**** **

**SDMX Implementation for UOE**

UNESCO-UIS / OECD / EUROSTAT  
data collection on formal education

DSD Guidelines

June 2019

*Table of Contents*

Contents

[1 Version History 3](#_Toc11964088)

[2 Introduction 3](#_Toc11964089)

[3 Concepts and Code Lists 4](#_Toc11964090)

[4 The DSD Matrix file 7](#_Toc11964091)

[4.1 Concept Scheme Sheet 7](#_Toc11964092)

[4.2 DSD-Concept Matrix Sheet 7](#_Toc11964093)

[4.3 Code List Sheet 10](#_Toc11964095)

[4.4 Comparison with previous DSD Matrix 11](#_Toc11964096)

[5 SDMX Artefacts 14](#_Toc11964102)

[5.1 UOE Education implementation model 14](#_Toc11964103)

[5.2 DSD Split 14](#_Toc11964104)

[5.3 Availability of SDMX Artefacts 15](#_Toc11964105)

[6 Annex 18](#_Toc11964107)

# 1 Version History

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Editor(s)** | **Remarks** |
| **1.0** | 07/02/2019 | UIS, OECD,Eurostat | Draft for public pilot |
| **1.1** | 21/06/2019 | UIS, OECD,Eurostat | Updates related to Education package v1.1 |
|  |  |  |  |

# Introduction

The aim of the SDMX implementation in the Education domain is to create a joint set of Data Structure Definitions (DSDs), harmonised concepts and code lists for use in joint UNESCO-UIS/OECD/EUROSTAT (UOE) data collection exercises.

Data collection on education at ESS level is based on [Regulation 912/2013](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2013.252.01.0005.01.ENG) and covers data from the 28 Member States, EFTA countries, Candidate and potential Candidate Countries.

Collection for OECD non-EU countries is governed by a gentleman's agreement.

The Manual "UOE data collection on formal education - Manual on concepts, definitions and classifications", introduces the concepts, definitions and classifications to be used for the UOE data collection. It sets up the conceptual and methodological framework of UOE data collection. The 2019 version of the Manual will be published on <http://uis.unesco.org/>.

UOE data are collected using 5 common questionnaires: "ENRL" – Enrolments, "ENTR" – Entrants, "PERS" – Personnel, "GRAD" – Graduates, "FIN" – Finance. Two questionnaires, "CLASS" – Class size and "DEM" – population, are used by OECD and UNESCO-UIS only. The questionnaires are embodied in Excel files which contain a number of worksheets with one 2-dimensional table for each sheet. The questionnaires are predefined Excel templates which are SDMX-compliant (i.e. SDMX codes are present in the template but not visible for data providers).

Since January 2017 there have been ongoing discussions among the international partners (Eurostat, UIS and OCED) on the possibility for data providers to use another format (either CSV or SDMX-ML) for data transmission. Using one of these formats would make it easier to automate some steps in the production process, such as the uploading of information to a database.

The SDMX sponsors (BIS, ECB, Eurostat, IMF, OECD, United Nations and World Bank) mandated UIS, Eurostat and OECD to propose Data Structure Definitions (DSDs) in order to implement SDMX in the Education domain based on the above-mentioned Regulation.

The SDMX implementation comprises in particular the definition of SDMX Data Structures for exchanging data between reporting countries and international organisations, and between these organisations. It also covers the maintenance of the DSDs over time, in accordance with the procedures agreed upon by the SDMX governance bodies.

The so-called “DSD Matrix” was used as an inventory for structuring the relevant data flows in the Education domain.

The DSD Matrix includes the list of concepts that are necessary to code the reporting requirements of the different SDMX sponsoring agencies. The codification of domain concepts in the DSDs has to be descriptive and comprehensive enough to fulfill any current and future reporting requirements within the scope of the education reporting framework.

This document will shortly describe the contents of a DSD. The aforementioned DSD Matrix will also be explained in more detail.

# Concepts and Code Lists

A Data Structure Definition (DSD) in SDMX consists of several statistical concepts: education level, field of education, institutional sector, expenditure type, etc. These concepts are broken down into dimensions and attributes. Dimensions are used to uniquely identify a certain data item or variable, and when joined together, they provide the “time series key” as the unique identifier for this item. In an SDMX dataset, a valid code must be assigned to each of the dimensions relevant for a data item, i.e. dimensions are mandatory. Attributes relating to, for example, compiling organisation, table identifier and number of decimals, are used to further describe the data. They can be attached at different levels of the data file: at the level of the whole file (attached at dataset level in SDMX terminology), at the level of a certain series (attached at series level), or at the level of single observations (attached at observation level). Attributes are either mandatory or conditional (i.e. they can be omitted).

Most concepts used in this DSD are coded, i.e. they are associated with code lists providing descriptors for the coded items. For example, the code list for the Education level contains an ordered set grouping education programmes in relation to gradations of learning experiences, as well as the knowledge, skills and competencies which each programme is designed to impart.

Code lists can be shared across concepts. For example, the same code list is used for identifying items of the reporting institutional sector and the counterpart institutional sector, as they both refer to the same list of sectors. All code lists in this package are presented as flat, non-hierarchical lists.

Some attributes are not coded but contain default values. This is the case for the REPYEARSTART "Reference year start" attribute used in every table (this attribute must contain a valid date value) and FIN\_STUD\_YEARSTART "School year start for FIN-STUD" and FIN\_STUD\_YEAREND "School year end for FIN-STUD" used in FIN questionnaire (these attributes must contain a valid date, with format YYYY-MM-DD).

Details on the concepts included, their role (dimension/attribute), the attribute attachment level, and further details on the concepts can be found in the matrix in the “**Concept Scheme**” sheet.

**Table 1** below provides a summary of the concepts used in the Education (UOE) data collection. As said earlier, each data item (observation) has to be identified by a combination of codes (for each dimension), which **uniquely** identify the relevant time series of observations within the reporting context.

##### Table 1: Concepts used in the SDMX UOE Education DSDs

|  |  |
| --- | --- |
| **Concept ID** | **Concept Definition** |
| EDUCATION\_TYPE | Determines whether education programmes cover formal initial and/or formal adult education |
| EDUCATION\_LEV | An ordered set which groups and classifies education programmes according to the knowledge, skills, competencies and qualifications which they are designed to impart |
| EDUCATION\_FIELD | Fields of education based on the ISCED 2013 international classification |
| GRADE | Indicates the specific stage, class or year of instruction within the ISCED level in initial education |
| INTENSITY | Intensity of participation |
| ORIGIN | Country of origin of statistical units |
| MOBILITY | Measures transfers between education systems (such transfers include the physical crossing of a border) |
| STAT\_UNIT | Entity for which information is sought and for which statistics are ultimately compiled |
| BREAKDOWN\_GROUP | Group of dimensions according to which the data are decomposed |
| TIME\_PERIOD | Timespan or point in time to which the observation actually refers |
| OBS\_VALUE | Value of a particular variable |
| FREQ | Time interval at which observations occur over a given time period |
| REF\_AREA | Country or geographic area to which the measured statistical phenomenon relates |
| REF\_SECTOR | Reference institutional sector |
| COUNTERPART\_SECTOR | Counterpart institutional sector |
| EXPENDITURE\_TYPE | Specifies the type of expenditure data whether direct, transfers, etc. |
| SEX | State of being male or female |
| AGE | Length of time that an entity has lived or existed |
| UNIT\_MEASURE | Unit in which the data values are expressed |
| TABLE\_IDENTIFIER | List of current data collection tables |
| REF\_YEAR\_AGES | Reference date for ages |
| ORIGIN\_CRITERION | Criterion used for country of origin |
| REPYEARSTART | Start of the reference period (e.g., start of school year or financial year) |
| REPYEAREND | End of reference period (e.g., end of school year or financial year) |
| FIN\_STUD\_YEARSTART | School year start for FIN-STUD (YYYY-MM-DD) |
| FIN\_STUD\_YEAREND | School year end for FIN-STUD (YYYY-MM-DD) |
| TYPE\_EXP\_REPORTED | Specifies whether the expenditure reported is actual or budget expenditure |
| OBS\_STATUS | Information on the quality of a value or an unusual or missing value |
| CONF\_STATUS | Information about the confidentiality status of the object to which this attribute is attached |
| COMMENT\_OBS | Comments to the observation value |
| DECIMALS | Number of digits of an observation to the right of a decimal point |
| TIME\_PER\_COLLECT | Dates or periods during which the observations have been collected for the target reference period |
| COMPILING\_ORG | Organisation collecting and/or elaborating the data being reported |
| UNIT\_MULT | Exponent in base 10 used for calculating the actual value in the unit of measure |

# The DSD Matrix file

The DSD Matrix file contains the following sheets:

* + - **Concept Scheme sheet**, listing the concepts used;
    - **DSD-Concept Matrix sheet**, showing the relationships between selected questionnaires from well-known transmission programmes on the one hand, and the concepts used on the other hand;
    - **Code list sheets**, showing the contents of each of the code lists used as well as some additional comments.

## 4.1 Concept Scheme Sheet

The Concept Scheme sheet summarises all concepts and code lists contained in the UOE reporting framework.

## DSD-Concept Matrix Sheet

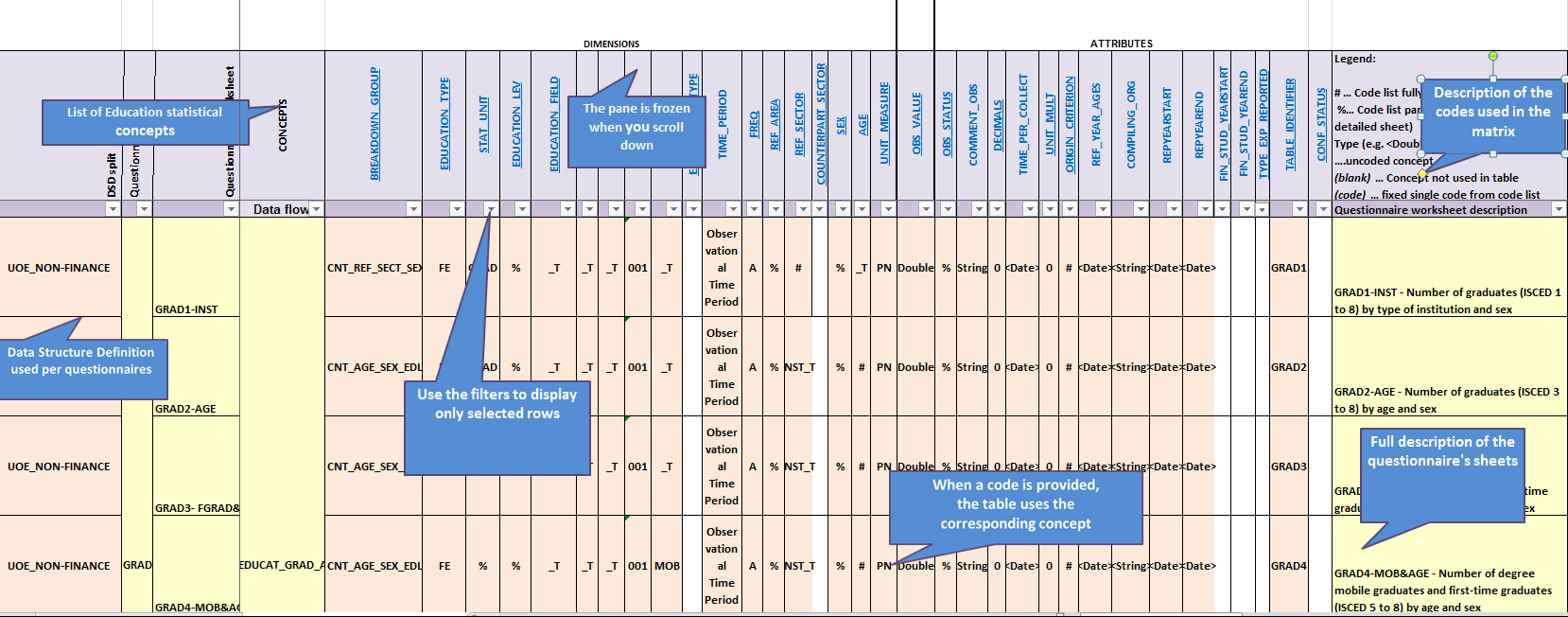
The list of concepts linked to all the tables is shown in the headline. Each coded concept has a hyperlink pointing to the corresponding code list sheet.

The cells are linking a transmission table (row) to a concept (column), and contain:

* + - a **#** sign if the code list associated to the concept is used in full;
    - a **%** sign if the code list associated to the concept is used only partially (in this case the code list is said to be "constrained");
    - a data type (e.g. <Double>, <Date>, <String>) if the concept is uncoded;
    - a ***code*** if the concept is fixed to a single code;
    - a ***Blank*** *if the* concept is not used in table.

**Example 1**: The cell at the intersection of FIN questionnaire🡪worksheet FIN1-SOURCE and concept STAT\_UNIT contains code EXP. This means that FIN questionnaire FIN🡪worksheet FIN1-SOURCE uses only code EXP (Expenditure) for the STAT\_UNIT (Statistical Unit) concept.

**Example 2**: The cell at the intersection of questionnaire ENRL🡪worksheet ENRL5-MOB&FIELD and concept MOBILITY contains a % sign. This means that several MOBILITY codes (MOB, MOB\_HOME) are linked to ENRL5-MOB&FIELD.



To see which dimensions are used in which DSD and which questionnaires are covered by the DSD, just select a DSD type in the "DSD Split" column filtering. Only tables using the selected data structure are displayed. The coloured columns correspond to the concepts used in the selected DSD.

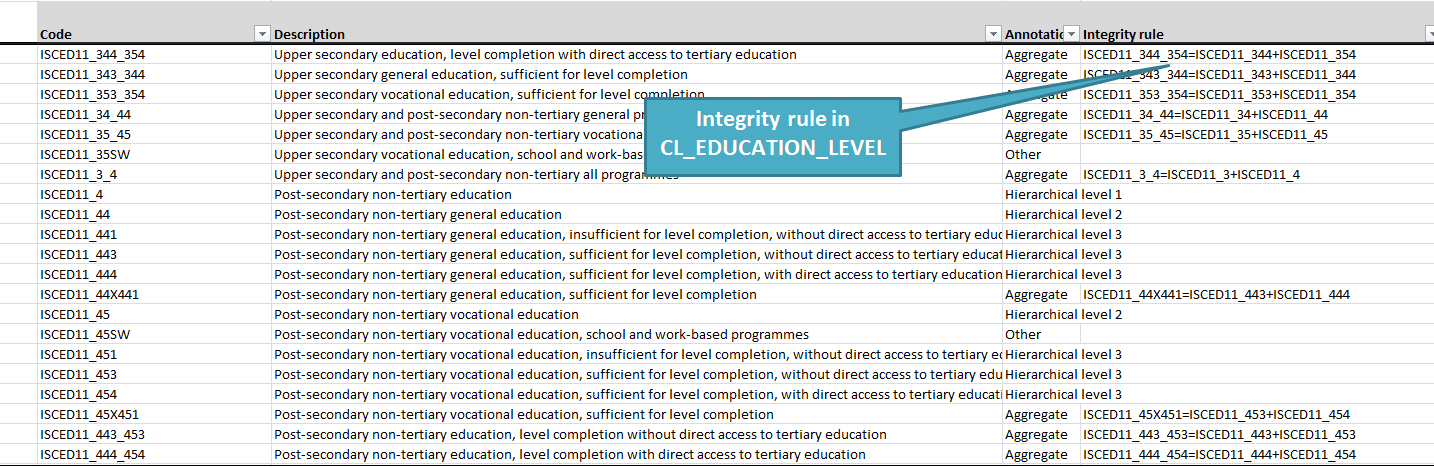
## 

## Code List Sheet

Each coded concept is associated to a code list as shown in the “**Concept Scheme**”. A distinct sheet is created for each code list:



Each code list contains the codes in the first column ("Code"), the labels in the second column ("Description") and, in some cases, additional comments in the third column ("Annotations"). In relevant cases, an additional column describing integrity rules can be added.



## Comparison with previous DSD Matrix

## Table 2 below provides an overview of the changes implemented in the new DSD Matrix (2019) compared to the 2014 Matrix.

## Table 2: Comparison with previous DSD Matrix

| **Concept or Code list  affected in new DSD Matrix** | **Changes compared to previous DSD Matrix** |
| --- | --- |
| EDUCATION\_TYPE | Concept EDU\_TYPE replaced with EDUCATION\_TYPE |
| CL\_EDUCATION\_TYPE | Code list CL\_EDU\_TYPE replaced with CL\_EDUCATION\_TYPE |
| EDUCATION\_LEV | Concepts ISCED11P\_LEVEL, ISC11P\_CAT, ISC11P\_SUB merged into new concept EDUCATION\_LEV |
| CL\_EDUCATION\_LEVEL | Code lists CL\_EDU\_LEVEL, CL\_ISC11P\_CAT and CL\_ISC11P\_SUB, merged into new code list CL\_EDUCATION\_LEVEL |
| EDUCATION\_FIELD | Concept FIELD replaced with EDUCATION\_FIELD |
| CL\_EDUCATION\_FIELD | Code list CL\_ISCED13\_FIELD replaced with CL\_EDUCATION\_FIELD |
| CL\_GRADE | * All codes in CL\_GRADE (e.g. 1, 2 ,3, 4 ….14) now prefixed with G (e.g.: G1, G2, G3, G4……G14) * Description of code \_T "Total: all grades" replaced with "Total" |
| CL\_INTENSITY | Description "Full- and part-time" for code \_T replaced with "Total" |
| ORIGIN | Concept COUNTRY\_ORIGIN replaced with ORIGIN |
| MOBILITY | Concept COUNTRY\_CITIZENSHIP replaced with MOBILITY |
| CL\_MOBILITY | Code list CL\_AREA attached to old concept COUNTRY\_CITIZENSHIP replaced with CL\_MOBILITY now attached to new concept MOBILITY |
| CL\_STAT\_UNIT | * Code AVG\_CLS deleted from code list STAT\_UNIT * Code PERS replaced with PS. * Code MAN replaced with MGMT * Codes BEG\_MINL, EXP\_MINL, HSAL\_HQUAL, EXP\_MQUAL moved to CL\_TEACH\_EXPER, but using new codifications * Codes CLS\_MULTIU, CLS\_SING, SCH, COMP, STAT\_COMPT, NRTEACH added by UIS |
| BREAKDOWN\_GROUP | * New concept |
| TABLE\_ID | * TABLE\_ID dimension transformed into attribute |
| CL\_SECTOR | * Code list CL\_SECTOR\_EDU replaced with CL\_SECTOR * code S1D "Total private" replaced with PRV\_NEDUC "Non-educational private sector" * code S1D\_ O "Other non-educational private entities" replaced with PRV\_NEDUC\_O "Other non-educational private entities" * Code OUT\_INST\_T "Outside educational institutions" replaced with \_TXINST "Sectors other than educational institutions" * Labels of codes S1311, S1312, S1313 aligned with National Accounts sector code list. Words "excluding social security" added to the label * Codes S1PE and GD\_SERV deleted |
| CL\_EXPENDITURE\_TYPE | * Codes DIR\_EXP, NET\_PAY and PAY merged into DIR\_EXP * Codes TRF, FAHS, TRF\_PAY and NET\_TRF merged into TRF * Codes RD\_HE, RD\_SEP, GRNT\_NTF, INT, PRIN deleted |
| REPYEARSTART/REPYEAREND FIN\_STUD\_YEARSTART /FIN\_STUD\_YEAREND | * Concepts REF\_PER\_START and FIN\_YEAR\_START merged into new concept REPYEARSTART * Concepts REF\_PER\_END and FIN\_YEAR\_END merged into new concept REPYEAREND * New concepts FIN\_STUD\_YEARSTART and FIN\_STUD\_YEAREND added in the Matrix. Concepts to be used when FIN-STUD is not aligned with financial year |
| CL\_OBS\_STATUS | * Code list CL\_OBS\_STATUS\_EDU replaced with SDMX cross-domain code list CL\_OBS\_STATUS   As a result, the following changes have been implemented in the questionnaires:   * Code M "Missing value" replaced with O "Missing value" * Code Z "Not available" replaced with M "Missing value; data cannot exist" * Code X "Data included in another category" replaced with K "Data included in another category" |
| UNIT\_MEASURE | CL\_UNIT\_EDU replaced with CL\_UNIT |
| CL\_AREA | IMF+CL\_AREA+1.0. replaced with SDMX+CL\_AREA+2.0 (which represents a combination of reference area codes in M49 and ISO-3611 classifications) |

## All aggregate codes for ORIGIN concept (e.g. Europe, Europe not specified, Asia, Asia not specified) have been replaced in GRAD and ENRL questionnaires, using the codes from the M49 classification.

## Table 3 below provides an overview of the changes implemented in the new GRAD and ENRL questionnaires for the aggregates associated to the ORIGIN concept.

## Table 3: Comparison with the previous codes for ORIGIN concept

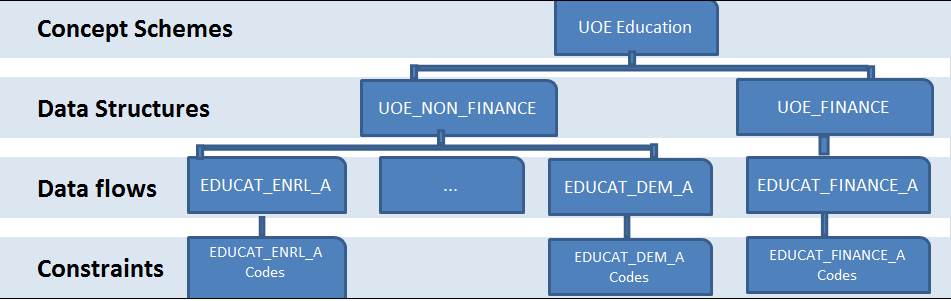
|  |  |  |
| --- | --- | --- |
| **Code description in the questionnaire** | **Previous code in GRAD and ENRL questionnaires for ORIGIN** | **New code in GRAD and ENRL questionnaires for ORIGIN according to M49 classification** |
| Africa not specified | F19 | 577 |
| Total: Africa | F1 | 002 |
| Northern America not specified | A29 | 293 |
| Total: Northern America | A2 | 021 |
| Latin America and the Caribbean not specified | A99 | 741 |
| Total: Latin America and the Caribbean | A9 | 419 |
| Asia not specified | S19 | 490 |
| Total: Asia | S1 | 006 |
| Europe not specified | E19 | 568 |
| Total: Europe | E1 | 007 |
| Oceania not specified | O39 | 527 |
| Total: Oceania | O3 | 009 |
| From abroad but country of origin not specified | W19 | 898 |
| Total: All countries of origin | W1 | 001 |

# SDMX Artefacts

The technical representations of the data structures and their related concepts and code lists are stored as so-called "SDMX artefacts" in the [SDMX Global Registry](https://registry.sdmx.org/FusionRegistry/). Once the SDMX reporting framework is used in production, the SDMX Global Registry serves as a central repository for all internationally agreed Data Structure Definitions and their related objects.

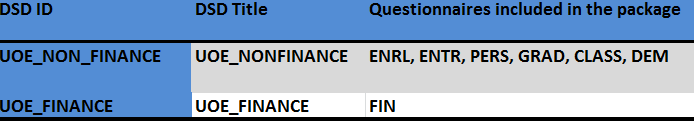
## 5.1 UOE Education implementation model

The approach in UOE Education was to start with the "Education Domain Concepts". In SDMX those are expressed in the Concept Scheme artefact from which the data structures have been derived. Data structures are then used to define multiple data flows. The relevant subset of codes applicable for a specific data exchange will be expressed using the Constraint artefact. The technical implementation model can be visualised as follows:



## 5.2 DSD Split

For optimal representation of the UOE Education reporting framework, the following Data Structure Definitions have been created by the technical group:



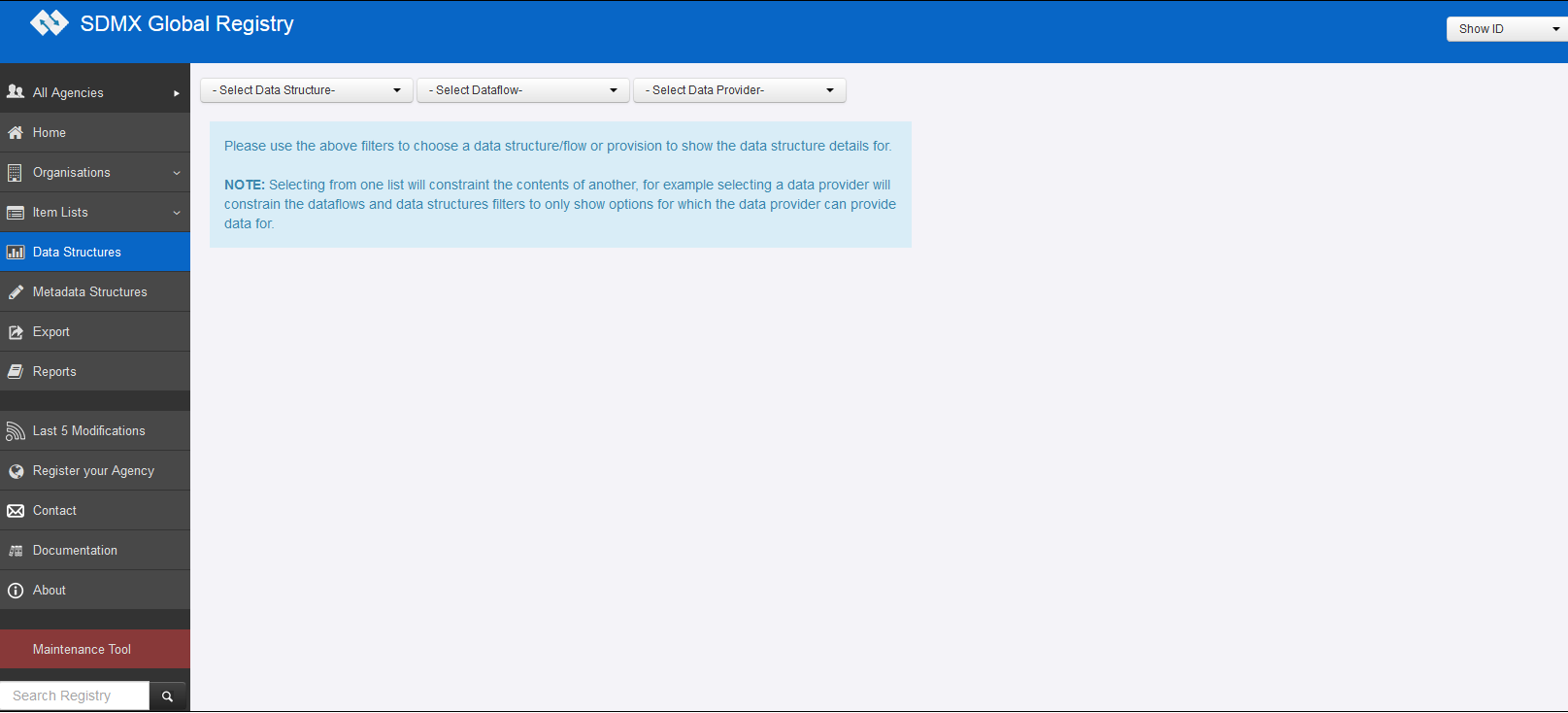
Further details on the dimensions linked to each DSD are shown in the above-mentioned matrix sheet .

## 5.3 Availability of SDMX Artefacts

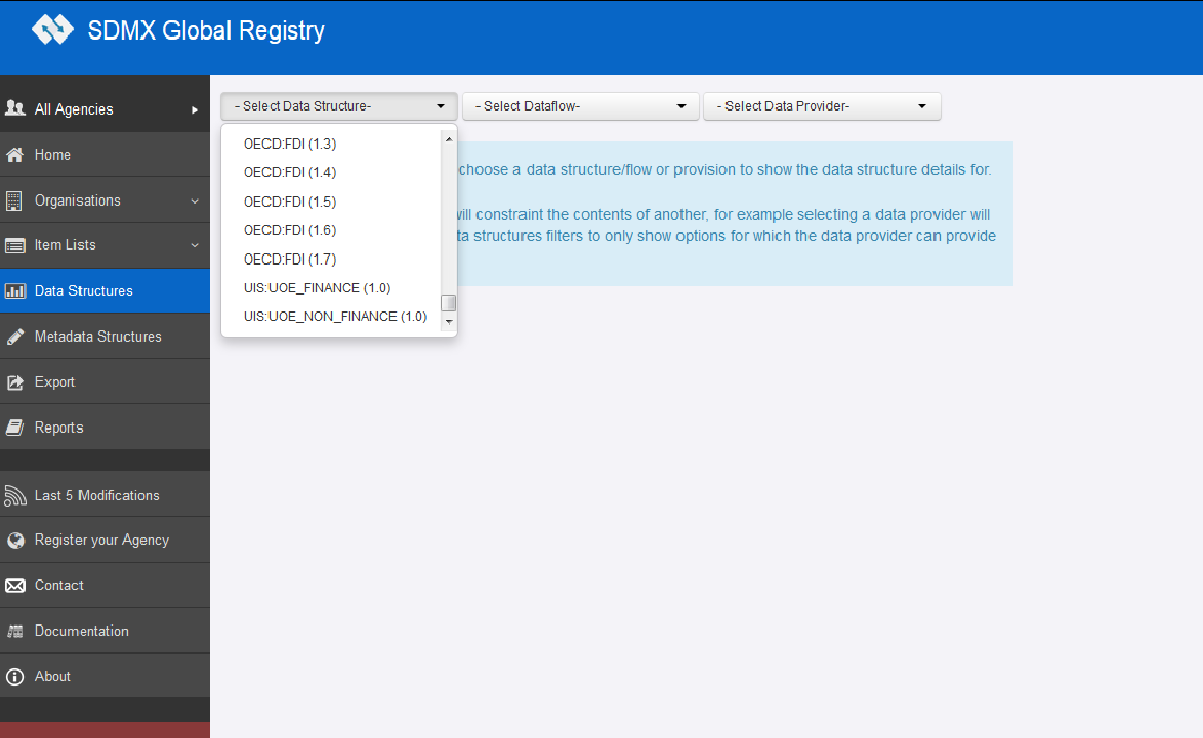
# The Data Structure Definitions and their related SDMX artefacts are available for download from the [SDMX Global Registry](https://registry.sdmx.org/FusionRegistry/).

The screenshots below show where to find the DSDs and related artefacts online and how to download them for local use.

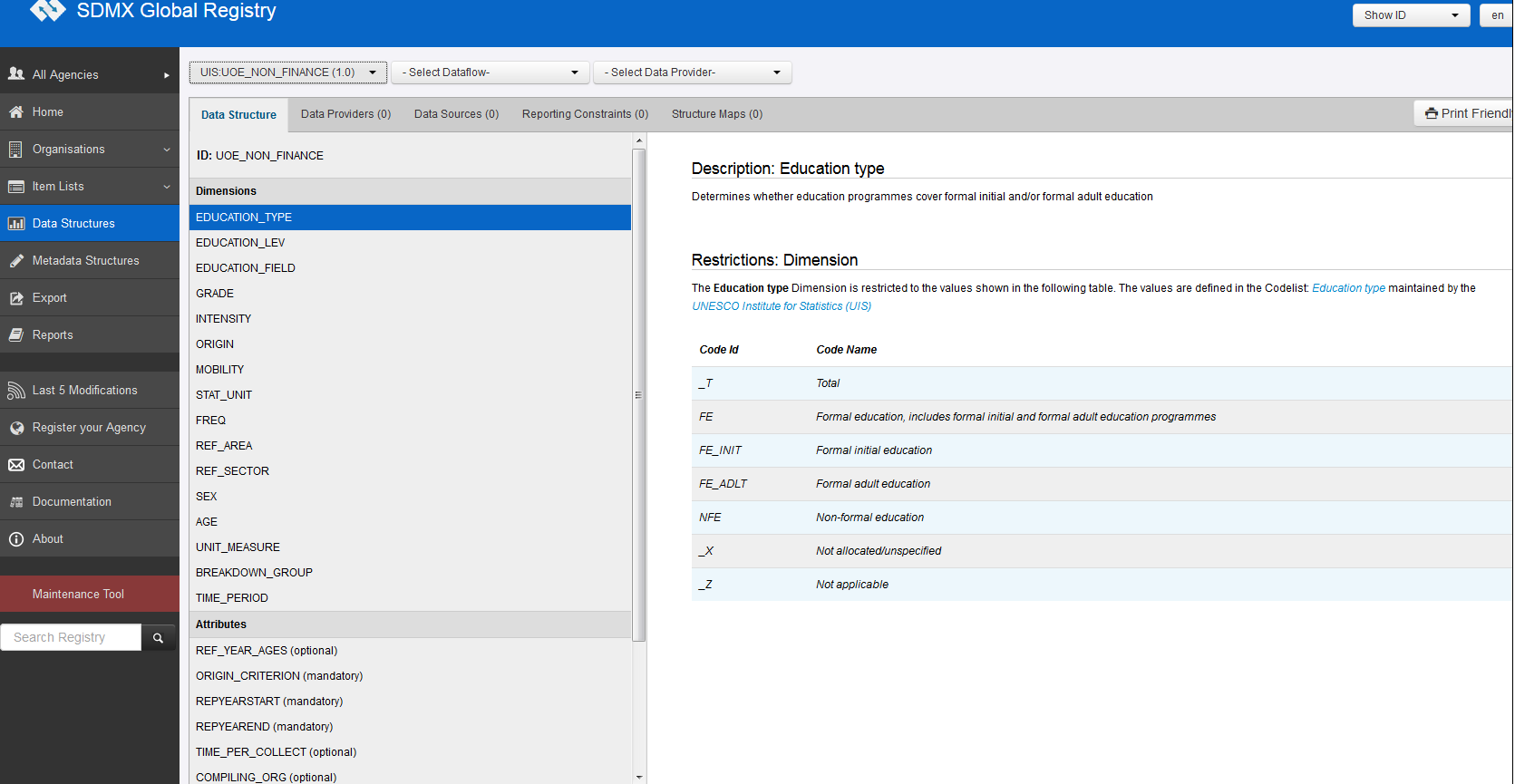
**TO VISUALISE THE ARTEFACTS ONLINE:**

****

Click here to view the DSD in the Registry

****

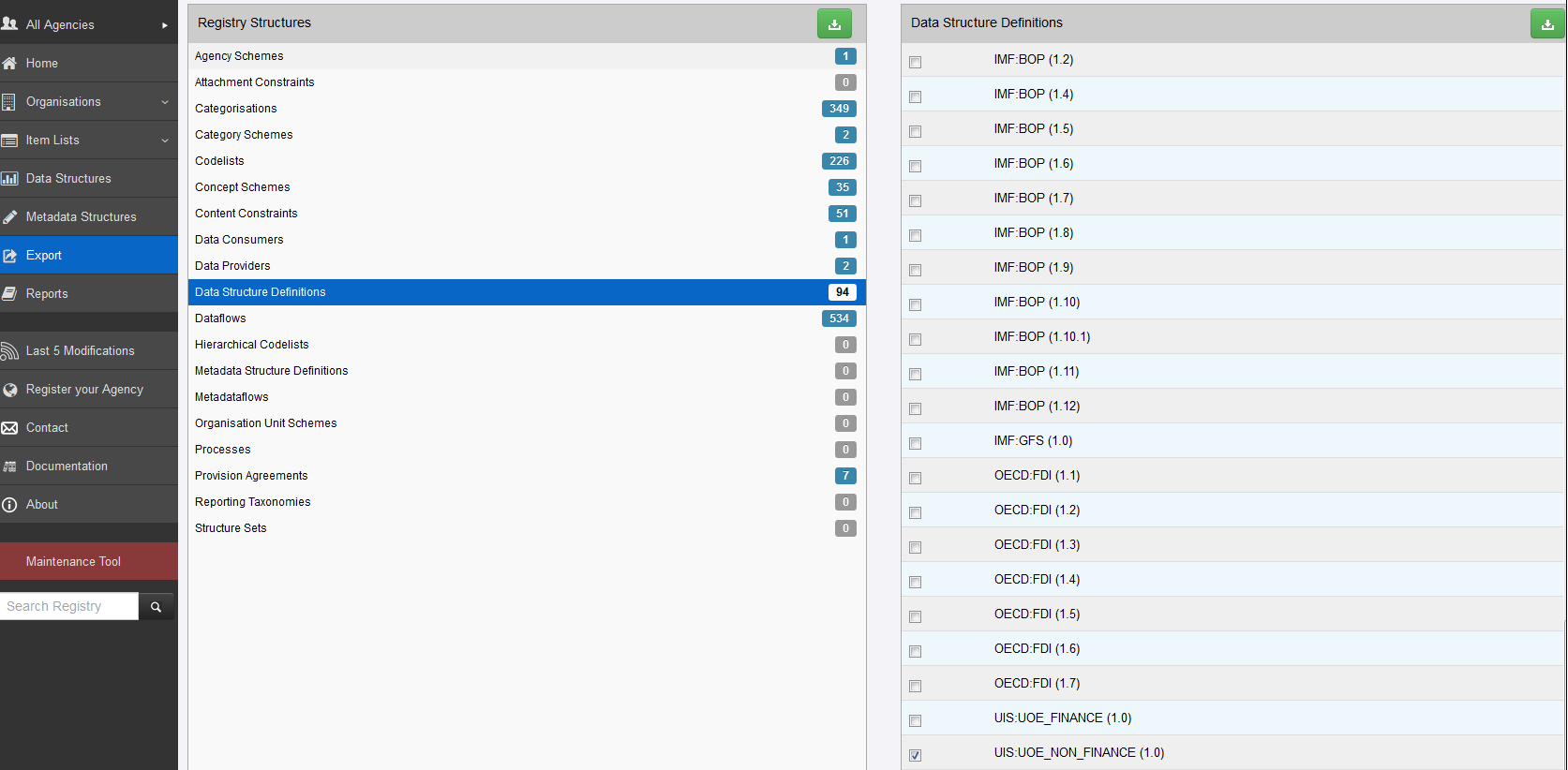
Select DSD from the dropdown list

****

Click on each attribute to see the DSD attributes

Click on each dimension to see the DSD dimensions

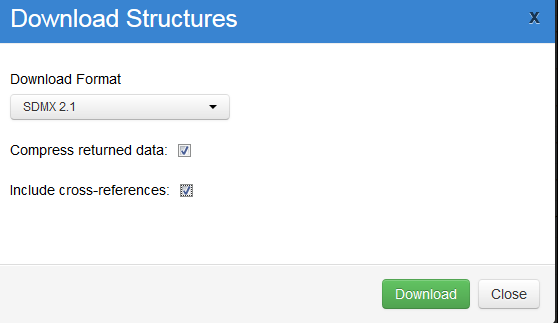
**TO DOWNLOAD THE ARTEFACTS:**

****

3. Click on the icon to download the selected structure (DSD)

1. Click on Export 🡪 Data Structure Definition to find the DSD for downloading

2. Tick this button to download the DSD UIS:UOE\_NON\_FINANCE(1.0)



Tick to include the code lists and concepts into the xml file

Select the SDMX version

Launch the download

# Annex

Three questionnaires (LANG Questionnaire, REGIO Questionnaire and CREDMOB Questionnaire) are specific to Eurostat, which means that the data exchange does not involve OECD nor UNESCO.

Table 4 below lists the concepts used in the Eurostat-specific data collection on education.

##### Table 4: Concepts used in the Eurostat DSDs for education

|  |  |
| --- | --- |
| **Concept ID** | **Concept Definition** |
| EDUCATION\_LEV | An ordered set which groups and classifies education programmes according to the knowledge, skills, competencies and qualifications which they are designed to impart |
| STAT\_UNIT | Entity for which information is sought and for which statistics are ultimately compiled |
| BREAKDOWN\_GROUP | Group of dimensions according to which the data are decomposed |
| TIME\_PERIOD | Timespan or point in time to which the observation actually refers |
| OBS\_VALUE | Value of a particular variable |
| FREQ | Time interval at which observations occur over a given time period |
| REF\_AREA | Country or geographic area to which the measured statistical phenomenon relates |
|
| SEX | State of being male or female |
| AGE | Length of time that an entity has lived or existed |
| UNIT\_MEASURE | Unit in which the data values are expressed |
| COUNTRY\_DESTINATION | Country of destination of statistical units |
| MOB\_SCHEME | Type of mobility scheme, i.e. under EU programmes (e.g. ERASMUS or other EU programmes), under other international/national programmes, other programmes; including/excluding degree mobile graduates |
| MOB\_TYPE | Type of mobility, i.e. study period and/or work placement |
| TABLE\_IDENTIFIER | List of current data collection tables |
| REF\_YEAR\_AGES | Reference date for ages |
| REPYEARSTART | Start of the reference period. For example, start of school year or financial year |
| REPYEAREND | End of reference period. For example, end of school year or financial year |
| OBS\_STATUS | Information on the quality of a value or an unusual or missing value |
| CONF\_STATUS | Information about the confidentiality status of the object to which this attribute is attached |
| COMMENT\_OBS | Comments to the observation value |
| DECIMALS | Number of digits of an observation to the right of a decimal point |
| TIME\_PER\_COLLECT | Dates or periods during which the observations have been collected for the target reference period |
| UNIT\_MULT | Exponent in base 10 used for calculating the actual value in the unit of measure |

Three additional concepts are used in Eurostat questionnaires but not in the common UOE questionnaires:

* COUNTRY\_DESTINATION
* MOB\_SCHEME
* MOB\_TYPE

## DSD Split for EUROSTAT data collection

For optimal representation of the EUROSTAT Education reporting framework, the following Data Structure Definitions have been created by the technical group:

* ESTAT+LANG+1.0
* ESTAT+REGIO+1.0
* ESTAT+CREDMOB+1.1

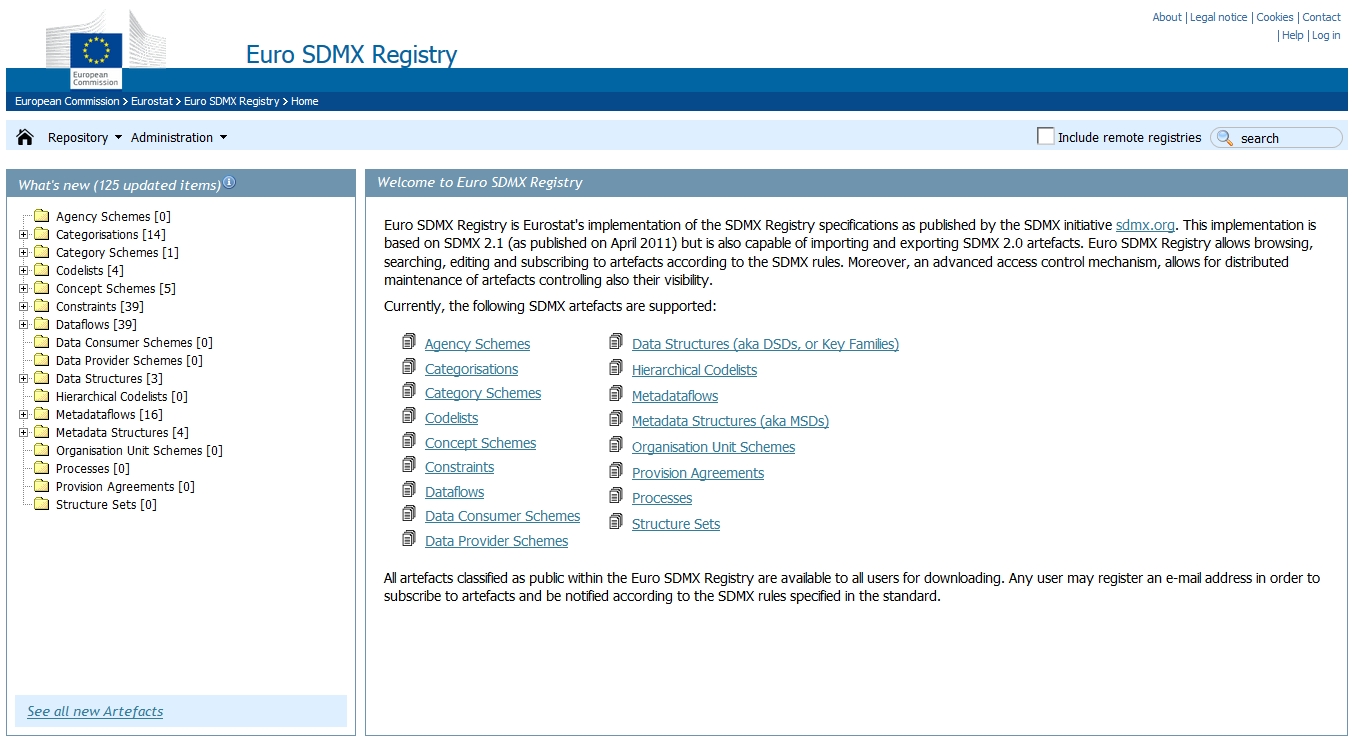
Further details on the dimensions linked to each DSD are shown in the matrix sheet.

## Availability of SDMX Artefacts for EUROSTAT data collection

# The Data Structure Definitions and their related SDMX artefacts are available for download from the [SDMX Euro Registry](https://webgate.ec.europa.eu/sdmxregistry).

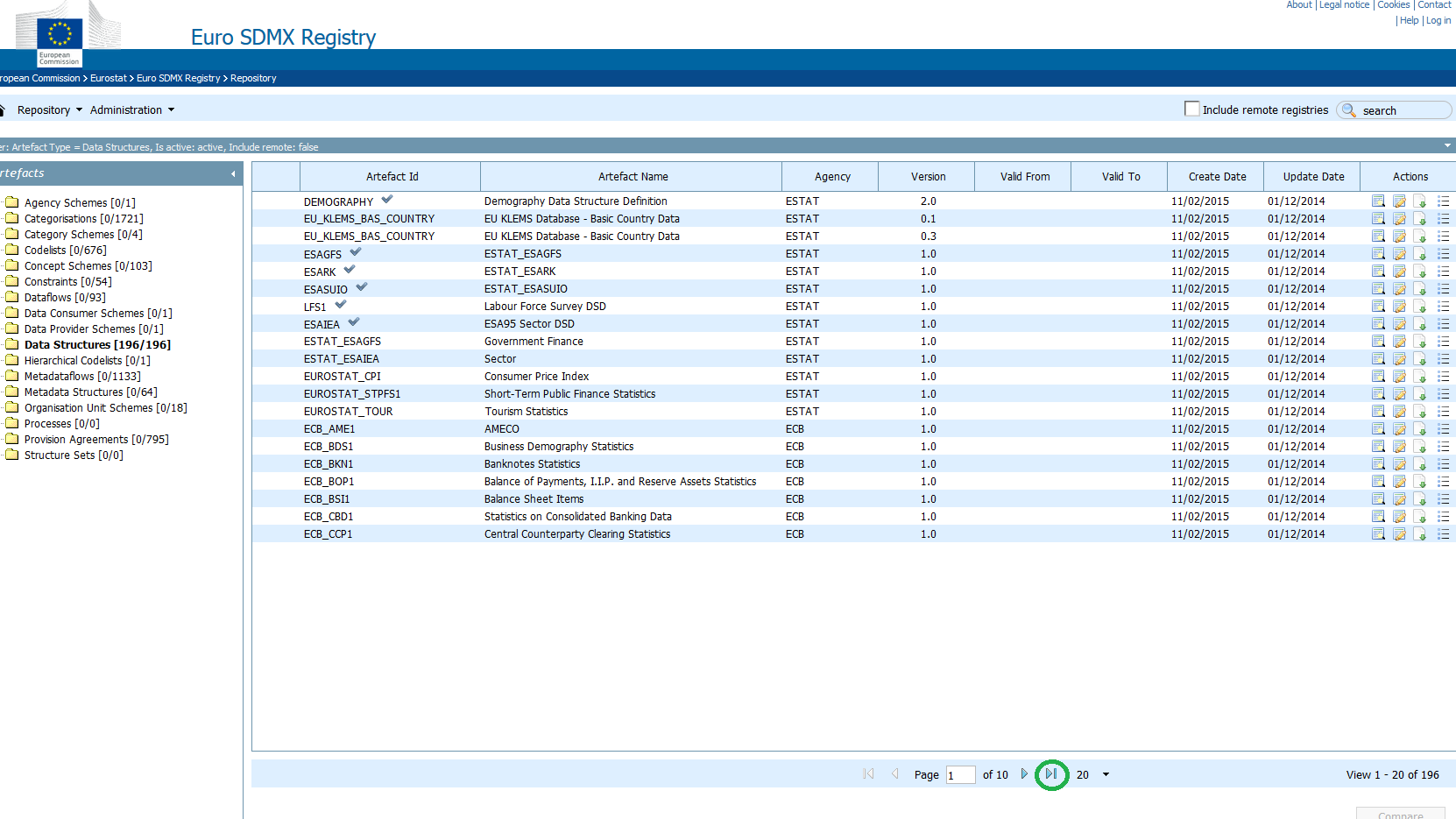
The screenshots below show how to find the DSDs and related artefacts online and how to download them for local use.

**TO VISUALISE THE ARTEFACTS ONLINE:**



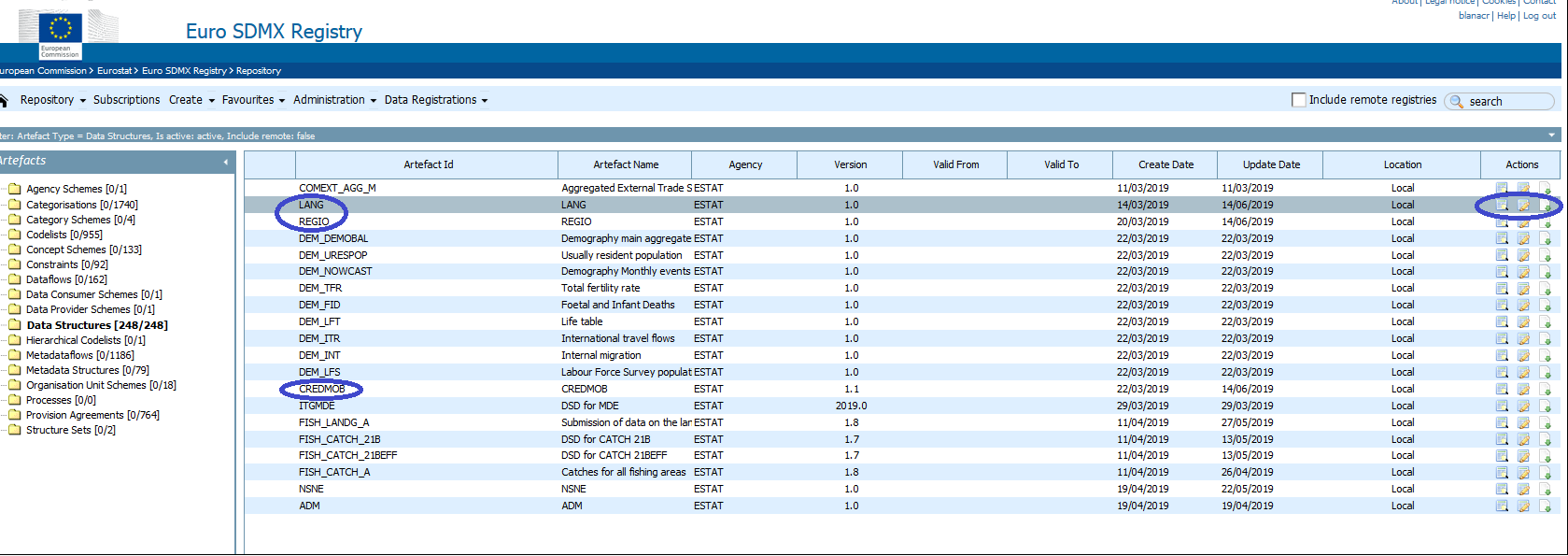
Click

All content in the Euro SDMX Registry

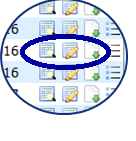


Click

**TO DOWNLOAD THE ARTEFACTS:**

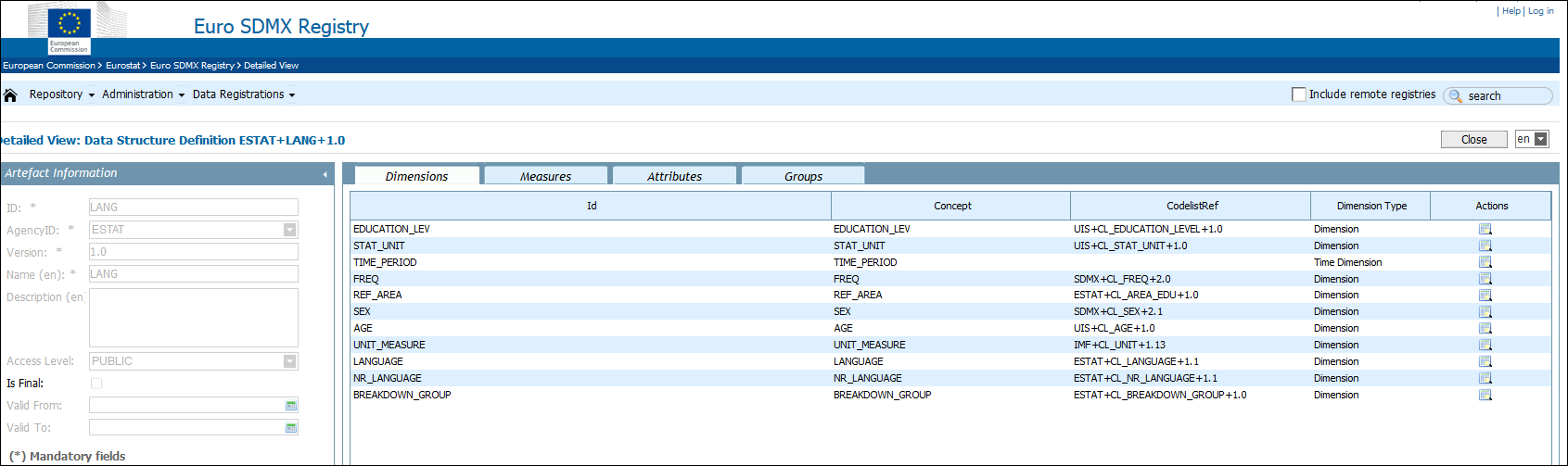


Detailed view

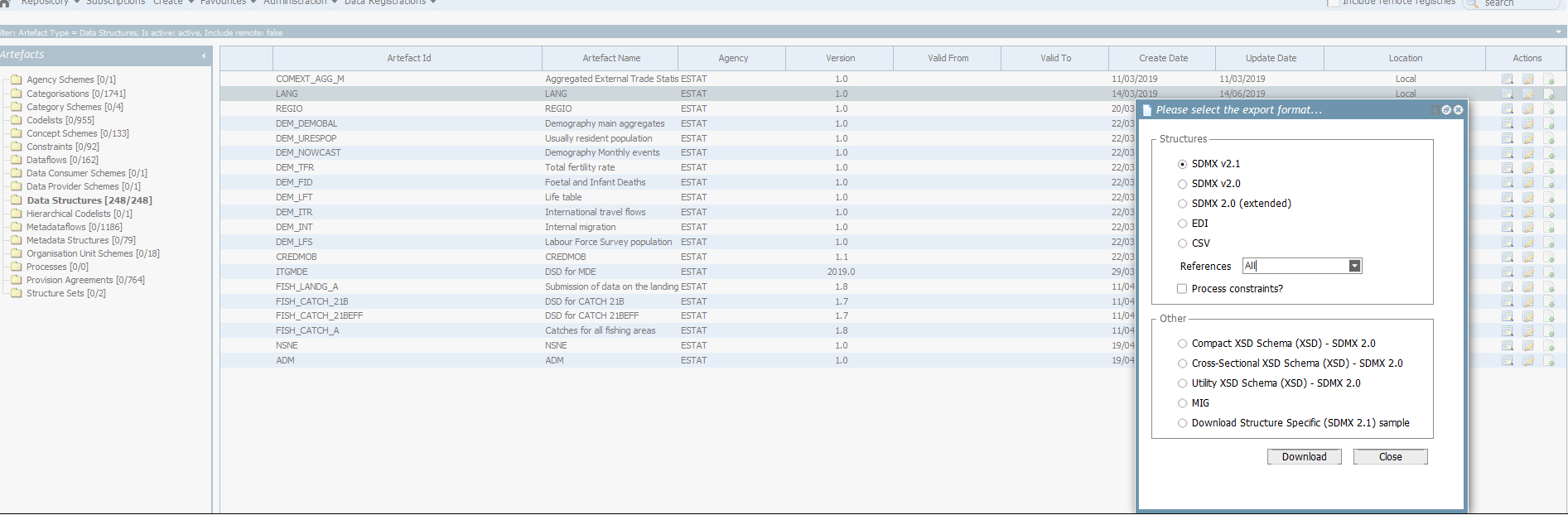


Download

Quick view

Detailed view of the DSD

Download of the DSD



Launch the download

Select SDMX v2.1 or SDMX 2.0 (extended)