

SwedPop Documentation

# SwedPop IDS

Authors: Annika Westberg and Maria Larsson, CEDAR  
Version August 25, 2022

# Contents

- 1 Introduction..... 4**
- 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 Extraction Tool database. .... 4
    - 2.1.3 Set values as described in *Reference guide to setting values in IDS tables*. .... 4
    - 2.1.4 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 ..... 5
    - 2.2.4 Multiple and different dates for the same event ..... 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’s and VALUE’s that appear in the local IDS databases are represented in the metadata table. 8
    - 3.1.4 All TYPE’s have a VALUE=DEFINITION. .... 8
    - 3.1.5 Multiple meanings of table column ID\_D..... 8
  - 3.2 *Table INDIVIDUAL* ..... 9
    - 3.2.1 Type ..... 10
    - 3.2.2 Value ..... 10
    - 3.2.3 Coded geographic locations ..... 10
    - 3.2.4 Uncoded geographic locations ..... 10
    - 3.2.5 All individuals must have values in both BIRTH\_LOCATION and BIRTH\_DATE..... 10
    - 3.2.6 Observation, arrival and departure ..... 10
    - 3.2.7 Occupations..... 11
    - 3.2.8 Cause of death..... 12
  - 3.3 *Table INDIV\_INDIV*..... 13
    - 3.3.1 Relation ..... 13
    - 3.3.2 Example of double directed relations ..... 13
  - 3.4 *Table CONTEXT* ..... 14
    - 3.4.1 Type..... 15
    - 3.4.2 Value ..... 15
  - 3.5 *Table INDIV\_CONTEXT*..... 16
    - 3.5.1 Relation ..... 16
  - 3.6 *Table CONTEXT\_CONTEXT* ..... 17
    - 3.6.1 Relation ..... 17

<b>4</b>	<b>Data delivery .....</b>	<b>18</b>
4.1	<i>Verify data consistency.....</i>	<i>18</i>
4.2	<i>Deliver data in proper format.....</i>	<i>18</i>
4.3	<i>Deliver correct tables.....</i>	<i>18</i>

# 1 Introduction

SwedPop is a national micro-level population infrastructure representing the coordination of existing population databases at five Swedish universities and institutes. The core databases contain individual-level data, longitudinal, cross-sectional as well as panel data. The overarching aim of this infrastructural initiative is to develop SwedPop as a national long-term infrastructure making the Swedish micro-level population data accessible for Swedish and international researchers. The core databases are presented at the SwedPop web-portal ([www.swedpop.se](http://www.swedpop.se)) which serves as a hub for the dissemination of data.

SwedPop data is structured by the principles of IDS – Intermediate Data Structure, a common format for databases containing longitudinal and non-aggregated information on persons, families and households. IDS has been developed by scholars in an ESF Research Networking Programme called European Historical Population Samples Network, EHPS-Net (see <https://ehps-net.eu/>). It forms an integrated and joint interface between databases in order to transcend regional and national differences.

The IDS structure is described in an article written by two of the founders of IDS – 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 to provide a complete description of the SwedPop IDS database, how to handle attributes and what demands need to be met before delivering data to the SwedPop IDS.

## 2 SwedPop IDS

The data is structured using the common database structure Intermediate Data Structure IDS. SwedPop have further developed the structure into a distinctive SwedPop IDS in order to fully cover the distinctive features that is found in the Swedish sources. Along with harmonization of the data structure, there are shared nomenclatures and coding systems as well as common system for metadata.

### 2.1 General rules

Some rules are general and applied in all tables.

#### 2.1.1 Only valid attributes are entered in the database.

Validity can be defined as correct spelling, using upper- and lower case letters correctly, entering attributes in correct tables, using date-related attributes together with correct TYPE's etc.

All valid attributes are described in the metadata table and further definitions are described in this documentation.

#### 2.1.2 When attributes are added, their values are defined in the Extraction Tool database.

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

#### 2.1.3 Set values as described in *Reference guide to setting values in IDS tables*.

The document *Reference Guide to setting values in IDS tables* is a guide written for to data providers and defines correct SwedPop IDS tables. It contains examples of what is included in correct tables.

2.1.4 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 does not contain rows that represent an empty or missing value, with the exception of the HISCO columns STATUS and RELATION.

## 2.2 Dates and timestamps

Dates are recorded in several tables as event- or period dates and 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

Date\_type describes how the date is recorded or observed, when an attribute is valid and if the database administrators have assigned the date from various sources. Since it refers to handling dates, and not other data, it is only used for TYPE's that contain dates. Further definitions can be found in *Reference guide to setting values in IDS tables*.

Rules for Date\_type include:

- Date\_type has value only for TYPE's that contain dates.
- Date\_type is blank cell when the TYPE does not handle a date.

### 2.2.2 Estimation

Database administrators can estimate dates based upon source information. Since it refers to handling dates, and not other data, it is only used for TYPE's that contain dates. Further definitions can be found in *Reference guide to setting values in IDS tables*.

Rules for Estimation include:

- Estimation has value only for TYPE's that contain dates.
- Estimation is blank cell when the TYPE does not handle a date.

### 2.2.3 Missing

Information on the reason why there are missing dates is described in this column. Further definitions can be found in *Reference guide to setting values in IDS tables*.

Rules for Missing include:

- Missing has value only for TYPE's that contain dates.
- Missing is blank cell when the TYPE should not contain dates.

Missing values for TYPE's are:

Missing	TYPE
Time_invariant	Time invariant TYPEs <ul style="list-style-type: none"><li>• CHILDBIRTH_ASSISTANT</li><li>• LEGITIMACY</li><li>• MULTIPLE_BIRTH</li><li>• SEX</li></ul>

Unavailable	Part of date is unavailable
	Date is not applicable to this type (missing=blank cell).

## 2.2.4 Multiple and different dates for the same event

Some individuals may have multiple and different dates for the same event in local databases, such as different birth dates. SwedPop IDS only has one date, and the local database choose which one is delivered.

## 2.3 Definition of family

SwedPop defines family as those living together in a conjugal family unit. Families are constructed from the sources in different ways depending on the type of source. In catechetical registers, family members are recorded on the same page during the same time. In censuses people are recorded within a household and the family relations are constructed from their role in the household.

Roles in the conjugal family unit are:

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

Further definitions are:

- A person can only be member of one family during the same time.
- Single persons has a family number of their own. Also siblings living together at the same time have a family number of their own.
- A person can be part of both a family and a household during the same time.

Including the selection family in data extractions mean that all family members to the sample individuals are added to the dataset, including the selection of variables.

*Family ID (ID\_C for LEVEL= Family)*

Identifying families in the databases is structured by the definition of twelve digits. The first position is database ID, the second position identifies that the context is family (=1). The third to twelfth position is a serial number.

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 household as:

- Members of a household live together at the same place during the same time.
- Members of a household can comprise related persons, employees or others.

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

At the moment, only SEDD and SweCens data contain data on households.

When including the selection household in data extractions, all household members to the sample individuals are added to the dataset, including the selection of variables.

*Household ID (ID\_C for LEVEL= Household)*

Identifying households in the databases is structured by a definition of twelve digits. The first position is database ID, the second position identifies that the context is family (=2). The third to twelfth position is a serial number.

Database	Household ID
GOPP	GOPP database ID=2 Household context=2 + 10 digit serial number
Rotemannen	Rotemannen database ID=3 Household context=2 + 10 digit serial number
SEDD	SEDD database ID=4 Household context=2 + 10 digit serial number
SweCens	SweCens database ID=5 Household context=2 + 10 digit serial number

## 2.5 Definition of relatives

The sources contain many different types of relatives. SwedPop defines relatives as:

All different kinds of relatives that are recorded in the sources are included; biological as well as non-biological. In contrast with family and household, relatives have no requirements of living together in time or space.

When including the selection relatives in data extractions, all relatives to the sample individuals are added to the dataset, including the selection of variables.

## 3 IDS tables

### 3.1 Table METADATA

The table METADATA provides complete descriptions 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.

No local changes are allowed in the METADATA table. If there are new attributes or values, they must be reported to the SwedPop IDS working group for inclusion in the metadata table and the Extraction Tool database.

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

The metadata table consists only of attributes that appear in the SwedPop IDS database.

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

All TYPE's must have a DEFINITION. The VALUE=DEFINITION is only represented in the metadata table.

#### 3.1.5 Multiple meanings of table column ID\_D

The meaning of ID\_D (ID of Database) is twofold depending on where it appears; in metadata or in the IDS database/extracted datasets. In the IDS database/extracted datasets, ID\_D is always the name of the data provider. In the metadata table, ID\_D can be either SWEDPOP or the name of the data provider.



Table 1. Explanation of appearance of ID\_D

	ID_D	In metadata table	In IDS database/ Extracted 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 database.  <i>Example 1:</i> the value 'Hermaphrodite' only appears in DDB data. ID_D=DDB.  <i>Example 2:</i> the type 'Has_left_the_swedish_church' only appears in SweCens data. ID_D=SweCens.  <i>Example 3:</i> 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

### 3.2 Table INDIVIDUAL

The table INDIVIDUAL contains all attributes that characterize an individual.

Table 2. Metadata table INDIVIDUAL

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	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(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 Type

The values of TYPE must always be defined in the metadata table and only valid TYPE's as defined in SwedPop-IDS metadata can be used. Values in TYPE must be written in upper case. Every TYPE has one row

### 3.2.2 Value

The IDS data is stored in the Extraction Tool Database (ET Database). The ET Database works generically with the online extraction tool, and values in VALUETYPE\_FK control parts of its functionality. Therefore, values must be set according to VALUE\_TYPE\_FK. Consult the document *Reference guide to setting values in the IDS database* if there are uncertainties.

Values in the column VALUE that are defined in the table metadata must correspond exactly to the value in the table metadata.

### 3.2.3 Coded geographic locations

The lowest level of coded geographic data in the column VALUE\_ID\_C is parish. Coded geographical data on lower levels than parish, such as location or residence, is handled in the table INDIV\_CONTEXT.

### 3.2.4 Uncoded geographic locations

Uncoded geographic locations are written in plain text in VALUE, without reference to VALUE\_ID\_C.

### 3.2.5 All individuals must have values in both BIRTH\_LOCATION and BIRTH\_DATE.

In the extraction tool the sample selections on birth location and birth date are ruled out if these fields are empty.

### 3.2.6 Observation, arrival and departure

Depending on what kind of observation or migration (arrival, departure) takes place they are handled in different ways:

- a) An individual's **geographical migration into or out from the sources** is recorded in ARRIVAL\_FROM and DEPARTURE\_TO in table INDIVIDUAL.
- b) An individual's **geographical migration between locations that are covered by the sources and without leaving observation** is recorded in table INDIV\_CONTEXT.
- c) **Observation in the sources can start by** birth, arrival, start source or unknown and is recorded in START\_OBSERVATION. Date of the value 'Arrival' corresponds with date of the type ARRIVAL\_FROM.
- d) **Observation in the sources can end by** death, departure, end source or unknown and is recorded in the type END\_OBSERVATION. Date of the value 'Departure' corresponds with date of the type DEPARTURE\_TO.

### 3.2.7 Occupations

Occupations in the SwedPop databases have been standardized and coded based on the international classification system HISCO (see *SwedPop Documentation: Principles of Coding Swedish Historical Occupations*).

Information about occupation should include the following variables:

OCCUPATION_STANDARD_PART_x
OCCUPATION_HISCO_PART_x
OCCUPATION_HISCO_STATUS_PART_x
OCCUPATION_HISCO_RELATION_PART_x
OCCUPATION_HISCO_PRODUCT_PART_x

When there is only one occupation on the record it will have the suffix PART\_1.

#### *Multiple occupational titles on the same record*

There can be multiple occupational titles on the same record, which is a problem when they also have identical date and source. In order to separate them and connect the correct titles and HISCO, they have a numeric suffix \_PART\_>1.

Table 3. 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

#### *Multiple sequential occupations with identical source and date*

When there are multiple occupations on the same record, and it is apparent that the first occupation on the record is a previous one by being crossed over, they can be separated in time by estimation. In these cases, the following information must be given:

DATE\_TYPE = Assigned

ESTIMATION= Middling

If occupations appear with information on previous status, such as ‘former plumber’, then no date is estimated because it is clear that the occupation is a previous one.

#### *Occupations in the extraction tool*

When selecting OCCUPATION\_HISCO in the extraction tool, three different variables are automatically added to the dataset: OCCUPATION\_HISCO, OCCUPATION\_HISCO\_STATUS and OCCUPATION\_HISCO\_RELATION. The variables STATUS and RELATION are attached as they define, among other things, ownership, artisan career steps, relations to occupation owner and former or future positions.

Example: The occupational title *shoemaker master's wife* is given HISCO-code for shoemaker (80110), STATUS for master (21) and RELATION for wife (11). These types of occupational titles need all three variables in order to cover the information.

### 3.2.8 Cause of death

SwedPop has harmonized data on cause of death based on the international historical system of coding causes of death ICD10h. For further information, see *SwedPop Documentation: Principles of Coding Historic Cause of Death*. There can be more than one cause of death on the same record, but there is not always information on primary and secondary causes of death. This is handled by adding a suffix (*\_PART\_n*) to the TYPE. The system of suffixes to separate multiple records is used also when only one cause of death is recorded.

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

TYPE	Explanation
DEATH_CAUSE_ICD10H_PART_1	The first death cause on the record is given the suffix <i>_PART_1</i> .
DEATH_CAUSE_ICD10H_PART_2	If there are multiple death causes, the second on the record is given the suffix <i>_PART_2</i> .

### 3.3 Table INDIV\_INDIV

The table INDIV\_INDIV shows how individuals are related to each other.

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, referring to ID_I in the INDIVIDUAL table.	I(8) not null	19
4	ID_I_2	Individual ID, relation to ID_I_1, referring 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 Relation

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

#### 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).

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 table CONTEXT lists all relevant contextual information, such as information about households and regions. Contextual information are the layers of environments where individuals live. They can be living environments, such as family or household members, but also in geographical environments, such as a parish. The CONTEXT table contains all associations between contexts displayed by level (family, household, locality, parish etc) and name of the level.

Values in ID\_C always contain twelve digits, starting with a three-digit prefix that identifies the context level.

Level	Context ID
Country	700 + country code, not used in this version of IDS except for foreign country 700999999998 – foreign country, i.e outside Sweden
Parish	701 + 10 digit parish code 701999999999 – unknown
County	702 + county code
Family	See, chapter 2.3 above
Household	See chapter 2.4 above
Polltax	SEDD database ID=4 Polltax context=3 + 10 digit serial number
Locality	DDB database ID=1 Locality context=4 + 10 digit serial number

Table 7. 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 Type

The values of TYPE must always be defined in the metadata table and only valid TYPE as defined in SwedPop-IDS metadata.xls can be used. Values in TYPE must be written in upper case.

### 3.4.2 Value

Values in the column VALUE are handled in one of the following ways:

- 1) Values of level Parish and County are defined in the harmonized SwedPop parish codelist.
- 2) Values are defined in the table metadata and spelling correspond to value in the table metadata.
- 3) Values for Location are written with first letter in every word in upper case and the rest in lower case.
- 4) Other values can be handled as every part wishes.

### 3.5 Table INDIV\_CONTEXT

The table INDIV\_CONTEXT places individuals into contexts. Data in this table associates individuals with a context, such as family or household, and a moment or period in time.

*Table 8. 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 Relation

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



### 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 of time. 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 9. 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 Relation

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

## 4 Data delivery

The data provider is responsible for verifying data consistency, to deliver data in proper format and to deliver the correct tables. Data that does not fulfill these requirements will not be added to SwedPop IDS and the extraction tool.

### 4.1 Verify data consistency

Local IDS database must be consistent with SwedPop IDS.

Run queries to verify consistency. If data is not consistent, do one of the following:

- i) Change the local database or
- ii) Report changes to the SwedPop IDS working group.

### 4.2 Deliver data in proper format

- i) Table columns are ordered according to the definitions in the tables metadata.
- ii) Data is stored in tab-separated text files.
- iii) Data is stored in UTF-8.

### 4.3 Deliver correct tables

The following tables must be delivered:

- i) INDIVIDUAL
- ii) INDIV\_CONTEXT
- iii) INDIV\_INDIV
- iv) CONTEXT
- v) CONTEXT\_CONTEXT