UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

Work Session on Statistical Metadata
(Geneva, Switzerland, 6-8 May 2013)
Topic (iii): Metadata in the statistical business process

BETTER DOCUMENTING STATISTICAL BUSINESS PROCESSES: THE EURO PROCESS METADATA STRUCTURE

Working Paper

Prepared by A. Consoli, A. Götzfried, H. Linden, B. Rychel (Eurostat)

I.   Introduction

1. Business process integration of statistical business processes and modernisation of official statistics is growing into a broad concern for many national and international statistical institutes around the world. This led to the shaping of national and international strategies for statistical business process integration. The most known ones are the strategy defined by the High-Level Group for the Modernization of Statistical Production and Services (HLG), and the European Statistical System Committee joint strategy implementing the so-called Vision defined for the European Statistical System.

2. Both strategies focus – inter alia – on the statistical exchange, production and dissemination processes with the aim of better integration of those processes in order to reduce the scarce resources available. In order to progress with this integration, statistical and technical standards as well as more commonly used IT applications are necessary.

3. Examples are SDMX as technical standard for data/metadata exchange, standard code lists for structural metadata or shared IT applications such as EDIT for data validation used in the European Statistical System. Another necessary precondition for statistical business process integration is a standard business process documentation describing many details of the statistical business executed in statistical institutes. Such business process documentation can be considered as metadata structure definition or reference metadata template.

4. In the following, we will describe how the Euro Process Metadata Structure (EPMS) as standard business process documentation was developed and implemented for Eurostat and how this documentation is or will be used for progressing towards more statistical business process integration.

II. The ESS standard reference metadata structures for ESS reference metadata and ESS quality reports

A. The Euro SDMX Metadata Structure (ESMS)

5. In 2009, the ESS standard for reference metadata has been defined. The so-called Euro SDMX Metadata Structure was recommended for use in the ESS with the Commission Recommendation 2009/498 of 23 June 2009 . Please see in annex 1 to this document the detailed structure of the ESMS.

6. The ESMS concepts mainly refer to the statistical concepts which are describing and documenting the statistical business processes and the outputs for data users. The concentration is on the description of the disseminated statistics (concepts 3 to 10), on the main quality information relevant for the users (concepts 11 to 18) and on basic information regarding the statistical data processing (mainly concept 20). In several user surveys the ESMS-files have been regarded as relevant and sufficiently accessible for data user.

7. Since 2009, the ESMS was broadly implemented for documenting the ESS reference metadata for Eurostat (more than 300 files are disseminated on the Eurostat webpage). Also more and more national reference metadata produced by Member States are standardised according to ESMS. Currently, more than 15 statistical domains use the ESMS with around 30 national ESMS files produced per domain and exchanged with Eurostat. We aim to finally standardise all national statistical reference metadata which are exchanged with Eurostat. In addition, National Statistical Institutes use the ESMS also more and more for national purposes.

B. The ESS Standard for Quality Reports Structure (ESQRS)

8. In addition to the ESMS, we also defined the ESS Standard for Quality Reports Structure (ESQRS). This was mainly done as some statistical domains produced more detailed producer oriented quality reports which also had to be standardised. Please see in annex 1 to this document the detailed structure of the ESQRS.

9. The basic ESQRS quality concepts are identical with the ESMS quality concepts. These basic ESMS or ESQRS quality concepts (such as V: accuracy and reliability or VI: timeliness and punctuality) are broken down into more producer oriented sub-concepts in the ESQRS (please see in annex 1 the further breakdown).

10. Since 2010, we also standardise more and more the Eurostat producer oriented quality reports and the national quality reports in converting non-standard structures to the ESQRS structure. More than 10 statistical domains are using already the ESQRS for the production of the Eurostat or of national user oriented quality reports. These reports also include quality indicators linked to the different sub concepts. Often these more detailed producer oriented quality reports are not disseminated on the Eurostat webpage, but kept within the ESS working groups or statistical domains.

C. Interrelationship between the ESS metadata structures

11. The interrelationship between the ESS metadata structures (i.e. the ESMS, the ESQRS and also the new Euro Process Metadata Structure EPMS) is as follows:

Figure 1: Interrelationship between the ESMS, the ESQRS and the EPMS
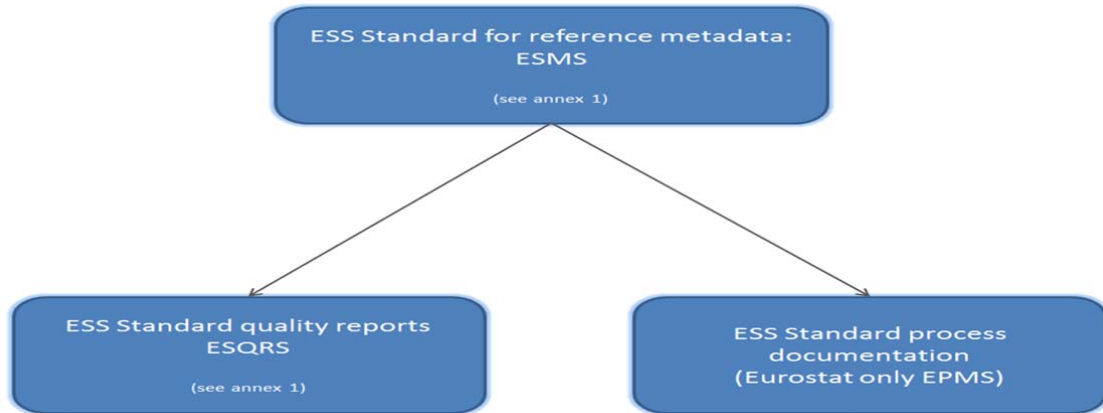
12. In general, the ESMS file documenting reference metadata can be considered as the master file to which the other ESS metadata structures (such as the ESQRS or the EPMS) are linked. Both additional structures take over a selection of basic ESMS concepts and detail them further for additional and more particular purposes (e.g. for the ESQRS: to provide more detailed producer oriented quality reports and for the EPMS: to provide a standardised documentation of statistical processes at Eurostat).

13. In any case, we are keeping all metadata reports using the abovementioned ESS reports structures fully consistent. This also implies that the Eurostat or ESS producer of those files only need to report once on each of the statistical concepts used in any of the ESS standard metadata reports.

IV. In addition: the Euro Process Metadata Structure (EPMS)

14. Based on the needs to faster progress with the integration of statistical business processes at Eurostat, we developed an additional metadata structure definition (or reference metadata template) aiming at better documenting the statistical business processes at Eurostat: the Euro Process Metadata Structure. The Euro Process Metadata Structure covers the following metadata concepts:

Euro Process Metadata Structure (EPMS)

	Concept name			Concept name
1	**Contact**		4.4	Data compilation
1.1	Contact organisation		4.4.1	Data compilation - variables
1.2	Contact organisation unit		4.4.2	Data compilation - weights
1.3	Contact name		4.4.3	Data compilation - aggregates
1.4	Contact person function		4.4.4	Data compilation - finalisation
1.5	Contact mail address		4.4.5	Data compilation - draftoutput
1.6	Contact email address		4.5	Data validation - final
1.7	Contact phone number		4.5.1	Data validation final - output
1.8	Contact fax number		4.5.2	Data validation final - explanation
2	**Summary process description**		**5**	**Confidentiality**
3	**Workflow**		5.1	Confidentiality - data treatment
4	**Statistical processing**		**6**	**Release policy**
4.1	Data collection		6.1	User access
4.2	Source data		**7**	**Dissemination format**
4.2.1	Source data - integration		7.1	Publications
4.2.2	Source data - coding		7.2	On-line database
4.3	Data validation		7.3	Micro-data access
4.3.1	Data validation in Member States		7.4	Other
4.3.2	Validation rules agreed with Member States		**8**	**IT applications**
4.3.3	Data validation - detection (Eurostat)		8.1	IT applications for data reception/collection
4.3.4	Data validation - correction (Eurostat)		8.2	IT applications for data processing
			8.3	IT applications for data validation
			8.4	IT applications for data confidentiality
			8.5	IT applications for metadata
			8.6	Other IT applications

15. Due to the needs and particularities of Eurostat, the EPMS concentrates on the GSBPM steps 4 to 7. Many concepts refer to the internal data validation processes executed at Eurostat and to the internal IT applications used for the validation. More transparency on this is definitely needed in order to integrate statistical processes in using standards and for reducing the diversity of internal IT applications.

16. We supported the production and delivery of the around 100 EPMS files at Eurostat as follows:

17. We integrated the EPMS into the existing Eurostat metadata production system and linked the EPMS to other metadata structures (i.e. the ESMS or the ESQRS) in order to avoid double production of information.

18. We provided detailed guidelines on what to fill in for the statistical concepts covered. We also provided extensive training to the process managers.

19. Business process documentation already existing beforehand could be reused and integrated into the standard EPMS structure.

20. In March 2013, more than 80 % of the EPMS files necessary for documenting all Eurostat statistical business processes were available.

21. Much of the discussions with the Eurostat process managers referred to questions on the appropriate level of granularity of the statistical business processes to be documented.

The general guidance on this was to integrate also the more detailed business process documentation in cases where many concepts are described identically.

22. Please see in annex 2 an example of an EPMS file for "Residence permits".

V.   How does the EPMS facilitate business process integration?

23. The EPMS was originally conceived and implemented with the aim of serving several purposes for information on the statistical production processes inside Eurostat, but it was designed based on an international standard, the GSBPM, to have the possibility to extend it to the ESS and beyond.

24. Its many purposes include to assure compliance with internal control standards of the Commission, to assure business continuity and in particular the handover of processes, to support IT developments and Eurostat's Quality Assurance Framework, and more and more important, to facilitate process integration and methodological standardisation, in line with Eurostat's Vision. In the following we focus on this last purpose and the way it is possible to fulfil it at Eurostat and in the ESS.

25. The first precondition to integrate processes is to have comparable descriptions of the way the processes work and this is possible if the different processes are described following the same standard. Using EPMS, processes can be decomposed using the same set of steps or concepts, which are described in a harmonised way. This enables us to compare two processes at an overall level, based on the EPMS concepts containing the summary process description and the workflow, and at a more detailed level of single processes phases. It is then possible to study if the statistical processes follow the same or similar business process and decide for possible further integration, or not. It must anyway be recognised that EPMS provides a basic process description, but once processes that are candidates for integration or redesign have been identified, a more detail description of each process step is needed (by use of Business Process Modelling tools) in order to proceed with the real integration or redesign. However, EPMS can already facilitate this, thanks to the possibility to attach documents describing the process phases more extensively.

26. Being able to identify processes with similar data treatment, allows benchmarks and so to detect process steps where gains are possible by modifying the way such steps are carried out or the IT tools that are used. A review of all processes with the same data life cycle makes it also possible to recognise and promote best practices. Knowing which functionalities are needed for a certain process step is also a precondition to develop or extend standard IT tools. Once standard methodologies and IT tools are put in place, it is possible to recommend or impose their usage for all those processes which do not show in their descriptions enough justifications to continue with ad-hoc solutions.

27. A clear example of this can be shown for data validation. This is a very time consuming activity, often duplicated, in Member States and at Eurostat, which may requires several iterations and the use of many heterogeneous ad-hoc IT tools. The EPMS contains a clear description of the types of validation checks that are performed, classified according to standard levels of validation, and of the IT tools which are used and as well of the agreement with Members States on who does what. For two processes treating the same types of data we can compare if the same validations are done, how long this step takes on average, how many iterations with Member States are needed and if standard IT tools are used.

28. For instance, if a standard IT tool solution works well for one process, it can be assumed that it could be applied to a second process, so avoiding the maintenance of specific IT tools. At the same time, the standard IT tool could be enhanced, if additional functionality required in the second process is not yet available.

29. If it appears then that in one process it is possible to improve the data validations as such or in another process the setting up of agreements with the data providers defining clear responsibilities on who does what, there are obvious cases for gaining in both performance and improved data quality

30. These principles are now being applied at Eurostat, where all statistical production processes have been described according to EPMS, starting with the rationalisation of the IT tools and with improvement in data validation. If this metadata structure would be extended, adding the necessary extra concepts, so to cover also the production in national statistical institutes, it could be applied at the ESS levels too.

31. From the comparison of similar processes in national statistical institutes and at Eurostat, with the possibility to benchmark and to describe best practices, statistical standards and IT tools could be applied and shared services could be developed and implemented for related business processes in the ESS.

VI.   Towards an ESS Process Metadata Structure

32. As already outlined above, the EPMS is used at Eurostat only and documents the statistical business process only for the GSBPM process steps Eurostat is mainly dealing with (i.e. process step 4 to 7). So the question arises how such as harmonised statistical process documentation would look like for the national statistical institutes. The latter are in general dealing with all main steps of the statistical business process (i.e. with the process steps 1 to 9 of the GSBPM: from 1: specify needs to 9: evaluate). Such a harmonised ESS business process documentation would then document the whole national statistical business process, including national dissemination and data/metadata exchange with Eurostat.

33. In the following we provide a first draft of such a metadata report structure which could be used for documenting the ESS statistical business process at national statistical institutes.

Table 1. Draft template for the documentation of national business processes in the ESS

Contact
Contact organisation
Contact organisation unit
Contact name
Contact person function
Contact mail address
Contact e-mail address
Contact phone number
Contact fax number
Summary process description
Specify needs
Micro data processing
Design (further details to be inserted)
Build (further details to be inserted)
Collect (further details to be inserted)
Statistical processing and analysis
Data collection
Source data
Source data: integration
Source data: coding
Data validation
Data validation: detection (Eurostat)
Data validation: correction (Eurostat)
Data compilation
Data compilation – variables
Data compilation – weights
Data compilation – aggregates

Data compilation – finalisation
Data compilation – draft output
Data validation – final
Data validation final – output
Data validation final – explanation
Confidentiality
Confidentiality – data treatment
Data release and data exchange
User access
Data exchange
Dissemination format
Publications
On-line database
Micro-data access
Other
Archive
Evaluate

34. The following needs to be considered when looking at this draft report structure for the ESS:

35. More details need to be provided for the GSBPM process steps Design, Build and Collect.

36. Some of the process steps are more structural (i.e. not specific to each process) and might not need to be detailed further for each process (for instance the GSBPM process step Archive).

37. We should deepen this report structure for those statistical concepts for which more integration with the ESS can be reached between national statistical institutes or with the national statistical institutes and Eurostat. This implies somehow the measuring of an increasing use of technical and statistical standards and shared services. One example for this could be the concept related to data exchange.

38. This process documentation certainly needs in-depth discussions with the ESS Member States also in order to accommodate their own needs and in order to compare this structure to already existing business process documentation in countries.

VII. Conclusions

39. This paper illustrates the ESS metadata structures in use such as the ESMS, the ESQRS or the newly implemented EPMS. It explains how the Eurostat standard process documentation is or will be used. Finally, the paper also outlines some ideas for an ESS process documentation which could be followed up within the ESS.

Annex 1

Picture 1: The ESS standard report structure for reference metadata: The Euro-SDMX Metadata Structure (ESMS)

The Euro-SDMX Metadata Structure (ESMS)

#	Concept Name	#	Concept Name	#	Concept Name
1	Contact	7	Confidentiality	15	Timeliness and punctuality
1.1	Contact organisation	7.1	Confidentiality - policy	15.1	Timeliness
1.2	Contact organisation unit	7.2	Confidentiality - data treatment	15.2	Punctuality
1.3	Contact name	8	Release policy	16	Comparability
1.4	Contact person function	8.1	Release calendar	16.1	Comparability - geographical
1.5	Contact mail address	8.2	Release calendar access	16.2	Comparability - over time
1.6	Contact email address	8.3	User access	17	Coherence
1.7	Contact phone number	9	Frequency of dissemination	17.1	Coherence - cross domain
1.8	Contact fax number	10	Dissemination format	17.2	Coherence - internal
2	Metadata update	10.1	News release	18	Cost and burden
2.1	Metadata last certified	10.2	Publications	19	Data revision
2.2	Metadata last posted	10.3	On-line database	19.1	Data revision - policy
2.3	Metadata last update	10.4	Micro-data access	19.2	Data revision - practice
3	Statistical presentation	10.5	Other	20	Statistical processing
3.1	Data description	11	Accessibility of documentation	20.1	Source data
3.2	Classification system	11.1	Documentation on methodology	20.2	Frequency of data collection
3.3	Sector coverage	11.2	Quality documentation	20.3	Data collection
3.4	Statistical concepts and definitions	12	Quality management	20.4	Data validation
3.5	Statistical unit	12.1	Quality assurance	20.5	Data compilation
3.6	Statistical population	12.2	Quality assessment	20.6	Adjustment
3.7	Reference area	13	Relevance	21	Comment
3.8	Time coverage	13.1	User needs		
3.9	Base period	13.2	User satisfaction		
4	Unit of measure	13.3	Completeness		
5	Reference period	14	Accuracy and reliability		
6	Institutional mandate	14.1	Overall accuracy		
6.1	Legal acts and other agreements	14.2	Sampling error		
6.2	Data sharing	14.3	Non-sampling error		

Picture 2: The ESS Standard for quality reports Structure: The ESQRS

The ESS Standard for Quality Reports Structure (ESQRS)

#	Concepts	#	Concepts	#	Concepts
I	Contact	V.3.4	Processing error	VIII	Comparability
I.1	Contact organisation	V.3.4.1	Imputation - rate	VIII.1	Comparability – geographical
I.2	Contact organisation unit	V.3.4.2	Common units - proportion	VIII.1.1	Asymmetry for mirror flows statistics - coefficient
I.3	Contact name	V.3.5	Model assumption error	VIII.2	Comparability – over time
I.4	Contact person function	V.3.6	Data revision	VIII.2.1	Length of comparable time series
I.5	Contact mail address	V.3.6.1	Data revision – policy	VIII.3	Comparability – domain
I.6	Contact email address	V.3.6.2	Data revision - practice	IX	Coherence
I.7	Contact phone number	V.3.6.3	Data revision – average size	IX.1	Coherence- cross domain
I.8	Contact fax number	V.3.7	Seasonal adjustment	IX.1.1	Coherence – sub annual and annual statistics
II	Introduction	VI	Timeliness and punctuality	IX.1.2	Coherence- National Accounts
III	Quality assessment	VI.1	Timeliness	IX.2	Coherence – internal
IV	Relevance	VI.1.1	Time lag – first result	X	Cost and Burden
IV.1	User needs	VI.1.2	Time lag – final result	XI	Confidentiality
IV.2	User satisfaction	VI.2	Punctuality	XI.1	Confidentiality – policy
IV.3	Completeness	VI.2.1	Punctuality – delivery and publication	XI.2	Confidentiality – data treatment
IV.3.1	Data completeness – rate	VII	Accessibility and clarity	XII	Statistical processing
V	Accuracy and reliability	VII.1	News release	XII.1	Source data
V.1	Overall accuracy	VII.2	Publication	XII.2	Frequency of data collection
V.2	Sampling error	VII.3	On-line database	XII.3	Data collection
V.2.1	Sampling error - indicators	VII.3.1	Data tables –consultations	XII.4	Data validation
V.3	Non-sampling error	VII.4	Micro-data access	XII.5	Data compilation
V.3.1	Coverage error	VII.5	Other	XII.6	Adjustment
V.3.1.1	Over-coverage - rate	VII.5.1	Metadata –consultations	XIII	Comment
V.3.2	Measurement error	VII.6	Documentation on methodology		
V.3.3	Non response error	VII.6.1	Metadata completeness – rate		
V.3.3.1	Unit non-response - rate	VII.7	Quality documentation		
V.3.3.2	Item non-response - rate				

Annex 2: Euro Process Metadata Structure EPMS file for residence permits

migr_res_epms preview

Residence permits
EPMS - Euro Process Metadata Structure

For any question on data and metadata, please contact: EUROPEAN STATISTICAL DATA SUPPORT

1. Contact	Top
1.1. Contact organisation	Eurostat, the Statistical Office of the European Union
1.2. Contact organisation unit	F2: Population
1.3. Contact name	xxxxxxxxx
1.4. Contact person function	xxxxxxxxxx
1.5. Contact mail address	xxxxxxxxxx
1.6. Contact email address	xxxxxxxxxx
1.7. Contact phone number	xxxxxxxx
1.8. Contact fax number	xxxxxxxx

2. Summary process description	Top
Residence permits data collection is based on the regulatory framework of Article 6 of the Regulation 862/2007 on migration and international protection statistics. Residence permits statistics are used for planning actions, monitoring and evaluating programmes in policies for European migration area, such as the policy and the allocation of money to EU Member States for Solidarity Funds (See DGHOME). Residence permits data is annually collected from administrative sources, having as statistical unit the number of persons (third-country nationals who received the residence permit). National data suppliers are sending the residence permits data to the following countries: EU Member States, Iceland, Norway, Liechtenstein and Switzerland. The main phases of the statistical process are: data collection, data processing, validation and dissemination. The data is validated through all processing steps. The same main phases are applied to Metadata information received from data providers.	

3. Workflow	Top
The residence permits data collection relates to reference periods of one calendar year and shall be completed by national data suppliers that are responsible at national level for the transmission of the residence permits data to Eurostat (National Statistic Offices, Ministries of Interior or Immigration Agencies).	

The workflow for residence permits statistics is described as follows:

Step 1. Data collection

The data sending obligations by EU Member States for each dataset are set within the eDamis application, in accordance with the residence permits collection framework. The obligations may differ between tables. Data on residence permits shall be provided by the EU Member States, Iceland, Norway, Liechtenstein and Switzerland in accordance with the legal framework (Article 6 of the Regulation 862/2007).

The residence permits data should be provided by countries within six months as of the end of the reference year. Data providers supply data to Eurostat in Excel templates via the Single Entry Point (SEP) with eDamis.

Data collection consists of 10 template collected from EU Member States (16 tables collected). The data transmission format is .XLS. Each file is considered a template and contains at least one table for data collection (a multidimensional table to be filled in with data provided by each EU Member State).

Detect errors: templates integrity (using the XLS templates provided by Eurostat for specific tables), use a specific name according to the reference period, totals consistency on the tables and between the tables, the relation between data and metadata information.

If needed, countries are asked to provide Eurostat with amended templates via Edamis.

Step 2. Data Validation

The data is validated through all processing steps (data collection, data processing and data dissemination), combining manual/visual validation with the functionalities of the IT tools used.

Step 3. Data format conversion

A specific CSV format should be created in order to import the data in the production database (EUROCUBE). Another XLS file is used for the conversion of several XLS files at the same time via a macro code. At this level, the transcoding process is possible between the codes collected and the codes used in EUROCUBE.

Detect errors: name, reference year, structure of tables.

If needed, some previous stages across the production workflow shall be repeated. There may be cases when countries are asked to provide Eurostat with amended templates via Edamis.

Step 4. Production database

The residence permits data is stored in EUROCUBE. It is a generic Oracle 10g application that can be used for managing the entire data collection.

The main operations performed in Eurocube are the following: data management (importing, storing, visualising, modifying and exporting the data), aggregation, validation and facility for sending the data to Eurostat Web Environment (EUROBASE).

EUROCUBE may be used to run some programs (sequences of actions) in order to produce different results (managing data or producing outputs). These programs are designed separately and are dedicated to a set of operations. Currently, two programs are designed for the residence permits domain: importing of CSV files and calculation of dissemination measure (Table RP0).

Detect errors: totals consistency within the dimensions hierarchy and data credibility (is checked for the time series).

If needed, some previous stages across the production workflow are repeated. There may be cases when countries are asked to provide Eurostat with amended templates via Edamis.

Step 5. Data dissemination

The residence permits data is published on the Eurostat Website (EUROBASE) at the end of the main validation process, usually within two months as of the end of the deadline for data transmission. Some data may be published at a later stage due to some supplementary validation steps. The publication on EUROBASE is done through EUROCUBE database.

Other publications outputs are compiled: SIF-s, Statistic Explained Articles when is required. Different tables are prepared for specific data users (e.g. DG HOME related to Solidarity Funds).

Detect errors: data correspond with the publication standards. If needed, some previous stages across the production workflow are repeated. There may be cases when countries are asked to provide Eurostat with amended data via Edamis.

4. Statistical processing	Top
4.1. Data collection	
Residence permits data is annually collected from administrative sources, having as statistical unit the number of persons (third-country nationals who received the residence permit). National data suppliers are sending the residence permits data for the following countries: EU Member States, Iceland, Norway, Liechtenstein and Switzerland. Data providers supply data to Eurostat in Excel templates via the Single Entry Point (SEP) with eDamis. The residence permits data collection relates to reference periods of one calendar year and shall be completed by national data suppliers that are responsible at national level for the transmission of the residence permits data to Eurostat (National Statistic Offices, Ministries of Interior or Immigration Agencies).	
4.2. Source data	
Administrative data sources only. Data are compiled from the administrative records of the national authorities, mainly the Ministries of Interior or Immigration Agencies.	
4.2.1. Source data - integration	
Not applicable.	
4.2.2. Source data - coding	
Not applicable.	

4.3. Data validation

Before publishing the data, consistent validation checks are performed. Residence permits data are validated through all processing steps, combining manual/visual validation with some tools integrated within the templates and the production environment.

4.3.1. Data validation in Member States

The statistical process is different in each country; however, all EU Member States shall provide the residence permits data based on the same methodology provided by Eurostat within the Technical Guidelines.

EU Member States shall apply the initial validation checks of the permit statistics before providing data to Eurostat. These validation checks shall ensure the internal consistency of the datasets at the level of each aggregation, the consistency of data between relevant tables, as well as the consistency of data between different reporting periods (the amount of changes between different periods). Therefore, each country shall develop its own data validation system, depending on the national infrastructure.

A series of validation rules should be applied at national level for Quality Assurance Program for the Solidarity Funds (QAPS), including validation checks for each table, validation checks for the same table over time, validation checks between related tables within the RESPER domain, validation checks from the templates.

A series of macros have been included in the Excel files in order to allow the data providers to perform the data validation before they are transmitted to Eurostat.

A report is produced by this macro with the indication of the cells with problems.

The validation checks implemented in the templates include:

the cells should contain only numeric values, without decimal figures or negative figures;

the vertical consistency: for each column in the table, the value provided for the total permits should sum to the values by individual citizenship;

the horizontal consistency: for each row in the table, the total permits should be consistent with the disaggregation provided in the different columns; moreover, if a sub-total is provided for any of the disaggregation, this should be consistent with the corresponding breakdown;

the number of permits by citizenship in each main table (RPx) should be identical to the number of permits by citizenship in the second table related to the desegregations by age and sex categories (RPx_AS);

the number of warnings regarding the relations between the data and data availability declared in the Metadata worksheet.

4.3.2. Validation rules agreed with Member States

Some data checks are included within the data collections templates and Residence Permits Technical Guidelines. All data providers are consulted before adopting these documents.

A series of validation rules are applied at national level for Quality Assurance Program for the Solidarity Funds (QAPS).

The residence permits data collection is recent and the process of developing a common validation framework is on-going.

4.3.3. Data validation - detection (Eurostat)

The data is validated through all processing steps, combining manual/visual validation with some tools integrated within the templates and the production environment. A series of validation rules are also applied for Quality Assurance Program for the Solidarity Funds (QAPS).

The residence permits data validation process is described by reference to the Eurostat VIP Validation levels:

Validation level 0

The files sent by Member States are cheeked in order to comply with the standards from the Residence Permits Technical Guidelines:

Different type of rights for data transmission are set into eDamis application;

Use only accepted format .XLS in accordance with the structure provided by Eurostat;

Use a specific name for the files transmitted to Eurostat. The name should have the following format: RESPER_RPX_A_CD_YYYY_0000_V000n_N ("CD" = country code, "YYYY" = reference year, "V000n" = the version number);

Assure the integrity of the template structure - Member States are not allowed to change the templates structure, the multidimensional structure and the formulas included in the excel worksheet. The structure of the templates is protected in MS Excel.

Validation level 1

Assure that the name of the files correspond with the values filled into the SUMARRY worksheet from the templates related to the reference year and the country involved;

Some checks are performed with a macro from the templates, including the number format of the cells and the totals consistency (see "Validation checks from the templates" from 4.3.1.). A report is produced by this macro with the indication of the cells with problems;

The templates are predefined with some formula for computing the totals on the rows and the columns. The option to reset the templates is also available within the SUMMARY worksheet. The templates include a macro for producing a number of warning regarding the relations between the data and data availability declared in the Metadata worksheet;

The totals consistency is verified also within the production environment (EUROCUBE) with Validation/Hierarchy option;

The data credibility is checked for the time series within the production environment (EUROCUBE Database). The data can be exported to Excel format before applying some validation checks (formula for checking the magnitude of changes overtime).

Validation level 2

Some checks are performed with a macro from the templates (see point 4.3.1. Data validation in Member States). The same validation process is done also at the level of Eurostat. Therefore, it should be checked if the number of permits by citizenship in each main table (RPx) is identical to the number of permits by citizenship in the second table related to the desegregations by age and sex categories (RPx_AS).

Validation level 3

Not applicable

Validation level 4

Residence permits data are tested together with Migration data in order to explore the possibility of validation across domains.

Validation level 5

The residence permits data from other institutions are not identified as being comparable with the data produced by Eurostat. Further investigations are needed.

4.3.4. Data validation - correction (Eurostat)

Not applicable.

If the data is considered uncertain (the data are not in line with the expectations) the data providers shall be asked to provide supplementary information.

If the data fail the critical validation checks, the statistics shall be recompiled and shall be resent by the data providers, in accordance with the Residence Permits Technical Guidelines.

4.4. Data compilation

Not applicable.

4.4.1. Data compilation - variables
Not applicable.
4.4.2. Data compilation - weights
Not applicable.
4.4.3. Data compilation - aggregates
Some totals (by rows and by columns) are computed within the excel files using the "SUM" formula. Some totals are computed in the production environment (EUROCUBE) due to the fact that the data is collected into multidimensional tables and some totals are shared by two or three categories. Therefore, the Validation/Hierarchy function from EUROCUBE is applied to five tables in order to assure the totals consistency within the same category. One table (RP0) is computed by retrieving the data from other production tables (RP1-RP4 tables). This operation is carried out by launching a specific program in EUROCUBE, which calculates/updates the RP0 table. Different aggregates (EU level, EFTA level, Citizenship level) are computed for publications purposes. These aggregates are simple sums of the values for the related dimensions (no estimation included).
4.4.4. Data compilation - finalisation
The data is published on the EUROBASE at the finalisation of the validation process. The data is published within the first two months after the transmission period (between six and eight months as of the end of the reference period) depending on the data availability. Data may be revised whenever the data providers consider necessary to correct the figures already transmitted. Another data processing cycle is done in case of any data revision sent by EU Member States.

4.4.5. Data compilation - draft output
Not applicable.
4.5. Data validation - final
4.5.1. Data validation final - output
Not applicable.
4.5.2. Data validation final - explanation
Not applicable.

5. Confidentiality	Top
5.1. Confidentiality - data treatment	
Not applicable. Not applicable	

6. Release policy	Top
6.1. User access	
In line with the Community legal framework and the European Statistics Code of Practice Eurostat disseminates European statistics on Eurostat's website (see item 10 - 'Dissemination format') respecting professional independence and in an objective, professional and transparent manner in which all users are treated equitably. The detailed arrangements are governed by the Eurostat protocol on impartial access to Eurostat data for users.	

7. Dissemination format	Top
7.1. Publications	
Eurostat free publication - Statistics in Focus on 'Residence permits issued to non-EU citizens in 2009 - Issue number 43/2011' released on 01 September 2011 and available online at: http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code= KS-SF-11-043	
7.2. On-line database	
Please consult free data on-line or refer to contact details.	
7.3. Micro-data access	
Not applicable.	
7.4. Other	
Not applicable.	

8. IT applications	Top
8.1. IT applications for data reception/collection	
MS Excel (XLS format) for data collection. eDamis application for data transmission from MS to Eurostat.	

8.2. IT applications for data processing
EUROCUBE

8.3. IT applications for data validation
MS Excel for different steps of data process (data collection, production, publications and post publication) EUROCUBE with data visualisation functionalities.

8.4. IT applications for data confidentiality
No specific applications are used.

8.5. IT applications for metadata
MS Excel (XLS format) for collecting some metadata information (Data Availability Questionnaire). MS Word is used for drafting metadata file. RAMON for the list of reference codes used for data collection.

8.6. Other IT applications
MS Access for testing different functionalities. "R" for testing cross domain validations (validation level 4). ImageMagic for working in relation with "R" WinSCP for EUROCUBE connection. Notepad++ for VBA programming.

Annex   Top
Residence Permits Statistics - workflow Residence Permits Statistics - internal data processing Residence Permits Statistics - format and templates for data transmission Residence Permits Statistics - internal data validation flow TECHNICAL GUIDELINES FOR RESIDENCE PERMITS