

Quality in Statistics

A Handbook of Quality Standards and Guidelines





Quality in Statistics

A Handbook of Quality Standards and Guidelines

Published by the Central Statistics Office, Ireland 2021

© Government of Ireland 2021

Material compiled and presented by the Central Statistics Office. Reproduction is authorised, subject to acknowledgement of the source.

Further information is available at: https://www.cso.ie/en/aboutus/whoweare/copyrightpolicy/



Table of Contents

Introduction	7
Quality through the Generic Statistical Business Process Model (GSBPM)13	3
Specify Needs Phase – "Understanding quality needs"16	6
Design Phase – "Planning for quality"19	9
Build Phase – "Creating the quality environment"	2
Collect Phase – "Manage quality in to deliver quality out"2	5
Process Phase – "Preparing data for analysis"28	8
Analyse Phase – "Confirming quality and interpreting information"	1
Dissemination Phase – "Delivering knowledge and understanding through information"	4
Evaluate Phase – "A commitment to continuous improvement"	7
Specific Components of Statistical Quality43	3
Classifications44	4
Use of Administrative Data4	5
Emerging Data Sources48	8
Questionnaire Design	9
Survey Frame and Sampling52	2
Measures to Reduce Non-Response54	4
Data Editing	5
Data Quality Evaluation (including Macro-Editing)57	7
Imputation	9
Weighting	0
Seasonal Adjustment	2
Statistical Confidentiality	4
Quality in the Presentation and Dissemination of Results67	7
Presentation of Results	8
Metadata & Quality Reports70	0
Revisions	3



CSO Process Mapping and Documentation76
CSO Quality Information System77
CSO Directory of Products and Services (DPS)78
CSO Standard Methods page79
CSO Data File Structure80
CSO Data Attestations81
CSO Data Inventory81
CSO Quality Review Self-Assessments (QRS)82
CSO Quality Training83
Statistical Quality Appraisals
Data Protection, Confidentiality and the Legislative/Regulatory Environment
Data Management and Security86
Guaranteeing Statistical Confidentiality
Access to Administrative Data in the CSO
Access to Administrative Data in the CSO
Access to Administrative Data in the CSO
Access to Administrative Data in the CSO
Access to Administrative Data in the CSO
Access to Administrative Data in the CSO
Access to Administrative Data in the CSO
Access to Administrative Data in the CSO. 89 Use of Statistical Statutory Instruments (SIs) 91 Statistical Legislative Environment 92 Appendices 95 Appendix 1. European Statistics Code of Practice (ES CoP) 96 Appendix 2. Quality Assurance Framework of the European Statistical System (ESS QAF) 97 Appendix 3. List of ESS Quality Performance Indicators (ESS QPIs) 100 Appendix 4: The UNECE Generic Statistical Business Process Model (GSBPM) 102
Access to Administrative Data in the CSO
Access to Administrative Data in the CSO. .89 Use of Statistical Statutory Instruments (SIs) .91 Statistical Legislative Environment .92 Appendices. .95 Appendix 1. European Statistics Code of Practice (ES CoP) .96 Appendix 2. Quality Assurance Framework of the European Statistical System (ESS QAF) .97 Appendix 3. List of ESS Quality Performance Indicators (ESS QPIs) .100 Appendix 4: The UNECE Generic Statistical Business Process Model (GSBPM) .102 Appendix 5: CSO policies, standards and key legislation .103 Sources and References .105





Introduction











Context

The Central Statistics Office (CSO) is Ireland's national statistical office with a mandate under the Statistics Act, 1993, for "The collection, compilation, extraction and dissemination for statistical purposes of information relating to economic, social and general activities and conditions in the State". CSO official statistical outputs are used to inform analysis and decision making by providing an accurate picture of Ireland's economic and social performance. As the provider of official national statistics, it is vital that the statistical outputs of the Office continue to be regarded as of the highest quality by our end users.

This handbook sets out the principles, practices and general approach to be adopted in the production of official statistics in the CSO using the European Statistics Code of Practice (ES CoP)¹ and the United Nations Economic Commission for Europe Generic Statistical Business Process Model (GSBPM)² as its key frames of reference. While the focus of this document is primarily on CSO processes, policies and work practices, the contents of this handbook also provide a framework for other statistical producers in the Irish Statistical System (ISS). Many of the internal CSO support structures referenced throughout this document are also accessible to other producers of official statistics in the ISS.

This handbook is designed to give a broad overview of the requirements for statistical quality management and aims to:

- Clarify what is meant by quality in statistics, across its different dimensions and production phases
- Set standards for statistical production in the CSO and the ISS in adherence to the principles of the ES CoP
- Confirm and set out standards for statistical production in the CSO in adherence to relevant Office Policies relating to Quality and Methodology
- Raise awareness of Office policies and national and international Quality standards
- Identify approaches to continuing improvement of the overall quality of statistical outputs
- Enhance awareness that everyone working in the production of Official Statistics has a responsibility for quality

This handbook supports the ongoing implementation of the CSO's Quality Management Framework (QMF) in its aim to ensure the quality of our statistics continue to meet international best practice and to maintain trust in the data we produce. The trust of the Irish people and all the users of official statistics is earned and maintained through transparent and responsible management of statistical information and of the statistical systems that produce it.

The CSO consistently strives to efficiently meet user needs for high quality statistics while working to reduce the burden on respondents. Effective quality management involves striking the appropriate balance between the resources available and the needs of users, while explaining how these decisions influence the quality of our statistical outputs.

Our commitment to quality, as represented in this handbook and the Office-wide compliance with the principles of the ES CoP, enable all staff in CSO to fulfil our role as a trusted provider of independent information and knowledge for effective debate and decision-making across government, business and society.

² See Appendix 4





¹ See Appendix 1

Definition of Quality and dimensions of Quality

Quality concerns the ability of organisations, processes and products to fulfil the purpose for which they have been established. Quality is now generally defined as **fitness for use/fitness for purpose** in terms of user perspectives – it is user needs that define quality. Different users may have different needs that must be balanced against each other to give the quality concept a concrete definition.

Statistical product quality is multi-dimensional and quality assurance techniques must be used to ensure that appropriate consideration is given to each relevant dimension. The dimensions of quality are interrelated and there are trade-offs between dimensions. Hence, applying protocols for appropriate management of each dimension is essential.

Everyone involved in statistical production needs to recognise and address competing demands in quality management and should understand that focus on one dimension of quality can sometimes lead to a deterioration in another dimension. For example, accuracy may be compromised when there is a drive to improve timeliness if this results in less data being collected or limiting time to complete all checks on the data. Efforts to improve the quality of official statistics must also take account of factors such as existing budgets, availability of specialised resources and response burden.

Quality Commitment

Institutional Level - Quality is at the core of everything we do

As an organisation, the CSO is committed to ensuring the quality of the statistics we produce. This Quality commitment encompasses the policies, practices and decisions undertaken throughout the CSO to ensure that statistical quality is at the forefront of the design and operation of each process. This is not a static commitment but reflects an ongoing effort to maintaining and improving statistical quality. The quality commitment means constantly assessing the impact of changes in the environment in which we operate and responding appropriately in how we manage quality. Advances in technology, methodology and society, combined with the availability of alternate or new sources of statistical information, contribute to changing expectations of relevancy, efficiency and expediency.

Individual Level - Everyone in the CSO has a role to play in assuring quality

Every person working in the CSO has a contribution to make to quality. Statistical production can be a complex operation and it is not possible to achieve quality by simply inspecting the final product or output. Quality needs to be built into every process from the outset because what we do at all stages of the statistical production process has an impact on the quality of the outputs. Success in assuring quality requires the sound application of knowledge and expertise by people at all levels within the CSO. The steps we take to ensure quality require constant evaluating, particularly by those who are directly involved in the processes.

Quality Reporting - The requirement to inform users about quality

Users require information about data quality so that they can evaluate its fitness for use. While timeliness can be easily observed from the user's perspective, other dimensions are not always as obvious. The user depends on the CSO for guidance in relation to accuracy and other dimensions of quality as they often have no other way of assessing it. To avoid misinterpretation, key information about quality should be prominent, easy to find, unambiguous and clearly signposted. Vital messages about quality – those that have a profound impact on what information can be drawn from the data – should be presented upfront. A comprehensive standard quality and methods report must be produced and made available to users at the dissemination stage for each statistical product. The content of each quality report should be reviewed as part of the regular dissemination process and must be updated at least once a year at a minimum. Inevitably, despite best efforts to maintain standards, there are variations in the quality of statistical outputs from iteration to iteration and quality reports should reflect these changes.





European Statistics Code of Practice

The ES CoP builds upon a common European Statistical System (ESS) definition of quality in statistics. The ES CoP sets out 16 key principles for the production and dissemination of European official statistics and the institutional environment under which national and Community statistical authorities operate. These principles in turn are built on the United Nations Fundamental Principles of Official Statistics, which sets out the basic framework under which all official statistical organisations operate internationally.

The Code of Practice requires official statistics producers to inform users about the quality of statistical outputs using the European Statistical System's Dimensions of Quality. These dimensions, presented in the table below, aim to give users a broad understanding of quality and fitness for purpose.

ES CoP Quality Dimensions for Statistical Outputs

Relevance:	The degree to which the statistical product meets the requirements of the user – how well the statistics meet current and emerging needs of the users.
Accuracy and	For survey data, accuracy is the degree of closeness between an estimate derived from the survey data and the true value.
Reliability:	Reliability refers to the closeness of the initial estimate to the subsequent estimates
Timeliness &	Timeliness is the time elapsed from the period to which the statistics relate to and the date of release of the statistics.
Punctuality:	Punctuality refers to the elapsed time between the actual and planned time and date of release
Coherence &	Coherence reflects the degree to which statistical information is logically consistent and can be brought together with information from other sources or different time periods.
Comparability:	Comparability is the degree to which data can be compared over time and domain.
Accessibility and	Accessibility refers to the ease with which users can access the data and the formats in which the data are available.
Clarity:	Clarity ensures that the statistics and their corresponding metadata are presented in a form that facilitates proper interpretation and meaningful comparisons.

Back



Office Policy

The Statistics Act 1993 and Regulation (EC) No. 223/2009 on European Statistics are the primary legal instruments which allow the Central Statistics Office to carry out its duties in the production of official statistics.

This handbook of Quality Standards and Guidelines is based on relevant governing legislation and on the ES CoP which sets the standard for developing, producing and disseminating European statistics. The handbook also incorporates elements of best practice outlined by international bodies such as the Organisation for Economic Co-operation and Development (OECD), United Nations Economic Commission for Europe (UNECE) and other National Statistical Institutes and their contribution is gratefully acknowledged.

This handbook updates Quality in Statistics, Standards and Guidelines, Volume 1, 2006. It is Office Policy that all staff familiarise themselves with these Standards and Guidelines and apply them to the areas of statistical production that they are responsible for, so that the guidelines are referenced and utilised appropriately during all phases of statistical production.

How to use this handbook

The handbook should not be treated as a checklist nor as a technical statistical manual. While examples of checks/indicators are provided throughout the handbook there is no minimum number of quality activities required for compliance. The lists of checks/indicators are included to prompt everyone involved in statistical production to constantly examine, re-evaluate and strive to improve the quality of the statistics they are responsible for producing. The handbook refers to and summarises various existing Office policies and standards relevant to statistical production and users should refer to these for greater detail, as required.

An electronic version of this handbook is available on the CSO's internal Quality Information System and on the CSO website (www.CSO.ie). The electronic version of the handbook will continue to be updated and expanded with relevant new Office policies and standards as they become available.

Acknowledgements

This handbook is built on quality guidance from other national and international statistical organisations, CSO Standards and Guidelines Volume 1 2006 and the policies and guidance produced by various internal CSO teams. The CSO would like to acknowledge the fact that this handbook is heavily built on their work and thank them for this.





Quality through the Generic Statistical Business Process Model (GSBPM)







Using the GSBPM to build quality into all phases of the statistical production process

The UNECE Generic Statistical Business Process Model (GSBPM) describes and defines the set of business processes needed to produce official statistics and can be used as an aid in understanding and managing statistical production. Internationally, the GSBPM provides a standard framework and harmonised terminology to help statistical organisations to modernise their statistical production processes, as well as to share methods and components.

The GSBPM is used in the CSO to provide a framework for integrating data and metadata standards, as a template for process documentation including process maps, for harmonising statistical computing infrastructures, and to provide a framework for process quality assessment and improvement. This handbook utilises the GSBPM as a reference frame for examining various quality dimensions and the application of indicators and measures related to the different stages of statistical production.

It is important to note that the GSBPM, consisting of 8 phases and various sub-processes, is not a linear model but should be viewed as a matrix through which there are many paths, where sub-processes are not necessarily followed in strict order. This chapter of the handbook summarises the phases of the GSBPM and describes checks and indicators that should be considered in examining quality through each phase. Some quality indicators and measures refer to survey data only, others refer to administrative data only and further indicators and measures refer to both survey and administrative data. The quality indicators described in this chapter are based on ESS guidelines on quality and performance indicators. There are 16 standard indicators in the ESS guidelines which are listed in the table below.

European Statistical System standard Quality and Performance Indicators (QPIs):

ESS QPI R1. Data completeness – rate	ESS QPI TP1. Time lag - first results
ESS QPI A1. Sampling error – indicators	ESS QPI TP2. Time lag - final results
ESS QPI A2. Over-coverage – rate	ESS QPI TP3. Punctuality - delivery and publication
ESS QPI A3. Common units – proportion	ESS QPI CC1. Asymmetry for mirror flows statistics - coefficient
ESS QPI A4. Unit non-response – rate	ESS QPI CC2. Length of comparable time series
ESS QPI A5. Item non-response – rate	ESS QPI AC1. Data tables – consultations
ESS QPI A6. Data revision - average size	ESS QPI AC2. Metadata - consultations
ESS QPI A7. Imputation – rate	ESS QPI AC3. Metadata completeness - rate

N.B Some ESS QPIs are calculated by the producer of the statistics and some are calculated by Eurostat

While the following sections of the handbook present examples of the various quality checks in a list format, there is no specific minimum number of quality measures or indicators that a statistical production manager must complete but each should be examined to identify which are the most appropriate or useful in delivering quality.











This first step in statistical production is initiated when a need for new statistics is identified, when feedback is received on current outputs, or when evaluation of existing statistics prompts a review. The first task in planning statistical production is to understand the research problem, regulatory need or information gap that gave rise to the need for information and then set objectives which will guide all subsequent steps.

Meeting objectives might require introducing a new survey, redesigning an existing survey or repurposing existing products/sources. Information needs as stated in the objectives must justify any response burden that will be generated.

The planning process for a new statistical activity or the redesign of an ongoing activity should include the definition of broad objectives, a targeted user population and the key questions or issues on which information is sought for analysis. To translate this initial planning into actual production, objectives and expected uses should be stated as precisely as possible to help ensure that the new or redesigned activity will meet specific user needs, including regulatory requirements. The ideas or information to be explored, together with clear, precise requirements of specific data needs, data quality expectations, budget, delivery dates, concepts, definitions, units of analysis and target population should be defined to a sufficient extent to allow data users to determine if the expected statistical output will meet their needs.

Sub-Processes



Back

Quality guideline checks for the Specify Needs Phase		
1.	The purpose for which the statistics are required is clearly stated (initially in a business case and later in GSBPM based survey documentation and associated manuals, background notes, quality reports etc.).	
2.	The key users/stakeholders (internal and external) have been identified and consulted.	
3.	The needs of key users and details on how they intend to use the data products, including specific regulatory requirements and planned output tables or analysis, have been discussed and documented.	
4.	The statistical producer understands users' needs, knows what the statistical product is expected to deliver, how users expect it to be delivered and their expectations around timing of delivery.	
5.	Existing data sources and products have been fully examined for their potential to meet user requirements. Gaps/constraints are documented to justify any new data collection which may result in additional response burden or costs.	
6.	Potential administrative sources have been examined in detail to fully comprehend the purpose for which the administrative data was originally intended, the structure and coverage of the data and how the data was compiled. An assessment has been made (and documented) of the quality and stability of the administrative data source for statistical purposes.	
7.	Where use of existing data sources (including administrative data sources) is proposed, an assessment has been undertaken of the legal framework, continuity of data supply, data limitations and any contingency plan requirements.	
8.	The concepts to be measured utilise existing statistical standards to the greatest extent possible and any use of non-standard concepts has been assessed, justified and documented.	
9.	The implications of all constraints on production of the proposed statistics (e.g. legal frameworks, resources, data collection options, data security) have been considered.	
10.	Consideration has been given as to how other National Statistical Institutions produce similar data and advice has been sought from relevant experts (e.g. Methodology, business experts).	
11.	A detailed assessment of the costs/benefits/constraints of the proposed statistical production and outputs has been prepared and the appropriate senior management/ governance body has confirmed that the resource requirements can be met and that these represent good value.	
12.	Formal approval to proceed to the design phase has been received from the appropriate senior management/governance body.	



Primary Quality Dimension(s): Relevance, Cost Effectiveness

Typical Quality Indicators for the Specify Needs phase		
Extent of user consultation	• What information is available on the classification of users/key users?	
Extent to which users' needs are met	 What statistical outputs are required by users and for what uses? What proportion of users' statistical needs are being met? What user needs are unmet? Why can these not be met? How important is timeliness of delivery for users? How recently have users confirmed satisfaction that their statistical needs are being met? 	
Extent of comparability and reusability	 What proportion of concepts and definitions meet existing or international standards? 	
Extent of potential of existing sources, including administrative data sources, to meet users' statistical needs	 What is the assessment of the existing sources in terms of: Security & continuity of data supply Accessibility and stability of systems for data supply Protocols for communication (queries) on data supplied Percentage of units not belonging to the target population Percentage of units missing from the target population Coverage of the data Absence of values for key variables Missing values in the source Total percentage of empty cells Clarity and completeness of metadata for all variables intended for use 	
Extent to which resource requirements for proposed statistical production and outputs represents good value	 What assessment has been made of the costs/benefits/constraints of proposed statistical production and outputs? Have all resourcing requirements and contingencies been outlined in a business case and signed off by the appropriate authority? How will costs be monitored and controlled on an ongoing basis? 	

Who should I Consult?

- Users
- CSO Methodology Unit
- CSO Administrative Data Centre
- Eurostat

- Other Subject Matter Experts internal/external
- CSO Quality Management Support and Assurance
- CSO Communications & Dissemination
- CSO Technology





Design Phase – "Planning for quality"

The design phase involves work undertaken to define the statistical outputs, concepts, classifications, methodologies, collection instruments and operational processes that will be applied in statistical production with the aim of designing quality in from the outset. This phase occurs prior to the first iteration of statistical production, and, for statistical outputs that are produced on a regular basis, it also occurs whenever improvement actions are identified in the Evaluate phase of a previous iteration.

The metadata and quality assurance procedures which need to be ready for use later in the statistical business process should be specified during the design phase. Quality considerations need to be at the forefront in this phase as quality is not something that can be "inspected in" at a later point but must be built into the process from the outset – Aim to ensure Quality by Design.

Survey design should utilise national and international standards wherever possible to reduce the length and cost of the design process and to enhance the comparability and usability of outputs. Efficiency can also be achieved by reusing or adapting elements from existing processes, systems and methodologies where appropriate. Relevant expertise, particularly methodological, technological and domain specific, should be consulted at this stage to ensure that the survey design is optimised. Sample design should incorporate steps to address sampling bias and error. Precision requirements for estimates from key variables for specified stratification levels are recommended where appropriate.

Sub-Processes:



Qua	lity guideline checks for the Design Phase	
1.	Statistical outputs have been designed to follow existing standards and classifications where possible and proposed deviations have been tested, justified and approved by an appropriate body (e.g. by CSO Classifications Unit).	
2.	The collection mode, methods and instruments to be used (including legal instruments) have been defined following detailed research and consultation.	
3.	Statistical databases are designed and arranged in a way that facilitates data linkage, using unique identifiers for statistical units as appropriate while ensuring data security and privacy considerations are addressed.	
4.	Questionnaire design takes account of user requirements, resources, administrative and processing requirements and the nature and characteristics of the target population.	
5.	The statistical variables to be collected, the variables that will be derived from them and the variables to be disseminated have been defined.	
6.	Sample design is based on detailed precision requirements, to the correct level of stratification, and has been validated by appropriate experts (e.g. Methodology Unit).	
7.	Sample design and planned operations contain procedures to check and to minimise coverage error.	
8.	The design ensures that the same frame is used for surveys with the same target population (to improve coherence, consistency, maintenance efficiency and to facilitate combination of estimates).	
9.	The design for statistical production clearly defines all technical and scheduling specifications for processing and production systems, technological platforms and workflows, following detailed consultation with appropriate expertise (e.g. CSO Technology).	
10.	Quality assurance procedures and controls have been built into the design of all phases of statistical production and formalised in Service Level Agreements (SLAs) where appropriate.	
11.	Design of the statistical production system includes specific steps to ensure compliance with Data Protection legislation, Data Management policies and confidentiality requirements, including potential requirements for a Data Protection Impact Assessment (DPIA) or a Transparency Notice at the collection stage. Data minimisation requirements have been considered, with justification for all proposed variables and requirements around use of special identifiers, such as PPSN, have been met (e.g. updating PPSN Register of Users).	
12.	The use of tools to facilitate reuse of questions and attributes – such as question libraries and questionnaire tools (e.g. Colectica) – has been fully considered and utilised where possible.	
13.	The collection design, systems design and methodological design, has been independently assessed, costed and validated (e.g. by Technology Division and Methodology Unit). Formal approval to proceed to the build phase has been received from the appropriate senior management/governance body.	

Back

Next 🕥

Primary Quality Dimension(s): Relevance, Coherence, Accuracy, Sound Methodology

Typical Quality Indicators for the Design phase		
Extent to which planned outputs will satisfy user requirements	 What proportion of users' statistical needs are being met/what proportion are unmet? Are all regulatory requirements being met? What is the ratio of data cells expected to be delivered to the number of data cells required – ESS QPI R1 	
Extent to which the survey population matches the target population	 What is the timeliness of updating of the survey frame? What is the level of over coverage in the frame? - ESS QPI A2 How will you identify potential under coverage? How is sampling error measured? - ESS QPI A1 	
Extent to which data will be revised	 How will the size and frequency of data revisions be monitored and reported? – ESS QPI A6 	
Extent to which each process step contributes to timing of delivery of final output	 What is the planned timing for each process step (time lags) and what is the timing for delivering the final output? – ESS QPI TP2 	
Extent to which quality will be measured and checked	 Number and type of Quality Performance Indicators that are calculated/reviewed and recorded for each GSBPM sub-process (When was the methodology last reviewed? Is this methodology used in other NSIs?) Number and type of edit checks? Expected edit failure range for an iteration? 	
Extent to which confidentiality and data access is protected	 What confidentiality rules apply to the data? What controls are in place to protect data? What procedures are in place for micro data access for researchers? Has data minimisation been applied? 	
Extent to which systems and data can be reused	 What proportion of concepts and definitions used meet/do not meet existing or international standards? Do the systems, processes and methodologies reuse established/ international standards? Can they be used by other surveys? 	

Who should I Consult?		
CSO Methodology Unit	Eurostat	
CSO Administrative Data Centre	CSO Questionnaire Design Unit	
CSO Technology Applications - Collect	Other Subject Matter Experts – internal/external	
CSO Data Office	CSO Quality Management Support and Assurance	

Back



This phase builds and tests the components specified in the design phase to ensure that the production solution is ready for use in a "live environment". The design phase determines what processes, instruments, information and services are to be assembled and configured in the build phase. New services are only built by exception where gaps occur and should be built to be reusable or shared in so far as possible. Typically, the build phase only occurs for the first iteration or when an evaluation/review prompts a redesign and rebuild.

The Build phase requires careful coordination between its different elements. Consideration must be given to the collection technique and potential usage of administrative data or registers. The collection instrument is generated or built based on the design specifications and the contents and functioning of that instrument must be tested thoroughly in this phase (e.g. testing the questions in a new questionnaire).

The Build phase should measure the functioning of the collection instrument in terms of response burden and ensure that, if possible, there should be a direct link between collection instruments and the statistical metadata system, so that qualitative information from respondents can be captured during the collection phase. It is important to safeguard against inadvertent, untested, deviations from the design to accommodate changes during the build phase.



Back

Next

Qua	Quality guideline checks for the Build Phase		
1.	The collection instrument is based on the exact design specifications as set out in the design phase (including multi-mode collection or extractions from existing statistical or administrative data sets).		
2.	A detailed project plan has been agreed by all those involved in the build phase (e.g. IT, Finance, subject matter experts), setting out deliverables, milestones, resources and schedules, with defined management and communications processes.		
3.	Data capture procedures and data collection tools and instruments such as electronic systems/questionnaires are tested using appropriate methods (e.g. questionnaire pre-test, pilot in real scenario, in-depth interviews, focus groups, interviewer supports) to ensure simplicity, effectiveness and data protection prior to implementation.		
4.	Collection systems for administrative and other data sources are tested before use, including procedures for linkage and integration of data.		
5.	All planned activities in the Process, Analyse and Dissemination phases have been tested.		
6.	All programs and routines in the production system have been individually tested and signed off, ensuring that data correctly enters and exits and that all functionalities have been executed as expected.		
7.	All testing is completed prior to use in the live environment, including testing of interactions between assembled and configured services to ensure coherence between processes, information and services.		
8.	Collection and processing systems are built to automate capture of paradata (data elements produced as a by-product of survey data collection which can be used to better understand survey errors and costs) and metadata to the greatest extent possible to facilitate ongoing evaluation and reuse in dashboards and other corporate information requirements.		
9.	The collection instrument/questionnaire has been reviewed independently by someone with expertise in questionnaire design.		
10.	The configured services and processes have been formally signed off (e.g. by Technology) as fit for purpose for use in the statistical production areas.		
11.	Any changes from the original design specifications for services, IT systems and processes made during the build phase have been documented, re-examined against user requirements and tested to ensure that the build remains fit for purpose.		
12.	New services or components that have been identified are built to be reusable within statistical production.		
13.	Comprehensive documentation/instructions on IT systems, work flows, quality checks and target measures are available for end user staff.		

Back

Typical Quality Indicators for the Build phase		
Extent to which collection instruments, systems, processes and workflows have been tested in the live environment prior to full production	 Has the survey questionnaire and delivery system been fully tested to ensure it meets expectations? Have all IT systems and procedures been tested and deemed acceptable? 	
Extent to which system testing procedures and test results have been documented	 Has all testing (and results of tests) been fully documented? Have all changes to systems and procedures been documented and justified? 	
Extent to which documentation/instructions on IT systems, work flows, quality checks and target measures are available for end user staff	 Are detailed and comprehensive instructions available for staff to enable them to operate efficiently and effectively? Are Service Level Agreements in place to ensure continuing support for systems maintenance and enhancement? 	
Extent to which processes, sub-processes and production of quality metrics/ metadata are automated	 Is statistical production automated to the greatest extent possible? How many processes are automated and how many are not? Is Metadata captured automatically? Are Quality Performance Indicators/Metrics calculated automatically to facilitate quality management? 	
Estimated time frame for all phases of production	• Has the timing of each step of processing been tested to see how these contribute to the time lag for final results? – ESS QPI TP2	
Extent to which user requirements are fulfilled by the build	• Have all key users expressed satisfaction regarding the proposed data and metadata outputs, information systems, dissemination formats & graphical supports?	

Who should I Consult?

- CSO Methodology Unit
- CSO Administrative Data Centre
- CSO Technology
- CSO Data Office

- Eurostat
- CSO Questionnaire Design Unit
- Other Subject Matter Experts internal/external
- CSO Quality Management Support and Assurance



This phase details the steps undertaken to obtain the information needed to create statistics using the most appropriate mode available e.g. through a survey, by extraction of information from statistical or administrative sources, or a combination of modes. The collect phase brings data and metadata into an environment that allows further processing but does not include any changes to the data itself as this occurs at the processing phase.

The population frame and the appropriate sample of the target population for inclusion in the data collection phase must be finalised during this phase. This involves utilising the appropriate frame or register and valid procedures for sample selection and coordination of samples to manage overlap and response burden. A data collection plan is required to ensure that people, processes, technology and legal/administrative agreements are in place to ensure data delivery.

This phase applies to data collection from administrative sources and from survey or mixed mode sources. Data collection manages the extracting or receiving of data from initial contact with data providers and any subsequent follow-up/reminder and respondent management actions. Basic validation of the structure and integrity of information received at an early point is advisable (e.g. checking that data is in the right format and contains the expected fields).

Quality controls should be used to manage the quality of the data collection and data capture processes. Meaningful feedback must be provided to those involved in data collection in a timely fashion. Performance metrics (e.g. numbers of interviews completed, appropriately stratified survey response rates) can be monitored against specified targets and any major deviation from these targets may require the provision of feedback or additional training for processing staff, interviewers or field staff. Discrepancies between planned versus actual performance, timelines and costs must be closely managed through analysis of collection process metadata (paradata) to ensure the collection activities meet the targeted requirements. The collect phase is completed when the collected data and metadata are loaded into a suitable electronic environment and validated as ready for further processing.

Sub-Processes:





Next

Quality guideline checks for the Collect Phase		
1.	Appropriate systems and staffing are in place to meet the targets set out in the sample design.	
2.	A plan is in place for recruitment, training and management of staff, including field staff.	
3.	Legal requirements for data collection have been fully considered and appropriate measures are in place (e.g. potential need for a Transparency Notice).	
4.	A data collection plan, with detailed scheduling and performance targets, is in place for all aspects of data collection, from the requesting of data to follow-up/reminder procedures.	
5.	The structure and integrity of information received is validated at the earliest point (e.g. checks that administrative or survey data is in the right format and contains the expected fields).	
6.	The same frame is used for surveys with the same target population, to avoid inconsistencies and to reduce costs of frame maintenance and evaluation. The frame is updated in a timely manner.	
7.	Appropriate sampling techniques are used to minimise sample sizes to achieve the target level of accuracy. Sample surveys are coordinated to distribute the burden on respondents.	
8.	The level of coverage from administrative sources has been determined. Agreement has been reached with the data owner of the source for monitoring and managing coverage and data supply, including handling of issues around data quality, timeliness, data structure, coverage, classifications and concepts etc.	
9.	Appropriate procedures are in place to protect data confidentiality during data collection.	
10.	Controls are in place to monitor the efficiency and effectiveness of data collection. Data from all data sources are reviewed and validated to identify potential problems, errors and discrepancies such as outliers, missing data and miscoding.	
11.	Systems are in place to ensure meaningful feedback is provided to processing staff, interviewers and field staff in a regular, timely, manner, based on analysis of collection process metadata.	
12.	For survey data, an acceptable response rate has been set – to an appropriate level of stratification, based on precision requirements. Measures are in place to reduce unit and item non-response and to manage this systematically.	
13.	Data collection process information (paradata) are gathered automatically to facilitate management and evaluation of processes.	
14.	Respondent burden and feedback is systematically managed, evaluated and documented.	

Primary Quality Dimension(s): Accuracy & Reliability, Relevance, Timeliness & Punctuality, Managing Respondent Burden

Typical Quality Indicators for the Collect phase	
Extent to which the survey frame is up to date	• What is the time lag between the creation of the frame and its use?
Extent of survey coverage issues	 What is the level of over-coverage/under-coverage? – ESS QPI A2
Extent of Sampling errors	 How will sampling error be measured - Standard error/coefficient of variation for key variables? – ESS QPI A1 Is there a level of sampling error beyond which data will be of insufficient quality to publish?
Extent of non-response in the survey	 Is non-response measured at an appropriate level of stratification? – ESS QPI A4 What is the level of item non-response in the survey? – ESS QPI A5 Is there a level of non-response above which data will be of insufficient quality to publish?
Extent of data capture errors	• How many units or items fail data capture checks and validations?
Extent to which administrative data supplement/replace direct collection	 What proportion of records/variables are sourced from administrative data?
Estimated time frame for all phases of production	• What is the timing required for each step of data collection and how does this contribute to the time lag for final results? – ESS QPI TP2

Who should I Consult?		
CSO Methodology Unit	• Eurostat	
CSO Administrative Data Centre	Other Subject Matter Experts – internal/external	
CSO Technology	CSO Quality Management Support and Assurance	

\bigcirc	Back





The process phase integrates, checks, cleans and prepares input data for analysis and dissemination as statistics. The "Process" and "Analyse" phases can be iterative and parallel as analysis can reveal a broader understanding of the data, which might make it apparent that additional processing is needed. Activities within the "Process" phase may commence before the "Collect" phase is completed to enable the compilation of provisional results where timeliness is an important concern for users or to increase the time available for analysis.

Integration of input data from various sources can occur during the process phase involving matching/record linkage, classification and coding of data. Data is reviewed and validated to identify potential problems, errors or discrepancies. Where data are incorrect, missing, or unreliable, new values may need to be inserted through the application of editing and imputation methods. New values which are written to the dataset must be flagged and documented with appropriate metadata so that any changes made can be understood and tracked over time. The process phase also involves the derivation of new variables needed for statistical outputs through the application of arithmetic formulae/models. Weights may be created to "gross up" results for sample surveys, to adjust for non-response or for other normalisation purposes. Finally, aggregates are calculated by applying weights and summing data to derive appropriate totals and appropriate datasets are created for macro level analysis.



Sub-Processes:

Back

Next

Quality guideline checks for the Process Phase		
1.	A fully documented system is in place to integrate, check and clean input data.	
2.	Standard classifications are used in classifying and coding input data and any use of non- standard classifications is justified and documented.	
3.	Information on type and cause of error (measurement error, coding error, etc.) is analysed to help reduce error rates.	
4.	Steps are in place to optimise the level of editing based on an assessment of the influence of editing on final results.	
5.	Edit parameters and other quality checks are reviewed systematically and on a regular basis.	
6.	Information on editing rates for each survey iteration is captured, retained and used for evaluation and comparative quality analysis.	
7.	The variance introduced by imputation is measured and the imputation methodology is regularly reviewed with the help of appropriate expertise to ensure it meets current best practice.	
8.	The methods/practices/systems used to process data, including data integration, editing, imputation, estimation and weighting, have been documented and are fully transparent.	
9.	Consideration has been given to the need for data anonymization at the most appropriate point to meet data protection requirements.	
10.	A plan is in place for the completion of each step of the processing phase to meet set targets for timeliness and quality.	
11.	Consideration has been given to developing and utilising appropriate quality measurements and targets such as minimum acceptable coefficients of variation (CVs) or standard error thresholds.	
12.	Procedures are in place to ensure that data held in processing datasets is harmonised with data used to produce the final output where survey returns are edited as part of the analysis phase. Appropriate commentary must be provided to explain changes.	
13.	Data Owners/Custodians responsible for processing personal data have clearly set out the justification/legal basis for processing this data as required under legislation (e.g. consent; contract; legal obligation; vital interests; public task; or legitimate interests). An appropriate Record of Processing Activity (RoPA) is prepared and maintained.	



Primary Quality Dimension(s): Accuracy & Reliability, Coherence, Timeliness & Punctuality

Typical Quality Indicators for the Process phase	
Extent of survey coverage issues	 What is the rate of over-coverage? - ESS QPI A2 What is the level of under-coverage and what actions are in place for its reduction and the associated bias risk?
Extent of Sampling errors	 How will sampling error be measured - Standard error/coefficient of variation for key variables? - ESS QPI A1 Is there a level of sampling error above which data will be of insufficient quality to publish?
Error rate	• How many units/items fail keying/scanning checks?
Extent of automatic coding	 Ratio between number of values automatically coded and the total number of values submitted for coding
Edit failure rate	• How many edits are generated for each iteration and how does this compare with previous iterations/other surveys?
Extent of imputation	 What proportion of units have had values imputed for a key variable (because value was missing/implausible/inconsistent)? - ESS QPI A7 What is the relative contribution of imputed values to the estimate/result for a particular statistic?
Extent to which administrative data supplement/replace direct collection	 What proportion of records/variables are common to the administrative data source and the survey? – ESS QPI A3
Estimated time frame for all phases of production	• What is the timing required for each step of processing up to delivery of a clean unit file and how does this contribute to the time lag for final results? – ESS QPI TP2

Who should I Consult?

- CSO Methodology Unit
- CSO Administrative Data Centre
- CSO Technology

- Eurostat
- Other Subject Matter Experts internal/external
- CSO Quality Management Support and Assurance



Analyse Phase – "Confirming quality and interpreting information"

This phase includes the sub-processes and activities that enable the preparation of statistical content while also ensuring that statistical outputs are "fit for purpose" prior to dissemination. Analysis is the principal tool for obtaining information from the data and this phase involves ensuring that the outputs are of sufficient quality for the user and that this quality can be explained to users so that the data can be interpreted appropriately.

This phase transforms the data inputs into statistical output by evaluating, summarising and preparing to communicate the results in a way that provides understanding for users. The analysis phase requires interpretation of the data and consideration of how best to explain the data to users. Data and metadata are examined in detail and compared with previous iterations and other relevant data sources to ensure data is coherent. Data analysis plays a key role in assessing data quality by identifying quality problems in each iteration of statistical production.

Appropriate checks to prevent primary and secondary disclosure of confidential information should be undertaken and the necessary controls applied, in line with Office Policy. Information on interpretation and on measures of consistency or uncertainty of the statistics should be discussed with appropriate internal subject matter experts before the statistical content is approved for release.

This phase occurs in every iteration. The "Analyse" phase and sub-processes are generic for all statistical outputs, regardless of how the data were sourced.



Sub-Processes:



Qua	ity guideline checks for the Analyse Phase	
1.	The results are analysed, compared with the results of previous releases and any significant differences are validated and documented.	
2.	Appropriate internal consistency checks are completed (e.g. longitudinal, regional, cross- sectional) and documented.	
3.	Validation of the data is completed in terms of meeting set expectations, (e.g. target population of data source matches the target for analysis, sufficiency of response, expected ratios between variables etc.).	
4.	Additional analytical measurements are produced (e.g. indices, trends or seasonally adjusted series) and macro-editing is performed where required. Paradata are retained for further review, analysis and evaluation.	
5.	Analysis considers the quality impact of unit/item non-response and imputation/grossing effects and how these should be explained to the user.	
6.	Appropriate disclosure control measures, in line with Office Policy, are applied to ensure that the data released does not breach confidentiality rules.	
7.	Coherence checks with other internal and external data sources are carried out and documented.	
8.	Procedures are in place to ensure that any editing of data during the analysis phase is reflected in processing systems with appropriate commentary to explain all changes.	
9.	Quality reports are prepared in advance of dissemination explaining the statistical output in language that is as clear and simple as possible, including explanations of the strengths, limitations or shortcomings of the data in each iteration.	
10.	There are standard procedures, measurements and controls in place for preparation and approval of statistical content for release.	

11. Discussions take place with relevant subject matter experts/managers to confirm the plausibility of results and appropriateness of commentary/presentation prior to release. The inputs and outputs from these discussions are documented.



Primary Quality Dimension(s): Relevance, Coherence, Accuracy, Sound Methodology

Typical Quality Indicators for the Analyse phase	
Sound Methodology	• Does the analyse process use standard methods that correspond to appropriate statistical procedures in ES CoP e.g. indices, trends, seasonal adjustment?
Coherence & comparability	 Is the output comparable with previous iterations? How coherent is the output with other relevant sources, across domains, across mirror flow statistics (e.g. export/import flows between countries)? What is the level of change between provisional and final results?
Relevance	 What proportions of required data cells are being disseminated? – ESS QPI R1
Accessibility & clarity	 How complete is the metadata? What proportion of outputs have provided relevant associated metadata? – ESS QPI AC3
Statistical confidentiality & security	 To what extent is the business process using standard or well known methods for identification and protection of sensitive information? To what extent is the data protected from the risk of disclosure of sensitive information? How much loss in usability of data has been caused by the application of statistical disclosure control procedures?
Accuracy	Are all calculations and totals correct?

Who should I Consult?

- CSO Methodology Unit
- Producers of related statistical outputs
- Head of Division

- Eurostat
- Other Subject Matter Experts internal/external
- CSO Quality Management Support and Assurance





This phase of the GSBPM covers the release of the statistical output to users through various media. The aim of dissemination is to enable timely and effective use of the information produced, thereby meeting the needs of the users.

Final checks to ensure that data are correctly presented and that all of the necessary metadata are in place occur in this phase. All statistical products disseminated must be accompanied by appropriate metadata and documentation of quality. This information must provide a complete and unambiguous record of the underlying concepts and statistical methodology to enable the user to understand the strengths and limitations of the statistical output.

Workflows and interactions with dissemination systems and processes should be documented fully and managed appropriately to ensure that statistical outputs are delivered through various dissemination modes in an efficient and timely manner. Expert or specialist support for dissemination and communications (e.g. Digital Communications, Graphic Design or Classifications sections) should be consulted well in advance of release of statistical outputs and fully utilised to help maximise accessibility and clarity.

The release of dissemination products should be managed closely to maximise the use and understanding of the statistics. Briefings for specific groups and options for the promotion of the statistical products to the widest possible audience should be considered. Errors or revisions to disseminated products must be managed systematically in line with established policies and users must be informed of changes in a full and transparent manner.

User interaction with statistical products should be monitored to ascertain the continuing relevance of the products. This information can be used to enhance customer service and to review and improve promotion of statistical outputs.

Sub-Processes



Back

Quality guideline checks for the Dissemination Phase		
1.	Data and Metadata for release are prepared and tested in advance of release.	
2.	Established standard classifications are used whenever possible in statistical outputs and new classifications are planned and cleared by the appropriate governance body well in advance of publication.	
3.	Updating of data and metadata into statistical output systems, including formatting, loading and linking is automated to the greatest extent possible to minimise error associated with manual intervention.	
4.	Information on the timeliness and punctuality of each statistical release is captured and analysed regularly. Users are kept informed of changes to dissemination schedules.	
5.	Statistical outputs are presented with a view to maximising usability and accessibility for users. Data is presented across multiple dissemination platforms using commentary, tables, charts, infographics and maps designed to make the information more accessible, interesting and more understandable for users.	
6.	Procedures are in place to monitor and analyse user experience.	
7.	Appropriately detailed information about the quality of the data is published at the time of release of the data, clearly relates to the current iteration and includes relevant comparisons to previous iterations. Quality reports meeting national and international standards and requirements are reviewed prior to release and made available on time.	
8.	Clear procedures are in place to deal with revisions/errors and any changes are highlighted to ensure that they are clearly communicated to users in line with CSO Errors and Revisions Policy.	
9.	Appropriate procedures are in place to comply with CSO Policy on Pre-Release Access and the underlying commitment that all users have equal access to statistical releases at the same time.	
10.	Procedures are in place to retain final data used for dissemination and related production code/programs (e.g. CSO Corporate Data Vault/SVN) so that results can be reproduced in the future.	
11.	Appropriate disclosure controls have been applied and checked to ensure confidentiality of data.	
12.	The content of the Methods page on the CSO website is kept up-to-date and meets current CSO guidelines around content and layout.	
13.	Arrangements are made for the provision of microdata for researchers, with accompanying metadata, in accordance with the CSO Data Management Policy.	



Primary Quality Dimension(s): Accessibility and Clarity, Relevance, Timeliness and Punctuality, Coherence, Statistical Confidentiality

Typical Quality Indicators for the Disseminate phase	
Accessibility & Clarity	 Is a release calendar available on the CSO website? When was metadata last updated? Is the metadata complete, available and accessible? - ESS QPI AC3 How many queries have been received from users who have been unable to access the data they require? Is a log maintained of query number and type?
Timeliness	 What is the time lag between the reference period and release of first results? - ESS QPI TP1 What is the time lag between the reference period and release of final results? - ESS QPI TP2
Punctuality	 What is the time lag between the actual delivery date of release of the output and the published target delivery date? – ESS QPI TP3
Quality Commitment	 Is a comprehensive Quality Report published with each statistical product? Are policies on dissemination, pre-release access, treatment of errors and revisions available to users?
Relevance	 How well do the statistical outputs meet current user needs? When was the last survey of users and their needs undertaken? When was a user satisfaction survey last undertaken? How many times have metadata files been viewed by users? - ESS QPI AC2 What website metrics are available for the statistical output? How many analytical and data products/tables have been accessed by users? - ESS QPI AC1 What proportion of required data cells are being output? - ESS QPI R1
Coherence & Comparability	 What is the length of the time series on which comparisons can be drawn/since the last break in the series? - ESS QPI CC2 Is information available on the number, frequency and scale of revisions? - ESS QPI A6 Is appropriate commentary provided around coherence with other outputs?
Statistical confidentiality & security	 What is the level of risk of disclosure of sensitive information? Are clear rules and controls in place around protection of confidentiality and access to microdata for researchers? How much loss in usability of data has been caused by the application of statistical control procedures?

Who should I Consult?

- CSO Methodology Unit
- CSO Classifications Unit
- Eurostat
- Users
- CSO Researcher Coordination Unit
- CSO Dissemination & Communications
- Other Subject Matter Experts internal/external
- Head of Division
- CSO Editorial Board






EVALUATE

Evaluation is central to the concept of continuous quality improvement. Change is a constant feature everywhere - in society, organisational environments and in statistical quality management. A regular, systematic and holistic evaluation of the efficiency and effectiveness of statistical production is required to ensure statistics meet changing quality requirements and remain relevant and fit for purpose. While evaluation of quality occurs throughout all phases of statistical production, different phases or sub-processes may be the responsibility of different managers – the Evaluate phase should bring together all the quality information collected in different phases and ensure this is examined in an integrated manner.

Statistical production is generally a repetitive cycle. Each iteration of this cycle should include an evaluation and feedback loop which enables the introduction of changes and improvements. Information on the quality of data and data processing, collected systematically throughout the statistical production lifecycle, is essential in undertaking a comprehensive assessment. Documented quality indicators and measures play an important role in detecting possible systematic errors or inefficiencies in the statistical processes. By using indicators, we can assess the quality of statistical data, critically evaluate the results obtained and gain insight into the operation of the statistical processes.

Any assessment of a statistical production cycle should ensure that feedback is gathered from all relevant people involved in each of the different phases, including interviewers/field staff, staff involved in collection and processing of administrative or survey data and staff involved in analysis and dissemination. Data management and security procedures should also be reviewed regularly. The review process and outcomes, including any changes to existing processes, should be documented.

When all relevant information and feedback has been examined, an evaluation report should be prepared noting any quality issues that have been encountered in a particular iteration and any recommendations for changes if appropriate. These recommendations should be discussed with relevant stakeholders leading to an action plan which will result in improvements to the overall quality of the statistical product.

The production and maintenance of documentation is a vital part of statistical production as it describes the statistical procedures and processes and ensures the reproducibility and sustainability of statistical production. Reviewing and updating documentation should be undertaken on a regular basis as part of the evaluation phase.

Sub-Processes:



Back

Next

Quality guideline checks for the Evaluate Phase				
1.	A formal, documented, process is in place to evaluate the strengths and weaknesses of survey/output quality for each iteration of the survey/output.			
2.	Each evaluation is carried out in conjunction with other staff who are involved in the production of the statistics (e.g. section staff, field staff, other managers/statisticians, ADC, etc.).			
3.	Dynamic quality performance information is available and is used to evaluate the quality of each survey/output iteration in comparison with previous iterations.			
4.	Staff are aware of their responsibility for quality; regular briefings are held to ensure quality issues are understood and addressed, and planning meetings are held prior to the start of each new production cycle where the previous iteration is reviewed for good practices and opportunities for improvement.			
5.	Users have provided input into the setting of objectives for data quality evaluation (e.g. through user satisfaction surveys or by other user consultations).			
6.	Data on the use of statistics (e.g. web analytics, number and types of downloads, subscriber reports) are collected and analysed to improve statistical outputs.			
7.	Quality Review Self-Assessments are carried out annually and action plans are in place to improve the quality of the statistical processes/outputs.			
8.	Production systems and programmes adhere to current best practice (e.g. SAS code meets CSO SAS Programming Standards and Quality Control guidance).			
9.	Current resources within the business area match the resources required to adequately produce these statistics. Staff within the business area are adequately skilled and knowledgeable in the tasks required to produce the statistics.			
10.	Support provided by other areas (e.g. Technology, Digital Communications, Printing, etc.) is adequate to ensure the efficient production and dissemination of outputs.			
11.	Plans are in place to address any resourcing or skills issues that have been identified.			
12.	Documentation for all stages of statistical production is comprehensive, up to date, fit for purpose and stored in accessible corporate systems so that knowledge is embedded in the organisation (e.g. CSO Quality Information System).			
13.	There are regular and systematic checks on the adequacy and appropriateness of data management and IT security procedures and the results of these examinations are documented. Additional training and procedural updates are implemented as needed.			
14.	Key aspects of statistical production have been validated by independent expertise (e.g. Internal Audit, Methodology, Quality) within an appropriately recent timeframe.			
15.	A risk assessment of statistical production is carried out on annually.			

Primary Quality Dimension(s): Relevance, Accuracy & Reliability, Timeliness, Coherence & Comparability, Accessibility & Clarity

Quality measures and quality indicators

Quality **measures** directly measure specific aspects of quality (e.g. the time lag from the reference date to the release of the output is a direct measure).

Quality **indicators** usually consist of information that is a by-product of the statistical process. As it can be difficult or costly to calculate many quality measures, we can use quality indicators to give insight into quality. For example it is very challenging to measure non-response bias as the characteristics of those who do not respond can be difficult to ascertain so response rate is used as an indicator to give an insight into the possible extent of non-response bias.

Typical Quality Indicators for the Evaluate phase (usually compiled in previous phases)					
Level of Documentation	Documentation - Integrated, up to date and comprehensive				
Performance Metrics	All Key Performance Indicators from earlier phases of the GSBPM				
Use of 16 ESS Quality Performance Indicators across various quality dimensions:	 R1. Data completeness - rate A1. Sampling error - indicators A2. Over-coverage - rate A3. Common units - proportion A4. Unit non-response - rate A5. Item non-response - rate A6. Data revision - average size A7. Imputation - rate TP1. Time lag - first results TP2. Time lag - final results TP3. Punctuality - delivery and publication CC1. Asymmetry for mirror flows statistics - coefficient CC2. Length of comparable time series AC1. Data tables - consultations AC2. Metadata - consultations AC3. Metadata completeness - rate 				
Regular self-assessments	Annual Quality Review Survey (QRS) & Action Plan				
Review meetings	Formal review meetings with staff, managers and other internal stakeholders				
User Consultation	User feedback				
Risk	Risk assessment & horizon scanning				



Who should I Consult?				
CSO Methodology Unit	Other Subject Matter Experts – internal/external			
Eurostat	Head of Division			
• Users	Respondents & non-respondents			
CSO Internal Audit	Field staff/interviewers/internal staff			
CSO Data Office	CSO Quality Management and Assurance			



Risk Assessment in the Evaluate Phase

An assessment of risk to statistical production should be carried out regularly. Risk assessment should consider various dimensions across all the phases of statistical production. The following table outlines an approach based on the UK Office of National Statistics' risk assessment methodology.

Risk Dimension	Sample Considerations	No issues	Some improvements possible	In need of attention	
Sources	Source data are of sufficient quality and reliability to deliver high quality statistics, analysis and insight.				
Methods	The current Methodology deals appropriately with key aspects of production such as: Data Acquisition/ Questionnaire design, Coverage of data (incl. Sample design & estimation), Processing, Editing & Imputation, Analysis, Disclosure Control, etc.				
Systems	IT systems/platforms used (e.g. for data collection, results processing and publication) meet business needs in terms of functionality, performance/stability, sustainability of the systems.				
Processes	Business processes in place (e.g. Data Collection & preparation processes; Results & Analysis processes) are appropriate and efficient.				
Quality of Outputs	The quality of the statistical output is of a sufficiently high standard across each of the following dimensions: Relevance, Accuracy, Timeliness & Punctuality, Accessibility & Clarity, Comparability & Coherence.				
Users	End Users are satisfied with the extent to which the output meets their current needs. The output meets all regulatory requirements and is capable of meeting future user needs and legislative or other changes.				
Reputation	There has been no negative or adverse media or other public commentary regarding the output impacting the CSO's reputation in the last 12 months and no increased likelihood of negative commentary is anticipated.				
People Resources & Skills	Sufficient numbers of staff are available to work on all aspects of statistical production (from collection through to publication). Staff have the necessary training, skills and capabilities to deliver high quality statistics, analysis and advice. Adequate provision has been made to cover the departure of key staff (are there any staff who might leave a critical skills gap if they left?).				



What to consult – suggested reading for GSBPM phases	Specify	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
European Statistics Code of Practice, 2017: https://ec.europa.eu/eurostat/web/quality	\checkmark							
ESS Quality Assurance Framework, 2019: https://ec.europa.eu/eurostat/web/quality	\checkmark							
UNECE. 2019. Generic Statistical Business Process Model (GSBPM), 2019, https://statswiki.unece.org/display/GSBPM/	\checkmark							
United Nations National Quality Assurance Frameworks Manual for Official Statistics (UN NQAF Manual), 2019: https://unstats.un.org/unsd/methodology/dataquality/un-nqaf-manual/	\checkmark							
UNECE Quality Indicators for GSBPM, 2015 https://statswiki.unece.org/display/GSBPM/	\checkmark							
Statistics Canada Quality Guidelines Sixth Edition: https://www150.statcan.gc.ca/n1/en/catalogue/12-539-X	\checkmark							
Republic of Slovenia Statistical Office, Guidelines for Quality Assurance, 2017: https://www.stat.si/StatWeb/en/Methods/	\checkmark							
DESAP The European Self-Assessment Checklist for Survey Managers, https://ec.europa.eu/eurostat/web/quality/quality-reporting	\checkmark							
European Statistical System handbook for quality and metadata reports (2020): https://ec.europa.eu/eurostat/web/quality/quality-reporting							\checkmark	\checkmark
Handbook on Data Quality Assessment Methods and Tools: https://ec.europa.eu/eurostat/web/quality/quality-reporting					\checkmark	\checkmark		\checkmark
UK Statistics Authority Administrative Data Quality Assurance Toolkit, 2019: https://osr.statisticsauthority.gov.uk/guidance/				\checkmark				
ESSnet Quality in Multisource statistics, (Quality in Multisource Statistics/Frames in Social Statistics): https://ec.europa.eu/eurostat/cros/content/essnet-quality-multisource-statistics-komuso_en		\checkmark	\checkmark	\checkmark				
UNECE Generic Statistical Data Editing Model (GSDEM): https://statswiki.unece.org/display/sde/					\checkmark	\checkmark		
A guide to good survey design, Stats New Zealand: https://www.stats.govt.nz/methods/a-guide-to-good-survey-design-fifth-edition		\checkmark						
Handbook on Improving Quality by Analysis of Process Variables: https://ec.europa.eu/eurostat/documents/64157/4373903/06-Handbook-on-improving- quality-by-analysis-of-process-variables.pdf/					\checkmark	\checkmark		
Statistics Netherlands' Quality Assurance Framework at Process Level: https://www.cbs.nl/en-gb/background/2014/12/quality-guidelines-2014			\checkmark	\checkmark	\checkmark	\checkmark		
CSO Data Management Policy (16/2019) – see CSO Data Office Resource page on Lotus Notes		\checkmark						
CSO Writing Style Guide, "Write Well, Write Clearly", 2019								
CSO Policy on Pre-Release Access: https://www.cso.ie/en/aboutus/lgdp/csodatapolicies/csopolicyonpre-releaseaccess/							\checkmark	

CSO Quality Information System: https://qualityinformationsystem.cso.ie/





Specific Components of Statistical Quality





Classifications

Key Learning Point(s):

Consult early when planning new outputs/tables (CSO Classifications Unit)

Utilise standard classifications

Consider implications for linkage and coherency

Standard Classifications

According to the OECD Glossary of Statistical Terms, 'Standard classifications are those that follow prescribed rules and are generally recommended and accepted. They aim to ensure that information is classified consistently regardless of the collection, source, point of time etc'. Ideally, standard classifications contain categories that are mutually exclusive and exhaustive i.e. any given response or datum can be matched to one category only and every possibility is anticipated and catered for. Classifications consist of codes and descriptors which allow survey responses to be put into meaningful categories to produce useful data.

The statutory role of the CSO in the co-ordination of official statistics in Ireland includes the authority to ensure adherence to statistical standards and the use of appropriate classifications.

In the CSO, a standard classification is one that has been approved as the official classification to be used for collecting, processing and outputting data on a particular topic. The CSO Classifications Unit has responsibility for managing standard classifications used in the production of statistical outputs in the CSO. These standards are developed following extensive research of the topic in close consultation with stakeholders. Any planned statistical production should include engagement with CSO Classifications Unit at an early stage to determine input/output requirements and to maximise reuse of existing classifications to the greatest extent possible.

The benefits of use of standard classifications include:

- Better Data Integration the use of statistical standards allows the integration of data from a variety of sources and between different periods of time.
- Improved Survey Efficiency using standard classifications reduces the resources and time requirements associated with many aspects of survey development and maintenance. They assist in planning new survey developments and in reviews of existing surveys. Once a standard exists it can be used in new surveys, thus eliminating the need to create afresh the conceptual definitions, code-files and coding rules that comprise the standard.
- Better Meet User Requirements standard classifications are created in close consultation with stakeholders who are involved in developing statistical surveys. This ensures that classifications used to collect data in surveys will meet users' needs.
- Greater Data Comparability standard classifications ensure that data is comparable over time and between different surveys, as they can be used to provide a link between different datasets. Classifications should be comparable with international standard classification where possible to maximise international comparability.
- More Efficient Use of Resources the use of standard code structures simplifies data processing and dissemination. Reuse of existing and standard classifications means only one version of the classifications coding needs to be stored. Greater use of standard classifications enriches the data produced by the Office.
- More Reliable Statistics The use of standard classifications gives increased confidence in the reliability of statistics.





Use of Administrative Data

Key Learning Point(s): Explore the potential of administrative data as a first option Develop procedures to effectively manage the use of administrative data Build and maintain close relationships with data suppliers Consult appropriate expertise (e.g. CSO Administrative Data Centre) Appropriately utilise key identifiers for location (Eircode), persons (PPSN) and business (Unique business ID) to maximise the value of the data

The term 'administrative data' refers to data holdings that contain information collected primarily for administrative rather than research or statistical purposes. This type of data is typically collected by government departments and other organisations for the purposes of registration, transaction and record keeping, usually during the delivery of a service. They include administrative registers with a unique identifier or other administrative data without a unique identifier.

Administrative data sources possess some advantages over survey sources - since they already exist, costs of direct data collection and further burden on respondents are avoided; they are often available for the complete target population, so may not be constrained by sampling error limitations. In addition, using administrative data to produce outputs can have timeliness benefits.

There may however be some limitations to be aware of in using administrative data for the production of official statistics. These might include the level of control over data quality (particularly in relation to variables that the data owner sees as peripheral) and the possibility of having missing items or missing records (an incomplete file). There may also be difficulties caused by differences in concepts used (which may contribute to bias or coverage problems) and timeliness requirements (because the data collection is outside the statistical agency's control, part or all of the data may not be received on time). There may also be cost implications in utilising administrative data, for instance, computer systems may be needed to clean and complete the data in order to make it useful.

Administrative data sources can be used in the production of statistical outputs in a number of ways such as the creation and maintenance of statistical registers and frames; the complete or partial replacement of statistical collection (via record linkage), editing, imputation and weighting of data from statistical collection and in the evaluation of statistical outputs.

When considering the use of administrative data, it is vital that the statistical producer:

- Understands the data and how it is collected
- Establishes and maintains strong, effective relationships with data suppliers
- Documents quality assurance processes and audit arrangements
- Keeps informed about the data collection systems and influences changes as appropriate
- Publishes clear information about quality for users of the outputs produced using administrative data





Guidelines

- Actively investigate and assess all potential sources of administrative data with a view to utilising these data sources even partially complete or partially inaccurate administrative data may still prove useful in reducing response burden in surveys and the improvement of survey results.
- Consult CSO Administrative Data Centre at the earliest opportunity for guidance and support
- Investigate potential administrative data sources to ensure that the data are of sufficient quality to be used in the production of statistics for the main intended use:
 - Why are the data collected? Will this add or detract from their usefulness for the planned statistical production?
 - How well does the data coverage in the administrative source match the target population?
 - Are there any timing factors that detract from the usefulness of the administrative source?
 - Is there clear information on how data are collected and processed? Are there any issues that would negatively impact the quality of the data for use in statistical production?
 - Are there other CSO users of the administrative data? What checks do these users carry out? What issues have they encountered?
 - Is there any other source of assurance/metadata about the quality of the data e.g. audit reports, data classification standards, definitions, instructions or governing regulations?
 - Are there any alternative related data sources, for comparison purposes?
- Investigate how data would be supplied and whether there are any associated issues with data supply
 - In what format is the data currently stored? How would the data be transmitted?
 - What data permissions might be required for access to the data?
 - What security classifications are attached to the data?
 - What arrangements are needed for handling of any personal data?
 - Arrangements for Data encryption, Data firewalls, etc.?
- Build and maintain working relationship with the data suppliers
 - Develop formal agreements with data suppliers setting out the format, content and timing for supply of data and associated metadata and how this data will be used in statistical production.
 - Formally agree conditions for supply of administrative data to researchers
 - Agree quality standards/protocols and mechanisms for addressing/querying any issues arising.
 - Regularly meet with and maintain contacts with the data suppliers to ensure they understand the statistical use of the administrative data and to ensure timely consultation on any developments that may impact the administrative data source. Work with the data supplier to identify and support potential enhancements to the quality of the data, including potential changes to IT systems and greater use of key identifiers for location (Eircode), persons (PPSN) and business (Unique business ID).

Back



- Develop and document appropriate procedures to validate data quality from administrative data sources and to address quality issues in a consistent and systematic manner:
 - Carry out validation of the administrative data at the earliest opportunity to ensure the data supplied is complete and in the expected structure and format (variables in expected order, of expected length etc.)
 - Examine data for outliers, limits, unusual trends or consistency within the data and with external sources where possible
 - Set up appropriate edit/imputation/weight adjustment procedures to deal with missing/problematic data
 - Maintain a log of any data concerns or issues, to highlight issues to the data supplier and as input into the quality information that will be published alongside the statistics.
- Consider the privacy implications of the publication of information based from administrative records and always ensure that the CSO guidelines for disclosure control are followed when preparing any data for release.
- Publish open, clear, messages about the quality of the underlying administrative source, any limitations that may exist and how this data was used in statistical production.



Emerging Data Sources

Key Learning Point(s):

Be informed of the potential of new data sources

Explore and evaluate the potential and the limitations of data sources over an extended period

Consult appropriate expertise

Technological, legislative and social change now present producers of statistics with new challenges but also new opportunities. New data sources offer the potential to supplement or replace existing data sources and can facilitate a unique insight that may not be available through existing sources. New data sources may present opportunities to significantly enhance the timeliness and scope of statistics. Statisticians should be open to exploring the potential offered by new data sources while taking extreme care to fully understand the constraints, data limitations and quality challenges associated with these. Rigorous testing over an extended period is an absolute requirement prior to any commitment to incorporate new data sources into official statistical purposes and the data and how these data are collected can be subject to rapid change. Much of the guidance provided for use of administrative data can also be applied to the exploration of new data sources.

When considering the exploration of new data sources, it is vital that the statistical producer:

- Sets out the statistical need that the new data source will address.
- Develops a project plan and seeks authorisation from the appropriate governance body in a timely fashion.
- Fully understands the data, how it is collected and the stability and reliability of the data.
- Understands the legal, technological, ethical and possible societal constraints around access to and use of the data.
- Ascertains the requirements for establishing effective relationships with data suppliers.
- Develops, implements and documents a comprehensive plan for testing and validating the new data source over an extended period.
- Documents planned quality assurance processes and audit arrangements.
- Reports clear and timely information about the quality of the outputs produced and risks associated with the new data sources and processes to the appropriate governance body to inform their decision-making around use of these data.

A large number of innovative data sources have emerged over recent years, for example:

- Big Data was defined by the European Commission in 2014 as 'large amounts of data produced very quickly by a high number of diverse sources.' Big Data has great potential as a data source but can involve structured and unstructured data that traditional systems and software find difficult to process because of the volume of data, the variety of sources and/or the speed of accumulation of data.
- Scanner data refers to an approach to data collection involving the collection of point-of-sale data directly from retailers. Similarly, loyalty card data and credit/debit card data may yield new information on personal consumption, prices, retail sales and tourism.
- Web-scraped data is the automated collection of data from the websites of retailers and other service providers.
- Geospatial data is the integration of geographical and statistical information to maximise the utility of data collected for statistical purposes.





Questionnaire Design

Key Learning Point(s):

Design questionnaires fit for purpose for the respondent population, the statistical requirements of data users and the administrative and data processing requirements of the survey organisation

Consult all stakeholders early - allow time for testing, consultation and changes

Utilise specialist expertise and tools (CSO Methodology, CSO Questionnaire Design Unit, Technology)

A questionnaire is a set of questions designed to collect information from a respondent. A questionnaire may be interviewer-administered or respondent-completed, using paper methods of data collection or electronic modes of completion. Questionnaires play a central role in the data collection process. They have a major impact on data quality, respondent behaviour, interviewer performance and respondent relations.

The design of questionnaires must take into account the statistical requirements of data users, the administrative requirements of the survey organisation, and the requirements for data processing, as well as the nature and characteristics of the respondent population. Questionnaire design begins by clarifying the objectives of the survey, determining the data which is to be produced by the survey and devising a list of questions to obtain this data.

The main functions of questionnaires are to attain information from respondents, so that it can be recorded and processed. If a questionnaire becomes too long or confusing, respondents may be unwilling to complete it, or they may make mistakes. Careful consideration should be given to a number of factors, including the type of questions to be used, the logical sequence and wording of questions, and the physical/electronic design of the questionaire. It is important to test each of these aspects of questionnaire design with a group of respondents in pilot testing before finalising the questionnaire. If necessary, the form can then be modified and retested until respondents can complete it accurately and quickly with a minimum of errors.

Question and Answer Formats

Questions can usually be described as either open or closed. In choosing between these alternatives, the questionnaire designer needs to consider the data requirements, accuracy requirements, ease of processing and coding requirements, the position of the questions on the form and the sensitivity of the question, amongst other factors. In general, closed questions are often better for both the interviewer and for coding and processing. Once the questions have been chosen, they should be tested and retested until the best solution is identified.

Open Questions: allow the respondents to answer the question in their own words e.g. 'What is your occupation?' These types of questions can collect exact values from a wide range of possible values but are more demanding than closed questions, both to answer and process. Open questions are often used in pilot tests to determine the range of likely responses.

Closed Questions: provide respondents with a set range of answers to choose from. Closed questions require more effort than open questions in the development and testing stages to ensure the range of answers is appropriate. Processing time of closed responses is much less than that of open-ended responses. Closed question sub-types include:

- Limited Choice: require a respondent to choose one of two mutually exclusive answers, e.g. yes/no;
- Multiple Choice: require a respondent to choose from a number of response options provided;
- Checklist: allow a respondent to choose more than one of the response options provided;
- Partially Closed: provide a set of responses where the last alternative is 'Other, please specify' (useful when it is difficult or impractical to list all possible choices, but can make data more difficult to interpret).







Other considerations in questionnaire design				
Question Types	Factual, Behavioural, Classification/Demographic, Opinion/Motivational, Knowledge, Hypothetical			
Structure & layout	Sequencing (logical, smooth flow), Filter questions, Order of Questions (Responses given to earlier questions can influence responses to later questions), Order of Response Options (The actual order of response options can also introduce bias.), Response Options and Respondent Difficulties (use of prompt cards for interviewer-based surveys), Length (number of pages, number of questions/mandatory questions, time taken to complete) Questionnaire Layout (poor layout can introduce response errors and processing errors), Physical Design (text formats, colour, graphics)			
Error Factors	Complex/Technical Language, Question Ambiguity, Double-Barrelled Questions (Multiple Concepts in one Question), Leading Questions, Unbalanced Questions (insufficient response options), Recall/Memory Error, Intrusive/Sensitive Questions, Acquiescence, Inadequate Response Categories (overlapping categories), Too many Response Options (Respondent fatigue), Tone, Impact of minor wording changes			

- Questionnaires in longitudinal surveys should be evaluated regularly.
- Questionnaires should be designed to be easily processed by both people and electronic systems.
- Use words and concepts that have the same meanings for both respondents and the questionnaire designers. In the case of business surveys, choose questions, terminology, time reference periods and response categories that are compatible with the respondent's expected record-keeping practices.
- In the introduction to all questionnaires:
 - Provide the title or subject of the survey
 - Explain the purpose of the survey
 - Request the respondent's co-operation
 - Indicate the authority under which the survey is taken, highlighting the confidentiality and data
 protection arrangements that are in place
 - Ensure that the value of providing information is made very clear to respondents. In addition, the importance of completing the questionnaire and how the survey data will be used must be highlighted.
- Opening questions should be applicable to all respondents, be easy and interesting to complete, and quickly establish whether the respondent is a member of the target population.
- Questionnaires should be designed to take account of mode of collection. Questionnaires that are to be administered in person or over the telephone must be made interviewer-friendly as well as respondentfriendly.
- Ensure that the instructions to respondents and or interviewers are short, clear and easy to find. Provide definitions at the beginning of the questionnaire or in specific questions, as appropriate.
- Ensure that time reference periods and units of response are clear to the respondent, specify inclusions or exclusions in the questions themselves and not in separate instructions.
- Ensure that response categories are mutually exclusive and exhaustive.
- Provide titles or headings for each section of the questionnaire and include instructions and answer spaces that facilitate accurate answering of the questions.
- Remember that minor changes in wording can have a significant effect on responses. Avoid use of negative words like "not" as they are easily missed by respondents.
- Seek design advice from appropriate expertise, as a poorly designed questionnaire can be the biggest source of non-sampling error (either directly or indirectly). The questionnaire can influence the response rate achieved in the survey, the quality of responses obtained and consequently the conclusions drawn from the survey results.







Advice and support for Questionnaire Design is available from experts throughout the CSO, including specialist Questionnaire Design Units in some directorates – early engagement with relevant expertise is strongly recommended as the lead-in time for development and testing can be significant.

Consult CSO Applications External Collect Unit at an early point regarding electronic data collection options in line with digital first policies.

Strong consideration should also be given to the use of tools that support standardisation and reuse of questions, e.g. **Colectica**.





Survey Frame and Sampling

Key Learning Point(s):

Use agreed centralised population frame & carefully monitor coverage

Develop objective sampling model

Review sampling model with input from suitable expertise (CSO Methodology)

Survey Frame

A survey frame is any list or register of units (e.g. persons, households, businesses, etc.) in the target population for the survey. The survey frame delimits, identifies, and allows access to the units/elements of the target population and is critical in the design of a survey. The target population is the set of elements about which information is wanted and about which estimates are required. The extent to which a survey frame includes all the units of the target population is referred to as coverage. Practical considerations may dictate that some units be excluded from some frames (e.g. companies with a small number of employees, institutionalised individuals). Since the selection of the sample is directly based on this list, the frame is one of the most important tools in the design of a survey. It determines how well a target population is covered and affects the choice of the data collection method.

The survey frame should conform to the target population and contain minimal under coverage and over coverage. Frame creation, use, maintenance and monitoring should be managed carefully.

Frame characteristics (e.g. classification, contact, address, size) need to be of high quality because of their use in stratification, collection, follow-up, estimation, record linkage, quality assessment and analysis. The frame must contain appropriate contact information for each of the units listed so that it can be used to access the population. This means that for postal surveys the frame should contain postal addresses; for electronic surveys the frame should contain email addresses; for interviewer-based surveys the frame should contain street addresses; and for telephone surveys the frame should contain telephone numbers.

Frame coverage errors (such as those caused by time lags between the moment the sample frame was created and actually used; failure to include new births in the frame; misclassification; failure to remove deaths or duplicates and similar out-of-scope elements) should be addressed carefully as these can complicate a survey, resulting in time delays, cost increases and loss of accuracy (through the introduction of bias and and variance).

- In designing new business surveys, or in the redesign of existing surveys, the CSO's Central Business Register should be used to construct the appropriate survey frame. Surveys in Social Statistics should look at utilising the Census of Population to construct the survey frame.
- Where possible, use the same frame for surveys with the same target population, to avoid inconsistencies and to reduce costs of frame maintenance and evaluation.
- Incorporate procedures to eliminate duplication and to update for births, deaths, out-of-scope units and changes in characteristics.
- Monitor the frame quality by periodically assessing its coverage and updating appropriately.
- For area frames, implement map checks to ensure clear and non-overlapping delineation of the geographic areas used in the sampling design (e.g. through field checks or the use of other map sources).
- For statistical production from administrative sources, determine and monitor coverage through contact with the manager of the data source. Where possible, negotiate with the supplier of the administrative source to address coverage issues.





- Whenever necessary, adjust the statistical results or use supplementary data to offset coverage differences between the frame and the target population.
- Include descriptions of the target population, frame and coverage in the survey documentation.

Sampling Procedures

Sampling is the selection of a set of units from a survey frame. This set of units is referred to as the sample. The choice of sampling method has a direct impact on data quality. It is influenced by many factors, including the desired level of precision of the information to be produced, the availability of appropriate frames, the availability of suitable stratification variables, the estimation methods that will be used and the available budget. The intention with any sample survey is to gather sufficient useful information from the sampled units to allow valid inferences about the target population.

Probability sampling is based on three principles: randomisation, known probability of inclusion and the ability to calculate these probabilities. Randomisation ensures that the units in the sample are selected at random. Survey units should be selected for inclusion in the sample with known probability or have a non-zero chance of being selected in the sample. Probability sampling means that measures of sampling variation can be obtained objectively from the sample itself which can be used to determine the level of precision in survey results.

Guidelines

- When determining sample size, consider the required levels of precision needed for the survey estimates, the type of design and estimator to be used, the availability of auxiliary information, as well as both sampling factors (e.g. stratification) and non-sampling factors (e.g. non-response).
- For highly skewed populations, include in the survey a stratum of large units that will be sampled with certainty.
- In determining sample allocation for stratified samples, account for expected rates of misclassification of units in the frame.
- For periodic surveys that use designs in which the sample size grows as the population increases, develop a method to keep the sample size stable.
- For periodic surveys, if efficient estimates of change are required or if response burden is a concern, use a rotation sampling scheme that replaces part of the sample in each period.
- For periodic surveys, develop procedures to monitor the quality of the sample design over time. Put in place an update strategy for selective redesign of strata that have suffered serious deterioration.

Quality Guidelines for Frames in Social Statistics

For specific guidance on frames in Social Statistics, users are advised to refer to "Quality Guidelines for Frames in Social Statistics", produced in 2020 by the ESSnet KOMUSO, Quality in Multisource Statistics Workgroup (see Appendices – Sources and references) which:

- delivers a building block for safeguarding compliance with the Code of Practice in terms of the construction, use and assessment of frames in social statistics.
- provides basic, generic guidance regarding all relevant processes for frames in social statistics, in a systematic way, based on agreed definitions and standards.
- broadens the perception of frames in social statistics so that they can be used as a possible direct source in a multi-source environment



Measures to Reduce Non-Response



Non-response has two effects on results: one contributing to bias of estimates when non-respondents differ from respondents in the characteristics measured; the other contributing to a decrease in the accuracy of the survey estimates resulting from the smaller effective sample size.

Non-response can affect the quality of the survey and must be managed as part of survey and quality management processes. Various types of non-response that affect surveys include individual or unit non-response (where a unit fails to respond), wave non-response (where a unit responds to some but not all waves of a survey) or item non-response (where at least one variable is missing in the response from a unit).

Survey management must seek to understand the factors contributing to non-response and the impact of non-response on survey results. The degree to which response is pursued in a survey is subject to budget and time constraints and the risk of non-response bias. Adjustments made to data to compensate for non-response are dealt with later in this chapter (e.g. weighting adjustments or imputation).

An effective respondent relations programme and a well-designed questionnaire are critical elements in maximising response. Survey managers should also consider how mode of survey delivery impacts response.

- Establish and maintain good relationships with respondents.
- Ensure interviewers are fully trained in interviewing techniques etc.
- When operational constraints permit, follow-up with the non-respondents either as a complete enumeration or on a sub-sample basis.
- Prioritise follow-up activities. For example, in business surveys, follow-up large or influential units first.
- Record and monitor reasons for non-response (e.g. refusal, non-contact, temporarily absent, technical problem).
- Use an appropriate method of imputation to compensate for non-response. Only as a last resort should grossing factor adjustment be used.
- Analyse how non-response has impacted survey quality and communicate this to users.
- Where applicable, ensure that those being surveyed are aware that the survey is statutory.





Data Editing

Key Learning Point(s):

Target and prioritise editing for maximum impact and identification of quality improvements

Consult appropriate expertise (CSO Methodology)

Data editing is the application of checks to detect missing, invalid, inconsistent or otherwise potentially erroneous records. Data editing can be broadly divided into data editing at micro and macro levels and involves the identification and alteration of erroneous/problematic records to ensure that the data is closer to the truth after editing than before. Edits must be carefully designed to detect and correct erroneous records while avoiding generating bias (e.g. by imposing implicit models) or the introduction of new errors. Editing that has a negligible impact on the final survey estimates can be described as over-editing and should be avoided.

The goals of editing are to:

- Provide the basis for future improvement of the survey
- Provide information about the quality of the survey data
- Tidy up the data.

Sometimes, a disproportionate amount of resources is concentrated on the third objective of 'cleaning up the data', with the danger that learning from the editing processes may play an undeservedly secondary role. This can mean that attention becomes focussed on addressing the symptoms of a problem rather than addressing the root cause.

Since the process of data editing is very demanding in terms of time and costs, editing procedures should be automated to the greatest extent possible while avoiding the temptation to increase the scope and volume of checks unnecessarily. While it is recognised that fatal errors (e.g. invalid or inconsistent entries) should be removed from the data sets to maintain credibility and to facilitate further automated data processing and analysis, caution should be exercised against the overuse of query edits (those pointing to all questionable records that may potentially be in error).

The impact of errors identified during editing can vary greatly and often a few errors can be responsible for the majority of changes in the survey results. Editing in a selective manner can maximise the potential efficiency gains without negatively impacting data quality. Priorities can be set according to the type or severity of error or according to the importance of particular variables or reporting units. Internationally, statistical institutes recognise that it is not necessary, or possible, to correct all data in every detail and that reliable estimates can be obtained without removing all errors from a data set. This has led to the development of more efficient editing approaches: in particular, selective editing, automatic editing and macro-editing.

In surveys of enterprises or business entities, selective data editing for the most influential units can often be applied. This involves setting data editing priorities for units based on the relevance of different units for survey results. Units that are selected as higher priority are given closer attention during editing, including a higher level of direct contact or follow-up, while lower priority units are edited mainly through automated editing.

The Generic Statistical Data Editing Model (GSDEM) has been developed by the UNECE High-Level Group for the Modernisation of Official Statistics (HLG-MOS). It is a useful reference for all official statisticians whose activities include data editing: https://statswiki.unece.org/display/sde/GSDEM

Back



- Editing should be designed to eliminate the errors that have a statistically significant impact on the quality of results, rather than aiming for the removal of all errors.
- The number, type and rules/parameters used in editing should be regularly examined to ensure the criteria are appropriate and that the appropriate balance is struck between efficiency and over-editing.
- Build edit checks in at the earliest, most efficient point (e.g. build checks into questionnaires in electronic data collection).
- Perform edit checks for missing values, invalid values, etc. as quickly and as expediently as possible in the processing cycle.
- Automate editing where possible to reduce costs, increase timeliness and ensure repeatability.
- Test all edits to ensure that edits are logical and internally consistent (i.e. not self-contradictory).
- Document editing procedures, criteria and controls to ensure repeatability.
- Re-apply edit checks to units to which corrections have been made to ensure that no further errors are introduced directly or indirectly.
- Ensure transparency and traceability of all changes to data arising from editing (e.g. using versioning with appropriate metadata/insertion of appropriate commentary in data systems).
- Consider editing to be an integral part of continuous improvement in statistical production. Use information gathered as part of editing to:
 - Sharpen definitions
 - Evaluate the quality of the data
 - Identify non-sampling error sources
 - Serve as a basis of future improvement of the whole survey process.



Data Quality Evaluation (including Macro-Editing)

Key Learning Point(s): Examine and validate data quality systematically Identify and address sources of error Document rationale around edit parameters being set Consult appropriate expertise

Data quality evaluation refers to the process of evaluating the final statistical output against the original objective of the statistical activity, in terms of the data's accuracy or reliability. While data editing identifies units for follow up on a record-by-record basis, data quality evaluation and macro editing considers all the data at once and subsequently identifies records for follow up or adjustment. Data quality evaluation allows users to make more informed interpretations of the survey results and can be used to improve surveys.

Two general types of data quality evaluation can be distinguished:

- Macro editing or quality validation is the process of reviewing the data before official release to ensure that grossly erroneous data are not released, or to identify data of marginal quality.
- Sources of error studies generally provide quantitative information on specific sources of error in the data. While timeliness is still important, the results of these studies often are only available after the official release of the data.

- Make planning of data quality evaluations part of the overall statistical process, as the information needed to conduct such evaluations often must be collected during the process itself.
- Involve users (internal and external) in setting the objectives for the data quality evaluation programme. Where circumstances permit, also involve them in the evaluation process itself.
- The following macro editing and quality validation methods should be used:
 - Checks of consistency with external sources of data, for example from other surveys or from previous instances of the same survey.
 - Internal consistency checks, for example calculation of ratios that are known to lie within certain bounds (e.g. gender ratios, average value of commodities, etc.).
 - Unit-by-unit reviews of the largest contributors to aggregate estimates (typically in business surveys).
 - Debriefings with staff involved in the collection and processing of the data.
- The following sources of error should be evaluated:
 - Coverage errors, which consist of omissions, erroneous inclusions, and duplications in the frame used to conduct the survey.
 - Non-response errors, which occur when the survey fails to get a full response.





- Measurement errors, which occur when the response received differs from the 'true' value, and can be caused by the respondent, the interviewer, the questionnaire, the mode of collection, or the respondent's record-keeping system.
- Processing errors, which can occur at the subsequent steps of data editing, coding, capture, imputation and tabulation.
- Sampling errors, which occur when the results of the survey are based on a sample rather than the entire population.





Imputation



Data collection invariably encounters problems with missing, invalid or inconsistent values. These problems occur both for data collected in traditional surveys and for administrative data. Imputation denotes the procedures used to assign replacement values for these missing records so that the data file contains plausible and internally consistent records that can be used for estimation of the population aggregates of interest.

Good imputation processes are automated, objective and reproducible, making efficient use of any available auxiliary information and should have a clear audit trail for effective evaluation. Checks should also be in place to ensure that imputed records are internally consistent.

To impute the missing values in a data set, several methods are available. Possible imputation methods include:

- deductive imputation,
- model-based imputation (including mean, ratio, and regression imputation), and
- donor imputation (including cold deck, random hot deck, and nearest-neighbour imputation as well as predictive mean matching).

A key reference is "CSO Imputation Guidelines - Dealing with Missing Data in the CSO", produced by CSO Methodology Unit https://methodology.cso.ie/statisticalmethodologies/. Eurostat's Collaboration in Research and Methodology for Official Statistics (CROS) Handbook on Methodology for Modern Business Statistics provides guidance on how different imputation methods may be useful in different contexts. https://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics_en

When determining the best approach to take to introducing or in reviewing an imputation methodology (whether single or multiple imputation), consultation with CSO's Methodology Unit and other subject matter experts with relevant experience (e.g. through the CSO Imputation Working Group) is advisable. The options tested, and results of these tests should be documented to verify the suitability of the chosen imputation model, to ensure repeatability and to facilitate any subsequent evaluation of the selected procedure.



Weighting

Key Learning Point(s):

Assess the data – Develop and test an appropriate model

Recalibrate using auxiliary data if required

Consult appropriate expertise (e.g. CSO Methodology)

In most surveys, data is obtained from only a certain proportion of the population (even with a census, as usually not all units respond). This means that calculation of survey weights may be required to estimate the entire population.

Survey weighting involves assigning a value to each unit in the data file to make the statistics computed from the data more representative of the population. The "weighted" value indicates how much each unit will count in a statistical procedure. Through weighting, some units in the sample are assigned greater importance because of issues of unequal probability of selection, non-response or need to adjust to population totals.

In a probability-based survey each unit included in the sampling frame has a known probability of being selected into the survey sample. This probability of inclusion in the sample takes account of various aspects of sample design such as stratification, clustering and multi-stage or multi-phase selection. The sampling weight reflects the inverse probability of inclusion in the sample – it denotes how many units in the sampling frame are represented by the selected sample unit.

In sample design each sampled unit has a design weight based on the probability of inclusion in the sample but if there is unit non-response in the survey the achieved sample is smaller than the original sample design. In this scenario it is necessary to adjust the design weights using an adjustment factor based on each unit's probability of response. Calculation of these adjustment factors should be done using an appropriate model which should be tested and recalibrated (using available auxiliary data) as required.

The two most common types of survey weights are:

- Design/Sampling Weights
- Estimation (Post-Stratification or Non-response) weights

Design Weights are normally used to compensate for over or under-sampling of specific cases or for disproportionate stratification. For example, in demographic population surveys it might be necessary to over-sample minority group members or in business surveys it may be necessary to over-sample businesses with particular characteristics (e.g. employment size category).

Estimation (Post-Stratification or Non-response) Weighting is used to compensate for the fact that units with certain characteristics are not as likely to respond to the survey or for other instances where auxiliary information is used to adjust data. For example, often general population surveys might have substantially more female than male respondents (often 60/40) although there are sometimes more males in the population. Because the survey response over-represents females and under-represents males in the population a weight is used to compensate for this bias. The likelihood that any unit will respond can be related to different characteristics such as age, education, ethnicity, gender, place of residence, socio-economic status etc., which needs to be taken into account in sample design and in weighting. Inclusion of auxiliary data from other data sources (or from previous survey iterations) can be used to improve the weighting and estimation process through use of calibration techniques.





Guidelines

- Weighting should be confined to probability-based sample surveys where statistical inference can be applied, and the precision of estimates assessed
- A weight needs to be associated with each sampled unit (either design weight or estimation weight)
- Non-response should be analysed to reduce non-response bias by establishing whether non-responding units differ from responding units and from ineligible units check that the structure of eligible and ineligible units in the population (and among non-responses) is equivalent to the structure of responding units. If the non-response rate is very high an additional study/survey of the non-responding units should be carried out to inform adjustment of the weights.
- Special attention must be given to outliers (units with extremely high or extremely low values compared to other units in the sample) so that weights are adjusted appropriately
- Weighting should be designed to compensate for over or under sampling, non-response bias and to utilise relevant auxiliary data sources in an appropriate and reproducible manner.

Consultation with CSO Methodology Unit is advised to assist in the development or validation of a weighting methodology.





Seasonal Adjustment



Seasonal Adjustment refers to the methods used to take account of, and remove, seasonal effects from statistical time series. By removing the seasonal effects, the long-term time series trends can be analysed more clearly.

Seasonal adjustment is of particular significance to the CSO since many seasonally adjusted series are of the highest public interest. Seasonal adjustment adds value to users since it allows them to gain a clearer picture of the underlying behaviour of particular time series.

Office Notice 07/2019 sets out the CSO Seasonal Adjustment Policy. The policy is based on Eurostat's 2015 Seasonal Adjustment (SA) Guidelines, on previous CSO best practice documentation and on input from the CSO Seasonal Adjustment User Group. A Seasonal Adjustment Guidelines document, which provides detailed guidance on the application of the Policy is available from CSO Methodology Unit and is intended to ensure compliance with Eurostat Best Practice as outlined in Eurostat's 2015 Guidelines.

It should be noted that the guidance is not a detailed instruction manual for seasonal adjustment and that the guidelines from Methodology are not policy. There may be cases where the guidelines are superseded by CSO or Eurostat policy, for example as in the specific Eurostat guidance on the seasonal adjustment of Quarterly National Accounts.

Seasonal Adjustment Software and Methods

Use of appropriate software and methods is critical. Details of current best practice software and methods should be sought from Methodology Unit. The use of inappropriate software and methods can lead to poor quality seasonal adjustment outcomes which, in turn, can generate misleading results and increase the possibility of disseminating false signals on the economy, e.g. the appearance of turning points in the business cycle and as a result undermine the CSO's reputation as a provider of high-quality statistics.

Back



Approved Methods and Software (*from CSO Seasonal Adjustment Policy, 2019)

The X-12/X-13 series of non-parametric Seasonal Adjustment methods are approved for the CSO.

X-11 approaches should not be used because they fail to account for outliers in a series, moving holiday effects, trading day effects, etc. and this in turn can have a big impact on the overall robustness and reliability of the seasonal adjustment.

TRAMO-SEATS is also deemed an acceptable approach in providing seasonal adjustment.

In terms of software implementation, the only approved Seasonal Adjustment software packages for conducting seasonal adjustment currently are Win X-13, JDEMETRA+ and Win X-12. SAS should not be used to conduct seasonal adjustment in the CSO.

Annual Reviews of Seasonal Adjustment

All business areas are required to carry out, at a minimum, annual reviews of their seasonally adjusted time series using one of the following options:

- Partial concurrent adjustment (*recommended approach in CSO Seasonal Adjustment Policy, 2019).
- Forward-factor adjustment
- Full concurrent adjustment

The following should be examined as part of the annual review:

- Are approved methods and software for seasonal adjustment being used?
- Is the ARIMA model specified still valid? (Assess using the relevant diagnostic output from seasonal adjustment software)
- Is there a need to add or remove outliers, level shifts, etc.?
- Has the nature of the unadjusted time series changed over the past twelve months? Is there any evidence of seasonal breaks in the series?
- Is any change in methodology from the previous year warranted, especially if the previous year's parameters still provide good quality seasonal adjustment?

Metadata on the seasonal adjustment process must be retained to ensure quality and to assist the annual review process (e.g. diagnostic output from the seasonal adjustment software for each period, information on the models chosen; information on pre-treatment, calendar effects and outliers; information on any deviations from the default X-12 parameter settings; information on the annual review process - when the review took place and a brief summary of any changes to the previous seasonally adjusted settings).

Training and knowledge of time series

Users performing annual reviews, or applying seasonal adjustment to new time series, should have sufficient training and knowledge of the subject to carry out their roles. Users who are not familiar with the topic must address this, by either receiving appropriate training or referring to recommended literature (see Central Statistics Office - Seasonal Adjustment Policy)

Central Statistics Office - Seasonal Adjustment Policy, 2019: https://methodology.cso.ie/media/methodology/ documentsuser/statisticalmethodologies/seasonaladjustment/CSO_Seasonal_Adjustment_Policy.pdf

Eurostat Handbook on Seasonal Adjustment, 2018: https://ec.europa.eu/eurostat/documents/3859598/8939616/ KS-GQ-18-001-EN-N.pdf





Statistical Confidentiality

Key Learning Point(s):

Individual data is strictly confidential and must be protected

Confidentiality is guaranteed by law

•••••

Apply appropriate statistical disclosure controls systematically to protect confidentiality

Consult appropriate expertise (e.g. CSO Data Office, CSO Methodology)

Statistical Confidentiality is a fundamental principle of the European Statistics Code of Practice, is a core value of the CSO, and is stipulated in legislation. All information supplied to the CSO is treated as strictly confidential and this commitment is guaranteed by law in The Statistics Act, 1993 and by EU legislation, such as Council Regulation (EC) No 223/2009 on European statistics for data collected for EU statistical purposes. Legislation sets stringent confidentiality standards, ensuring that information collected may be used only for statistical purposes, and no details that might be related to an identifiable person or business undertaking may be divulged to any other government department or body. The General Data Protection Regulation (GDPR) has also placed new requirements on statistical producers to ensure that appropriate safeguards are in place for statistical data. These safeguards include methods to ensure the protection and confidentiality of data during the storage and processing stage as well as to ensure that published data is non-disclosive.

All CSO staff are Officers of Statistics and must sign a Declaration of Secrecy under the Statistics Act 1993. Staff must adhere to regulatory requirements and must familiarise themselves with relevant CSO policies and procedures designed to protect statistical confidentiality and data security (e.g. CSO Data Management Policy).

- All individual information obtained by the CSO is treated as strictly confidential. The information is used solely for statistical purposes.
- Results are published in aggregate form and great care is taken to ensure that details relating to an identifiable person or undertaking are not inadvertently divulged.
- Confidentiality is guaranteed by law.

Statistical Disclosure Control (SDC)

Statistical Disclosure Control (SDC) methods allow the dissemination of statistical information without revealing confidential information about respondents. The ES CoP states that producers of official statistics should: 'Ensure that arrangements for confidentiality are sufficient to protect the privacy of individual information, but not so restrictive as to limit unduly the practical utility of official statistics.' In the CSO, Office Notices 04/2019 and 12/2019 set out how SDC should be applied.

The final decision on what constitutes acceptable levels of disclosure risk remains with data custodians. The rationale for SDC decisions within statistical production areas must be recorded and retained. Clear rules for checking confidentiality (e.g. minimum frequency/dominance rules) must be established and appropriate controls must be applied. The SDC risk from unauthorised matching should also be considered. CSO Methodology Unit has published detailed guidance on Statistical Disclosure Control for both tabular and microdata output which can be accessed on the the CSO Methodology portal: https://methodology.cso.ie/statisticalmethodologies/















Quality in the Presentation and Dissemination of Results







Presentation of Results

Key Learning Point(s):

Present timely, accurate and relevant information clearly, concisely and in formats that meet the needs of the maximum number of users

Presentation and dissemination are among the final stages in the statistical production process and strongly influence the perception of statistical quality by users. Multiple modes of dissemination should be explored to ensure maximum accessibility for all users.

The forms of presentation and dissemination chosen should aim to facilitate ease of access and interpretation by all users. Data should be presented in appropriate formats and must be accompanied by appropriate metadata. Tables, whether in publications or in electronic formats, should have clear headings, identifiable units and a logical layout. Charts or infographics should give a clear and accurate representation of the phenomenon being studied.

In official statistical releases and publications, a suitable analysis highlighting the more important and interesting features of the published statistics should be provided. This should be based on the statistical results available and should never contain any political or similar judgements. In general, the larger and more infrequent the release or publication, the more detailed the analysis should be.

CSO Dissemination and Communications Unit provide in-house guidance and support on the presentation of statistical results in multiple formats and should be consulted at an early stage for advice on how to improve the presentation and dissemination of results. CSO Editorial Board provide expert advice on finding and communicating the story within the data.

The CSO guide on writing style, "Write Well, Write Clearly" should be fully utilised in preparing content for CSO releases and publications (see: https://editorialboard.cso.ie/).

Guidelines

Commentary

- Commentary should be clear, understandable, interesting and unbiased
- Key findings of the statistics should be highlighted
- Commentary should contain relevant comparisons and explanations of significant changes

Tables

- Table layout should be clear and easy to follow, and labels should clearly identify the content
- All units of measurement should be displayed clearly
- Use summary tables in releases/publications and provide detailed data in databases
- Ensure all rounding to significant digits is mathematically correct
- Ensure that footnotes are clearly marked and that the text is clear and readable

Infographics

Utilise infographics to make complex or detailed statistics more understandable and interesting





Charts

- Chart titles must explain what phenomenon is represented and the time periods covered
- Legends, labels and data points should all be clear and readable all elements on the chart should be clearly identified (all axes must be clearly labelled, and units of measure included)
- Any apparent discrepancies/anomalies represented by data in the chart should be highlighted and explained

Maps

- Use maps to visualise the spatial distribution of statistical data
- Ensure maps are titled and labelled in a clear and understandable way

Social Media

- Consider the potential of social media to disseminate statistical information to a wider audience
- Consult with Communications & Dissemination Unit for guidance and support





Metadata & Quality Reports

Key Learning Point(s):

Provide all users with the information that they need to understand the data, its quality limitations and how data quality changes over time.

••••••

A key requirement for producers of statistics is the need to provide users with information about the quality of the statistics. Good quality reporting should give the user a clear understanding of the strengths and limitations of the statistical output. No statistical estimate can be perfect, there are always limitations to the quality of the data produced and every effort should be made to give the user a clear picture of what those limitations are so that they can correctly interpret the data. Provision of good metadata and informative quality reporting is critical to enable users understand and appropriately utilise the data.

The key reference for producers of Quality Reports and Metadata in the European Statistical System is the ESS Handbook for Quality and Metadata reports: <u>https://ec.europa.eu/eurostat/web/quality/quality-reporting</u>

European Statistical System Metadata Reporting Standards

Eurostat guidance describes two distinct types of statistical metadata: structural metadata and reference metadata. It is important to understand that statistical output is meaningless without accompanying metadata - both structural and reference metadata are required to ensure users understand the data. For example, a number such as such as 4,761,865 is totally meaningless unless you are provided with its accompanying structural metadata, namely that it concerns the total population of Ireland in April 2016.

Structural metadata are used to identify statistical data, e.g. titles, subtitles, short descriptions, dimension names, variable names, dimensions, etc. Structural metadata include the titles of the variables and dimensions of statistical datasets, as well as the units of measurement used (e.g. EUR), code lists (e.g. for territorial coding), data formats, potential value ranges, time dimensions, value ranges of flags, classifications used, etc.

Reference metadata describe statistical concepts and methodologies used for the collection and generation of data and provide information on data quality. They thus assist users with the interpretation of the data, thereby being strongly content oriented. Unlike structural metadata, reference metadata can be decoupled from the data, i.e. they can be generated, collected or disseminated separately from the statistics to which they refer. They can also be associated with different levels of data: entire collections, data sets from a given country, or for a data item concerning one country and one year.

METADATA

STRUCTURAL

Data and Structural Metadata

must go together. Without column names, dimensions, etc. it is impossible to understand the statistical data

REFERENCE

Reference Metadata

provide a quality assessment of the data set as a whole (general description of the dataset, classifications used, evaluation of quality, etc.)

Can be decoupled from data sets

- Conceptual metadata
- Methodological metadata
- Quality metadata

Back



European Statistical System Quality Reporting Standards

Eurostat has prescribed standards to be applied when completing quality reports to accompany data transmitted from national statistical institutes. While these different reporting standards currently co-exist within the ESS Quality Reporting Framework, the ESS has now integrated the separate standards into the Single Integrated Metadata Structure (SIMS). In SIMS all the statistical concepts from previous standards (e.g. ESMS and ESQRS) have been included and streamlined. This ensures that concepts are reported upon only once (direct re-usability of existing information). SIMS is designed to be the dynamic inventory and conceptual framework for all ESS quality and reference metadata concepts.

SIMS incorporates the European Statistical System (ESS) recommendation that quality reporting addresses the five quality dimensions of statistical outputs (Relevance; Accuracy and Reliability; Timeliness and Punctuality; Accessibility and Clarity; Coherence and Comparability). As SIMS is embedded across the CSO it will become the foundation for all quality reporting and will replace other current reporting standards over time. Further information is available on the Eurostat website at https://ec.europa.eu/eurostat/web/quality/quality/quality-reporting



CSO Standard Quality Reports

A standard template for quality reports in the CSO is available and must be published at the same time as the associated release or publication. Quality reports must be comprehensive, informative and transparent in relation to the limitations of data quality for each iteration. The content of each quality report should be reviewed as part of the regular dissemination process and must be updated at least once a year at a minimum. Currently, the standard CSO Quality Report provides information to users under the following headings: (a) Overview; (b) General Information; (c) Statistical Concepts and Methods; (d) Production of the Statistics, Data Processing, Quality Assurance; (e) Quality. The structure of the CSO Standard Quality report will be reviewed and updated so that it is aligned to the SIMS standard. Further information is available on the CSO Quality Information System at: https:// qualityinformationsystem.cso.ie/qualitydocumentation/

- Publish a relevant and easily accessible quality report with every statistical output
- Revise quality reports at least once a year.
- Ensure that background notes, including descriptions of methodology and processing, are up to date.





- Explain relevant statistical concepts and maintain codebooks of variables present in the release.
- Provide users with adequate and accessible quality information to facilitate appropriate interpretation and use of statistics. Clearly inform users of any significant variations in quality and promptly inform users of revisions.


Revisions

Key Learning Point(s):

Apply procedures to clearly inform users of revisions in line with policies

It is recognised internationally that implementation of a sound revisions policy helps maintain credibility in official statistics. ES CoP explicitly references the need for a clear and transparent revisions policy and this is also underpinned by the CSO Revisions Policy 2021.

Statistics are subject to revision, either scheduled or unscheduled, for a variety of reasons. Users require statistics that are fit for purpose and robust which means statisticians often have to reconcile timeliness with accuracy. Frequently, the first publication of a data series may be provisional and subject to some degree of uncertainty. Revisions to initial estimates are an inevitable consequence of the trade-off between timeliness and reliability. Revisions to estimates that are considered final may also occur for a variety of reasons. The following table outlines the three types of revisions that are typically encountered.

Types of Revisions

Planned Routine	Changes in published data related to the regular data production process e.g.				
Revisions	updates to preliminary/advance estimates				
	updates to seasonal factors				
Planned Major	Substantial changes in published data to cater for:				
Revisions	 changes to definitions/classifications/methodology 				
	 updating/rebasing/re-weighting of sub-components/insertion of new items 				
	 updating of key statistical series (e.g. Census of Population) 				
	 accommodating new data sources 				
	cessation of existing data sources				
Unplanned	Unplanned revisions generally occur when:				
Revisions	• published statistics, that are considered final, are revised due to receipt of updated data that significantly impacts figures				
	existing data sources are suddenly lost				
	errors are detected				
	 improved data becomes available from other sources 				

Producers of statistics should familiarise themselves with the latest guidance and put in place procedures to ensure adherence to current policy. Key elements of the CSO revisions policy deal with the need to notify users of revisions, the need to consult with users in relation to planned revisions and the general approach to revisions in terms of timeliness, record-keeping, discontinuities, provisional data, PxStat updates, researcher micro data etc.

CSO policies and guidance on revisions are available on the CSO Quality Information System.

Guidelines

- Producers of statistics in the CSO should be familiar with the latest policies and guidance
- Procedures should be put in place to ensure that:
 - Results that are subject to revision are highlighted in background notes and/or PxStat tables





- Users are given advance notice of future planned revisions (where relevant) and mechanisms are in place to facilitate user consultation
- The CSO Revisions Calendar is kept up to date.
- Revisions practices for outputs are documented in survey background notes, quality reports and supporting methodological documents.
- All disseminate data is clearly labelled so that data is stored securely and is clearly identifiable and retained (prior and post revision).
- Errors discovered in published statistics are corrected at the earliest possible date and publicised appropriately.
- Revisions are regularly analysed so that users can be informed on the scope and impact of the revisions and to systematically improve source data and statistical processes.





CSO Quality Management Framework







CSO Quality Management Framework

The CSO Quality Management Framework (QMF) is an extensive and long-term programme introduced to ensure that the statistical production standards applied in the CSO achieve the highest standards of quality and efficiency. This commitment to meeting the highest standards has always been a core value for the CSO but has become even more relevant in the context of increasing and more formal scrutiny of official statistics at EU and at a wider international level.

The overall goal of the QMF is meeting the required standards as set out in ES CoP. The QMF foundations are based on establishing the UNECE's GSBPM as the operating statistical production model in the CSO. The GSBPM allows the CSO to advance a more standardised approach to quality management and is the central plank on which each of the QMF projects are built. The CSO QMF projects target building robust standardised systems, maintaining comprehensive documentation to ensure repeatability and generally applying structures that support systematic quality management.

While each statistical area is responsible for managing the quality of their statistical processes and outputs, the CSO's QMF provides common structures, tools and supports to enhance and standardise the steps required both to achieve the highest standards and to be able to demonstrate that these standards are being met consistently and systematically. Statistical production managers should be proactive in implementing each of the elements of the QMF in their statistical production. Elements of the QMF that have been rolled out to date which will be described in this section include:

- Process Mapping & Documentation
- Quality Information System
- Directory of Products & Services
- Standard Methods Page
- Data File Structure

- Data Attestations
- Data Inventory
- Quality Review Self-Assessments
- Supported Quality Appraisals
- Quality Training

Guidance and support on the QMF is available from the CSO Quality Unit qualityunit@cso.ie

CSO Process Mapping and Documentation

Key Learning Point(s):

Document all statistical processes, systems and controls

Review and update regularly

Provide a legacy of knowledge embedded in the organisation

The QMF demands high quality, comprehensive and up to date documentation of statistical processes and systems. This commitment involves dedicating time and resources to producing and maintaining high quality end-to-end survey documentation, including process maps aligned to the GSBPM.

Process maps are often easier to follow and read than a standard written document and are an excellent tool for providing a high-level overview of the workflow of processes in a section. They give a good insight into roles and interactions in the statistical production process as well as where individual roles fit into the overall process. Process maps should be supported by high quality, comprehensive, survey documentation to ensure that knowledge is embedded in the organisation. Together, process maps and documentation clarify duties and responsibilities and enable smooth transfer of responsibility without inadvertently compromising quality or efficiency.





Process Maps and Survey Documentation act as a baseline and information repository for existing statistical processes in the CSO, contributing to the stability and reproducibility of statistical production. Process maps and documentation aligned to the GSBPM are useful in training of new staff and can also assist with standardisation of similar processes across statistical areas. Good process maps and survey documentation contain structural metadata indicating hard deadlines (e.g. Eurostat transmission dates) and metrics informing basic quality targets for the process (e.g. Sample sizes, Acceptable Response Rates, Acceptable edit rates).

Guidelines

Staff and managers in the CSO should:

- Ensure that process maps and other survey documentation materials are kept up to date.
- Ensure that the Quality Information System (QIS) contains the most up to date version of survey documentation and process maps notify Quality Unit of any changes in process ownership or organisational structure so records can be kept up to date.
- Self-certify the accuracy and relevance of the maps and survey documentation by carrying out periodic reviews of maps and documentation. Notify their Head of Division of compliance with office policy regarding Process map and Survey documentation status.
- Carry out walkthroughs of their maps as part of the certification process at appropriate intervals

CSO Quality Information System

Key Learning Point(s):

Lodge key information on statistical production in corporate systems for greater transparency and security

Check and update regularly

The CSO Quality Information System (QIS) is a central repository of quality documentation accessible to all CSO staff. The QIS underpins the QMF by ensuring that best practice, guidance and support is shared across the organisation.

The QIS contains survey documentation and process maps from each statistical area so improvements in quality and efficiency that are made in one area can be quickly made available to other areas.

Staff should regularly review the QIS to ensure that the content related to their own areas is fully up to date and to familiarise themselves with the latest guidance updates.

Guidelines

Staff and managers in the CSO should:

- Ensure that the Quality Information System (QIS) contains the most up to date version of survey documentation and process maps
- Notify Quality Unit of any changes required (email: qualityunit@cso.ie)



Quality Information System

The QIS is available on the CSO Lotus Notes homepage and at: https://qualityinformationsystem.cso.ie/





CSO Directory of Products and Services (DPS)



The Directory of Products and Services is an application that lists all the statistical products the CSO produces, together with information on the corporate services that support these products.

This information represents the most up to date information on products and services offered by the CSO. For administrative purposes, all products and services have been assigned a single owner, who acts as the contact point for internal queries on that product or service. Reports generated from the DPS are used for publishing official reports on what the CSO produces (e.g. Annual Statistical Work Programme).

The DPS application is used to link various QMF initiatives and to support data matching, Privacy Impact Assessment and other requests through the Data Office register.

Guidelines

Staff and Managers in the CSO should:

- Ensure that all statistical outputs and services are included in the DPS.
- Ensure that the information held in the DPS is accurate and relevant by reviewing and updating, particularly after any reorganisation.
- Notify QMSA of any discontinued products, the reasons for the discontinuation and whether they've been replaced by a different product.

To add a new product or service, or to update any information, please email qualityunit@cso.ie. The DPS can be accessed at https://dps.cso.ie/



CSO Standard Methods page

Key Learning Point(s):

Provide key information to users consistently across all outputs – Check and update regularly

The CSO has developed a standard for the provision of quality information attached to releases and publications on the CSO website. The three main parts to the Standard Methods page outline survey information, methodological and other quality information and contact details related to the statistical output.

Guidelines

Staff and Managers in the CSO should:

- Ensure that the layout and content of the Methods page associated with your statistical products meets current office standards.
- Ensure that information provided on the CSO website is accurate and up to date.
- Check Methods page details whenever statistics are released or updated build this into standard procedures around dissemination.
- Ensure that metadata indicating when information was last updated provides assurance to users that the information is current.

CSO Methods Page Standard Elements				
Survey Information	Survey Name Legislative basis National Reference Area Time coverage Survey Size	Survey Purpose Legislative basis European Sampling Frame Base Period Principal Variables	Periodicity Statistical Population Sector Coverage Principal External Users	
Methodological & Other Quality Information	Process Summary Quality report	Survey Background Notes Other Methodology documents	Survey Forms Process Model (GSBPM)	
Contact details	Contact Organisation Contact Person Function Contact e-mail	Business Area Contact Person Other Business Area Specialists	Contact phone number	



CSO Data File Structure

Key Learning Point(s):

Manage data storage to a corporate standard

Data Management is facilitated by taking a standardised approach to the organisation of data holdings in the CSO. The QMF stipulates that data files are stored in accordance with the CSO Data File Structure which is aligned to the GSBPM and focuses on the appropriate storage of data through the Collect, Process, Analyse and Disseminate phases of statistical production.



Back

Guidelines

- Ensure that the file structure used in the production of statistics follows the model prescribed for data holdings and programs
- Ensure that an appropriate data retention/security level classification is assigned to data holdings.
- Follow appropriate naming conventions, giving files and folders meaningful names so that content can be easily identified.

CSO Data Attestations

Key Learning Point(s):

Manage and monitor data access rights

Data management in the CSO also requires that access to data holdings is restricted to only those who have a legitimate requirement for access. The right to access data holdings is subject to stringent governance procedures overseen by Heads of Division who are designated as data owners. Data owners assign and remove access privileges to staff and are required to monitor this access on an ongoing basis.

The QMF provides for regular monitoring of data locations and access whereby data owners are required to review the location of data for their Division, the people who can access this data and the level of access (edit/read only/ etc.). They are required to "attest" that their lists of data holdings are correct and that this data is accessible only by those staff who are legitimately entitled to such access.

Guidelines

Heads of Division must:

- Review their data holdings and confirm that the list of data holdings is complete and correct in respect of location and content.
- Ensure that access to data holdings is limited to staff who require access for statistical production.
- Review records of access permissions when requested and attest that these permissions are correct and appropriate.
- Take immediate action to remove access permission from staff who no longer have a legitimate requirement for access to particular data holdings.

CSO Data Inventory



The CSO Data Inventory is an application that captures information on the various types of data used in statistical production, where this data is located and what programmes are used to process, analyse and disseminate statistical outputs. The Data Inventory records key information on these data assets of the Office so that data can be maintained and governed.



The overall purpose of the Data Inventory is to provide full transparency around the CSO's data holdings and data flows. It reduces the risk of knowledge being lost due to staff mobility and provides clarity on how data from one section is often reused in the production of outputs from other business areas in the Office.

The Data Inventory also allows measurement of the contribution of administrative data in each statistical output and allows better management of data assets over time to facilitate the repeatability of statistical production.

The Data Inventory:

- Describes the data used during the statistical lifecycle of a product from start to finish
- Increases transparency and aids repeatability thereby reducing risk and improving consistency
- Assists in managing GDPR obligations by recording file locations and business rules around how data is
 retained and managed
- Tracks how data inputs in statistical production change over time
- Assists in the training and upskilling of new staff

Guidelines

Managers in the CSO should:

- Record file locations and details of all data sources for each iteration/survey period (input & output variables) in the Data Inventory
- Detail which SAS/R Programs/Projects are used in statistical production
- Describe, at a high level, the methodologies used, along with the data classifications and retention periods for all data sources for the entire statistical product life cycle.
- Ensure that any changes to data sources, flows or work responsibilities are recorded in the Data Inventory when prompted by the system when rolling over to a new survey period. This will ensure the Data Inventory is kept up to date.

CSO Quality Review Self-Assessments (QRS)

Key Learning Point(s):

Self-assess quality in statistical production

Use standardised self-assessment questionnaire

Develop improvement plans

The CSO QRS is a self-assessment tool based on the European Self-Assessment Checklist for Survey Managers (DESAP) and the Quality Assurance Framework of the European Statistical System (ESSQAF). The QRS prompts survey owners to review the quality of their statistical processes against ES CoP principles for each phase of the GSBPM they are using. It allows them to rate their process performance, highlight good practice and set out improvement actions for areas which need improvement.

The QRS is issued annually by QMSA to all statisticians. Results are collated and analysed by QMSA and reviewed at divisional level to facilitate the development of improvement action plans. The outputs from the Quality review can be used to access targeted support from Quality and Methodology divisions where required.





Guidelines

Managers in the CSO should:

- Complete a Quality Review Self-Assessment annually covering each statistical product they have responsibility for.
- Develop an action plan to implement any improvement opportunities that have been identified.

CSO Quality Training

Key Learning Point(s):

Ensure that all staff understand their role in Quality

Avail of training opportunities as part of continuous professional development

The QMF requires that all CSO staff have a high degree of awareness of the importance of quality, an understanding of the role they play in managing quality, and an understanding of the various dimensions of quality and how these must be balanced and managed. Training is essential to ensuring that staff have the knowledge and skills to address quality in their work in statistical production.

A range of training supports have been developed and are offered to staff on a regular basis by QMSA including:

- An introductory "Why quality" training programme is available to all CSO staff including seconded staff in the ISS providing an overview of Quality Management and the QMF
- Mandatory training courses on GSBPM and the ES CoP for all staff
- Introductory training on quality for new entrants as part of CSO "Smart Start" programme
- Intermediate training on quality for managers on Quality management systems and frameworks, including Total Quality Management systems, ISO 9001:2015, EFQM, CAF, ESS QAF, etc.
- Training on Quality Reporting based on the ESS Handbook for Quality and Metadata Reports (2020)
- Training on Statistical Code of Practice and Principles for staff in the Irish Statistical System

Guidelines

Managers and staff should:

- Engage in continuous personal & professional development
- Complete a profile of skills in the CSO Skills Register
- Identify skills gaps and learning needs each year
- Avail of quality related training and complete any mandatory training within a reasonable time frame
- Share learning and good practice with other staff.





Supported Quality Appraisals

Key Learning Point(s):

Engage with independent statistical quality assurance supports for producers of statistics

Independent oversight forms a part of any quality assurance system. The CSO Management Board has established a programme of Supported Quality Appraisals (SQA) as a mechanism for independent statistical quality assurance. In each SQA, multidisciplinary teams with appropriate expertise from Quality, Methodology and Technology work to support statistical business areas in statistical quality assurance.

Using a multidisciplinary team in the design or review of statistical production helps ensure that the widest range of quality considerations around different components of production are examined. Methodologists bring expertise in statistical methods and standards; Technological experts provide insight into the efficiency and effectiveness of IT systems, while the subject-matter experts from the business area bring their in-depth knowledge of existing systems, user needs and domain content. The SQA team examines the full statistical production cycle, using the GSBPM as a frame of reference, working closely with business area managers to ensure an appropriate balance is sought between concern for quality and the considerations of cost, resources and burden.

While statistical quality remains primarily the responsibility of those working in each business area, an SQA provides additional quality assurance through the conduct of independent reviews of the implementation of quality standards and evaluation of statistical production systems. An SQA identifies priorities for improvement and supports the business area in the implementation of these improvement activities. SQAs, therefore, provide an independent, objective analysis of risks to statistical quality as well as providing practical steps that may be undertaken to mitigate such risks.

Guidelines

Managers involved in statistical production should undertake regular self-assessments of quality, identifying areas of concern that may be prioritised for inclusion in the annual programme of SQAs.







Data Protection, Confidentiality and the Legislative/Regulatory Environment





Data Protection - a core value

These Quality Guidelines are focussed primarily on the ESCoP principles of quality related to statistical processes and statistical outputs, nevertheless, there is an increased need for those involved in statistical production to have a sound knowledge of the wider framework of data management and the legislative environment. This section of the handbook introduces the reader to some key features of this environment, but users are expected to consult and familiarise themselves with the more detailed advice and and guidance provided elsewhere by the relevant areas of expertise.

Data is the lifeblood of every statistical organisation and the CSO, as Ireland's National Statistical Institute, relies on ready access to public, personal and business data to produce official statistics. Our ability to access this data, our organisational reputation and our ability to function is dependent on how we handle this data.

The CSO operates under the Statistics Act, 1993 collecting information directly from people, households and businesses in both statutory and voluntary surveys while also utilising the statistical potential of administrative records maintained by other public authorities. The credibility and reputation of the CSO depends on maximising the value and utility of the data available to it, while at the same time protecting the confidentiality of the individual persons and businesses supplying the data.

The information gathered by the CSO is used by the Office for statistical purposes only. This is a core value of the Office. The manner in which information should be protected is set out in legislation, through Office policies and in operational procedures.

The data environment in which we operate is rapidly evolving in terms of legal and societal requirements and all those involved in statistical production need to be aware of their responsibilities. The Office is committed to respecting and observing the latest Data Protection legislation and has invested in a comprehensive framework of policies, systems and supports to ensure that staff are cognizant of changing requirements.

Data Management and Security

Key Learning Point(s): Protect and manage data Understand data classifications Apply appropriate procedures, controls and checks Consult appropriate expertise (CSO Data Office)

CSO Data Management Policy

Data management is critical in the production of statistics and in the operation of any statistical agency. There are stringent obligations set out in national and European legislation with which the CSO and other producers of statistics must comply. The relevant acts, which apply in Irish law to Statistics and Information Systems, include (but are not limited to):

86

- The Statistics Act (1993)
- EU Regulation 223/2009 on European statistics as amended by Regulation 2015/759
- General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679)
- Data Protection Acts 1988 to 2018





Legislation in the area of data management and protection is constantly evolving. The CSO Data Management Policy is the overarching policy framework for the management of data in the CSO and is regularly updated to reflect changing legislative or policy requirements. The CSO Data Management policy is a living document which:

- Creates a data management and governance framework for the CSO.
- Defines minimum controls required to reduce risk and protect the CSO's data and information assets.
- Provides guidelines and instructions for the ownership, management, storage, retention, protection, destruction and security of CSO data.
- Outlines key related business and Technology processes and procedures.

The policy encompasses all information systems owned or operated by the CSO; all data and information held by the CSO, irrespective of how stored; paper records as well as IT storage (where applicable); all employees of the CSO and all contractors engaged by the CSO.

The Data Management Policy is founded on the appropriate classification of data held by the CSO which is set out in the CSO Data Classification Scheme, with associated detailed rules on the treatment of data in line with CSO's legal obligations.

CSO Data Classification Scheme – Categories and sub-categories		
Category A	All confidential statistical micro-data	
A1	All personal confidential statistical micro-data	
A2	All business confidential statistical micro-data	
A3	Research Micro-data Files (RMFs)	
Category B	Personal or commercially sensitive records used in the course of office administration	
Category C	Dissemination data	
C1	Releases, Tables and Publications in preparation	
C2	Published Releases, Tables and Publications	
C3	Tables or analysis provided in response to ad-hoc queries	
Category D	All other information recorded in the course of running the Office	

The Data Management Policy also defines various roles and responsibilities applicable to governance structures, management duties, policy and security supports and the duties of staff at all levels in the CSO.

All CSO staff must comply with the Data Management Policy and must adhere to the security, confidentiality and data protection standards which are implemented in their area for all stages of the data lifecycle, including

- Implementation of CSO business rules in relation to data
- Security, transport and storage of the data in line with CSO rules
- Management, control, ownership, security and integrity of the data
- Procedures to ensure compliance with the Data Management Policy and all other associated policies for statistical and non-statistical data
- Management of the data lifecycle, including secure destruction of statistical and non-statistical data when no longer required





The CSO Data Office provides advice on data management policies, working with other corporate support and statistical areas to drive compliance and provide training and promotion of CSO policies. Topics that the Data Office provide advice on include:

- Legislation
- Statistical confidentiality
- Principles of Data Protection
- General Data Protection Regulation (GDPR)
- Personal Data
- Privacy by Design and Default
- Data Protection Impact Assessments (DPIA)
- Anonymisation and pseudonymisation
- Role of staff in data protection
- Dealing with Data Protection Breaches

Various policies have been developed to deal with specific aspects of Data Management in the CSO operating environment. CSO staff must familiarise themselves with these policies and how they relate to their work. Policies under the aegis of the Data Management Policy currently include:

- Password Security Policy
- Data Retention and Destruction Policy
- Data Archiving/Long-term storage Policy
- Data Backup and Restore Policy
- Laptop/Tablet Security Policy
- Remote working access Policy
- CSO Mobile Device Security Policy
- Physical Security Policy

- Administrative Data Centre (ADC) Policy
- Research Microdata Files (RMFs) Policy
- Cryptography
- Work Hand-over Procedures and Responsibilities Policy
- Redaction Policy
- Statistical Disclosure Control (SDC) for microdata
- Statistical Disclosure Control (SDC) for Tabular Data

CSO staff should familiarise themselves with the CSO Data Management Policy (see Data Office Resource page on Lotus Notes). Detailed guidance, support and training is available from CSO Data Office: Dataoffice@cso.ie

Transparency about how we utilise data is an important aspect in maintaining public trust. CSO's Data Office has developed a web application facility allowing users to view details from various registers of data handling activities, including data matching, privacy impact assessments and data protection assessments. This can be accessed at: https://doregisters.cso.ie/



Guaranteeing Statistical Confidentiality



All information supplied to the CSO is treated as strictly confidential and is used for statistical purposes only. This commitment is guaranteed by law in The Statistics Act, 1993 and by EU legislation, such as Council Regulation (EC) No 223/2009 on European statistics for data collected for EU statistical purposes.

Further information on requirements around Statistical Data Confidentiality is provided in these guidelines under Section 3: Specific Components of Statistical Quality.

Access to Administrative Data in the CSO

Key Learning Point(s):

Explore the potential of administrative data as a first option

Develop procedures to effectively manage use of administrative data

Build and maintain close relationships with data suppliers

Consult appropriate expertise (e.g. CSO Administrative Data Centre)

In exploring the potential of administrative data sources within the CSO, it is important that those involved follow appropriate procedures to ensure compliance with legislative requirements. This can be done by consulting CSO's Administrative Data Centre (ADC) for the latest guidance. ADC manage administrative data flows and provide users with the appropriate level of access to relevant data.

The ADC Data Policy sets out data governance procedures in the Administrative Data Centre unit, and CSO user access procedures and responsibilities with respect to data holdings in the ADC data warehouse, which is the central repository of unit-level administrative data in the CSO.

In addition to administrative data, the ADC data warehouse holds certain non-administrative data such as data acquired commercially from non-public bodies (e.g., GeoDirectory data, Companies Registration Office accounts data) and some CSO survey data for which there is a legitimate access demand by CSO users outside the source section/division (e.g., Census data).





Some key points in relation to the ADC Data Warehouse include:

- All data held in the ADC data warehouse are organised by data flow. Most data flows relate to a single source type.
- Each ADC data flow is organised by reference period or instance. The instance refers to a time period usually relevant to some aspect of the data (i.e., date of registration, date of occurrence).
- All data flows in the ADC data warehouse are listed in the Administrative Data Register
- All data holdings in the ADC data warehouse are subject to ADC policy on data retention and formal agreements with external data providers.

The ADC data warehouse is structured in four tiers with data categorised according to the CSO Data Classification Scheme:

Data Tier	Data Classification	Purpose	Access Controls
Migration Tier	A1	Maintained primarily in case data are lost from other tiers and need to be reproduced in those other tiers. Data is held as received from the external body (e.g. in SAS, ASCII, CSV, HTML/XML, or other formats)	Access restricted to ADC personnel and essential IT Service Management (ITSM) personnel
Source Tier	A1	SAS dataset equivalents of data flows in the Migration Tier with no changes to data values records/ variables other than formatting/ labelling etc. Often contains identifying information relating to persons and businesses.	Access restricted to CSO users who have received approval from Data Office and a Director, Assistant Director General or the Director General.
Analysis Tier	A2	Exists to facilitate CSO user access to less-sensitive Source Tier data, and to hold certain value-added data flows created by ADC personnel. The Analysis Tier contains (as SAS datasets) de-identified subsets, or a combination, of Source Tier data flows	Access available to CSO users upon request, i.e., without the need for third-person approval or a PIA.
Research Microdata File Tier	A3	Contains Research Microdata File data flows produced as the result of RMF requests made directly to ADC unit, or other RMF requests requiring ADC unit input. No non-CSO identifier, whether relating to a business or a person, should appear on any RMF	Access for researchers to data held in the ADC data warehouse comes under the rules and protocols of the CSO's Policy on Access to Research Microdata Files

CSO's Administrative Data Centre can be contacted at: adc@cso.ie



Use of Statistical Statutory Instruments (SIs)



The bulk of the statistics produced by the CSO are mandatory under EU Statistical legislation and while every effort is made to minimise the statistical reporting burden in surveys through the use of sampling, the CSO must ensure that the samples used are representative and that the derived results are accurate and reliable. The CSO utilises the Statistics Act, 1993 to undertake both statutory (i.e. obligatory) and voluntary surveys. For some critical CSO inquiries, the Taoiseach can make a Statutory Order under Section 25 of the Statistics Act, requiring persons or undertakings to provide particular information to the CSO. Participation is compulsory under law in these statutory inquiries. Preparation and maintenance of Statutory Instruments must be closely managed.

Detailed procedures which must be followed in the preparation and approval of Statutory Orders (also called Statutory Instruments) are set out in Office Notice 10/2008. Preparation of Statutory Instruments is a complex, time-consuming process and early engagement with CSO Data Office and CSO Governance Compliance & Legal Unit is strongly advised. Some of the key points to note in relation to the process include:

- Allow sufficient time for all phases of the process, including consultations and interactions with appropriate internal (e.g. DG Office, Data Office) and external bodies (e.g. Data Protection Commissioner, Department of the Taoiseach, Office of the Attorney General).
- Put proposed statutory instruments through the "Quality Regulation Checklist" (detailing considerations and assessment around necessity, impact, compatibility, burden, timelines and consultations).
- Request drafting of the statutory instrument by the Office of the Attorney General (providing a suitable draft statutory instrument, copy of any previous SI if applicable, copy of the Statistics Act 1993, copies of the Inquiry forms, copies of any EU Acts to be put into effect, and any other relevant information).
- Liaise with appointed Parliamentary Counsel to arrange the settling of the SI (i.e. clearing), approval or amendment of the SI and stamping and approval by the Office of the Parliamentary Counsel.
- Ensure English and Irish versions of the SI are sent both electronically and in hard copy to the contact person in the Department of the Taoiseach, with an explanatory note outlining the purpose of the order. Give as much prior notice as possible to the Department of the Taoiseach to ensure that all proofreading is completed in advance of signature and to ensure the efficient production of the SI.
- Register the SI and publish the signed and numbered SI on the CSO website as soon as possible after signature. Provide all required details to the relevant Parliamentary Counsel.

GDPR and Statistical Statutory Instruments

The General Data Protection Regulation imposes requirements on those preparing statistical Statutory Instruments (SIs) involving the processing of personal data to engage in prior consultation with the Data Protection Commission (DPC).

The Data Office acts as the liaison between the CSO and the DPC and should be consulted at an early point to ensure that the appropriate documentation, including a Data Protection Impact Assessment (DPIA), is compiled for engagement with the DPC. Early engagement with Data Office is advised to help ensure compliance with legislative requirements. For more information, see: https://www.dataprotection.ie/en/organisations/know-your-obligations/ data-protection-impact-assessments/prior-consultation





Statistical Legislative Environment

Key Learning Point(s):

Understand legal basis for statistical activities (National & EU legislation)

Evolving legal environment – Consult appropriate expertise (e.g. CSO Legal Unit, CSO Data Office)

The operation of the CSO as Ireland's National Statistical Institute and the collection and management of data is governed by a wide body of legislation which is being added to on a regular basis. Those involved in statistical production should familiarise themselves with the legal basis and the legal obligations that apply to their work. CSO Governance Compliance & Legal Unit and CSO Data Office can provide detailed information on the legislative environment and should be consulted on a regular basis. Some of the key legislation that underpins statistical production is briefly described here.

Statistics Act, 1993

The Statistics Act 1993 provides the primary legal basis for the CSO. Prior to 1993, Irish statistics were compiled under the Statistics Acts, 1926 and 1946. The Statistics Act, 1993 provides a legislative basis for the compilation and dissemination of official statistics incorporating the following key provisions:

- the establishment on a statutory basis of the CSO as an independent Office (under the aegis of the Taoiseach) for the collection, compilation, extraction and dissemination for statistical purposes of information relating to economic, social and general activities and conditions in the State.
- the functions of the CSO, including the obligation to co-ordinate official statistics compiled by public authorities, the right to assess the statistical potential of the records maintained by public authorities and to ensure that this potential is realised in so far as resources permit;
- the appointment on a formal basis of the Director-General of the CSO who shall also be independent on statistical matters (i.e. sole responsibility for the statistical methodologies and professional standards to be followed, the content of statistical releases and publications, and the timing and methods of dissemination of the statistics compiled);
- the right of access, subject to some limitations and conditions, of the CSO to administrative records held by public authorities for statistical purposes; and
- the obligation on the CSO to treat all individual information relating to persons or concerns as strictly confidential and to use such information solely for statistical purposes.

The Act fully reflects the Fundamental Principles of Official Statistics adopted by the UNECE in 1992.

EU Statistical Regulations

The CSO is legally bound to contribute statistical data on a European level. The CSO's legal obligations are contained in a variety of European legislation/regulations. EU Regulation 223/2009 (amended by Regulation 2015/759) constitutes the legal basis for the preparation of the European statistical programme and provides the framework for the development, production and dissemination of European statistics.

The full listing and text of European legislation, including Commission Regulations and Decisions, are available at: https://eur-lex.europa.eu/statistics/statistics.html

Back



Data Protection Acts 1988 - 2018

The Data Protection Act, 2018, updates the Data Protection Acts 1988 and 2003. Together with the General Data Protection Regulation, the Data Protection Acts govern how, when and why any organisation can process personal data.

The Data Protection Act 2018:

- Establishes a new Data Protection Commission as the State's data protection authority
- Gives further effect to the GDPR in areas where member states have flexibility

General Data Protection Regulation (GDPR), 2018

The GDPR harmonises, simplifies and streamlines data protection rules across the European Union, with the following general objectives:

- strengthening the protection of people's fundamental right to privacy;
- adopting a risk-based approach when assessing a data processing operation;
- putting data protection compliance at the centre of an organisation's governance.

The GDPR sets out the key principles, rights and obligations for the processing of personal data of data subjects (with some limited exceptions). It imposes responsibilities on organisations to have a legal basis for, maintain the privacy of, and have protections in place for the personal data they process. It provides for data subject rights and stipulates powers of investigation and enforcement for the supervisory authority (Data Protection Commission). All statistical organisations must be able to demonstrate accountability to these rules.

CSO performs its data management responsibilities under the GDPR in accordance with the following seven principles:

- Lawfulness, fairness and transparency processes are lawful, fair and transparent
- Purpose limitation collected for specified, explicit and legitimate purposes.
- Data minimisation adequate, relevant and limited to what is necessary.
- Accuracy accurate and, where necessary, kept up to date.
- Storage limitation kept for no longer than is necessary
- Integrity and confidentiality ensure appropriate security of the data,
- Accountability not only be required to comply with data protection requirements, but also be able to demonstrate compliance

Privacy by Design and Default

Data Protection must be factored into each new process, product, service or application from the outset, and, by default, only minimum amounts of personal data as necessary for specific purposes are collected and processed.

















Appendix 1. European Statistics Code of Practice (ES CoP)

The ES CoP is the cornerstone of the common quality framework of the European Statistical System. It is a selfregulatory instrument based on 16 Principles covering the institutional environment, statistical processes and statistical outputs. The ES CoP was adopted by the European Statistical Programme Committee in 2005 and was revised in 2011 and 2017.

Each of the Principles has an associated set of indicators of best practice and standards that provide guidance for reviewing the implementation of the Code of Practice. Detailed information on the Code of Practice can be found on the Eurostat website at: https://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice

The principles of the ES CoP are grouped under three headings:

Institutional environment

Institutional and organisational factors have a significant influence on the effectiveness and credibility of a statistical authority developing, producing and disseminating European Statistics. The relevant Principles are: Professional Independence; Coordination and Cooperation; Mandate for Data Collection; Adequacy of Resources; Quality Commitment; Statistical Confidentiality; Impartiality and Objectivity.

Statistical Processes

European and other international standards, guidelines and good practices are fully observed in the statistical processes used by the statistical authorities to develop, produce and disseminate European Statistics, while constantly striving for innovation. The credibility of the statistics is enhanced by a reputation for good management and efficiency. The relevant Principles are: Sound methodology; Appropriate statistical procedures; Non-excessive burden on respondents; Cost effectiveness

Statistical Output

Available statistics meet users' needs. Statistics comply with the European quality standards and serve the needs of European institutions, governments, research institutions, business concerns and the public generally. The relevant Principles are: Relevant; Accurate and Reliable; Timely and Punctual; Coherent and Comparable (across regions and countries); Readily Accessible by Users i.e. the Principles of Statistical Output Quality.

European Statistics Code of Practice – 16 Principles



Back

Appendix 2. Quality Assurance Framework of the European Statistical System (ESS QAF)

Common Quality Framework of the European Statistical System

The 2019 edition of the Quality Assurance Framework of the European Statistical System (ESS QAF) is aligned with the European Statistics Code of Practice (ES CoP) and together are key parts of the common quality framework of the ESS. These provide a base on which high quality European Statistics are developed, produced and disseminated.

The ESS QAF supports the implementation of the ES CoP and applies to statistical authorities of the ESS, including Eurostat, the National Statistical Institutes (NSIs) and Other National Authorities (ONAs) which are responsible for the development, production and dissemination of European Statistics. The ESS QAF represents a collection of detailed methods, tools and good practices that are suggested for use across the ESS in the application of the high-level Principles and Indicators of the ES Code of Practice.

This appendix highlights the indicators of quality associated with ES CoP Principles for Statistical Processes and Statistical Outputs that have been outlined in the ESS QAF. Users should refer to the ESS QAF document for the full list of indicators, methods and good practices (The ESS QAF is available from Eurostat at https://ec.europa.eu/eurostat/web/quality and is also available on the CSO QIS).

Statistical Processes - Principles and Indicators

Principle 7: Sound Methodology.

Sound Methodology underpins quality statistics. This requires adequate tools, procedures and expertise.

Indicators:

- The overall methodological framework used for European Statistics follows European and other international standards, guidelines, and good practices, while constantly striving for innovation.
- Procedures are in place to ensure that standard concepts, definitions, classifications and other types of standards are consistently applied throughout the statistical authority.
- The registers and frames used for European Statistics are regularly evaluated and adjusted if necessary in order to ensure high quality.
- Detailed concordance exists between national classifications systems and the corresponding European systems.
- Graduates in the relevant academic disciplines are recruited.
- Statistical authorities implement a policy of continuous vocational training for their staff.
- Statistical authorities maintain and develop cooperation with the scientific community to improve methodology, the effectiveness of the methods implemented and to promote better tools when feasible.

Principle 8: Appropriate Statistical Procedures.

Appropriate statistical procedures, implemented throughout the statistical processes, underpin quality statistics.

Indicators:

- When European Statistics are based on administrative and other data, the definitions and concepts used for non-statistical purposes are a good approximation to those required for statistical purposes.
- In the case of statistical surveys, questionnaires are systematically tested prior to the data collection.
- Statistical processes are routinely monitored and revised as required.
- Metadata related to statistical processes are managed throughout the statistical processes and disseminated, as appropriate.
- Revisions follow standard, well-established and transparent procedures.





- Agreements are made with holders of administrative and other data which set out their shared commitment to the use of these data for statistical purposes.
- Statistical authorities co-operate with holders of administrative and other data in assuring data quality.

Principle 9: Non-excessive Burden on Respondents.

The response burden is proportionate to the needs of the users and is not excessive for respondents. The statistical authorities monitor the response burden and set targets for its reduction over time.

Indicators:

- The range and detail of European Statistics demands is limited to what is absolutely necessary.
- The response burden is spread as widely as possible over survey populations and monitored by the statistical authority.
- The data sought from businesses is, as far as possible, readily available from their accounts and electronic means are used where possible to facilitate its return.
- Administrative and other data sources are used whenever possible to avoid duplicating requests for data.
- Data sharing and data integration, while adhering to confidentiality and data protection requirements, are promoted to minimise response burden.
- Statistical authorities promote measures that enable the linking of data sources in order to minimise response burden.

Principle 10: Cost effectiveness

Resources are used effectively.

Indicators:

- Internal and independent external measures monitor the statistical authority's use of resources.
- The productivity potential of information and communications technology is being optimised for the statistical processes.
- Proactive efforts are made to improve the statistical potential of administrative and other data sources and to limit recourse to direct surveys.
- Statistical authorities promote, share and implement standardised solutions that increase effectiveness and efficiency.

Statistical Output - Principles and Indicators

Principle 11: Relevance

European Statistics meet the needs of users.

Indicators:

- Procedures are in place to consult users, to monitor the relevance and value of existing statistics in meeting their needs, and to consider and anticipate their emerging needs and priorities. Innovation is pursued to continuously improve statistical output.
- Priority needs are being met and reflected in the work programme.
- User satisfaction is monitored on a regular basis and is systematically followed up.

Principle 12: Accuracy and Reliability.

European Statistics accurately and reliably portray reality.





Indicators:

- Source data, integrated data, intermediate results and statistical outputs are regularly assessed and validated.
- Sampling errors and non-sampling errors are measured and systematically documented according to the European standards
- Revisions are regularly analysed in order to improve source data, statistical processes and outputs.

Principle 13: Timeliness and Punctuality.

European Statistics are released in a timely and punctual manner.

Indicators:

- Timeliness meets European and other international release standards.
- A standard daily time for the release of European Statistics is made public.
- The periodicity of statistics takes into account user requirements as much as possible.
- Divergence from the dissemination time schedule is publicised in advance, explained and a new release date set.
- Preliminary results of acceptable aggregate accuracy and reliability can be released when considered useful.

Principle 14: Coherence and Comparability.

European Statistics are consistent internally, over time and comparable between regions and countries; it is possible to combine and make joint use of related data from different data sources.

Indicators:

- Statistics are internally coherent and consistent (i.e. arithmetic and accounting identities observed).
- Statistics are comparable over a reasonable period of time.
- Statistics are compiled on the basis of common standards with respect to scope, definitions, units and classifications in the different surveys and data sources.
- Statistics from different data sources and with different periodicity are compared and reconciled.
- Cross-national comparability of the data is ensured within the European Statistical System through periodical exchanges between the European Statistical System and other statistical systems. Methodological studies are carried out in close co-operation between the Member States and Eurostat.

Principle 15: Accessibility and Clarity.

European Statistics are presented in a clear and understandable form, released in a suitable and convenient manner, available and accessible on an impartial basis with supporting metadata and guidance.

Indicators:

- Statistics and the corresponding metadata are presented, and archived, in a form that facilitates proper interpretation and meaningful comparisons.
- Dissemination services use modern information and communication technology, methods, platforms and open data standards.
- Custom-designed analyses are provided when feasible and the public is informed.
- Access to microdata is allowed for research purposes and is subject to specific rules or protocols.
- Metadata related to outputs are managed and disseminated by the statistical authority according to the European standards.
- Users are kept informed about the methodology of statistical processes including the use and integration of administrative and other data.
- Users are kept informed about the quality of statistical outputs with respect to the quality criteria for European Statistics.





Appendix 3. List of ESS Quality Performance Indicators (ESS QPIs)

The following table lists and briefly describes the 16 standard ESS QPIs. Information on how to calculate these is available in ESS Guidelines for the Implementation of the ESS Quality and Performance Indicators (QPI), available from Eurostat at https://ec.europa.eu/eurostat/web/quality/quality-reporting

R1. Data completeness - rate	The ratio of the number of data cells (entities to be specified by the Eurostat domain manager) provided to the number of data cells required by Eurostat. The ratio is computed for a chosen dataset and a given period.
A1. Sampling error - indicators	The sampling error can be expressed:
	a) in relative terms, in which case the relative standard error or, synonymously, the coefficient of variation (CV) is used. (The standard error of the estimator is the square root of its variance) The estimated relative standard error (the estimated CV) is the estimated standard error of the estimator divided by the estimated value of the parameter.
	b) in terms of confidence intervals, i.e. an interval that includes, with a given level of confidence, the true value of a parameter. The width of the interval is related to the standard error.
A2. Over-coverage - rate	The rate of overcoverage is the proportion of units accessible via the frame that do not belong to the target population (are out-of- scope).
A3. Unit non-response - rate	The ratio of the number of units with no information or not usable information (non-response, etc.) to the total number of in-scope (eligible) units. The ratio can be weighted or un-weighted.
A4. Item non-response - rate	The item non-response rate for a given variable is defined as the (weighted) ratio between in-scope units that have not responded and in-scope units that are required to respond to the particular item.
A5. Imputation - rate	Imputation is the process used to assign replacement values for missing, invalid or inconsistent data that have failed edits. This excludes follow-up with respondents and manual review and correction (if applicable). Thus, imputation as defined above occurs after data collection, no matter from which source or mix of sources the data have been obtained, including administrative data.
A6. Common units - proportion	The proportion of units covered by both the survey and the administrative sources in relation to the total number of units in the survey.
A7. Data revision - average size	The average over a time period of the revisions of a key item. The "revision" is defined as the difference between a later and an earlier estimate of the key item.
T1. Time lag - first results	The timeliness of statistical outputs is the length of time between the end of the event or phenomenon they describe and their availability as first results.
T2. Time lag - final results	The timeliness of statistical outputs is the length of time between the end of the event or phenomenon they describe and their availability as final results.

Next

T3. Punctuality - delivery and publication	Punctuality is the time lag between the delivery/release date of data and the target date for delivery/release as agreed for delivery or announced in an official release calendar, laid down by Regulations or previously agreed among partners.
AC1. Data tables - consultations	Number of consultations of data tables within a statistical domain for a given time period. By "number of consultations" it is meant number of data tables views, where multiple views in a single session count only once.
AC2. Metadata - consultations.	Number of metadata consultations (ESMS) within a statistical domain for a given time period. By "number of consultations" it is meant the number of times a metadata file is viewed.
AC3. Metadata completeness – rate	The ratio of the number of metadata elements provided to the total number of metadata elements applicable.
CC1. Asymmetry for mirror flows statistics – coefficient	Discrepancies between data related to flows, e.g. for pairs of countries.
CC2. Length of comparable time series	Number of reference periods in time series since last break. If there has not been any break, the indicator is equal to the number of the time points in the time series.



Appendix 4: The UNECE Generic Statistical Business Process Model (GSBPM)

Building quality into all phases of the statistical production process

The Generic Statistical Business Process Model (GSBPM) describes and defines the set of business processes needed to produce official statistics and can be used as an aid in understanding and managing statistical production. Internationally, it provides a standard framework and harmonised terminology to help statistical organisations to modernise their statistical production processes, as well as to share methods and components.

The GSBPM can also be used for integrating data and metadata standards, as a template for process documentation, for harmonising statistical computing infrastructures, and to provide a framework for process quality assessment and improvement.

Specify need	s Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame and sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluations
1.3 Estabilsh output objectives	2.3 Design collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit & impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system		5.5 Derive new variables and units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business	2.6 Design production	3.6 Test statistical		5.6 Calculate weights			
workflows proces		process	ocess	5.7 Calculate			
		3.7 Finalise production system		5.8 Finalise data files			



Appendix 5: CSO policies, standards and key legislation

Statistical
Statistics Act, 1993
Consolidated text of Regulation 223/2009 on European Statistics
Data Protection & Freedom of Information
Data Protection Act 2018
General Data Protection Regulation
Informal Compendium of Data Protection Acts 1988 & 2003
Freedom of Information Act 2014
Civil Service
Official Secrets Act, 1963
Key CSO Policies and Guidelines
Statutory Instruments
Production of Statutory Instruments (Office Notice 10/2008)
Confidentiality
CSO Statistical Confidentiality Code of Practice
Data Classification Scheme - Office Notice 24/2014
CSO Data Management Policy - Office Notice 16/2019
Official Statistics and Data Protection - Office Notice 10/2018
CSO Policy on Retention and Secure Disposal of Statistical Returns and Storage Media (Office Notice 10/2016)
 CSO Policy on the Disposal of Confidential Paper Records (Office Notice 9/2009)
 CSO Policy on the Disposal of Confidential Paper Records (Office Notice 9/2009) CSO Policy on Protected Disclosure Reporting in the Workplace (Office Notice 04/2016)
 CSO Policy on the Disposal of Confidential Paper Records (Office Notice 9/2009) CSO Policy on Protected Disclosure Reporting in the Workplace (Office Notice 04/2016) Declaration regarding statistical confidentiality and data protection – to be signed by Employees and Agents of Contracting Companies (Office Notice 26/2014)
 CSO Policy on the Disposal of Confidential Paper Records (Office Notice 9/2009) CSO Policy on Protected Disclosure Reporting in the Workplace (Office Notice 04/2016) Declaration regarding statistical confidentiality and data protection – to be signed by Employees and Agents of Contracting Companies (Office Notice 26/2014) Statistical Disclosure Control
 CSO Policy on the Disposal of Confidential Paper Records (Office Notice 9/2009) CSO Policy on Protected Disclosure Reporting in the Workplace (Office Notice 04/2016) Declaration regarding statistical confidentiality and data protection – to be signed by Employees and Agents of Contracting Companies (Office Notice 26/2014) Statistical Disclosure Control Statistical Disclosure Control for Microdata (Office Notice 04/2019)
 CSO Policy on the Disposal of Confidential Paper Records (Office Notice 9/2009) CSO Policy on Protected Disclosure Reporting in the Workplace (Office Notice 04/2016) Declaration regarding statistical confidentiality and data protection – to be signed by Employees and Agents of Contracting Companies (Office Notice 26/2014) Statistical Disclosure Control Statistical Disclosure Control for Microdata (Office Notice 04/2019) Statistical Disclosure Control for Tabular data (Office Notice 12/2019)

Next

 \bigcirc

Research Microdata Files

Revised Office Policy on Access to Microdata (Office Notice 16/2017)

Administrative Data

The Administrative Data Centre Data Policy

Data Security

CSO Policy on the Reporting and Management of Data Security Breaches (Office Notice 11/2017)

Seasonal Adjustment

CSO Seasonal Adjustment Policy

Technology

- Information and Technologies Acceptable Use Policy (Office Notice 15/2015)
- CSO Policy on the Use of Portable Electronic Media (Office Notice 13/2013)
- User Advice Note on Security Best Practice and Policies
- End User Computing Standards

Risk

■ CSO Risk Management Policy (Office Notice 07/2016)

Further details are available at:

CSO website: https://www.cso.ie/en/aboutus/

- CSO Methodology Homepage: https://methodology.cso.ie/
- CSO Editorial Board homepage: https://editorialboard.cso.ie/
- CSO Policies homepage on Lotus Notes



Sources and References

Eurostat
European Statistics Code Of Practice, 2017: https://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice
European Statistical System, Quality Assurance Framework of the European Statistical System, 2019 https://ec.europa.eu/eurostat/web/quality
ESS Handbook for Quality and Metadata reports, 2020: https://ec.europa.eu/eurostat/web/quality/quality-reporting
ESS Quality and Performance Indicators, 2014: https://ec.europa.eu/eurostat/web/quality/quality-reporting
Handbook on Data Quality Assessment Methods and Tools, 2007: https://ec.europa.eu/eurostat/ramon/ statmanuals/files/Handbook_on_data_qual_assess_tools.pdf
Handbook on improving quality by analysis of process variables: https://ec.europa.eu/eurostat/documents/64157/4373903/06-Handbook-on-improving-quality-by-analysis-of- process-variables.pdf/b0006e09-1708-4f8e-97e0-6a54d840b92b
Single Integrated Metadata Structure and its Technical Manual, 2014: https://ec.europa.eu/eurostat/web/quality/quality-reporting
European Self-Assessment Checklist for Survey Managers (DESAP): https://ec.europa.eu/eurostat/web/quality/quality-reporting
ESS guidelines on seasonal adjustment, 2015 edition: https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-15-001
ESS guidelines on temporal disaggregation, benchmarking and reconciliation, 2018: https://ec.europa.eu/eurostat/documents/3859598/9441376/KS-06-18-355-EN.pdf/fce32fc9-966f-4c13-9d20- 8ce6ccf079b6
Quality Guidelines for Multisource Statistics. ESSnet KOMUSO, 2019: https://ec.europa.eu/eurostat/cros/content/quality-guidelines-multisource-statistics-qgmss_en
Quality Guidelines on Frames for Social Statistics, ESSnet KOMUSO, 2019, V1.51: https://ec.europa.eu/eurostat/cros/system/files/qgfss-v1.51.pdf
Quality Declaration of the European Statistical System: https://ec.europa.eu/eurostat/documents/4031688/8188985/KS-02-17-428-EN-N.pdf/116f7c85-cd3e-4bff- b695-4a8e71385fd4
Eurostat Portal on Imputation: https://ec.europa.eu/eurostat/cros/content/imputation_en
State of the art concerning the auditing activity in NSIs, Leadership Group on Quality (LEG), 2003: https://ec.europa.eu/eurostat/documents/64157/4374310/35-AUDITING-ACTIVITY-IN-NSIs-2003- %28LEG+0N+QUALITY%29.pdf/0f34409a-1559-4bce-95b8-dcd93e145c86

UN/UNECE

UN Fundamental Principles of Official Statistics, 2014: https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx



UN National Quality Assurance Frameworks Manual for Official Statistics (UN NQAF Manual), 2019: https://unstats.un.org/unsd/methodology/dataquality/un-nqaf-manual/

UNECE Generic Statistical Business Process Model (GSBPM), Version 5.1, 2019: https://statswiki.unece.org/display/GSBPM/

UNECE Quality indicators for GSBPM: https://statswiki.unece.org/display/GSBPM/

Nationally and Internationally Developed Data Quality References, UNSD: https://unstats.un.org/unsd/methodology/dataquality/quality-references/

Guidelines on producing leading, composite and sentiment indicators, 2019: https://unstats.un.org/unsd/nationalaccount/docs/ECECESSTAT20192.pdf

UNECE Generic Statistical Data Editing Model GSDEM (Version 2.0, June 2019): https://statswiki.unece.org/display/sde/GSDEM

Guidelines on the use of statistical business registers for business demography and entrepreneurship statistics, UNECE, 2018: https://unece.org/DAM/stats/publications/2018/ECECESSTAT20185.pdf

Recommendations for Promoting, Measuring and Communicating the Value of Official statistics, UNECE, 2018: https://unece.org/DAM/stats/publications/2018/ECECESSTAT20182.pdf

Guidelines on the use of registers and administrative data for population and housing censuses, UNECE, 2018: https://unece.org/fileadmin/DAM/stats/publications/2018/ECECESSTAT20184.pdf

OECD

Quality Framework for OECD Statistical Activities, 2011: https://www.oecd.org/sdd/qualityframeworkforoecdstatisticalactivities.htm

OECD Good Statistical Practice Toolkit: http://www.oecd.org/statistics/good-practice-toolkit/goodpractices/

Australia

Australian Bureau of Statistics Quality Framework: https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Quality:+The+ABS+data+quality+framework

ABS Data Quality Framework, 2009: https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Quality:+The+ABS+data+quality+framework

Bulgaria

Guidelines on Quality Criteria in the National Statistical System of Bulgaria: https://www.nsi.bg/en/content/475/basic-page/quality

Guidelines for quality assessment in the NSI, National Statistical Institute of Bulgaria: https://www.nsi.bg/en/content/475/basic-page/quality

Handbook on Internal Quality Audit in the NSI, National Statistical Institute of Bulgaria: https://www.nsi.bg/en/ content/475/basic-page/quality





Canada

Statistics Canada Quality Guidelines. Sixth Edition, 2019: https://www150.statcan.gc.ca/n1/en/catalogue/12-539-X

Statistics Canada Quality Assurance Framework, 2017: https://www150.statcan.gc.ca/n1/en/catalogue/12-586-X

Statistics Canada, Policy on Informing Users of Data Quality and Methodology, 2000: https://www.statcan.gc.ca/eng/about/policy

Statistics Canada. Policy on Standards, 2004: https://www.statcan.gc.ca/eng/about/policy/standards

Statistics Canada Data Quality Toolkit, 2017: https://www.statcan.gc.ca/eng/data-quality-toolkit

Statistics Canada Survey Methods and Practices, 2010: https://www150.statcan.gc.ca/n1/pub/12-587-x/12-587-x2003001-eng.pdf

Denmark

Objectives and Frameworks for the Development of Statistics Denmark, Statistics Denmark, 2015: http://www.dst.dk/pukora/epub/upload/14870/strat.pdf

Estonia

Quality Policy of Statistics Estonia, Statistics Estonia, 2016: https://vana.stat.ee/quality-policy

Dissemination Policy of Statistics Estonia, Statistics Estonia,: https://vana.stat.ee/dissemination-policy

Finland

Quality Management, Statistics Finland: http://www.stat.fi/org/periaatteet/laadunhallinta_en.html

Ireland

CSO Standards and Guidelines, Quality in Statistics, Volume 1, Central Statistics Office, Ireland, 2006: https://www.cso.ie/en/aboutus/lgdp/corporatepublications/

CSO Code of Practice on Statistical Confidentiality: https://www.cso.ie/en/aboutus/lgdp/csodatapolicies/

CSO Statement of Strategy 2020-2023: https://www.cso.ie/en/aboutus/lgdp/corporatepublications/

National Statistics Board, A World Class Statistical System for Ireland - Strategic Priorities for Official Statistics 2015 – 2020, Ireland, 2015: https://www.nsb.ie/media/nsbie/pdfdocs/NSB_Strategy_2015-2020.pdf





National Statistics Board, The Irish Statistical System: The Way Forward and Joined Up Government Needs Joined Up Data, 2011:

https://www.nsb.ie/media/nsbie/pdfdocs/NSB_ISS_Position_Papers.pdf

Revenue Ireland, Using Behavioural Insights to Increase Survey Response Rates, Revenue Statistics & Economic Research Branch, 2019:

https://www.revenue.ie/en/corporate/documents/research/behavioural-insights.pdf

Revenue Ireland, Statistical Disclosure Control Protocol for Revenue, Revenue Statistics & Economic Research Branch, 2019:

https://www.revenue.ie/en/corporate/documents/statistics/about/statistical-disclosure-control.pdf

Italy

Quality Guidelines, Istituto Nazionale di Statistica (Istat), Italy: https://www.istat.it/en/methods-and-tools/tools-for-data-quality/guidelines

Tools for Data Quality, Istituto Nazionale di Statistica (Istat), Italy: https://www.istat.it/en/methods-and-tools/tools-for-data-quality/guidelines

Information System on Quality (SiQual), Istituto Nazionale di Statistica (Istat), Italy: https://www.istat.it/en/methods-and-tools/tools-for-data-quality/siqual

Netherlands

Quality Guidelines 2014, Statistics Netherlands' Quality Assurance Framework at Process Level: https://www.cbs.nl/en-gb/background/2014/12/quality-guidelines-2014

Statistics Netherlands, Checklist Quality of Statistical Output, 2009: https://www.cbs.nl/en-gb/background/2009/07/checklist-quality-of-statistical-output

New Zealand

Stats NZ (2019). A guide to good survey design: Fifth edition: https://www.stats.govt.nz/methods/a-guide-to-good-survey-design-fifth-edition

Statistics NZ, Data quality for 2018 Census: https://www.stats.govt.nz/2018-census/data-quality-for-2018-census

Norway

Statistics Norway, Quality work in Statistics Norway, 2019: https://www.ssb.no/en/befolkning/artikler-og-publikasjoner/quality-work-in-statistics-norway

Portugal

General Guidelines of Official Statistical Activity 2018-2022, Statistics Council, Portugal: https://www.ine.pt/ine_novidades/LGAE0_2018_2022_EN/




Slovenia

Republic of Slovenia Statistical Office, Guidelines for Quality Assurance, 2017: https://www.stat.si/StatWeb/en/Methods/

Republic of Slovenia Statistical Office, Quality Indicators, 2017: https://www.stat.si/StatWeb/en/Methods/

South Africa

South African Statistical Quality Assessment Framework (SASQAF), 2nd ed./ Statistics South Africa – Pretoria: Statistics South Africa, 2010 https://unstats.un.org/unsd/dnss/docs-ngaf/SouthAfrica-SASQAF_Edition_2.pdf

Spain

Arribas, C., Lorca, D., Salinero, A., Colmenero, A , Measuring statistical quality at the Spanish National Statistical Institute – Presentation to European Conference on Quality in Official Statistics, 2008: http://q2008.istat.it/sessions/9.html

Sweden

Biemer, Paul, Trewin, Dennis, Bergdahl, Heather and Japec, Lilli. "A System for Managing the Quality of Official Statistics" Journal of Official Statistics, vol.30, no.3, 2014, pp.381-415. https://doi.org/10.2478/jos-2014-0022

United Kingdom

Bremner, C., Davies, C., A New Process for Assessing the Quality of an Output Makes Better Quality Statistics, Office for National Statistics, United Kingdom – Presentation to European Conference on Quality in Official Statistics, 2016 :

https://www.ine.es/q2016/docs/q2016Final00094.pdf

Government Statistical Service, GSS Quality Strategy 2019 to 2021: https://gss.civilservice.gov.uk/wp-content/uploads/2019/06/GSS-Quality-Strategy.pdf

Guidelines for Measuring Statistical Output Quality, Version 4.1, Office for National Statistics, United Kingdom, 2013:

https://webarchive.nationalarchives.gov.uk/20160106003751/http://www.ons.gov.uk/ons/guide-method/method-quality/quality/guidelines-for-measuring-statistical-quality/index.html

NISRA, Quality Assurance and Quality Review, 2016: https://gss.civilservice.gov.uk/wp-content/uploads/2016/01/NISRA-Quality-Assurance-and-Quality-Reviews.pdf

Office for National Statistics, UK, Quality reviews – Explanations and lists of quality reviews completed by ONS: https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/qualityinofficialstatistics/ qualityreviews

Back



Office for National Statistics, UK, Quality and methodology information – Reports covering the relevance, accuracy, timeliness, accessibility, and coherence of our statistics: https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/qualityinofficialstatistics/ qualityandmethodologyinformation

Scottish Government Improving Quality Subgroup, A Guide to Basic Quality Assurance in Statistics, 2010: https://www.webarchive.org.uk/wayback/archive/20150218144849/http://www.gov.scot/Topics/Statistics/About/QAguide

Sharp, G., The application of "Value Engineering" tools to risk assess the outputs of an NSI, – Presentation to European Conference on Quality in Official Statistics, 2014 : http://www.q2014.at/papers-presentations.html

UK Statistics Authority, Code of Practice for Official Statistics, Edition 2.0, 2018 https://www.statisticsauthority.gov.uk/code-of-practice/

UK Statistics Authority, Quality Assurance of Administrative Data - Setting the Standard, 2015: https://osr.statisticsauthority.gov.uk/wp-content/uploads/2019/02/qualityassurancetoolkit_updated_Feb19_2.pdf

Welsh Government, Top Tips for Checking: https://gss.civilservice.gov.uk/wp-content/uploads/2016/01/Welsh-Government-Top-Tips-for-Checking.pdf

Whilst every effort has been made to ensure that links to documents or websites provided in this handbook are correct, it should be noted that these links are subject to change. If a link is not functioning as expected, try a general search using the title of the relevant document and please contact **quality_unit@cso.ie** to let us know about the issue.



Glossary	
DPC	Data Protection Commissioner
DPIA	Data Protection Impact Assessment
DPS	CSO Directory of Products and Services
ES CoP	European Statistics Code of Practice
ESS	European Statistical System
EUROSTAT	European Statistics organisation (statistical office of the European Union)
GDPR	General Data Protection Regulation
GSBPM	Generic Statistical Business Process Model
ISTAT	Italian National Institute of Statistics
NSI	National Statistical Institute
OECD	Organisation for Economic Cooperation and Development
ONA	Other National Authority (producing official European Statistics)
ONS	Office of National Statistics, UK
PIA	Privacy Impact Assessment
QAF	Quality Assurance Framework
QIS	CSO Quality Information System
QMF	Quality Management Framework
QPI	Quality Performance Indicator
QRS	Quality Review Survey
ROPA	Record of Processing Activities
SI	Statutory Instrument
SLA	Service Level Agreement
SQA	Supported Quality Appraisal
UNECE	United Nations Economic Commission for Europe
UNECE HLG	United Nations Economic Commission for Europe High-Level Group

A detailed glossary of acronyms is available on the Quality Information System at: https://qualityinformationsystem.cso.ie/

