that can be used for intra- and intergenerational analyses. From its beginning, the PSID has followed all 1968 family members and their descendants. When family members split-off and create their own separate family units, the PSID interviews these new family units as well as the original family units. It is important to note that though some families do not have any descendants in the PSID, others have children who currently make up the seventh generation. The numbers of sibling pairs, parent-child pairs, and grandparent-grandchild pairs are substantial. For example, tens of thousands of sibling pairs exist in the data archive. Of course, not all siblings are alive or reporting data in each and every wave of the PSID, therefore the number of siblings available for any particular analysis will be smaller. However, in the more recent waves, there are thousands of sibling pairs who are Reference Persons or Spouses/Partners, which means that the full set of PSID data is collected on these siblings. These large samples support a wide range of analyses.

With the goal of facilitating the use of these data to support complex models of family and life course development, the PSID offers the Family Identification Mapping System tool (FIMS). FIMS creates customized map files that contain the identification variables of the relatives an analyst wishes to examine. FIMS also has the ability to integrate relationships from the Child Development Supplement (CDS), which is a supplemental study within the PSID that provides researchers with extensive data on children and their extended families for the study of the dynamic process of early human capital and social capital formation. FIMS also supplies code (for SAS, SPSS, and Stata) that uses the map file and a file containing the variables the analyst wishes to examine to create a new data file that includes the variables for the relatives of interest. For example, an analyst may be interested in "biological grandparents" as the relative of interest. FIMS would then generate a file that contains the identification variables of all PSID sample members and each of their grandparents if they were ever observed or reported on in the PSID. FIMS provides code that will create a data file in the shape desired by the

analyst. The available maps are "wide" – i.e., one observation per individual – and "long" – i.e., one observation per relationship pair. The user would then create an individual level data file that contains all of the necessary analysis variables and merge it with the FIMS file to match them to individuals and their grandparents. FIMS data may also be downloaded as an Excel Spreadsheet.

FIMS offers three distinct types of maps. The intra-generational (SIB) map allows the identification of various types of siblings (full siblings, half siblings). In the 2021 wave, there were 5,251 unique adult sibling pairs, where both individuals were reference persons/spouses/partners of their own Family Unit. Since there is one individual in each pair who was a reference person or spouse/partner, in depth information is known about the individuals in these pairs from the Core PSID interviews. Please see Table 1 below for the number of these relationship pairs present in each wave of data collection.

The inter-generational (GID) map matches PSID individuals to their predecessors, going back up to three generations, i.e. parents, grandparents, and great-grandparents. This intergenerational map is retrospective in nature, that is, it starts with an individual and goes back along in his or her family lineage. In 2014, a prospective intergenerational map (GID PRO) was added to FIMS. Here, the starting generation (G1) is the original sample from 1968 (person numbers between 1 and 19). Descendants of original PSID households form subsequent generations, again up to three generations down (G2 - child, G3 – grandchild, G4 – great-grandchild). The prospective generation map format is "long", i.e., each row is a distinct set of individuals observed in the PSID. The map also provides biological relation, father (F) or mother (M), between individuals and generation position of all individuals listed on each row. In the 2021 wave, there were 4,383 parent-child family pairs and 1,085 grandparent-grandchild family pairs in the sample, where both individuals are reference persons/spouses/partners of their own Family Unit. Since there is one individual in each pair who was a reference person/spouse/partner, in depth information is known about the individuals in these pairs from the Core PSID interviews. Please see Table 2 below for the number of these family pairs present in each wave of data collection.

Please note that as of the 2015 Core PSID data release in May 2017, additional individuals from various family roster modules, including the 1988 and 2013 roster modules, were added to the

4

retrospective GID map. These individuals are related family members who were identified by the PSID respondent and may or may not have ever been part of a responding PSID family unit. Individuals who were never present in a PSID family unit may have limited information collected about them. These individuals have Person Numbers of 800 and above and can be excluded from your data extract if you are not interested in these additional relationships. Due to the prospective nature of the GID PRO, only individuals ever observed in a PSID family unit, i.e. those with Person Numbers below 800, are included as the first generation (G1).

Table 1. - Number of Sibling Family Pairs Observed 1968-2021
(where each individual is a Reference Person or Spouse/Partner)

Year	Sibling Pairs
1968	14
1969	56
1970	157
1971	289
1972	446
1973	675
1974	909
1975	1,126
1976	1,359
1977	1,571
1978	1,787
1979	2,085
1980	2,308
1981	2,484
1982	2,670
1983	2,864
1984	3,008
1985	3,204
1986	3,288
1987	3,406
1988	3,546
1989	3,603
1990	3,712
1991	3,804
1992	4,008
1993	4,096
1994	4,502
1995	4,461
1996	4,341
1997	3,081
1999	3,281
2001	3,580
2003	3,840
2005	4,262
2007	4,577
2009	4,947
2011	5,219
2013	5,328
2015	5,403
2017	5,594
2019	5,536
2021	5,251

Table 2. - Number of G1-G2 (Parent-Child) and G1-G3 (Grandparent-Child)
Family Pairs Observed 1968-2019 (where each individual is a Reference Person or Spouse/Partner)

| Year	G1-G2 Families	G1-G3 Families
1968	38	0
1969	196	1
1970	438	4
1971	659	8
1972	902	11
1973	1,160	19
1974	1,446	24
1975	1,671	34
1976	1,902	41
1977	2,128	40
1978	2,338	44
1979	2,612	56
1980	2,825	64
1981	2,970	69
1982	3,150	77
1983	3,333	85
1984	3,460	96
1985	3,654	110
1986	3,696	114
1987	3,803	132
1988	3,893	144
1989	3,927	162
1990	4,147	263
1991	4,306	298
1992	4,585	350
1993	4,776	461
1994	5,190	561
1995	5,113	606
1996	4,873	622
1997	3,337	459
1999	3,551	572
2001	3,860	723
2003	4,139	893
2005	4,357	1,082
2007	4,581	1,238
2009	4,864	1,336
2011	4,971	1,421
2013	4,989	1,451
2015	4,822	1,370
2017	4,846	1,410
2019	4,730	1,345
2021	4,383	1,085

2. Background

FIMS is based on information stored in the Parent Identification File (PID), which is a cumulative file compiled from several PSID data sources. The PID synopsizes information collected about parent-child relationships from various sources since the 1983 wave of the Panel Study of Income Dynamics (PSID). This file consists of identifier variables that link children with their parents, containing identifiers for the child, his or her birth- and adoptive parents, and information source indicators. The file is intended to be used to facilitate linking children's and parents' data records from 1968 to the present wave Individual File. Linkages can be done from either the child's or parent's standpoint.

The first source of the PID information began in 1983 when the PSID started asking interviewers to mark birth mothers on family listings. In 1985, birth history collection began and has continued to be updated each wave. The PID contains indicator variables summarizing from where PID information is derived, starting with 1983/84 birth mother information mentioned above and continuing through the Family Roster and Transfers Module in 2013, and beyond. These data are then used to create the FIMS map files. For more information on the PID and its structure, please see the PID File Description and Codebook.

Not all familial relationships are included in the PID, however, for varying reasons. Individuals who were once in a PSID family unit, but whose parents attrited before the collection of birth history information are not included. These parents must have attrited, never returned to a PSID family unit, and never been mentioned in a family roster effort (e.g., 1988 Time and Money Transfers, 2013 Family Roster and Transfers Module). It may also be the case that an individual's parents did not verify their parental relationship (biological or adoptive) in their birth history listing and/or that the individual did not verify their relationship to their parents during the course of the study. In these instances, the PID will not have information for an individual's biological or adoptive parents, and thus they will not have information in FIMS.

3. Intra-Generation Mapping

The intra-generation map currently supports the mapping of siblings. Below is an explanation of the selections to be made before downloading your map file, as well as a file description.
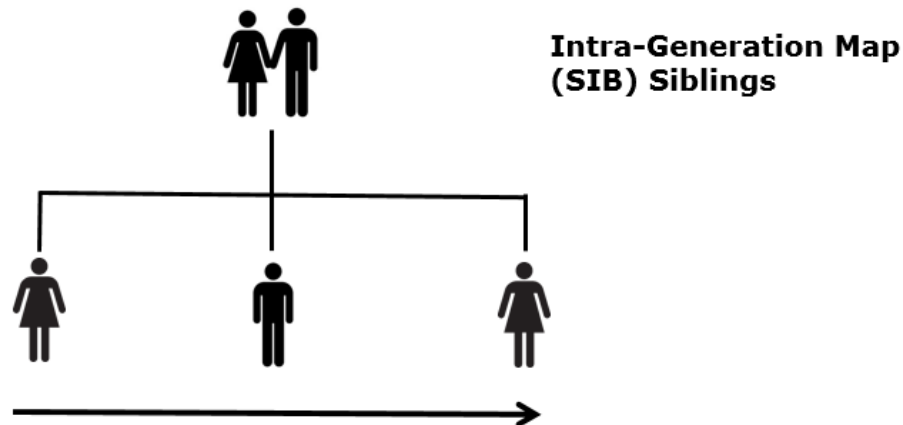
3.1 – Map Specifications: Intra-Generational (SIB) Mapping

SIB Options – These selections will appear after you have chosen the Intra-generational (SIB) map option. The description of each choice is explained below.

Sibling Type: Select a value of sibling type from the following options.

1. "Full siblings" – Biological mother and father are known for both siblings, and both parents are in common

2. "Half siblings, mother in common" - Biological mother and father are known for both siblings, and only the mother is in common

3. "Half siblings, father in common" - Biological mother and father are known for both siblings, and only the father is in common

4. "At least half siblings, mother in common" - Biological mother known for both siblings (but biological father not known for both siblings), and the siblings have the same biological mother

5. "At least half siblings, father in common" – Biological father known for both siblings (but biological mother not known for both siblings), and the siblings have the same biological father

6. "All others" – Adoptive parents, biological mother/father and adoptive mother/father, and other combinations of parental types

7. "All of the above" – All types of siblings will be included in the map file

Map Format: Available file format options are: "long" and "wide". In a long file format, each relationship pair (e.g., focal person-sibling pair) is a separate row. The wide file format provides the data with one row per focal person.

> **Wide Format Options:** "Merge with CDS Map" - When this check box is checked, a customized SIB map and CDS map file are merged for CDS researchers. This is an individual or child level file available in the "wide" format only. The number of rows in this file are equal to the total number of unique children from the Original CDS (1997 cohort) and the Ongoing CDS (begun in 2014).

Output File Type: Following output file options are available:

SAS System File
ASCII with read-in statements in SAS (.sas), SPSS (.sps), or Stata (.do)
EXCEL spreadsheet

Submit: A custom map file is created for user download.

3.2 – Variable Definitions and Codes

The file structure and the code definitions of the Sibling "long" file are below:

Variable Name	Variable Label
ID68	1968 INTERVIEW # OF INDIVIDUAL
PN	PERSON # OF INDIVIDUAL
SEX	GENDER OF INDIVIDUAL
SIBNUM	SIBLING NUMBER
ID68S	1968 INTERVIEW # OF SIBLING
PNS	PERSON # OF SIBLING
SIBSEX	GENDER OF SIBLING
SIBTYPE	TYPE OF SIBLING
PSTRING	PARENTS STRING
SIBMAKER	SIBLING MAKER

SEX, SIBSEX: These variables define the sex of the individual (SEX) and their specified sibling (SIBSEX).

Code	Value
1	Male
2	Female

PSTRING and SIBMAKER: These variables are defined using the codes below:

Code	Value
1	Biological Mother
2	Biological Father
3	Adoptive Mother
4	Adoptive Father

PSTRING: This column describes the "known" parents of the individual and sibling pair. The string has two parts separated by a "-"; where the left portion is parents' information of the individual and the right portion is parents' information of the sibling. For example, a value of "1200-0200" means that biological mother and father of the individual are known and only biological father of sibling is known for the given pair.

SIBMAKER: This column reports the "common" parents of the individual and sibling pair. A value of "(1-1)" means that biological mothers are common for the pair.

SIBTYPE: Sibling Type values are determined by inspecting the values from these two columns.

The resulting output file will be a customized data set fit to your specifications. The sibling pairs will be in duplicate form, where 'Sibling A' and 'Sibling B' will be listed as two observations, once as A-B and again as B-A. This allows each researcher to make analytical decisions as to which individual is the focal individual, and which is the sibling of the focal individual. These decisions may be different for each researcher based on age, sex, birth order, or a variety of other options.

4. Retrospective Inter-generation Mapping

The retrospective inter-generation map currently supports the mapping of an individual back to their parents, grandparents, and great-grandparents. Below is an explanation of the selections to be made as you create your inter-generational map file.
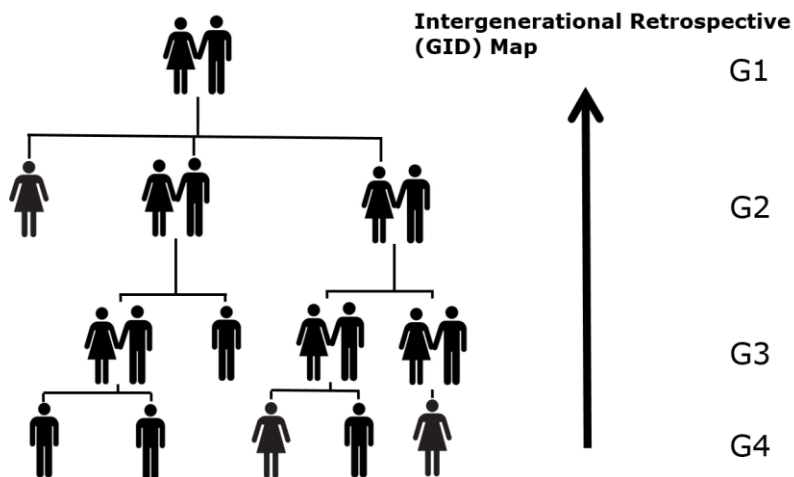
4.1 – Map Specifications: Inter-generational (GID) Retrospective Mapping

GID Options – These selections will appear after you have chosen the Inter-generational (GID) map option.

Parents: Available choices are: "Biological only," and "Biological and adoptive." When "Biological and adoptive" is selected and both parents are the same, then the adoptive parent identification variable is set to zero; this situation exists for less than one dozen cases.

Generation Map: "Individual to parents," "Individual to grandparents" and "Individual to great grandparents" are available choices. Based on the selection, each row in the output file is a set of individuals (individual's ID68 and PN) from selected generations.

Map Type: Balanced and unbalanced map types are available.

> **Balanced:** A generation map with non-missing data from any generation is called a balanced generation map. For instance, a balanced, two-generational map only includes PSID individuals who have at least one parent who was ever observed or reported on in the PSID. A balanced, three-generational map only includes PSID individuals who have at least one parent and at least one grandparent observed or reported on in the PSID.

> **Unbalanced:** A generation map with completely missing data on the prior generation(s) is called an unbalanced generation map. For instance, an unbalanced, two-generational map includes PSID individuals included in the balanced map plus those who do not have any parent who was ever observed or reported on in the PSID. An unbalanced, three-generational map

includes PSID individuals included in the balanced map plus those who do not have any grandparent who was ever observed or reported on in the PSID.

Map Format: Available file format options are: "long" and "wide." In a long file format, each relationship pair (e.g., focal person-parent pair) is a separate row. The wide file format provides the data with one row per focal person.

>   **Wide Format Options:** "Merge with CDS Map." When this check box is checked, a customized GID map and CDS map file are merged. This is an individual- or child-level file available in the "wide" format only. The number of rows in this file are equal to the total number of unique children from the Original CDS (1997 cohort) and the Ongoing CDS (2014 and beyond).

Output File Type: The following output file options are available:

SAS System File
ASCII with read-in statements in SAS (.sas), SPSS (.sps), or Stata (.do)
EXCEL spreadsheet

Submit: A custom map file is created for users to download.

4.2 – Variable Definitions and Codes

Label Abbreviations:

Label Abbreviation	Definition
P	Parent
GP	Grandparent
GGP	Great-Grandparent
M	Mother
F	Father
AM	Adoptive Mother
AF	Adoptive Father
MM	Mother's Mother (maternal grandmother)
MF	Mother's Father (maternal grandfather)
FM	Father's Mother (paternal grandmother)
FF	Father's Father (paternal grandfather)

These abbreviations are used together to define the types of parents, grandparents, and great-grandparents, and their relationships to the focal individual. As an example, ER30001_GGP_FMM would refer to the great-grandparent (GGP), and more specifically the Father's Mother's Mother (FMM), or the mother of the paternal grandmother. These suffix labels describe each individual and their relationship to the focal individual.

5. Prospective Inter-generation Mapping

The prospective intergeneration map currently supports the mapping of an individual down to their children, grandchildren, and great-grandchildren. Below is an explanation of the selection to be made before downloading your map file, and what you can expect to see in your downloaded map.
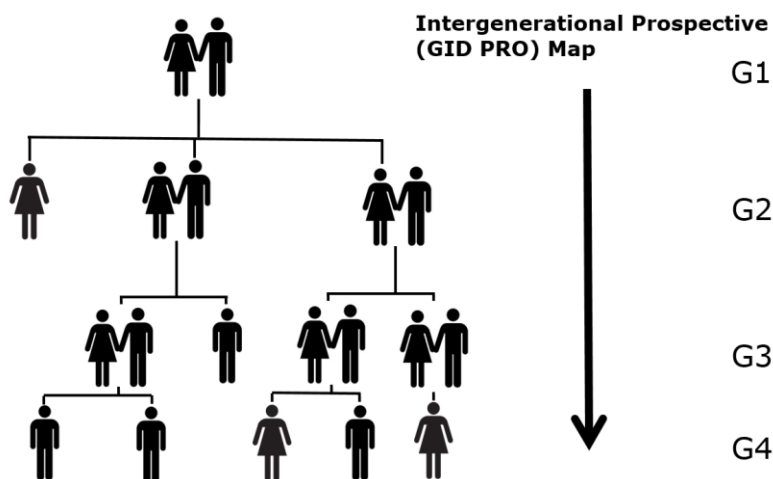
5.1 – Map Specifications: Inter-generational (GID PRO) Prospective Mapping

Prospective GID Options – These selections will appear after you have chosen the Inter-generational (GID) Prospective map option.

Parents: Available choices are: "Biological only," and "Biological and adoptive." When "Biological and adoptive" is selected and both parents are the same, then the adoptive parents identification variable is set to zero; this situation exists for less than one dozen cases.

Prospective Generation Map: "Parent to child," "Grandparent to child" and "Great grandparent to child" are available. Based on the selection, each row in the output file is a set of individuals (individual's ID68 and PN) from selected prospective generations.

Map Type: Balanced and unbalanced map types are available.

> **Balanced:** A generation map with non-missing data from any generation is called a balanced generation map. For instance a balanced, two-generational prospective map includes PSID individuals who have at least one child who was ever observed or reported on in the PSID. A balanced, three-generational prospective map includes PSID individuals who have at least one grandchild who was ever observed or reported on in the PSID.

> **Unbalanced:** A generation map with completely missing data on the following generations(s) is called an unbalanced generation map. For instance, an unbalanced, two-generational prospective

map includes PSID individuals included in the balanced map plus those who do not have a child observed or reported on in the PSID. An unbalanced, three-generational prospective map includes PSID individuals included in the balanced map plus those who do not have a grandchild observed or reported on in the PSID.

Map Format: The file format option for the prospective map is "long." In a long file format, key variables (1968 interview number and person number) from each relationship set (row) is unique and non-missing.

Output File Type: The following output file options are available:

SAS System File
ASCII with read-in statements in SAS (.sas), SPSS (.sps), or Stata (.do)
EXCEL spreadsheet

Submit: A custom map file is created for users to download.

5.2 – Variable Definitions and Codes

The file structure and the code definitions of the "long" file are below:

| Variable Name	Variable Label
G1ID68	1968 INTERVIEW # OF G1 INDIVIDUAL
G1PN	PERSON # OF G1 INDIVIDUAL
G1TYPE	PERSON TYPE OF G1 INDIVIDUAL
G1POS	GENERATION POSITION OF G1 INDIVIDUAL
G2ID68	1968 INTERVIEW # OF G2 INDIVIDUAL
G2PN	PERSON # OF G2 INDIVIDUAL
G2TYPE	PERSON TYPE OF G2 INDIVIDUAL
G2POS	GENERATION POSITION OF G1 INDIVIDUAL

G#TYPE: This set of variables defines whether the individual is known to have children, whether they are a biological/adoptive mother/father, and whether they have not been verified to have had children.

Code Value	Definition
M	Mother
F	Father
AM	Adoptive Mother
AF	Adoptive Father
I	Individual (not known to be a parent)

G#POS: These columns refer to the generational position of the individual. For example, if an individual is the child of an original sample member, and also has a child of their own, they will appear in a row as their parent's child (G2POS=2), as well as in a row as their child's parent (G1POS=2). With each data release, new information is added to these files to display the most up to date information available.

6. Merging the FIMS Map File with Data Files

Once you have downloaded your FIMS map file, you will continue to the PSID Data Center to download data files containing the content you wish to merge onto the FIMS map file (i.e., "content data"). For this example, let us assume we are mapping children to their parents. There would be two content data files, one for the children, and one for the parents. These data files may contain any number of variables from any wave(s) of the PSID. It is important, however, to include at least one individual-level variable in the data cart, which then allows the automatic addition of individual-level unique identifiers to be added to the Data Center download (i.e., 1968 ID (ER30001) and Person Number (ER30002)). These are the unique identification variables that are used to merge the child dataset to the children in the FIMS map, and the parent dataset to the parents in the FIMS map. The 1968 ID (ER30001) and Person Number (ER30002) are used together to identify an individual across waves. Though you can combine them uniquely in many ways we find that many researchers use the following method - (ER30001 * 1000) + ER30002. Once the two content data sets include the unique identification variables, they can be merged, in this example, with the FIMS map file, into one.

Separate Files
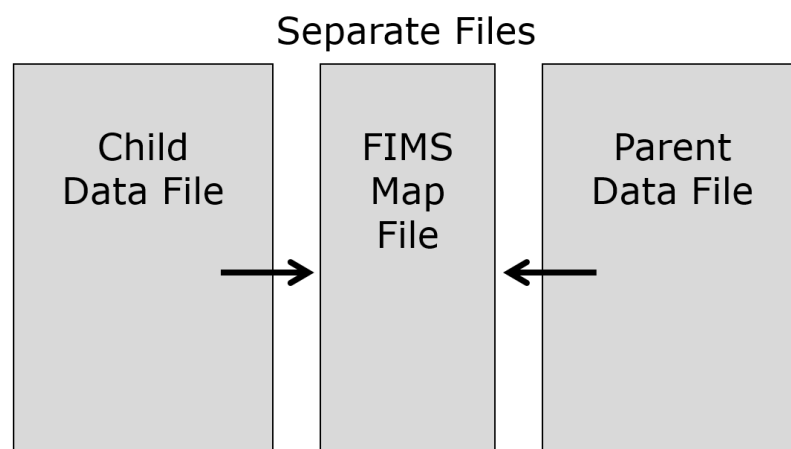
It is important to note that the resulting FIMS data merge will include all known relationships in the PSID, but not all individuals will be present in each wave. This is because in earlier waves, an individual might not have been born yet, or included in the study for various reasons, and in later waves,

the individual may have died or attrited. In either case, not all family members will be in each wave of the PSID. For this reason, when you merge your data file onto your FIMS map file, in most cases, you will not see a one-to-one merge, but only complete merges for those who are in the waves you have included in your content data file, and for those for whom we have verified relationship information in FIMS.

For more information on the code to use to merge these two types of files together (i.e., FIMS Map file with content data), please see the supporting FIMS documentation, located on the FIMS webpage.

Fully Merged File

Child Data	FIMS Map	Parent Data
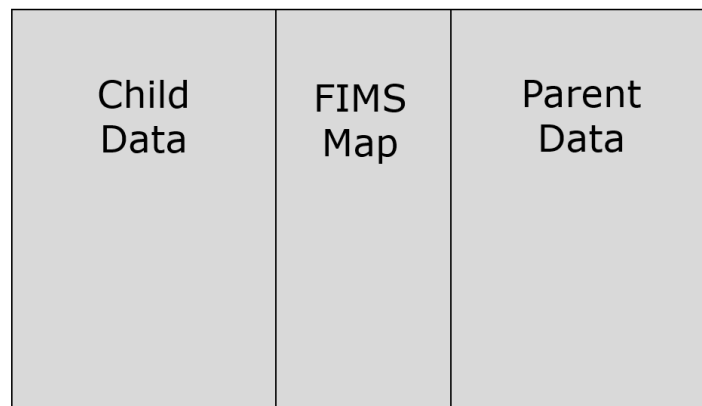
Now that you have completed your FIMS map selections, created your content data files (i.e., Child data and Parent data), and merged your files together, you will have one fully merged dataset containing content for all of the relatives of interest. For an overview and examples of the structure of the FIMS dataset, please see the web tutorial 'The Family Identification Mapping System.'