

Time Use Survey data processing and dissemination: - Development of a database with Time Use microdata

Grant Agreement No. 07141.2013.001-2013.553,

Amendment No 1 to Grant Agreement No. 07141.2013.001-2013.553

Reply to Request for Amendment No 2 to Grant Agreement No 07141.2013.001-2013.553

Objectives

The contract was signed on 13 November 2013. The objective of the project was to build a harmonised microdata database for the 2010 wave of the Time Use surveys and document the data quality. The database includes 18 countries. The participating countries are: Austria, Belgium, Finland, Germany, Estonia, Greece, Spain, France, Hungary, Italy, Luxembourg, Netherlands, Norway, Poland, Romania, Serbia, Turkey and United Kingdom.

Updating the detailed specifications of the contents and format of the files

The detailed specifications of the contents and format of the database developed during the 2000 wave were reviewed and updated. The draft specifications for Individual and household variables, Episode file variables, Activity classification and Metadata information were sent to Eurostat on 27 February 2014 by email (list of the documents in Annex 1).

Agreement with Eurostat on the contents and format of the database to be delivered

The updated detailed specifications of the contents and format of the database were discussed and agreed with Eurostat at the meeting in Luxembourg on 4 March 2014 (Minutes in Annex 2).

It was agreed that only data from the 2010 wave will be added to the database and it is not complemented with missing data from the 2000 wave because Eurostat saw no use for the data from the 2000 wave.

It was decided that the database will comprise three SAS data files:

- 1) Individual and household information file (person as the unit),
- 2) Diary day file (day as the unit), and
- 3) Episode file (day as the unit).

It was decided that instead of compiling Word and PDF documents (as in the Grant Agreement), metadata and quality reports are collected online with the European Statistical System Metadata Handler (ESS-MH) tool.

Development of the database and loading of a pilot country

After the meeting with Eurostat, the final (draft) detailed specifications of the contents and format of the files were produced. Meta information concerning the applied statistical methods needed for understanding the structure of the sample design, estimation procedures and data collection were defined. (List of the documents in Annex 1.)

File formats were tested using Finland as the pilot country. The background variables in the Individual and household information file were aggregated using the same procedure that was used in the 2000 wave.

The pilot data files were sent to Eurostat via eDAMIS on 30 September 2014. The specifications for the Pilot Individual and household information file, the Diary day file and the Episode file, Questions and answers about the codes, the Control tables, and the Meta-information document were sent by email (list of the documents in Annex 1).

Agreement with the NSIs delivering data to the multinational database on the detailed specifications of the contents and format of the files to be delivered

Discussions with the NSIs delivering data to the multinational database were held at the Working Group meeting in Luxembourg on 17 July 2014 (Minutes in Annex 3).

The format, types and structure of the files were approved as proposed by Statistics Finland.

The variables for “country of birth and citizenship” were specified.

It was agreed that the variables on the start time of the first activity in the diary and the end time of the last activity will be provided to Finland but their quality will need to be well documented.

It was concluded that, whenever possible, the variables on computer and Internet use will be provided and coded separately.

Both main and secondary activities will be collected using 3-digit codes.

The proposed 144 “with whom” variables with 32 categories were decided to be replaced by separate 144 binary variables for each of the six “with whom” diary columns.

The proposed diary activities’ codes were accepted.

Delivering the detailed specifications of the contents and format of the files

After the Working Group meeting, the Guidelines were revised according to the decisions of the meeting. The final Guidelines were transmitted to the countries via email on 6 August 2014. The Guidelines were loaded in CIRCABC by Eurostat (list of the documents in Annex 1).

Reception of national data files

Data transmission was done via eDAMIS.

| Country | The first version of the national files received | | Data format |
|---------|--|------|-------------|
| Finland | September | 2014 | SAS |
| Serbia | October | 2014 | SPSS |
| Spain | October | 2014 | SAS |
| France | January | 2015 | SAS |
| Estonia | January | 2015 | SPSS |

| | | | |
|----------------|-----------|------|------|
| Hungary | February | 2015 | SPSS |
| Italy | March | 2015 | SAS |
| Romania | March | 2015 | SPSS |
| Greece | July | 2015 | CSV |
| Norway | November | 2015 | SAS |
| Germany | December | 2015 | SAS |
| Luxembourg | April | 2016 | SAS |
| Austria | May | 2016 | SPSS |
| Netherlands | July | 2016 | POR |
| Belgium | September | 2016 | SPSS |
| Poland | September | 2016 | CSV |
| Turkey | October | 2016 | SAS |
| United Kingdom | October | 2016 | SPSS |

Data checking and harmonisation of data where necessary

The SPSS data files of eight countries and the CSV data files of two countries were converted into SAS format. The files transmitted by the countries were checked with a SAS program designed for checking permissible variable values (e.g. missing values) and lists of errors were printed out. If the checking revealed deviations from recommendations, these were corrected in cooperation with the NSIs concerned.

Consistency of the individual and household information files and the diary day files were checked by comparing the frequencies of the variables between countries

The background variables in the Individual and household information file were aggregated using the same procedure that was used in the 2000 wave (Annex 4).

The 3-digit primary and secondary activity HETUS codes and location codes were converted into the aggregate classification at Statistics Finland (Annex 10). The time use estimates were compared between countries based on the control tables.

The phases of the data checking process:

- 1) Receiving the data files via the eDAMIS system.
- 2) Saving the data files to the production folder (own folder for each country).
- 3) Conversion of the files if another program than SAS had been used.
- 4) Producing frequencies from the Individual and household information file and the Diary day file.
- 5) Comparing the frequency distributions with other countries and commenting.
- 6) Viewing the files and commenting the Excel file.
- 7) Checking specified variables (Life cycle, Day of the week, Net monthly income band, "With whom" variables when the main activity is sleeping).

- 8) Preparing the checking program for the country in question.
- 9) Running the checking program (the SAS program with macros).
- 10) Viewing the error lists saved in the working folder (own folder for each country).
- 11) Sending an email to the country with questions about errors and confusions.
- 12) Checking the corrections made by the country and possible additional queries.
- 13) Asking for classification of activities from the country in question.
- 14) Reviewing the classification of activities answers. Standardisation of the classification of activities with the country in question.
- 15) Aggregation and harmonisation of primary and secondary activities and location, and the individual and household variables. Saving the data files in SAS 32 bit format.
- 16) Producing three control tables:
 1. Time use of population by life cycle and sex,
 2. Location of population by sex, and
 3. Time use of population by age.
- 17) Comparing time-use estimates between countries:
 1. Women in a relationship with a youngest child aged zero to six years,
 2. Location of population, both sexes,
 3. Time use of population, both sexes,
 4. Time use of population, aged 10 to 14 years
- 18) Sending the control tables in Excel format to the NSIs for approval.

The individual and household information file (INDFILE) should contain only information of members who have kept diaries. Six countries had extra observations. They were removed from the files. The episode file (EFILE) should contain the same number of observations as the Diary day file (DDFILE). The number of observations in the files of three countries did not match. They were corrected in cooperation with the countries in question.

The number of received file versions varied according to the following:

INDFILE: 1 to 9 versions

DDFILE: 1 to 6 versions

EFILE: 1 to 4 versions.

All in all, the number of file versions delivered by the countries varied from 4 to 19.

The DDFILE and INDFILE had different kinds of missing values that were required. Often these were misunderstood: there was -1 instead of -9, or -6 was used with other missing values. If there was no other mistakes than these in the missing codes, Statistics Finland asked the country for

permission to make corrections and checked the file again before aggregation.

If national activity codes were used in the EFILE instead of the HETUS codes, the country had to correct the file. Wrong or missing codes were often corrected by Statistics Finland after asking the country for permission.

Checking the variable INC3 Lifecycle was made by cross tabulation of the variables INC3, INC2 Age and HHC2 Age of the youngest person in the household. Many countries had to correct the variable INC3 after checking.

Stepwise delivery of an updated multinational database to Eurostat

Harmonised SAS (32 bit) data files were delivered stepwise after the data validation with the documentation in English to Eurostat. The data transmission was done via eDAMIS. The data files were sent to Eurostat as datasets TUS_EFILE_O, TUS_DDFILE_O, TUS_INDFILE_O by eDAMIS.

In addition to data files, descriptions of deviations in the variables were sent to Eurostat as Excel files with the eDAMIS system. Check tables and the Questions about diary codes, as well as documents concerning the activity classifications were sent to Eurostat by email.

| Country | Delivery of the first version of harmonised files to Eurostat | |
|----------------|--|------|
| Finland | October | 2014 |
| Spain | February | 2015 |
| France | May | 2015 |
| Serbia | June | 2015 |
| Estonia | July | 2015 |
| Romania | June | 2015 |
| Italy | June | 2015 |
| Hungary | September | 2015 |
| Greece | April | 2016 |
| Germany | June | 2016 |
| Luxembourg | November | 2016 |
| Poland | December | 2016 |
| Austria | January | 2017 |
| Norway | February | 2017 |
| Turkey | February | 2017 |
| Netherlands | February | 2017 |
| Belgium | February | 2017 |
| United Kingdom | February | 2017 |

Problems in comparability

Background variables

Most of the deviations were included in the following variables. Several countries did not ask them, or there were deviations in the categories.

- Child care variables

- Type of accommodation
- How many rooms
- Appliances
- Income variables
- Help variables
- Age (population)
- Lifecycle
- Fixed start and finishing time at work
- Usual weekly working hours in all second jobs
- Looking for work
- Level of education currently receiving and completed
- Have a chronic physical or mental health condition
- How often feel rushed
- Children under 18, contact
- Car or motorbike driving licence
- Country of birth
- Country of citizenship
- Cohabiting
- Spouse variables

Diary day variables

- Number of diary days
- Did diarist feel rushed on the diary day
- What kind of day was the diary day
- Starting time of the first activity (valid data from only three countries)
- Ending time of the last activity (valid data from only three countries)

Time use variables

Main activity

- Lunch break
- Unspecified study
- Unspecified household and family care
- Other or unspecified making and care for textiles
- Tending domestic animals
- Other or unspecified gardening and pet care
- Other help of a dependent adult household member
- Help to a non-dependent adult household member
- Help in employment and farming
- Care of own children living in another household
- Other childcare as help to another household
- Socializing with family
- Collecting
- Travel related to changing locality
- Binary code (main activity)

Secondary activity

- Lunch break
- Unspecified study
- Unspecified household and family care

- Other or unspecified making and care for textiles
- Tending domestic animals
- Other or unspecified gardening and pet care
- Other help of a dependent adult household member
- Help to a non-dependent adult household member
- Help in employment and farming
- Care of own children living in another household
- Other childcare as help to another household
- Collecting
- Travel related to changing locality
- Binary code (secondary activity)

Location

- Weekend home or holiday apartment
- Workplace or school
- Other people's home
- Restaurant, cafe or pub
- Shopping centres, markets, other shops
- Hotel guest house, camping site
- Travelling by moped, motorcycle or motorboat

With whom

- Parent, living in the household
- Household member up to 9 years

Summaries of variable deviations can be found in Annexes 4 to 6.

Other issues

Some countries (Germany, Turkey) provided several background variables in the INDFILE and the DDFILE grouped for data confidentiality reasons.

Management of the quality of the data collection and processing, including the drafting of the quality report for the harmonised multinational Time Use data, based on the ESS Standard for quality reports structure

For the structure of the metadata, Statistics Finland studied different reporting recommendations for data collection and sampling designs. Based on these, instructions for the content of the metadata were compiled and delivered to Eurostat for commenting.

The metadata of Finland's pilot data was compiled in accordance with the EURO-SDMX Metadata Structure (release 3, March 2009) format and delivered to Eurostat in September 2014.

Eurostat created ESS-MH workflows for the participating countries. Statistics Finland was granted domain manager rights in the tool that allowed editing and summarising the national reports.

Checking of the microdata started when the countries filled the ESS-MH workflow or when sufficient descriptions of the research and sampling design had been received in some other manner.

The ESS-MH reports were checked and checks were requested for missing sections. Necessary quality descriptions were not received from several countries. One country interrupted filling in the ESS-MH and no completion was received despite several requests. Even though microdata had been delivered, the ESS-MH was not filled in, and the process only started after several requests.

Greece delivered the quality report according to the SDMX model as a Word file. Greece was asked to transfer the data into an ESS-MH report and correct the inaccuracies in the document. Eventually, Statistics Finland corrected the errors and the corrected version was delivered to Eurostat.

The following procedures were carried out for all countries:

- nonresponse checking
- calculating quality measures/indicators of diary data
- analyses of weighting.

Germany, the Netherlands, Austria, Hungary, Italy and Spain have deviated from the two days of diary keeping recommended by the HETUS Guidelines. The weights of these countries are day-specific weights. Instructions of the use of these weights are given in ESS-MH. By following the weighting correctly, the allocation of the diaries and the number of diaries do not affect comparability.

A summary of the metadata information in Annex 7 as a PDF copy.

Meetings

Two meetings with Eurostat were held. The first meeting was held in Luxembourg in March 2014, where specification of the contents and file formats were discussed and agreed (Minutes in Annex 2).

The second meeting with Eurostat was held in Luxembourg in January 2016, where it was agreed that Eurostat merges the data files that Statistics Finland sends to Eurostat with a country code (Minutes in Annex 8).

The Working Group meeting “Time Use Survey” was held in Luxembourg in July 2014. The focus of the meeting was on the technicalities related to the process of TUS data collection at European level (Minutes in Annex 3).

The situation of the project was reviewed at the Working Group meeting “Time Use Survey” in Luxembourg in January 2016. The remaining countries were asked to prepare their data as soon as possible, no later than during summer 2016 so that Statistics Finland would have time to process them during the validity of the agreement by the end of 2016. The countries were encouraged to fill in the Metadata Handler (Minutes in Annex 9).