SwedPop Documentation

SwedPop IDS

Author: Annika Westberg, CEDAR

Version: SwedPop IDS, version 4.
January 26, 2026

Contents

1 Introduction 3

2 SwedPop IDS 4

2.1 General rules 4

2.1.1 Only valid attributes are entered in the database. 4

2.1.2 When attributes are added, their values are defined in the Data Retrieval Tool Database. 5

2.1.3 Only values that are represented and valid are entered in the database. If an individual lack data, there is no row in the table. 5

2.2 Dates and timestamps 5

2.2.1 Date_type 5

2.2.2 Estimation 5

2.2.3 Missing 6

2.3 Definition of family 6

2.4 Definition of household 7

2.5 Definition of relatives 8

3 IDS tables 8

3.1 Table METADATA 8

3.1.1 There is one common SwedPop IDS metadata table 8

3.1.2 Changes in metadata table are made by the SwedPop IDS working group. 8

3.1.3 TYPE and VALUE that appear in the local IDS databases are represented in the metadata table. 9

3.1.4 All TYPE's have a VALUE=DEFINITION. 9

3.1.5 Interpretation of table column ID_D 9

3.2 Table INDIVIDUAL 10

3.2.1 Rules for Type Attributes 11

3.2.2 Representation of coded geographic locations 11

3.2.3 Representation of uncoded geographic locations 11

3.2.4 Sample selection: All individuals must have values on both BIRTH_LOCATION and BIRTH_DATE. 11

3.2.5 Observation, arrival and departure 11

3.2.6 Occupations 12

3.2.7 Cause of death 13

3.3 Table INDIV_INDIV 14

3.3.1 Rules for Relation Attributes 14

3.3.2 Example of double directed relations 15

3.4 Table CONTEXT 15

3.4.1 Rules for Type 17

3.4.2 Rules for Value 17

3.5 Table INDIV_CONTEXT 18

3.5.1 Rules for Relation attributes 18

3.6 Table CONTEXT_CONTEXT 19

3.6.1 Valid RELATION values 19

1 Introduction

This data has been retrieved from the Swedish research infrastructure SwedPop. SwedPop constitutes a national platform for micro-level population data, designed to integrate and coordinate existing population databases maintained by five Swedish universities and archival institutions. The core databases provide individual-level information in longitudinal, cross-sectional, and panel formats. Through harmonization and standardization, SwedPop offers accessible datasets with unified variable definitions and a common data structure, covering a broad spectrum of historical demographic and socio-economic indicators.

The databases are accessible via the SwedPop web portal (www.swedpop.se), which functions as a central hub for data dissemination and comprehensive documentation, including details on shared coding systems and metadata standards. The overarching aim of SwedPop is to establish a sustainable, long-term national resource that facilitates systematic access to Swedish micro-level population data for both domestic and international research communities.

The SwedPop data architecture adheres to the principles of the Intermediate Data Structure (IDS)—a standardized format designed for databases containing longitudinal and non-aggregated data on individuals, families, and households. IDS was developed within the framework of the ESF Research Networking Programme *European Historical Population Samples Network (EHPS-Net)* (see https://ehps-net.eu/). This structure provides an integrated interface that harmonizes data across regional and national boundaries, thereby enabling interoperability among diverse datasets. A detailed description of IDS is provided in an article authored by two of its principal developers, George Alter and Kees Mandemakers (see https://ehps-net.eu/article/intermediate-data-structure-ids-longitudinal-historical-microdata-version-4).

The purpose of this documentation is twofold. It addresses data delivery to the SwedPop infrastructure, including guidelines for attribute handling and compliance requirements prior to data submission. When demographic data from multiple sources are combined into a single database, harmonization is critical to ensure consistency, accuracy, and usability. This process involves aligning data elements, ensuring that they share a common structure, meaning, and format.

The second purpose of this documentation is to introduce and define key concepts employed within the database framework.

Data files

The data files are delivered in a zip-folder.

Data are csv-files, semicolon (;) separated, UTF-8 format.

Citation of SwedPop data

By retrieving data from the SwedPop database, the user agrees to acknowledge SwedPop as the data provider and the Demographic Database (DDB) as the repository holder.

Correct citations must be included in any publication, report, conference paper, or other material in which SwedPop data are used. The Citation page provides recommended citations for the current database version in multiple citation styles. Information required to cite earlier versions of the database is available at the Revision History page.

2 SwedPop IDS

The SwedPop data is organized according to the standardized Intermediate Data Structure (IDS). Within SwedPop, this structure has been further refined into a specialized variant, referred to as the SwedPop IDS, developed to accommodate the unique characteristics and complexities inherent in Swedish historical and demographic sources.

In addition to structural harmonization, the SwedPop IDS incorporates shared nomenclatures, standardized coding schemes, and a unified metadata management system. These components ensure semantic consistency across datasets, enabling accurate interpretation and comparability of variables originating from heterogeneous sources.

The harmonized IDS data is subsequently stored in the SwedPop Data Retrieval Tool Database, which serves as an intermediary platform for managing the linkage between the underlying database architecture and the online data extraction interface. This design supports efficient query execution, controlled data retrieval, and ensures that harmonized datasets remain accessible for advanced analytical workflows.

2.1 General rules

Some rules are general and applied in all tables.

2.1.1 Only valid attributes are entered in the database.

In the context of database design and data quality assurance, validity refers to the adherence of stored attributes to predefined syntactic and semantic rules. This includes, but is not limited to, the following criteria:

- **Orthographic correctness**: Accurate spelling and appropriate use of upper- and lower-case characters.
- **Structural integrity**: Placement of attributes within their designated tables according to the scheme specifications.
- **Temporal consistency**: Correct association of date-related attributes with their corresponding data types (e.g., DATE, TIMESTAMP) and format standards.
- **Type conformity**: Ensuring that attribute values comply with the expected data type definitions and constraints.

All attributes deemed valid are documented in the metadata table.

2.1.2 When attributes are added, their values are defined in the Data Retrieval Tool Database.

All attributes in the SwedPop IDS database are defined in the Data Retrieval Tool Database. When adding attributes, they must be defined there.

2.1.3 Only values that are represented and valid are entered in the database. If an individual lack data, there is no row in the table.

The SwedPop IDS is designed to minimize the presence of null or empty records. As a general principle, the structure does not include rows that represent missing or undefined values. An exception to this rule applies to specific columns within the HISCO classification system, namely STATUS, RELATION, and PRODUCT.

2.2 Dates and timestamps

Dates are recorded in several tables as event- or period dates and are further defined by the columns Date_type, Estimation and Missing. Depending on values, contents and combinations of these, there are a set of rules how to handle them. Note that the date-related columns only have values when the TYPE refers to a date, for example birth date or arrival from.

2.2.1 Date_type

The Date_type column specifies the manner in which a date is recorded or observed, determines the validity period of an attribute, and indicates whether the date has been assigned by database administrators from various sources. Since this column exclusively pertains to the handling of temporal data, it is applied only to attributes where the TYPE represents a date.

Rules governing Date_type include:

- The Date_type column contains a value only when the associated TYPE refers to a date attribute
- When the TYPE does not correspond to a date, the Date_type column remains empty

2.2.2 Estimation

Database administrators can estimate dates based upon source information. Since this column exclusively pertains to the handling of temporal data, it is applied only to attributes where the TYPE represents a date.

Rules governing Estimation include:

- Estimation contains a value only when the associated TYPE refers to a date attribute
- When the TYPE does not correspond to a date, the Estimation column remains empty

2.2.3 Missing

The Missing column provides explanatory information regarding the absence of date values within the dataset.

Rules governing Missing include:

- The Missing column contains a value only when the associated TYPE refers to a date attribute.
- When the TYPE does not correspond to a date, the Missing column remains empty

Multiple and different dates for the same event

In certain local databases, an individual may have multiple and conflicting dates recorded for the same event, such as different birth dates derived from various historical sources. SwedPop IDS enforces the rule whereby only one date per event is stored. The responsibility for selecting which date to deliver rests with the local database administrators, who determine the most appropriate or authoritative value prior to integration into the SwedPop IDS.

2.3 Definition of family

SwedPop defines a family as individuals living together in a conjugal family unit. Families are constructed differently depending on the source type. In catechetical registers, family members are recorded on the same page during the same period. In censuses, individuals are listed within a household, and family relationships are inferred from their roles in that household.

Roles within the conjugal family unit include:

- Married couple without children
- Married couple with unmarried children (biological-, step-, foster- or adoptive children)
- Lone parent with unmarried children (biological-, step-, foster- or adoptive children)
- Single persons

Additional definition:

- A person can only belong to one family at a given time
- Single persons are assigned their own family number
- Siblings living together at the same time share a family number
- A person can belong to both a family and a household simultaneously

Including the family selection in a data retrieval means that all family members of the sampled individuals are added to the dataset, along with the selected variables.

Family ID (ID_C for LEVEL= Family)

Families in the databases are identified using a 12-digit code. The first digit represents the database ID, the second digit indicates that the context is family (value=1), and digits three through twelve form a unique serial number.

Table 1. Databases and family ID

Database	**Family ID**
DDB	DDB database ID=1 Family context=1 + 10 digit serial number
GOPP	GOPP database ID=2 Family context=1 + 10 digit serial number
Rotemannen	Rotemannen database ID=3 Family context=1 + 10 digit serial number
SEDD	SEDD database ID=4 Family context=1 + 10 digit serial number
SweCens	SweCens database ID=5 Family context=1 + 10 digit serial number

2.4 Definition of household

SwedPop defines a household as:

- Members living together in the same place at the same time.
- Members may include related persons, employees, or others.

In catechetical registers, household members are recorded together on the same page during the same period. In censuses, household members are listed as a group.

Including the household selection in a data retrieval means that all household members of the sampled individuals are added to the dataset, along with the selected variables.

Household ID (ID_C for LEVEL= Household)

Households in the databases are identified using a code. The first digit represents the database ID, the second digit indicates that the context is household (value=2), and the subsequent digits provide a unique serial number.

Table 2. Databases and household ID

Database	Household ID
GOPP	GOPP database ID=2 Household context=2 + serial number
Rotemannen	Rotemannen database ID=3 Household context=2 + serial number
SEDD	SEDD database ID=4 Household context=2 + serial number
SweCens	SweCens database ID=5 Household context=2 + serial number

2.5 Definition of relatives

The sources include many different types of relatives, both biological and non-biological. All recorded kinship relations in the sources are included. Unlike family and household, kinship relations do not require shared residence in time or space.

Including the relatives selection in a data retrieval means that all relatives of the sampled individuals are added to the dataset, along with the selected variables.

3 IDS tables

3.1 Table METADATA

The table METADATA serves as a comprehensive description of values as well as explanations of all attributes in the database.

3.1.1 There is one common SwedPop IDS metadata table

The one and unique METADATA table secures that its content is always up to date.

3.1.2 Changes in metadata table are made by the SwedPop IDS working group.

All modifications to the metadata table are centrally managed by the SwedPop IDS working group. Local edits are strictly prohibited to maintain schema consistency and data governance standards. Any introduction of new attributes or values must be formally submitted to the SwedPop IDS working group for validation and integration. Upon approval, these elements are incorporated into both the metadata table and the Retrieval Tool database, ensuring harmonized documentation and interoperability across the system architecture.

3.1.3 TYPE and VALUE that appear in the local IDS databases are represented in the metadata table.

All TYPE and VALUE elements present in local IDS databases are systematically represented within the METADATA table to ensure semantic consistency and standardized interpretation. The METADATA table is restricted to attributes that exist in the SwedPop IDS database, thereby maintaining alignment with the central schema and preventing uncontrolled schema divergence.

3.1.4 All TYPE's have a VALUE=DEFINITION.

Variable descriptions are maintained in the DEFINITION field. Each TYPE must be associated with a corresponding DEFINITION to ensure semantic clarity and structural consistency. VALUE-to-DEFINITION mappings are exclusively represented within the metadata table.

3.1.5 Interpretation of table column ID_D

The attribute ID_D (Identifier of Database) has a dual role depending on its context of use:

- **In the IDS database and retrieved datasets:**

  ID_D consistently represents the name of the data provider, ensuring traceability of the source for each record.

- **In the metadata table:**

  ID_D may take one of three possible values:

  - SWEDPOP – indicating definitions standardized by the SwedPop framework.
  - STANDARD – referring to generic or system-wide standard definitions.
  - The name of the data provider – specifying source-specific metadata.

Table 3. Explanation of appearance of ID_D

	ID_D	**In metadata table**	**In IDS database/ Retrieved datasets**
i)	STANDARD	TYPE or VALUE is found in one or more SwedPop-databases and is approved by the International IDS clearing committee.	Not applicable
ii)	SWEDPOP	TYPE or VALUE is found in two or more SwedPop-databases. ID_D=SWEDPOP	Not applicable
iii)	Name of database	TYPE or VALUE is only found in one SwedPop-database. ID_D= Name of database. Example: the TYPE *Vaccination* is only found in DDB data. ID_D=DDB	Not applicable
iv)	Name of database	TYPE or VALUE is only found in more than one database. Example: the value *Female* appears in SweCens, SEDD and DDB data for three different individuals. Individual 1: ID_D=SweCens Individual 2: ID_D=SEDD Individual 3: ID_D=DDB	Applicable

3.1.6 Description of ID_I – identify unique individuals

Individuals are identified by ID_I, which is unique within each separate database. However, since the databases are not yet linked, the same ID_I may appear in more than one database.

To identify unique individuals across databases, ID_I must be combined with ID_D.

3.2 Table INDIVIDUAL

The table INDIVIDUAL contains all attributes that characterize an individual.

Table 4. Metadata table INDIVIDUAL

ORDINAL POSITION	**COLUMN**	**DESCRIPTION**	**DATATYPE**	**COLUMN SIZE**
1	ID	Table row serial number	I(8) not null	19
2	ID_D	Identifier of the data provider, indicates the database from which the data originates	VC(50) not null	50
3	ID_I	Unique identifier for the individual	I(8) not null	19
4	SOURCE	What type of source the data is extracted from	VC(50) not null	50
5	TYPE	Type	VC(32) not null	32
6	VALUE	Value of the type	VC(150)	150
7	VALUE_ID_C	Context ID	I(8)	19

8	DATE_TYPE	Type of date	VC(50)	50
9	ESTIMATION	Type of estimation of date	VC(50)	50
10	MISSING	Reason why a date or a part of a date is missing (and had to be estimated)	VC(50)	50
11	YEAR	Year	I(4)	4
12	MONTH	Month	I(4)	2
13	DAY	Day	I(4)	2

3.2.1 Rules for Type Attributes

The TYPE attribute in the SwedPop IDS must adhere to strict validation rules to ensure consistency and interoperability across datasets:

- **Metadata Compliance**: Every TYPE must be explicitly defined in the SwedPop-IDS metadata table. Only values listed as valid in this metadata schema are permissible.
- **Case Sensitivity**: All TYPE's must be written in uppercase letters to maintain uniformity and prevent discrepancies during data processing and extraction.

These constraints guarantee semantic clarity, facilitate automated validation, and support harmonized integration of heterogeneous historical data sources.

3.2.2 Representation of coded geographic locations

Values for geographic location in the table INDIVIDUAL are stored in the column VALUE_ID_C. The lowest level of coded geographic data in this column is parish. Coded geografical data on lower levels than parish, such as location or residence, is handled in the table INDIV_CONTEXT.

3.2.3 Representation of uncoded geographic locations

Uncoded geographic information is stored as plain text in the VALUE column. In these cases, no corresponding reference to VALUE_ID_C is provided, as the location lacks an associated standardized geographic code.

3.2.4 Sample selection: All individuals must have values on both BIRTH_LOCATION and BIRTH_DATE.

The sample selection functionality within the Data Retrieval Tool relies on two critical attributes: birth location and birth date. Both elements are required to define and execute a valid sample selection. If either value is missing, the sample selection process cannot be performed.

3.2.5 Observation, arrival and departure

There are distinct rules for recording observations and migrations, depending on the nature of the event:

1. **Migration into or out of the source**
   An individual's geographical movement into or out of the sources is recorded in the INDIVIDUAL table using the attributes ARRIVAL_FROM and DEPARTURE_TO.
2. **Migration within the source without leaving observation**
   Movements between locations that are covered by the sources, while the individual remains under observation, are recorded in the INDIV_CONTEXT table.
3. **Start of observation**
   Observation in the sources can begin due to birth, arrival, start of source coverage, or unknown reasons. This is recorded in START_OBSERVATION.
   - If the value is *Arrival*, its date corresponds to the date of the type ARRIVAL_FROM.
4. **End of observation**
   Observation can end due to death**,** departure, end of source coverage, or unknown reasons. This is recorded in END_OBSERVATION.
   - If the value is *Departure*, its date corresponds to the date of the type DEPARTURE_TO.

3.2.6 Occupations

Occupational titles are coded using the international classification system HISCO. Methods and principles are described in *SwedPop Documentation: Principles of Coding Swedish Historic Occupations*. A complete list of standardized occupational titles, codes and descriptions of unit groups can be found in the document *HISCO codes and description*.

Documentation is stored at https://swedpop.se/data-description/

In certain cases, a single record may contain multiple occupational titles associated with the same individual. This scenario introduces ambiguity in linking the correct occupational titles to their corresponding HISCO classifications.

To resolve this issue, the SwedPop IDS employs a mechanism by appending a suffix in the format _PART_*n* to the attribute name. This suffix differentiates multiple titles within the same record, ensuring accurate mapping between occupational titles and their respective HISCO codes.

Table 5. Example of a record containing two occupational titles with identical source and date ("HOUSE SERVANT, MASTER SHOEMAKER'S WIFE").

TYPE	**VALUE**
OCCUPATION_STANDARD_PART_1	HOUSE SERVANT
OCCUPATION_ STANDARD _PART_2	MASTER SHOEMAKER'S WIFE
OCCUPATION_HISCO_PART_1	54020
OCCUPATION_HISCO_PART_2	80110
OCCUPATION_HISCO_STATUS_PART_1	-9
OCCUPATION_HISCO_STATUS_PART_2	21
OCCUPATION_HISCO_RELATION_PART_1	-9
OCCUPATION_HISCO_RELATION_PART_2	11
OCCUPATION_HISCO_PRODUCT_PART_1	-9
OCCUPATION_HISCO_PRODUCT_PART_2	-9

Occupations in the Data Retrieval Tool

When selecting OCCUPATION_HISCO in the data retrieval tool, four different variables are automatically added to the dataset: OCCUPATION_HISCO, OCCUPATION_HISCO _STATUS, OCCUPATION_HISCO_RELATION and OCCUPATION_HISCO_PRODUCT as they define crucial information about the occupation.

3.2.7 Cause of death

SwedPop provides harmonized data on causes of death, based on the International Historical Classification System for Coding Causes of Death, version 2 (ICD10h). For detailed information, refer to SwedPop Documentation: Principles of Coding Historic Causes of Death.

The documentation is available at: https://swedpop.se/data-description/

To access the complete ICD10h coding and classification scheme, see:
Reid, A., Garrett, E., Hiltunen Maltesdotter, M., & Janssens, A. (2025). Historic Cause of Death Coding and Classification Scheme for Individual-Level Causes of Death – Codes. Apollo – University of Cambridge Repository. https://doi.org/10.17863/CAM.109961.2.

In historical sources, a single record may contain more than one cause of death for an individual. However, source material does not always specify whether these causes represent primary or secondary conditions. To manage this complexity, the SwedPop IDS employs a systematic approach by appending a suffix in the format _PART_*n* to the TYPE attribute. This suffix differentiates multiple causes of death within the same record, ensuring accurate representation and linkage. Notably, this suffixing system is applied consistently—even in cases where only one cause of death is recorded.

Table 4. Example of a record containing two causes of death.

TYPE	**VALUE**	**Explanation**
DEATH_CAUSE_ICD10H_PART_1	I51.400	ICD10h classification code for the cause of death. Part 1 refers to the first recorded cause of death for the individual
DEATH_CAUSE_STANDARD_PART_1	MYOCARDITIS	Standardized cause of death description in Swedish or Latin, preserving historical terminology. Part 1 refers to the first recorded cause of death for the individual
DEATH_CAUSE_ICD10H_PART_2	J18.900	ICD10h classification code for the cause of death. Part 2 refers to the second recorded cause of death for the individual
DEATH_CAUSE_STANDARD_PART_2	PNEUMONIA (ACUTA)	Standardized cause of death description in Swedish or Latin, preserving historical terminology. Part 2 refers to the second recorded cause of death for the individual

3.3 Table INDIV_INDIV

The table INDIV_INDIV is designed to represent interpersonal relationships between individuals

Table 5. Metadata table INDIV_INDIV

ORDINAL POSITION	COLUMN	DESCRIPTION	DATATYPE	COLUMN SIZE
1	ID	Table row serial number	I(8) not null	19
2	ID_D	Name of data provider; which database data originates from.	VC(50) not null	50
3	ID_I_1	Individual ID, relation to ID_I_2, refering to ID_I in the INDIVIDUAL table.	I(8) not null	19
4	ID_I_2	Individual ID, relation to ID_I_1, refering to ID_I in the INDIVIDUAL table.	I(8) not null	19
5	SOURCE	What type of source the data is extracted from: Death records, Marriage records, Migration records, Poll-tax registers, Birth records, Income registers, Parish registers, Censuses.	VC(50) not null	50
6	RELATION	The relationship from the first individual to the second. In case of biological relationships a timestamp is not necessary	VC(50) not null	50
7	DATE_TYPE	Type of date	VC(50)	50
8	ESTIMATION	Type of estimation of date	VC(50)	50
9	MISSING	Reason why a date or a part of a date is missing (and had to be estimated)	VC(50)	50
10	START_YEAR	Start year of period	I(4)	4
11	START_MONTH	Start month of period	I(4)	2
12	START_DAY	Start day of period	I(4)	2
13	END_YEAR	End year of period	I(4)	4
14	END_MONTH	End month of period	I(4)	2
15	END_DAY	End day of period	I(4)	2

3.3.1 Rules for Relation Attributes

Valid values of RELATION are definded in the table metadata and spelling must correspond to the spelling in the metadata table.

The RELATION attribute in the SwedPop IDS must comply with strict metadata-based validation requirements:

- **Metadata Definition**: All permissible values for RELATION are explicitly defined in the metadata table. Only these predefined values are considered valid.
- **Exact Spelling**: The spelling of values in the RELATION column must correspond exactly to the spelling specified in the metadata table. This includes case sensitivity and character formatting.

3.3.2 Example of double directed relations

Table 6. Example of double directed relations for three individuals: 3299455 (marked green as child, husband father), 7287800 (marked orange as child, wife, mother), 3717922 (marked blue as child).

Table 6. Illustration of double directed relations

ID_I_1	ID_I_2	RELATION	Explanation
3299455	9231326	Child	ID_I_1 is child to 9231326.
3299455	2425875	Child	ID_I_1 is child to 2425875.
3299455	3717922	Father	ID_I_1 is father to 3717922, who also appears as relation=child to this ID_I_1. This father has six children in total.
3299455	4803874	Father	ID_I_1 appears as father to six children.
3299455	6923780	Father	ID_I_1 appears as father to six children.
3299455	4829302	Father	ID_I_1 appears as father to six children.
3299455	9572150	Father	ID_I_1 appears as father to six children.
3299455	6545077	Father	ID_I_1 appears as father to six children.
3299455	7287800	Husband	ID_I_1 is husband to 7287800, who also appears as relation= wife to this husband.
3717922	7287800	Child	ID_I_1 is child to 7287800, who also appears as relation= mother to this child
3717922	3299455	Child	ID_I_1 is child to 3299455, who also appears as relation= father to this child.
7287800	2955094	Child	ID_I_1 is child to 2955094.
7287800	3335148	Child	ID_I_1 is child to 3335148.
7287800	3717922	Mother	ID_I_1 is mother to 3717922, who also appears as relation=child to this ID_DI_1. This mother has six children in total.
7287800	4803874	Mother	ID_I_1 appears as mother to six children.
7287800	6923780	Mother	ID_I_1 appears as mother to six children.
7287800	4829302	Mother	ID_I_1 appears as mother to six children.
7287800	9572150	Mother	ID_I_1 appears as mother to six children.
7287800	6545077	Mother	ID_I_1 appears as mother to six children.
7287800	3299455	Wife	ID_I_1 is wife to 3299455, who also appears as relation= husband to this wife.

3.4 Table CONTEXT

The CONTEXT table in the SwedPop IDS stores all relevant contextual information, which represents the environmental layers in which individuals exist. These contexts include:

- **Living environments**: such as family units or household memberships.
- **Geographical environments**: such as localities, parishes, and regions.

The table captures associations between contexts by specifying both the context level (e.g., family, household, locality, parish) and the name of the level, thereby enabling hierarchical representation of social and spatial structures.

ID_C Attribute

Values in ID_C are standardized to contain a series of digits. Country, county and parish begin with a three-digit prefix that identifies the context level, followed by a unique serial number. Locality

and polltax are represented by database ID as the first digit followed by context code. The subsequent digits provide a unique serial number.

Table 7. ID_C Attributes

Level	**Context ID**
Country	700 + country code 700999999998 – foreign country, i.e outside Sweden
County	702 + 2 digit county code + serial number
Parish	701 + 9 digit parish code 701999999999 – unknown
Locality	DDB database ID=1 Locality context=4 + serial number
Polltax	SEDD database ID=4 Polltax context=3 + serial number
Family	See chapter 2.3 above
Household	See chapter 2.4 above

Table 8. Metadata table CONTEXT

ORDINAL POSITION	COLUMN	DESCRIPTION	DATATYPE	COLUMN SIZE
1	ID	Table row serial number	I(8) not null	19
2	ID_D	Name of data provider; which database data originates from.	VC(50) not null	50
3	ID_C	Context ID	I(8) not null	19
4	SOURCE	What type of source the data is extracted from: Death records, Marriage records, Migration records, Poll-tax registers, Birth records, Income registers, Parish registers, Censuses	VC(50) not null	50
5	TYPE	Type	VC(32) not null	32
6	VALUE	Value of the type	VC(50) not null	50
7	DATE_TYPE	Type of date	VC(50)	50
8	ESTIMATION	Type of estimation of date	VC(50)	50
9	MISSING	Reason why a date or a part of a date is missing (and had to be estimated)	VC(50)	50
10	START_YEAR	Start year of period	I(4)	4
11	START_MONTH	Start month of period	I(4)	2
12	START_DAY	Start day of period	I(4)	2
13	END_YEAR	End year of period	I(4)	4
14	END_MONTH	End month of period	I(4)	2
15	END_DAY	End day of period	I(4)	2

3.4.1 Rules for Type

The TYPE attribute in the SwedPop IDS must adhere to strict validation rules to ensure consistency and interoperability across datasets:

- **Metadata Compliance**: Every TYPE must be explicitly defined in the SwedPop-IDS metadata table. Only values listed as valid in this metadata schema are permissible.
- **Case Sensitivity**: All TYPE's must be written in uppercase letters to maintain uniformity and prevent discrepancies during data processing and extraction.

3.4.2 Rules for Value

Attributes in the VALUE column of the CONTEXT table are managed according to the following principles:

1. **Parish and County Levels**
   Values corresponding to Parish and County must adhere to the standardized entries defined in the harmonized SwedPop parish codelist.
2. **Metadata-Defined Values**
   Values specified in the metadata table must match the spelling and formatting exactly as documented in the metadata schema.

3. **Location Formatting**
   For Location values, the first letter of each word must be written in uppercase, while all subsequent letters are in lowercase.
4. **Other Values**
   Values not covered by the above rules may be handled according to the preferences of each contributing data source.

3.5 Table INDIV_CONTEXT

The INDIV_CONTEXT table in the SwedPop IDS is designed to associate individuals with specific contexts—such as families, households, or other social and spatial environments—along with a corresponding moment or period in time. This structure enables the representation of dynamic relationships between individuals and their living environments, supporting longitudinal analyses of household composition, family structures, and residential patterns.

Table 9. Metadata table INDIV_CONTEXT

ORDINAL POSITION	COLUMN	DESCRIPTION	DATATYPE	COLUMN SIZE
1	ID	Table row serial number	I(8) not null	19
2	ID_D	Name of data provider; which database data originates from.	VC(50) not null	50
3	ID_I	Individual ID	I(8) not null	19
4	ID_C	Context ID	I(8) not null	19
5	SOURCE	What type of source the data is extracted from: Death records, Marriage records, Migration records, Poll-tax registers, Birth records, Income registers, Parish registers, Censuses	VC(50) not null	50
6	RELATION	The relationship between individual and context.	VC(50) not null	50
7	DATE_TYPE	Type of date	VC(50) not null	50
8	ESTIMATION	Type of estimation of date	VC(50)	
9	MISSING	Reason why a date or a part of a date is missing (and had to be estimated)	VC(50)	50
10	START_YEAR	Start year of period	I(4)	10
11	START_MONTH	Start month of perio	I(4)	10
12	START_DAY	Start day of period	I(4)	10
13	END_YEAR	End year of period	I(4)	10
14	END_MONTH	End month of period	I(4)	10
15	END_DAY	End day of period	I(4)	10

3.5.1 Rules for Relation attributes

The RELATION attribute in the SwedPop Intermediate Data Structure (IDS) must adhere to strict metadata-based validation requirements:

- **Metadata Definition**: All valid values for RELATION are explicitly defined in the metadata table. Only these predefined values are permissible.

- **Exact Spelling**: The spelling of values in the RELATION column must correspond exactly to the spelling specified in the metadata table, including case sensitivity and character formatting.

3.6 Table CONTEXT_CONTEXT

The CONTEXT_CONTEXT table defines connections between different layers in a hierarchy. Contexts are hierarchical, but several contexts may be defined for the same layer. The system of contexts may change over time and several systems may exist for the same period. A database may include multiple context hierarchies, such as address-neighborhood-municipality and page-volume-district.

With the context hierarchy, attributes of more inclusive layers can be linked directly or indirectly to lower layers. This means that only one record is necessary in the INDIV_CONTEXT table to grasp all contextual situations for an individual in a particular context hierarchy.

Table 10. Metadata table CONTEXT_CONTEXT

ORDINAL POSITION	COLUMN	DESCRIPTION	DATATYPE	COLUMN SIZE
1	ID	Table row serial number	I(8) not null	19
2	ID_D	Name of data provider; which database data originates from.	VC(50) not null	50
3	ID_C_1	Context ID, relation to ID_C_2	I(8) not null	19
4	ID_C_2	Context ID, relation to ID_C_1	I(8) not null	19
5	SOURCE	What type of source the data is extracted from: Death records, Marriage records, Migration records, Poll-tax registers, Birth records, Income registers, Parish registers, Censuses	VC(50) not null	50
6	RELATION	The relationship from the first context to the second. The relationship is not necessarily hierarchical.	VC(50) not null	50
7	DATE_TYPE	Type of date	VC(50)	50
8	ESTIMATION	Type of estimation of date	VC(50	50
9	MISSING	Reason why a date or a part of a date is missing (and had to be estimated)	VC(50)	50
10	START_YEAR	Start year of period	I(4)	4
11	START_MONTH	Start month of period	I(4)	2
12	START_DAY	Start day of period	I(4)	2
13	END_YEAR	End year of period	I(4)	4
14	END_MONTH	End month of period	I(4)	2
15	END_DAY	End day of period	I(4)	2

3.6.1 Valid RELATION values

Valid RELATION values are explicitly defined in the METADATA table, and their spelling must strictly correspond to the standardized representation specified therein. This constraint ensures lexical consistency across all IDS databases, prevents semantic ambiguity, and supports automated validation processes within the SwedPop IDS framework.