

# CONSISTENT QUALITY REPORTING WHILE REDUCING REPORTING BURDEN: A CASE STUDY OF SIMS IMPLEMENTATION.

Susana Portillo Cruz<sup>1</sup>

<sup>1</sup>Central Statistics Office, Ireland

## Abstract

Transparency in statistical production is a fundamental principle to foment trust in official statistics. As such the use of Quality Reporting is a core requirement for all producers of official statistics to inform users on the methodologies, processes and key quality indicators associated with the disseminated outputs. Through these reports users can evaluate and make informed decisions on the correct use and assess the quality of the disseminated statistics. With the advent of Eurostat's Single Integrated Metadata Structure (SIMS) in 2015 this obligation was standardized across Europe.

This paper discusses the efforts of the Central Statistics Office (CSO) in Ireland to harmonise different types of quality reports and ensure a unified message across various requirements. In addition, it highlights the technical aspects of the migration to SIMS using Colectica as a central repository of reports to achieve a more streamlined quality reporting workflow for all our disseminated statistics, while simultaneously decreasing the reporting burden on our Statisticians.

**Keywords:** Trust, Standards, Reference metadata, Quality reporting, Efficiency, Consistency

## 1 Introduction

Quality reporting is a European requirement for the transmission of statistics that is taking greater traction at national level thanks to Eurostat's policy of 'No data without metadata'. This policy is of extreme importance since we need to give the users of our statistics the information necessary for them to evaluate whether the disseminated data is fit for the purpose for which they require it. And this statement does not only apply to Eurostat transmissions: it should be a tenet for all publications of official statistics by a government body, whether European, national or other statistics.

The Single Integrated Metadata Structure (SIMS) was introduced by Eurostat in 2015 with a view to harmonise quality reporting between all member states of the EU. The SIMS standard

comprises a superset of concepts that englobes the existing ESMS (user oriented reports) and ESQRS (producer oriented reports) into a single structure.

But within a national statistical institute the situation can be somewhat more complex than a single report.

## **2 Why harmonise quality reports**

Almost all production of European official statistics in the CSO that are disseminated to Eurostat have a counterpart national publication. With regards quality reports this translates into a particular producer in a domain having to create several slightly different quality reports:

- A report based on the SIMS or ESQRS standards
- And or a User oriented report for Eurostat
- A national report that is maintained and disseminated as a pdf document
- Any other type of quality report, e.g. for the IMF, OECD etc. depending on what bodies data is transmitted to.

The need to generate several different reports places unnecessary burden on the producer of the statistics. In addition, it carries a risk of introducing discrepancies in reports for the same iteration of a statistic if the reports are kept separate. The harmonization of reports into a single structure at national level will address both the burden and the risk mentioned above.

## **3 The harmonization process**

The process of harmonising quality reports at national level may not be as simple as it sounds. Oftentimes our producers are already overstretched with schedules and deadlines, and we cannot expect for them to, in addition, sit back and analyse their work and harmonise different sources of information even if the benefits will be manifold in the end. For this reason, a big part of harmonisation has been based on having a dedicated team within the office to carry out the harmonisation using an incremental approach towards implementing the “write once for all purposes” principle and carry out the initial migration to harmonised reports.

Our harmonisation process is based on principles similar to any other process improvement project, with four main implementation steps:

1. Harmonize the content and structure of our national reports to fit around Eurostat's SIMS.
2. Migrate the text contained in national reports still using the old technology of word and pdf.

3. Development of a new software tool to allow for the electronic management of quality reports.
4. Populate the database or repository that will host the quality reports and roll out to other users.

An initial project timeline was established that allowed for the concurrent operation of steps 2 and 3 above after completion of step 1. Once a stable version of the software was achieved in step 3, we were also able to commence with the implementation of step 4, more so the required set up of reporting standards than the population of the repository itself.

### **3.1 National harmonisation of report structure**

To harmonise the content of all our quality reports we carried out a pilot study with one of the larger and most complex statistical divisions as a proof of concept to determine the possible structure of our harmonised reports.

We started by analyzing our existing report at national level as this was the most comprehensive report to be completed (other than those required by Eurostat) by that specific statistical unit and identified synergies and differences with the SIMS concepts and found around a 90% level of overlap in the concepts between both standard reports.

The remaining 10% concepts were a combination of SIMS concepts that were not present at national level and vice-versa.

A further analysis showed that the National concepts that were not present in SIMS were actually present by default as part of the official Eurostat guidelines describing the content of certain concepts. Since SIMS has been designed to be an extendable standard, we exploited this functionality it by adding these extra sub concepts under the adequate SIMS concept thus ensuring that information required by the guidelines would not be forgotten. An example of these additions can be seen in the figure below.

Figure 1: Example of extended concepts for CSO SIMS

CONCEPT_NAME	SIMS ID	ESQRSID	ESMSID	CSO SIMS ID	usage (SIMS/ Domain specific)	parent	NCEPT
DATA_REV_AVGSIZE	S.17.2.1	6.6.1	-	19.2.1	SIMS		REV_PRACTICE
STAT_PROCESS	S.18	3	18	20	SIMS		N/A
SOURCE_TYPE	S.18.1	3.1	18.1	20.1	SIMS		STAT_PROCESS
<b>POP_FRAME</b>					<b>CSO SIMS</b>		<b>SOURCE_TYPE</b>
<b>SAMP_DESIGN</b>					<b>CSO SIMS</b>		<b>SOURCE_TYPE</b>
<b>SURVEY_SIZE</b>					<b>CSO SIMS</b>		<b>SOURCE_TYPE</b>
<b>SURVEY_TECHNIQUE</b>					<b>CSO SIMS</b>		<b>SOURCE_TYPE</b>
FREQ_COLL	S.18.2	3.2	18.2	20.2	SIMS		STAT_PROCESS
COLL_METHOD	S.18.3	3.3	18.3	20.3	SIMS		STAT_PROCESS
<b>SURVEY_TYPE</b>					<b>CSO SIMS</b>		<b>COLL_METHOD</b>
<b>QUESTIONNAIRE</b>					<b>CSO SIMS</b>		<b>COLL_METHOD</b>
<b>SURVEY_PARTICIPATION</b>					<b>CSO SIMS</b>		<b>COLL_METHOD</b>
<b>DATA_CAPTURE</b>					<b>CSO SIMS</b>		<b>COLL_METHOD</b>
DATA_VALIDATION	S.18.4	3.4	18.4	20.4	SIMS		STAT_PROCESS
DATA_COMP	S.18.5	3.5	18.5	20.5	SIMS		STAT_PROCESS
<b>IMPUTATION</b>					<b>CSO SIMS</b>		<b>DATA_COMP</b>
IMPUTATION_RATE	S.18.5.1	6.3.4.1	-	20.5.1	SIMS		DATA_COMP
<b>GROSS_WEIGHT</b>					<b>CSO SIMS</b>		<b>DATA_COMP</b>
ADJUSTMENT	S.18.6	3.6	18.6	20.6	SIMS		STAT_PROCESS
SEASONAL_ADJ	S.18.6.1	6.4	-	20.6.1	SIMS		ADJUSTMENT
COMMENT_DSET	S.19	12	19	21	SIMS		N/A

The harmonization concluded with a set of quality reporting concepts for CSO that is fully SIMS compliant which we call the 'Base CSO SIMS'.

### 3.2 Migration of reports to new structure

Once the set of CSO SIMS concepts was determined we commenced to harmonize the content of the national reports. To achieve this, a dedicated team was initially created as part of the quality team and subsequently trained on the full SIMS, including the extra CSO concepts. This team were tasked with migrating all national reports into the new CSO SIMS format and structure and liaise with the report owners (the producers of the statistics) to ensure that there was no loss of information and that the information in the new report was up to date.

To carry out the harmonization of content the team looked at several sources:

- Eurostat reports in any standard (SIMS, ESMS, ESQRS) as stored in Eurostat's metadata handler.
- CSO Standard Report on Methods and Quality (national level report).

- Survey background notes (released by CSO with every iteration of disseminated results for a statistic).
- CSO 'Methods' metadata pages on cso.ie, these are mandatory pages on the website that available for all the statistics produced by CSO and contain:
  - a) a basic level of metadata, including quality related
  - b) links to any used paper instruments
  - c) links to a full quality report
  - d) links to metadata documentation where available.
- CSO internal Directory of Products and Services application, a repository of all products and services produced by CSO with includes certain reference metadata for each of the products it contains.
- Any other quality report created by sections that the team were made aware of.

The main purpose of this part of the project is to ensure that all published Quality Reports have the same look and feel and can easily be identified by our customers as being disseminated by CSO. In addition, the structure of the reports is consistent therefore the correct information will always be found under the same headings (concepts).

The harmonization also translates into the same information being conveyed to the user regardless of the type of report disseminated (national, Eurostat, etc.) ensuring consistency in reporting.

By having this process completed by a dedicated team, we are also reducing the reporting burden place on our statisticians, who can then focus on ensuring that the final detail in the report is accurate.

### **3.3 Report Management tool**

To ensure that the harmonisation is successful and embedded into the business as usual the solution provided for managing reports needs to be simple, easy to use and provide immediate benefits for our users.

The most complex step in the project referred to the design and development of a software application to manage the lifecycle of the reports.

Our office are already heavy users of Colectica for questionnaire design, which is based on the Data Documentation Initiative (DDI). Colectica software offers functionality to create quality standards that are implemented via quality statements. In view of this we decided to use our existing Colectica repository as the tool for implementation of this SIMS project. However, the solution is not an off-the-shelf one for our requirements, since we wanted to integrate the management of quality reports with our internal Directory of Products and Services application as well as create a management workflow for their regular update and dissemination.

The report management application is designed as front-end web-based application that enables users to interact with the Colectica repository in an easy-to-use manner.

The application contains the following features:

- Create domain specific extensions to the SIMS standard as required.
- Implement a SIMS standard for a data collection within a DDI study.
- Enable the marking of mandatory items via user identifiers, carry out validation of the standard prior to dissemination to ensure all mandatory items are included, display the number of concepts completed and tag any missing mandatory concepts.
- Populate concepts with audience specific text should they be required. This is to cater with situations such as the report mentioning specific tables for national dissemination that differ from those in Eurostat transmissions.
- Retrieve metadata from our Directory of Products and services to ensure consistency in the content of certain concepts common to both systems.
- Roll over a quality statement from one iteration of the data collection to the next for ease of completion. It allows the marking of concepts that possibly do not change with every data collection and automatically inserts the same text into the rolled over iteration.
- Compare quality reports between two consecutive iterations, highlighting the text changes.
- Extract quality reports in the different formats currently required:

- CSO branded Pdf files for publication in the CSO website
- SDMX formatted files for either the SIMS, ESMS or ESQRS ready to be transmitted to Eurostat via eDamis.
- Integrate with Active directory to implement a workflow system for signing off reports: from “Draft” to “Quality Assurance Review” to “Quality Assured” to signed off and “Published” so the consistency of all reports against the SIMS standards can be managed centrally and compliance with quality reporting monitored regularly.

### **3.4 Embedding the new structure into “business as usual”**

The final step in this standardization process consists of embedding the usage of the SIMS standard in the day-to-day work of our statistical experts producing and disseminating official statistics. This involves populating the repository of quality reports using the Quality Report Manager tool.

An “administrator” role type will create the adequate SIMS standard to be implemented by each domain. Each of these different standards need to take into account the CSO Base SIMS and any further Eurostat extensions to SIMS which are domain specific. The administrator also will create the DDI Series/Studies/Data Collection placeholder structure in the Colectica system under which the quality reports will reside.

For each domain, the most recent report harmonized in word format will be initially completed by a member of the quality team (QA role) and subsequently allocated to the subject matter expert for revision.

The subject matter expert will then be able to review and edit the text and revert it back to QA for approval. Upon final inspection QA will be able to either ‘Reject’ with comments or “Approve” the report. Both cases will allocate the report back to its owner.

In the final step the owner will publish the report and it will be closed for that data collection. This last step will write the completed Quality Standard to the Colectica repository, and the metadata will be available via a web portal for discovery.

Additionally, once the Quality Standard has been moved to a 'Published' state the owner will be able to extract a version in pdf format to be disseminated in parallel to the statistics. The required SDMX file (if any) for mandatory transmission of metadata to Eurostat via the eDamis system can also be extracted from this quality report manager.

Any CSO specific sub concepts in the local standard will be automatically bundled with the parent Eurostat SIMS concept on generation of the SDMX file, ensuring no loss of information.

## **4 Main challenges encountered**

Each of the four steps taken in the approach outlined in this document comes with their own distinct set of obstacles. This section details the three major types encountered and the solutions adopted to remediate them.

### **4.1 Structure**

The initial migration of the reports started in full force with statistical products that were clearly identifiable as requiring their own quality report. As time progressed, we found more products that the office classified as individual products but that were a by-product of a production process generating a dissemination dataset from which multiple products are created. Identifying those proved more difficult than the team anticipated, and it considerably slowed down the migration process as we negotiated with product owners, provided guidance, compiled extra information and agreed on the final report.

### **4.2 Change management**

To date, the creation of reference metadata and quality reporting in the CSO has operated as a secondary process following the dissemination of statistics. Oftentimes, unless a deadline is prescribed by a specific authority for the submission of reference metadata, there has been a large gap of time between the dissemination of the statistics and the creation and update of a quality report. In addition, every domain manager within the CSO has had flexibility to manipulate their own national reports and make changes to the content and



structure. This siloed approach has resulted in an unstandardized mismatch of publications not easily recognizable as disseminated by the organization.

The proposed move to harmonised quality reports not only involves taking away that previous flexibility but has also introduced the requirement to report on concepts majorly ignored to date, for example the calculation of complex quality performance indicators that are part of SIMS. It has also introduced the need to re-write existing national reports to conform to the new standard.

To get statisticians on board the quality unit undertook a communications exercise to inform statistical managers of the upcoming development, its roadmap, the implementation plan and how they would be affected.

Once the report manager solution is fully delivered and populated, the timeframes for ensuring that all reports are updated regularly will be integrated within the work of a quality unit, and the dedicated team will be accountable for ensuring that quality reports are updated at regular intervals taking into account their frequency of dissemination and generally released to coincide with the dissemination of statistics.

### **4.3 Knowledge**

The harmonisation used throughout this project requires a good working knowledge of the SIMS standard: its structure, concepts and definitions of these concepts. The team carrying out the initial migration was mainly formed from administrative staff who are not familiar with statistical terminology and it was difficult for them to identify where and how to record information within the new format. This was managed by providing on the job training with a statistician and developing and using a template containing descriptions, guidance and, in some cases examples, of the information required under a particular concept. This template was subsequently used during the migration and is currently issued to statisticians who might not be producing official statistics but wish to disseminate a comprehensive quality report for their users.

In addition to this, for more complex technical information such as key performance indicator concepts which involve statistical knowledge, a series of short informative videos were created and released on our internal media platform.

This knowledge or lack thereof can also be attributed to our statistical experts who, in many cases, complete Eurostat reports based on the name of the concept rather than the description. To mitigate this issue a full training course has been developed in house, with specific CSO examples, to clarify the content of the SIMS. This course is based on the ESTP Advanced course on Quality Reporting. It will be delivered for national producers of statistics once the reporting tool is available and will be open to other national authorities producing official statistics, not only CSO staff.

With regards the quality report manager tool, training to users will also be made available by using a combination of:

- On the job training of the quality administrators
- The creation of manuals and process maps
- A short series of usage videos for report managers/producers of statistics
- Ad hoc assistance for report managers

#### **4.4 Resources**

The availability of human resources at the initial stages of planning and development of this project posed a negative impact on the timeframe for delivery. These affected mainly two of the deliverables: the migration of existing reports to SIMS and the development of the software tool.

The rollout of the migration was planned in a manner that similarly themed statistics were migrated together, thus ensuring continuity of the information and topics being migrated and making the information to migrate easier to follow for a person with limited knowledge on the topics.

The in-house development of the software tool would have required a steep learning curve on both the DDI standard and the Colectica Software Development Kit to be used so that the

software also integrates with the existing repository. In this case development was outsourced to Algenta Colectica technologies to mitigate those two elements and ensure a final system that is robust and in compliance with DDI Standards.

## **5 Future developments**

The wealth of information contained in the SIMS report can be used not only to inform our users of the quality of disseminated statistics but also to elicit management conversations internally when evaluating, interpreting and explaining outputs prior to publication and start the modernisation towards a more metadata driven system. The integration of a management dashboard to monitor key performance indicators that will inform those conversations has already been identified for the next phase of development.

One of the disadvantages of the current development is the need for national quality reports to be extracted in pdf format and any subsequent editing of those files necessary for disseminating them to the public, as well as requiring managing the publication via yet a third software system. It is envisaged to use existing technologies to ingest the information required at national level directly from the repository into specified structure within the CSO website without the need for interim steps.

Finally, in CSO we avail of several different applications all of them using similar types of both reference and structural metadata. A very long-term view of the direction of travel is being able to integrate the information in all these different databases into a single source of truth as appropriate, and where reference metadata will only be stored in one database and reused in the rest of them.

## **6 Conclusion**

This project was started with great energy and impetus, and it was still in the early stages when we realised that it was more ambitious than we originally planned for. The completion timelines have been adjusted several times to cater for unexpected eventualities, both from a human resource and technical perspectives. Two years from the start we are finally in a position to confidently give a timeframe for completion of the development plans with testing of the web application due to be completed during the first half of 2024.

Undoubtedly there will be issues arising once the system is in full production, especially since one of the deliverables from this project refers to the bespoke extension of a software product. This will involve the need to adapt some of the envisaged functionality to conform to the already existing standard (DDI) that the system is built upon, and the integration with existing tools.

Once the rollout is underway we expect to start generating efficiencies in processing reports. The users of our statistics will avail of a repository of quality reports that will allow for explanation of any statistics released by the office at any given point in time. In any case there have already been some immediate benefits which apply in particular to the national users of our statistics given that every quality report of official statistics has now been standardized, and information related to any specific quality topic will always be found under an adequate concept heading.

From an organisational trust point of view, the harmonization of the branding look and feel in our reports makes them easily identifiable as being disseminated by CSO, an official producer of factual statistics at a time of large competition from private sources in a busy data and information market. Together with the establishment of a process to maintain reports up to date it will give our users the assurance that the disseminated report is official and can be used to determine whether the published statistics are suitable for the purpose for which they are needed.

## **Acknowledgment**

Our initial research was assisted by our colleagues in statistics Denmark who have pioneered the automation of the SIMS within the European statistical system and were able to point us in the direction most suitable for the needs of our own national statistical institute.

The original pilot that allowed us to determine the full set of CSO SIMS concepts was carried out with the CSO's Balance of Payments team. The team helped us identify many of the "special situations" that could be encountered in quality reporting across our office, thus helping us from early on to find a standardized solution to our issue.

Finally, we want to thank the Algenta Colectica team for the development of the software application that will make the integration and management of quality reporting in CSO that bit easier, and for their guidance and expertise on ensuring that the production system built is kept within the parameters set by the DDI standard.

## **References**

European Statistical System handbook for quality and metadata reports (2021 re-edition), <https://ec.europa.eu/eurostat/documents/3859598/13925930/KS-GQ-21-021-EN-N.pdf/143394de-e5a0-31ac-2c90-2aa9c15803f0?t=1639042312202>  
DDI Lifecycle 3.3, <https://ddialliance.org/Specification/DDI-Lifecycle/3.3/>  
Colectica Repository - Quality Statements, <https://docs.colectica.com/designer/manage-content/quality-statements/>