International reporting – a fully automated process

Jacob Chr. Solander, Web & Statbank, jcs@dst.dk

Abstract

This paper describes how Statistics Denmark is using automated processes to ensure optimal workflow and a minimum of difficulties for the statistical divisions when sending data to Eurostat.

Like the other EU member states, Denmark and Statistics Denmark are obliged to send data series to Eurostat in SDMX format. Statistics Denmark has always emphasised that sending data or reports to Eurostat must not be difficult for the individual internal statistical divisions. That is why we at Statistics Denmark have always worked to ensure that there is a single source for our data, meaning that the statistical divisions only need to update their data in one single place. The single source is a database called "the Sum Database". Multiple fully automated and internally developed tools and services are using this database as their only source of statistical data. Some of these send data to Eurostat in the desired format using our existing dissemination tools and Statbank.dk. The Statbank API constitutes the basic tool for the data transfer from Statistics Denmark and will therefore be a significant focus point for future development areas.

This presentation will review the following:

- *Automatically generated SDMX files to Eurostat*
- *Saved queries from Statbank.dk to the IMF converted into SDMX format*
- *Queries for CENSUS 2021*
- *Use of API for the automatic processes*

Keywords: Eurostat, International reporting, online communication, IMF, CENSUS 2021

1. Introduction

One of the main goals from the early days of the development of Statbank Denmark is to store data only in one place and only once. The fact that data is stored in one place, from where you can extract data for any task – be it tables or newsletters, ensures consistency in the dissemination of our statistics. In other words, the source of the data is always the same. The users will experience this whether they are looking at a newsletter or a Statbank table.

Statbank Denmark

Statistics Denmark has a well-functioning Statbank based on the common Nordic data model. We currently store our data in Oracle databases and as a whole; we call the system "the Sum Database". It consists of both internally and externally available databases and a host of tools and applications. The Statbank is the public part of the Sum Database and is located in a separate database from where external users have access to see and extract published data. A part of the Sum Database is used for specific types of dissemination for example newsletters, and it contains data for the Statbank as well as for international reporting. The Statbank API founds the base for the transmission of data and has an internal component called a Sum API that works on the internal database.

The tables in Statbank Denmark

Statbank Denmark is the primary release channel for all data from Statistics Denmark. The work with tables for the Statbank is divided into two. A fixed part and a variable part. The fixed part consists of the creation of tables and maintenance of metadata done by the Web and Statbank team. The variable part consists of updating the tables with new data. The statistical divisions are responsible for the variable part including the preparation of data and discretion of data before publication. The official release time at Statistics Denmark is daily at 08.00. All data is released and available to everyone at the same time.

Loading data

The statistical divisions need to load an SDF file (SumDataFile) or a table in Oracle to upload data to the Sum Database. An SDF file is a text file consisting of rows with data enriched with variable values, which represent the codes used to display metadata in the tables.

Image 01 (BCSPVIM - SDF file)

There are three ways of loading data to the Sum Database:

- **SumTools** is an internal program to create and load data. From where the statistical divisions can search for their tables and load the SDF file.

- **SumInitializer** is a batch function where it is possible for the statistical divisions to load data without using SumTools. The batch function has the advantage that many data sets can be loaded simultaneously from a local folder.

- **FOTO** (**F**rom **O**racle **T**o **O**racle) is a functionality supported both by SumTools and by SumInitializer. It enables the statistical divisions to load data from their Oracle database tables directly into the Sum Database, without having to create a separate SDF file.

All data sets that are loaded into the Sum Database are set to a future release date that is one or several days ahead in time. If metadata needs to be updated, it will always happen the following day, not before and not after. Therefore, it is only for the data sets that we have the option to set a future release date.

Using the tables from the Statbank

In addition to Statbank Denmark being Statistics Denmark's online database where everyone can find our published data, the Statbank is also used to form tables and graphs for our website www.dst.dk. For example, on all our subject pages, the tables and figures are generated directly from the Statbank using saved queries (identified by a so-called "PXS ID"). This helps us to ensure that there is full consistency in our dissemination of data as stated previously.

2. Tables for international reporting

Many of my colleagues in the statistical divisions are to a greater or lesser degree used to working with Statbank tables and SDF files. That is why we decided to use the Statbank for creating data sets for international reporting.

The tables for international reporting are created just like any other table in the Statbank. Loading data is done the precise same way as all other tables. In order to distinguish the tables in Statbank from each other, the ordinary tables have an organisation code ("DST") and the tables for Eurostat have a different organisation code ("DSTEU").

The challenge with the tables for Eurostat is that the metadata and codes are defined in Eurostat's DSD matrix and are not necessarily the same as those we use for the general Statbank tables. Therefore, the tables for Eurostat must be created from scratch.

That is why we have a separate statistical subject called "International reporting" in the Statbank, which can only be accessed with a specific and strictly internal user profile. We do not want to mix the tables for international reporting with the general tables in the Statbank.

From the internal user profile, almost 100 saved queries (identified by the "PXS ID" as previously mentioned) have been generated and stored, each of which covers a data set to be sent to Eurostat.

DSD matrix

Once we have received the DSD matrix from Eurostat, we can start to create the table that will form the basis for the data set to Eurostat. The basic structure of the table is the same as the general Statbank tables. However, the table construction differs based on the number of variables and attributes in the data set as well as the use of codes and naming of variable headings that must be complied with.

Case: Sending the data set BCS_PVI_M to Eurostat

This case will be a short presentation of how the data set BCS_PVI_M to Eurostat is formed.

We want to create a table with as few variables as possible to make it simpler for the statistical divisions to create data sets. At the same time, we also have to make the table future-proof, meaning that it needs to be able to accommodate any necessary metadata updates. A new base year can be added to the table, without having to delete the table and start over.

From the DSD, we can see that the data set BCS_PVI_M has 14 concepts. The concepts contain both variables and attributes.

Image 02 (DSD Matrix for BCS_PVI_M)

Below we have made a table BCSPVIM in the Statbank containing five variables: ACTIVITY, SEASONAL_ADJUST, BASE_PER, CONF_STATUS and OBS_STATUS.

Image 03 (Table BCSPVIM from Statbank Denmark)

The underlying codes and associated texts for the five variables are the official codes, which can be found in Eurostat's DSD. In image 04 you can see how we make metadata for the Statbank table BCSPVIM from the DSD and in image 01 you can see how the SDF file for BCSPVIM looks like. From the table BCSPVIM we can make saved queries to use for generating an SDMX file to Eurostat.

Image 04 (BCSPVIM - Metadata)

Kolonne 1: CL_ACTIVITY (VM3_ACTIVITY_BCS_PVI)

Værdikode	Værditekst (bruges *ikke* i datasættet!)
B	Mining and quarrying
B05	Mining of coal and lignite
B06	Extraction of crude petroleum and natural gas
B07	Mining of metal ores
B08	Other mining and quarrying
B09	Mining support service activities
BTD	Mining and quarrying; manufacturing; electricity, gas, steam and air conditioning supply
B_C	Mining and quarrying; manufacturing
C	Manufacturing
C10	Manufacture of food products
C11	Manufacture of beverages
C12	Manufacture of tobacco products
C13	Manufacture of textiles
C14	Manufacture of wearing apparel
C15	Manufacture of leather and related products
C16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
C17	Manufacture of paper and paper products
C18	Printing and reproduction of recorded media
C19	Manufacture of coke and refined petroleum products
C20	Manufacture of chemicals and chemical products
C21	Manufacture of basic pharmaceutical products and pharmaceutical preparations
C22	Manufacture of rubber and plastic products
C23	Manufacture of other non-metallic mineral products
C24	Manufacture of basic metals
C25	Manufacture of fabricated metal products, except machinery and equipment
C26	Manufacture of computer, electronic and optical products
C27	Manufacture of electrical equipment
C28	Manufacture of machinery and equipment n.e.c.
C29	Manufacture of motor vehicles, trailers and semi-trailers
C30	Manufacture of other transport equipment
C31	Manufacture of furniture
C32	Other manufacturing
C33	Repair and installation of machinery and equipment
D	Electricity, gas, steam and air conditioning supply
D35	Electricity, gas, steam and air conditioning supply
MIG_CAG	MIG - capital goods
MIG_DCOG	MIG - durable consumer goods
MIG_ING	MIG - intermediate goods
MIG_NDCOG	MIG - non-durable consumer goods
MIG_NRGXE	MIG - energy (except section E)

Kolonne 2: SEASONAL_ADJUST (VM3_ADJUSTMENT_BCS)

Værdikode	Værditekst (bruges *ikke* i datasættet!)
N	Neither seasonally nor working day adjusted
W	Working day adjusted, not seasonally adjusted
Y	Working day and seasonally adjusted

Kolonne 3: CL_BASE_PER (VM3_BASE_YEAR_BCS)

Værdikode	Værditekst (bruges *ikke* i datasættet!)
0000	Unspecified
2010	2010
2015	2015
2020	2020
ABS0	Absolute value

Kolonne 4: FL_CONF_STATUS (VM3_CONF_STATUS_BCS)

Værdikode	Værditekst (bruges *ikke* i datasættet!)
C	Confidential statistical information
F	Free
N	Not for publication, restricted for internal use only

Kolonne 5: FL_OBS_STATUS (VM3_OBS_STATUS_BCS)

Værdikode	Værditekst (bruges *ikke* i datasættet!)
A	Normal
E	Estimated value
P	Provisional value
U	Low reliability

Creating an SDMX file

From our point of view, there are two ways of creating an SDMX file:

- Using Eurostat's converter, available on Eurostat's website. The converter requires a CSV file with all the concepts from the DSD. The CSV file is uploaded to the converter and when the SDMX files is created, you can manually upload it to Eurostat through Edamis4.

- Using saved queries from the Statbank and our internally developed application "SdmxMerge" that automatically uploads the SDMX file to Edamis4. This application provides a minimum of work for the statistical divisions, as they only need to form an SDF file and nothing else to be able to send their data to Eurostat.

Converter

The converter from Eurostat is not used for all the data sets sent to Eurostat. Internal studies at Statistics Denmark have shown that the vast majority of statistical divisions have their data stored in an Oracle database. The use of excel files or csv files is not that common for data sets. Data is often stored in the database from where the SDF files are created directly.

SdmxMerge

SdmxMerge is an internal application and only used by the Web and Statbank team members, not in the statistical divisions. The application's main purpose is to create an XML file based on a saved query from the Statbank and to merge the DSD attributes into the XML file by using the PXS ID to retrieve data from the saved query.

It requires some work from our side in Web and Statbank team to modify a saved query and load it into the SdmxMerge application, but once everything has been set up correctly and Eurostat has validated the SDMX file, it remains unchanged unless Eurostat makes corrections.

The next step in the process for BCS_PVI_M is to use a saved query from the table BCSPVIM (see image 05). The PXS ID is modified and "*" is manually added to "Value1" on the individual variables. This indicates that only the data that has been

updated in the Sum Database is to be sent to Eurostat and not the entire data set each time. This option makes it more flexible for the statistical divisions to decide how many time periods they want to send to Eurostat. Once the saved query has been formed, it is not modified until Eurostat has essential updates for the data set.

SdmxMerge uses the PXS ID to transcode the data into an SDMX file and the attributes are added to the XML file. There are seven attributes in this case: REF_AREA, INDICATOR, PRODUCT, TRANSFORMATION, DECIMALS, UNIT_MEASURE and UNIT_MULT (see image 06). The XML files can now be transformed into an SDMX file.

Image 05 (PXS-ID from BCSPVIM)   Image 06 (SdmxMergeEditor)


```
\\srvsbprod2\pxs$\225519.pxs
[query]
TimeOpt=0
Subject=34
Table=BCSPVIM
Varant=5
SubTab=S0
Heading=5
TimeOrder=5
[Contents]
Cont1=Indhold
[Var0]
Kod=FL_OBS_STATUS
Aggreg=no
Prestext=text
Elim=
Value1=*
[Var1]
Kod=FL_CONF_STATUS
Aggreg=no
Prestext=text
Elim=
Value1=*
[Var2]
Kod=CL_BASE_PER
Aggreg=no
Prestext=text
Elim=
Value1=*
[Var3]
Kod=SEASONAL_ADJUST
Aggreg=no
Prestext=text
Elim=
Value1=*
[Var4]
Kod=CL_ACTIVITY
Aggreg=no
Prestext=text
Elim=
Value1=*
[Time]
Kod=Tid
Aggreg=no
Value1=2021-05

Gem
```

```xml
SdmxMergeEditor - SavedQuery:225519
Dokument  Indsæt  XML  SDMX  Handlinger  Indstillinger  Om...
<?xml version="1.0" encoding="utf-8"?>
<SdmxMerge xmlns="http://api.statbank.dk/xmlns/sdmxmerge">
  <Transcode>
    <Obs>
      <Concept name="Tid">
        <Name to="TIME_PERIOD" />
      </Concept>
    </Obs>
  </Transcode>
  <Merge>
    <KeyFamily>
      <Id>STSALL</Id>
      <AgencyId>ESTAT</AgencyId>
      <Version>1.0</Version>
    </KeyFamily>
    <Header>
      <Id>BCS</Id>
      <Truncated>true</Truncated>
      <Senders>
        <PartyTypeBean index="0">
          <Id>DK1</Id>
        </PartyTypeBean>
      </Senders>
      <Receivers>
        <PartyTypeBean>
          <Id>4D0</Id>
          <Names>
            <TextTypeBean>
              <Text>EUROSTAT</Text>
            </TextTypeBean>
          </Names>
        </PartyTypeBean>
      </Receivers>
      <DataSetAgency>ESTAT</DataSetAgency>
      <DataSetID>BCS_PVI_M</DataSetID>
    </Header>
    <Series>
      <SeriesKeys>
        <Concept name="REF_AREA" value="DK" />
        <Concept name="INDICATOR" value="PROD" />
        <Concept name="PRODUCT" value="_Z" />
        <Concept name="TRANSFORMATION" value="N" />
        <Concept name="DECIMALS" value="1" />
        <Concept name="UNIT_MEASURE" value="IX" />
        <Concept name="UNIT_MULT" value="0" />
      </SeriesKeys>
    </Series>
  </Merge>
  <Actions>
    <InsertFreq />
    <InsertTimeFormat />
  </Actions>
</SdmxMerge>
```

Image 07 (The complete SDMX file – ready to be sent)

```xml
<?xml version="1.0" encoding="UTF-8"?>
<CompactData xsi:schemaLocation="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/message SDMXMessage.xsd urn:sdmx:org.sdmx.infomodel.keyfamily.KeyFamily=ESTAT:STSALL(1.0):compact ESTAT_STSALL_Compact.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:sts="urn:sdmx:org.sdmx.infomodel.keyfamily.KeyFamily=ESTAT:STSALL(1.0):compact" xmlns="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/message">
  <Header>
    <ID>BCS_20220509092222_cb64b06b-6d23-4b6f-936d-659c3ff9c176</ID>
    <Test>false</Test>
    <Truncated>true</Truncated>
    <Prepared>2022-05-09T09:22:22+02:00</Prepared>
    <Sender id="DK1">
      <Name xml:lang="en">Statistics Denmark</Name>
      <Contact>
        <Name xml:lang="en">Bo Henry Eriksen</Name>
        <Telephone>+45 3917 3552</Telephone>
        <Email>bhe@dst.dk</Email>
      </Contact>
    </Sender>
    <Receiver id="4D0">
      <Name>EUROSTAT</Name>
    </Receiver>
    <DataSetAgency>ESTAT</DataSetAgency>
    <DataSetID>BCS_PVI_M</DataSetID>
    <Extracted>2022-05-09T09:22:22+02:00</Extracted>
  </Header>
  <sts:DataSet>
    <sts:Series TIME_FORMAT="P1M" FREQ="M" UNIT_MULT="0" UNIT_MEASURE="IX" DECIMALS="1" TRANSFORMATION="N" PRODUCT="_Z" INDICATOR="PROD" REF_AREA="DK" ACTIVITY="B" SEASONAL_ADJUST="N" BASE_PER="2015">
      <sts:Obs CONF_STATUS="F" OBS_STATUS="A" OBS_VALUE="29.2" TIME_PERIOD="2021-05"/>
      <sts:Obs CONF_STATUS="F" OBS_STATUS="A" OBS_VALUE="44.5" TIME_PERIOD="2021-06"/>
      <sts:Obs CONF_STATUS="F" OBS_STATUS="A" OBS_VALUE="44.5" TIME_PERIOD="2021-07"/>
      <sts:Obs CONF_STATUS="F" OBS_STATUS="A" OBS_VALUE="45.9" TIME_PERIOD="2021-08"/>
      <sts:Obs CONF_STATUS="F" OBS_STATUS="A" OBS_VALUE="43.1" TIME_PERIOD="2021-09"/>
```

The above SDMX file is the final result of the transcoding of a saved query from the table BCSPVIM.

Now this file can be sent manually to Eurostat through Edamis4, but Statistics Denmark has chosen to go one step further and make this process fully automated.

Automated transmission of data

The whole idea of a fully automated process for international reporting is that the statistical divisions only have to concentrate on loading new or updated data and validating that the data is correctly loaded into the Sum Database. This is exactly what they do with general tables in the Statbank so the learning curve for international reporting is extremely limited for the statistical divisions.

To further simplify the process of sending data sets to Eurostat, a service has been set up to check daily whether tables with the organisation code DSTEU have been updated with new data. If so, the SdmxMerge process is activated autonomously and the SDMX file is created. At 09:40, the SDMX file is placed in a specific folder (called "FilesToEurostat"). At 10:00, the SDMX file is automatically uploaded to EDAMIS4 by sFtp using an application called Edamis4Communicator. Statistics Denmark's official release time is at 08:00, but historical reasons transmissions to Eurostat are set to 10:00.

The fully automatic process runs every day, but it is also possible to use the system manually. A few of my colleagues in the Web and Statbank team have access to a separate application that can generate the SDMX file after the official release time at 08:00. It only requires a saved query (PXS ID) to be inserted into the application and the whole process is started. The SDMX file is now automatically placed in a folder

called SdmxBatch and the file needs to be manually moved to the folder FilesToEurostat. The service is set up to check FilesToEurostat every hour from 10:00 until 17:00 and again at 21:00 and send new files to EuroStat. For example, if a file is uploaded to the folder FilesToEurostat after lunch at 12:20, it is automatically sent at 13:00.

OECD – Monthly SDMX file

The aforementioned application that makes it possible to start the SDMX creation process manually is also used to generate an SDMX file for the OECD. Each month, the Balance of Payments uploads a data set to the table BOPOECD in the Sum Database in the same way as all other general Statbank tables. After the release at 08:00 on the scheduled release date, we manually create an SDMX file from BOPOECD and send it to the statistical divisions and they send it to the OECD.

IMF - SDMX files for SDDS plus

Like many other countries, Denmark is obliged to send data to the SDDS plus data categories. Again, the main idea is that the statistical divisions need to have a minimum or a non-burden to meet and comply with the requirements for data to SDDS plus. On our website www.dst.dk/imf we present both an SDMX file and a link to human-readable data. The link to the human-readable data is solved by saved queries (PXS ID) from already existing Statbank tables and these links are inserted on the website. The link for the Balance of Payments is www.statbank.dk/pxs/229785. The reason we use existing Statbank tables for the IMF is that the data series we have to send already exists in our Statbank tables.

The SDMX files for the IMF is generated in the same way as the files for Eurostat. The centre of the process is a saved query (PXS ID), and typically, the same PXS ID is used for both the readable file and the SDMX file. We use the SdmxMerge application again to transcode the saved queries into an SDMX file. In this case, the PXS ID is the same (229785).

Some challenges in relation to codes and texts arise when saved queries from existing tables in the Statbank are used. The variable headings and codes cannot be used as they are, as they are not named as specified in the Concepts Schemes from the IMF.

We need to transcode the existing codes into codes that can be used for the SDMX file.

The SdmxMerge application has functionality to transcode the existing codes to the ones from Concepts Schemes and merge them into the final SDMX file. Image 08 shows how the values from the PXS ID: 229785 from the Statbank table BBM are transcoded into the correct concept names. Link for the SDMX file for the Balance of Payments is https://api.statbank.dk/v1/s1/saved/229785/SdmxCompact?lang=en and image 09 shows the SDMX file for the IMF.

Image 08 (Transcode and Replace Concepts)

```xml
<Transcode>
  <DataSet>
    <Concept name="COMMENT_DSET">
      <Name to="COMMENT" />
    </Concept>
  </DataSet>
  <Series>
    <Concept name="INDUDBOP">
      <Name to="ACCOUNTING_ENTRY" />
      <Value from="D" to="D" />
      <Value from="K" to="C" />
      <Value from="N" to="B" />
    </Concept>
    <ReplaceConcepts>
      <From>
        <Concept name="LAND" value="W1" />
        <Concept name="POST" value="1" />
        <Concept name="SÆSON" value="1" />
        <Concept name="ENHED" value="93" />
      </From>
      <To>
        <Concept name="INT_ACC_ITEM" value="CA" />
      </To>
    </ReplaceConcepts>
    <ReplaceConcepts>
      <From>
        <Concept name="LAND" value="W1" />
        <Concept name="POST" value="1.A" />
        <Concept name="SÆSON" value="1" />
        <Concept name="ENHED" value="93" />
      </From>
      <To>
        <Concept name="INT_ACC_ITEM" value="GS" />
      </To>
    </ReplaceConcepts>
    <ReplaceConcepts>
      <From>
        <Concept name="LAND" value="W1" />
        <Concept name="POST" value="1.A.A" />
        <Concept name="SÆSON" value="1" />
        <Concept name="ENHED" value="93" />
      </From>
      <To>
        <Concept name="INT_ACC_ITEM" value="G" />
      </To>
    </ReplaceConcepts>
    <ReplaceConcepts>
```

Image 09 (The SDMX file for the Balance of payments to the IMF)

```xml
<?xml version="1.0" encoding="UTF-8"?>
- <CompactData xsi:schemaLocation="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/message SDMXMessage.xsd urn:sdmx:org.sdmx.infomodel.keyfamily.KeyFamily=IMF:BOP(1.1):compact IMF_BOP_Compact.xsd"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:bop="urn:sdmx:org.sdmx.infomodel.keyfamily.KeyFamily=IMF:BOP(1.1):compact" xmlns="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/message">
  - <Header>
      <ID>SDDS_20220509101358_944ec0c2-30b1-4d5c-8360-d060d4ac469a</ID>
      <Test>false</Test>
      <Prepared>2022-05-09T10:13:58+02:00</Prepared>
    - <Sender id="DK1">
        <Name xml:lang="en">Statistics Denmark</Name>
      - <Contact>
          <Name xml:lang="en">Mads Møller Liedig</Name>
          <Department xml:lang="en">Balance of Payment</Department>
          <Telephone>+45 3917 3478</Telephone>
          <Email>mml@dst.dk</Email>
        </Contact>
      - <Contact>
          <Name xml:lang="en">Jacob Chr. Solander</Name>
          <Department xml:lang="en">Communication</Department>
          <Telephone>+45 3917 3158</Telephone>
          <Email>jcs@dst.dk</Email>
        </Contact>
      </Sender>
      <Receiver id="1C0"/>
      <KeyFamilyRef>ECOFIN_DSD</KeyFamilyRef>
      <KeyFamilyAgency>IMF</KeyFamilyAgency>
      <Extracted>2022-05-09T10:13:58+02:00</Extracted>
    </Header>
  - <bop:DataSet>
    - <bop:Series TIME_FORMAT="P1M" INT_ACC_ITEM="CA" FREQ="M" TIME_PER_COLLECT="E" COMP_METHOD="N" VALUATION="_X" CURRENCY_DENOM="_T" UNIT_MEASURE="XDC" MATURITY="_Z" INSTR_ASSET="_Z" FUNCTIONAL_CAT="_Z"
        FLOW_STOCK_ENTRY="T" COUNTERPART_SECTOR="_Z" REF_SECTOR="S1" ADJUSTMENT="_Z" REF_AREA="DK" COUNTERPART_AREA="_Z" DECIMALS="0" UNIT_MULT="6" ACCOUNTING_ENTRY="C">
        <bop:Obs CONF_STATUS="F" OBS_STATUS="A" OBS_VALUE="74395.0" TIME_PERIOD="2005-05"/>
        <bop:Obs CONF_STATUS="F" OBS_STATUS="A" OBS_VALUE="80645.5" TIME_PERIOD="2005-06"/>
        <bop:Obs CONF_STATUS="F" OBS_STATUS="A" OBS_VALUE="76454.5" TIME_PERIOD="2005-07"/>
        <bop:Obs CONF_STATUS="F" OBS_STATUS="A" OBS_VALUE="78268.9" TIME_PERIOD="2005-08"/>
```

Census2021

For Census2021, a number of hypercubes are to be created, which in Denmark's case will be stored in our Sum Database. From the Sum Database, we can create saved queries and use them to create the mapping required by Eurostat. As with the tables for Eurostat, we create separate metadata for the Census2021 tables and use SumTools to upload data. A special subject is also created in the Sum Database where all the tables for the hypercubes are located. I am responsible for updating metadata, and the statistical division for Population is responsible for the data part and for uploading the data sets to the tables.

3. The Challenges

The main challenges are that even though the statistical divisions experience a low burden when sending data to Eurostat, setting up the dataflow for the SDMX files involves a lot of work for the Web and Statbank team. Reading a DSD matrix can be difficult even after having seen and read many of them. In order to be able to come up with a good presentation for the statistical divisions on what their data set should look like to make an SDMX file, we need to have some knowledge of the individual statistical areas to be able to understand the data sets and what challenges they might have.

As we primarily send data to Eurostat from the Sum Database and only the Web and Statbank team can create metadata for the tables, we can ensure that metadata are reused for the individual statistical areas. In the case with the table BCSPVIM, which is part of many new tables in the BCS area, I needed to have the full overview of the

DSD matrix to guide the statistical divisions on which tables to create first and who in the individual statistical divisions to give access to upload data to the tables.

4. The future focus areas

Some areas that we will have focus on in the future are validation and reuse of metadata and documentation. More specifically, how to improve the validation of the data loaded via SumTools into the Sum Database. Alternatively, Eurostat could make local validation available in an easy way!

A more ambitious dream of ours is to look into whether the DSD matrix can be used to generate metadata for the Eurostat tables automatically.

Data validation

When loading data into the Sum Database via SumTools there are currently only a few general, but important validation checks of codes and data.

In order to load data to the Sum Database, the release date needs to be +1 day (the day after or later). If a data set has already been loaded for a future date, the system asks if data should be overwritten. It is not possible to load an earlier date than the one already used for release, but a later date can be used.

The time format of the SDF file is validated to match the table's metadata. It is not possible to load a quarterly time format into a table with a monthly time format.

If data is loaded via SumTools, the delimiter used in the data set (semicolon or tab) is validated. The decimal separator in the data set must be a comma (,) and not a period (.). In relation to Eurostat terminology, we can call these validations STRUVAL (Structural validation), but we do not check for CONVAL (Content validation), that is whether there can be or must be negative data, zeros or dots (..) in the data set. The system does not check if there are outliers in the data set. It is only when data has been sent to Eurostat that the data set becomes subject to CONVAL.

All the new SDMX files to Eurostat are uploaded for validation through Eurostat's test portal Acceptance before they are sent into production through Edamis4.

The SDMX files for the IMF are only sent for validation to the IMF once via the IMF SDMX Central, when SDMX Central has validated the link, it is inserted into our website at www.dst.dk/imf

Metadata and documentation

Metadata and statistical documentation is also an area where we want there to be single source. The statistical divisions update their statistical documentation in a program called Colectica.

The program has a template for each statistics with a number of fields to be filled in. The same template is used for statistical documentation to send to Eurostat.

The IMF use a similar template, but with a few additional information fields in the template. I am currently working on a task, together with the division for Methodology and Analysis, to upload the template directly to the IMF. We want the template with the statistical documentation to be converted in the SDMX format, so it will be easy to upload to the IMF. I hope that we will succeed in this by the end of the year.

Again, our focus is on single source, as the statistics divisions only have to fill in one template for their documentation, and the completed template is used for different end users.

5. Discussion

What is the broader future perspective with the SdmxMerge program?

The program has been developed internally at Statistics Denmark, we have had the program for several years, and it has now become an integral part of creating SDMX files. Statistics Denmark gives high priority to minimising the burden for the individual statistical divisions, especially when it comes to sending data to Eurostat. Therefore, the program has a future and it will be further developed. A future development could be rewriting the program into a web-based program to make it easier for programmers to maintain.

Can other statistical organisations use the program?

The answer is no, unfortunately! The way the system is made, makes it impossible to use out-of-the-box for other organisations beside Statistics Denmark. The program is tailored to Statistics Denmark and it is based on saved queries from our Statbank and is designed to use our input programs. However, it could be investigated whether it is possible to generalise the system so other statistical organisations can benefit from a fully automated process.

The common Nordic data model

Statbank Denmark is based on the common Nordic data model however, we have not used the data model 2.3 and the possibilities that comes with it, so far. The Web and Statbank team are working together with our IT on a big project to rewrite and update SumTools and as a part of that project we will look into the use of some of the components from data model 2.3. Maybe in the future, the Nordic data model will play a greater role in our program SumTools and in the Statbank Denmark.

In any event, I hope that this presentation has given inspiration to how sending data to international reporting can be made easy.