How to decode information and data included in the SDMX-CSV files

1. Structure of the SDMX-CSV files: header, records, variables and codes.

The available files include ESSPROS data by scheme on expenditure and receipts for each country.
For example, the file EXPEND_IT.csv includes the data on expenditure (EXPEND) for Italy (IT);
the file RECEIPT_FR.csv includes the data on receipts (RECEIPT) for France (FR).
In addition, SCHEMES files include information on the list of schemes for each country.

File names usually include also the (last) reference year of data.

See here below an extract from a file on expenditure for Italy and Ireland.

**time_period;ref_area;counterpart_area;flow;counterpart_sector;exp_type;function;means_test;
benefit_payment;detailed_benefit;custom_breakdown;measure;freq;unit;unit_mult;obs_value;
obs_status; comment_obs**
1990;IT;_T;EXP;S14;SPR;FAM;NMT;CASH_LUMP;_T;SCH021;EXP_GRS;A;EUR;6;0;;
1990;IT;_T;EXP;S14;SPR;FAM;NMT;CASH_LUMP;_T;SCH022;EXP_GRS;A;EUR;6;0;;
1990;IT;_T;EXP;S14;SPR;FAM;NMT;CASH_LUMP;_T;SCH023;EXP_GRS;A;EUR;6;0;;
1990;IT;_T;EXP;S14;SPR;FAM;NMT;CASH_LUMP;_T;SCH024;EXP_GRS;A;EUR;6;0;;

**time_period;ref_area;counterpart_area;flow;counterpart_sector;exp_type;function;means_test;
benefit_payment;detailed_benefit;custom_breakdown;measure;freq;unit;unit_mult;obs_value;
obs_status; comment_obs**
2014;IE;_T;EXP;S1;ADM;_T;_T;_T;_Z;SCH002;EXP_GRS;A;EUR_FX;6;0.517974639;;
2014;IE;_T;EXP;S1;ADM;_T;_T;_T;_Z;SCH007;EXP_GRS;A;EUR_FX;6;17.28707372;;
2014;IE;_T;EXP;S1;ADM;_T;_T;_T;_Z;SCH015;EXP_GRS;A;EUR_FX;6;3.097207728;;
2014;IE;_T;EXP;S1;ADM;_T;_T;_T;_Z;SCH024;EXP_GRS;A;EUR_FX;6;1095.060649;;

In a CSV files the pieces of information are distinguished by using a "separator": in these files the standard separator is a "semicolon" (";").

The **first row of the file is the header** (includes the short description of the "variables" included in each column).

Some of the variables are intuitively understandable: for example the first and second column ("time_period" and "ref_area") indicate the reference year of the data and the declaring country. Indeed, all the four records above from the Italian file refer to the year 1990 and to Italy ("IT").

An important variable is "obs_value": "observation value". In the example of the fourth record from the Irish file, it corresponds to the value of expenditure (1095.060649) for a specific item of the ESSPROS questionnaire (more specifically, the administration costs of scheme 24). The standard format used for this numeric field ("obs_value") is as follows:
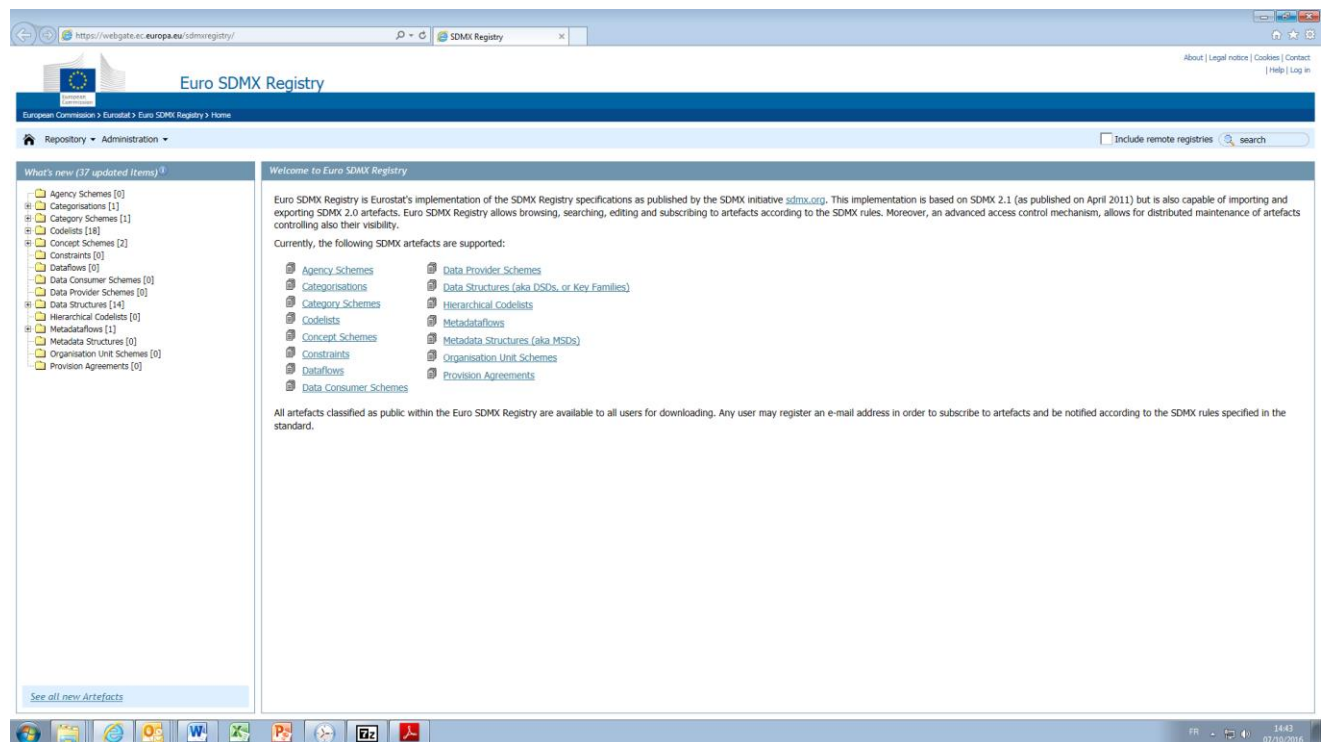- the decimal separator is a "point" (".").
- there is no "thousand separator".

For other "variables" and related codes the interpretation is less intuitive. For example, what's the meaning of the third variable "counterpart_area" and the corresponding code "_T" in the records shown above?

In order to get complete (and updated) information of the meaning of each "variable" and the related code list, please use the SDMX registry:

https://webgate.ec.europa.eu/sdmxregistry

Figure 1: SDMX Registry home page

However, please find here below a quick overview of the meaning of the ESSPROS "variables". For general concepts, a more complete reference can be found in the SDMX glossary:

https://sdmx.org/?sdmx_news=new-sdmx-glossary-available .

Finally, in the following section (section 2) you will find instruction concerning the related code lists and specific codes.

The only variable used in the csv files for which can not be found in the SDMX Registry code lists is named "item". This variable corresponds to the original ESSPROS classification of expenditure and receipts, as published in the Appendix 1 of the "***ESSPROS Manual and user guidelines – 2016 edition***":
http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-16-010 .

TIME PERIOD (CONCEPT ID: TIME_PERIOD)

Definition   Timespan or point in time to which the observation actually refers.

REFERENCE AREA (CONCEPT ID: REF_AREA)

Definition   Country or geographic area to which the measured statistical phenomenon relates.

COUNTERPART REFERENCE AREA (CONCEPT ID: COUNTERPART_AREA)

Definition   Secondary area, as opposed to reference area, to which the measured data are in relation.

Context   The "counterpart area" (also known as "vis-a-vis area") is related to statistics on foreign trade, migration or other domains. It determines, from the point of view of the reporting country, the corresponding area to which the economic or other flows are related to (for instance, in statistics on imports, the counterpart reference area is the area of origin of the goods). For example in ESSPROS, some receipts may originate from non-resident institutional units. In this case the counterpart area of these receipts is he "rest of the world"

FLOW DIRECTION (CONCEPT ID: FLOW)

Definition   Direction of economic flow (Expenditure vs. Receipts).

COUNTERPART SECTOR (CONCEPT ID: COUNTERPART_SECTOR)

Definition   Counterpart institution sector (according to National Accounts definitions) of economic flows (for example: Sector of origin of scheme's receipts).

EXPENDITURE TYPE (CONCEPT ID: EXP_TYPE)

Definition   Type of expenditure.

RECEIPT TYPE (CONCEPT ID: REC_TYPE)

Definition   Type of receipts.

FUNCTION (CONCEPT ID: FUNCTION)

Definition   Social Protection Function.

MEANS TESTING (CONCEPT ID: MEANS_test)

Definition   Means-testing as defined in ESSPROS.

BENEFIT PAYMENT (CONCEPT ID: BENEFIT_PAYMENT)

Definition   Type of social benefits.

DETAILED BENEFIT (CONCEPT ID: DETAILED_benefit)

Definition   Specific type of social benefits.

ESSPROS SCHEME NUMBER (CONCEPT ID: CUSTOM_BREAKDOWN)

Definition   Integer number identifying the ESSPROS scheme.

ESSPROS SCHEME NAME IN ORIGINAL LANGUAGE(S) (CONCEPT ID: CUSTOM_BREAKDOWN_LBL)

Definition   Name of the ESSPROS scheme in original language(s).

ESSPROS SCHEME NAME IN ENGLISH (CONCEPT ID: CUSTOM_BREAKDOWN_LBL_EN)

Definition   Name of the ESSPROS scheme in English.

ESSPROS SCHEME STARTING DATE (CONCEPT ID: ACTIVE_FROM)

Definition   Year during which the ESSPROS scheme was introduced.

ESSPROS SCHEME ENDING DATE (CONCEPT ID: ACTIVE_TO)

Definition   Last year during which the ESSPROS scheme was active before being discontinued.

MEASURE (CONCEPT ID: MEASURE)

Definition   Statistical concept for which data are provided in a data set.

FREQUENCY OF OBSERVATION (CONCEPT ID: FREQ)

Definition   Time interval at which observations occur over a given time period.

UNIT OF MEASURE (CONCEPT ID: UNIT)

Definition   Unit in which the data values are expressed.

UNIT MULTIPLIER (CONCEPT ID: UNIT_MULT)

Definition   Exponent in base 10 specified so that multiplying the observation numeric values by 10^UNIT_MULT gives a value expressed in the unit of measure.

OBSERVATION VALUE (CONCEPT ID: OBS_VALUE)

Definition   Value of a particular variable.

OBSERVATION STATUS (CONCEPT_ID: OBS_STATUS)

Definition   Information on the quality of a value or an unusual or missing value.

COMMENTS (CONCEPT ID: COMMENT_OBS)

Definition   Comments related to the observation value.

2. How to access the ESSPROS code lists in the SDMX registry

The information on the "variables" used in each ESSPROS SDMX-CSV file is included in the "data structures" (DSD= "data structure definition") developed for ESSPROS.

In the SDMX Registry there are 5 DSD for ESSPROS: one for expenditure, one for receipts, one for number of beneficiaries, one for net benefits and one for the list of schemes.
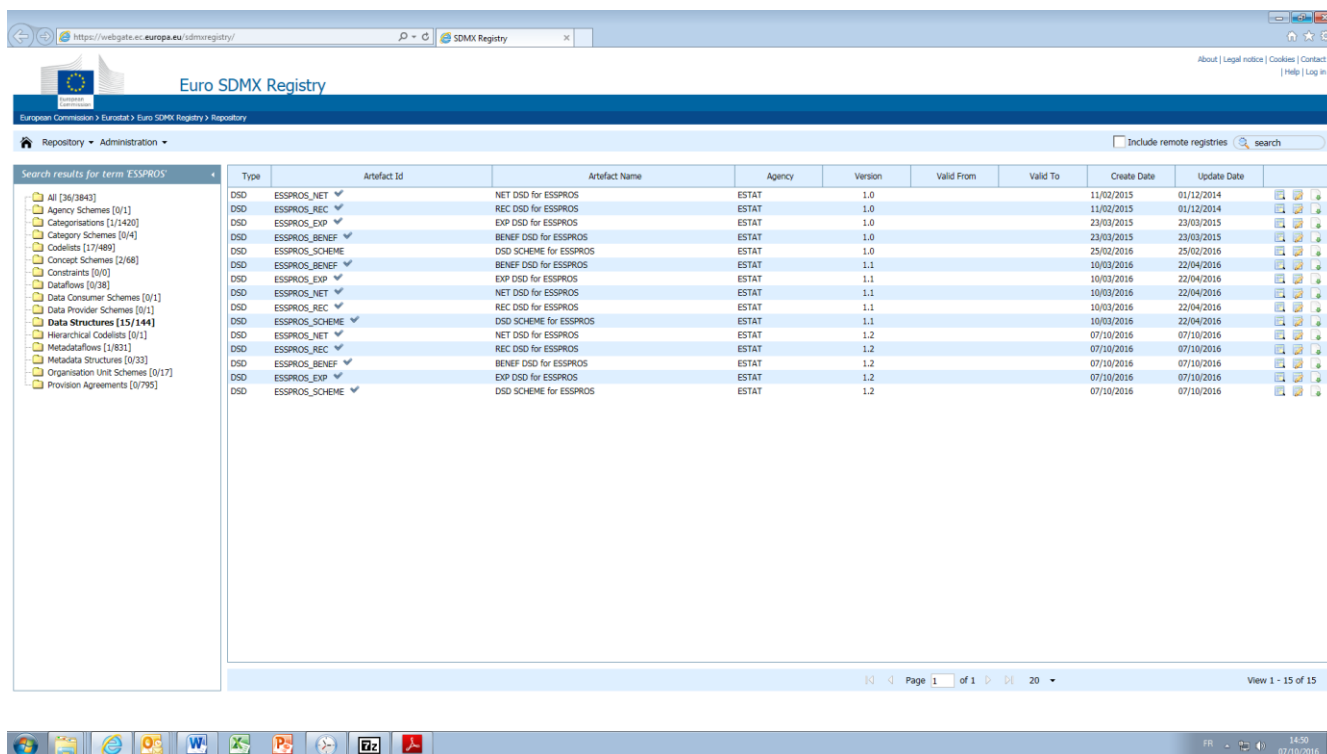
In order to decode the content of, for example, the "expenditure" file, one has to find the DSD for "ESSPROS expenditure"

From the registry home page:
1. search for "ESSPROS" (in the up right corner) and
2. select "Data structures" (on the left side menu).

The result is the list of all the ESSPROS DSD:

Figure 2: List of ESSPROS DSD

Since the DSDs are regularly updated, there are different versions of them (see column "version"): if you do not know the version corresponding to the file you need to decode, please choose the most recent version. In our example please choose the DSD ESSPROS_EXP, version 1.2 created on 7/10/2016.

DSD can be viewed on the screen or downloaded (click on the icons on the right side, next to each DSD).

Let's download ESSPROS_EXP version 1.2.
To download the DSD in CSV format, please select "CSV".

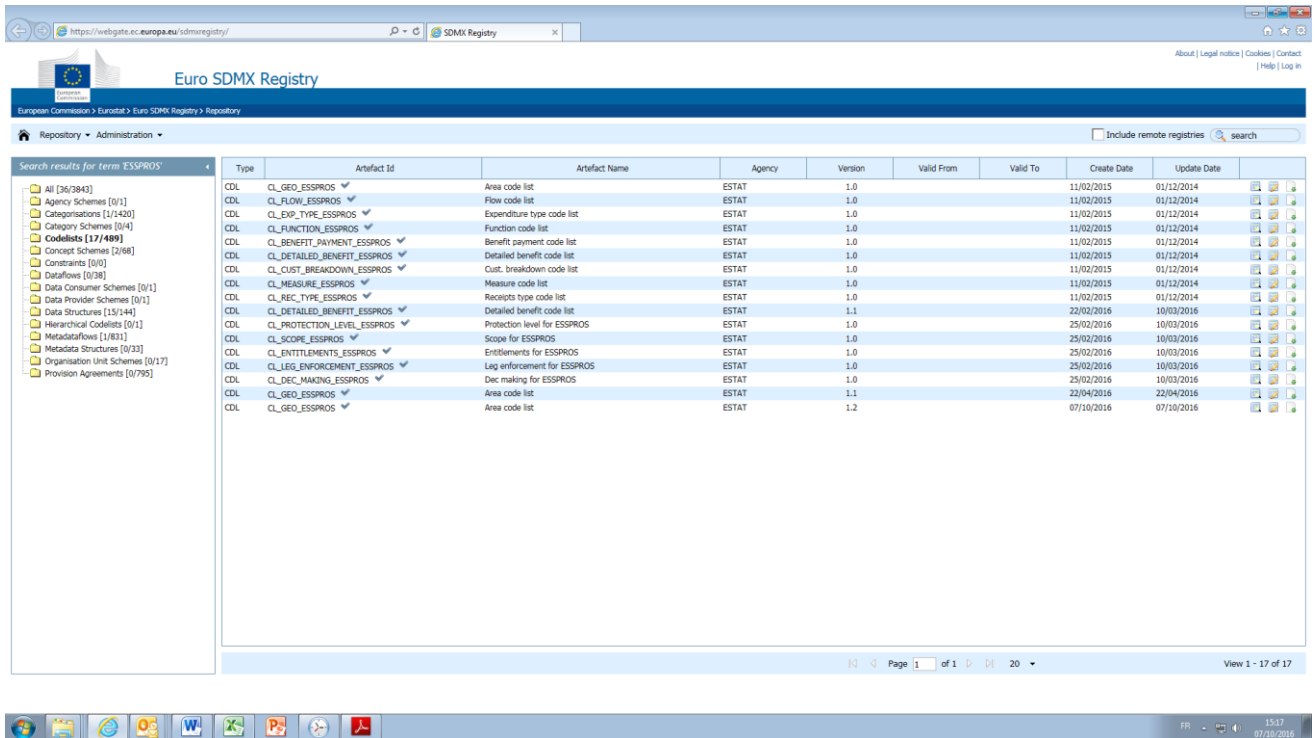See here below the content of the DSD ESSPROS_EXP version 1.2, in CSV format:

concept_id;format;codelist_id;role;conditional;position
FREQ;;SDMX+CL_FREQ+2.0;Dimension;;1
REF_AREA;;ESTAT+CL_GEO_ESSPROS+1.2;Dimension;;2
COUNTERPART_AREA;;ESTAT+CL_GEO_ESSPROS+1.1;Dimension;;3
FLOW;;ESTAT+CL_FLOW_ESSPROS+1.0;Dimension;;4
COUNTERPART_SECTOR;;ESTAT+CL_SECTOR+1.2;Dimension;;5
EXP_TYPE;;ESTAT+CL_EXP_TYPE_ESSPROS+1.0;Dimension;;6
FUNCTION;;ESTAT+CL_FUNCTION_ESSPROS+1.0;Dimension;;7
MEANS_TEST;;ESTAT+CL_MEANS_TEST+1.0;Dimension;;8
BENEFIT_PAYMENT;;ESTAT+CL_BENEFIT_PAYMENT_ESSPROS+1.0;Dimension;;9
DETAILED_BENEFIT;;ESTAT+CL_DETAILED_BENEFIT_ESSPROS+1.1;Dimension;;10
CUSTOM_BREAKDOWN;;ESTAT+CL_CUST_BREAKDOWN_ESSPROS+1.0;Dimension;;11
MEASURE;;ESTAT+CL_MEASURE_ESSPROS+1.0;Dimension;;12
TIME_PERIOD;OBSERVATIONAL_TIME_PERIOD;;Time Dimension;;13
OBS_VALUE;FLOAT;;Primary Measure;;14
OBS_STATUS;;SDMX+CL_OBS_STATUS+2.0;Attribute;Mandatory;
CONF_STATUS;;SDMX+CL_CONF_STATUS+1.0;Attribute;Mandatory;
UNIT;;ESTAT+CL_UNIT+1.3;Attribute;Mandatory;
UNIT_MULT;;SDMX+CL_UNIT_MULT+1.0;Attribute;Mandatory;
DECIMALS;;SDMX+CL_DECIMALS+1.0;Attribute;Conditional;
OBS_STATUS2;;SDMX+CL_OBS_STATUS+2.0;Attribute;Conditional;
TIME_FORMAT;;SDMX+CL_TIME_FORMAT+1.0;Attribute;Mandatory;
OBS_STATUS3;;SDMX+CL_OBS_STATUS+2.0;Attribute;Conditional;
COMMENT_OBS;STRING;;Attribute;Conditional;

The DSD includes for each « variable » (« **concept_id** » in the **header**) the associated code list ("codelist_id").
For example the code list associated to the variable "COUNTERPART_AREA" is the code list (CL) "GEO_ESSPROS (version 1.1)".

In order to download the code list "GEO_ESSPROS 1.1", please select "Code lists" on the left side menu of the SDMX Registry ….

Figure 3: Code lists specific for ESSPROS (=key word used for "searching" = "ESSPROS")

… and then download (or view) the code list GEO_ESSPROS 1.1.

Here below an extract of this code list. The code "_T" ("All countries") means that the expenditure is calculated for beneficiaries resident in "all countries" of the world (indeed, in general "_T" means "Total" in the code lists).

```
id;name_en;description_en;parent
LV;Latvia;;
LU;Luxembourg;;
LT;Lithuania;;
HR;Croatia;;
(…)
PL;Poland;;
RS;Serbia;;
CH;Switzerland;;
RU;Russian Federation;;
LI;Liechtenstein;;
WRL_X_REP;Rest of the world (total except country in REF_AREA);;
_T;All countries;;
```

You can repeat this operation for all the code lists you need (for example those used in the SDMX-CSV files you need to "decode").

Please note that some codelists are not specific to ESSPROS only but are in common with other statistical domains. For example the code lists for counterpart sector and for means-testing (respectively "CL_SECTOR" and "CL_MEANS_TEST") require a separate "search" operation (see top right corner of the SDMX homepage), since you will not find them using the keyword "ESSPROS". Please use for example "SECTOR" and "MEANS" respectively as keywords for the searching engine.

NB: If you want to download both the DSD and **all the associated code lists** (both ESSPROS-specific and generic ones) in one operation, keep the selection "resolve references?" when downloading the DSD.

3. Special advice when using the file SCHEME

The information included in the file SCHEME may appear not to be very well structured or clear (especially the last two columns including the name of the scheme in the original language and in English).

This may depend on the editor used to open/manipulate the file and its ability to correctly decode possible special characters included in the file.

Some useful links about SDMX

Eurostat SDMX infospace: https://webgate.ec.europa.eu/fpfis/mwikis/sdmx/index.php/Main_Page

SDMX organisation: http://sdmx.org/