The ASPIRE Review 2023

Susan Linacre, Stephen Penneck, Laurie Reedman,

and the Central Quality Team at Statistics Sweden

September 18, 2023

Contents

1 Introduction

In lieu of product evaluations during 2023, Statistics Sweden (SCB) has reviewed the ASPIRE approach together with the ASPIRE expert team – Stephen Penneck, Susan Linacre and Laurie Reedman.

The review was done between April and September 2023. The report was written by the ASPIRE expert team with support from the central quality team. It is primarily directed to SCB senior management but can be of interest for the Swedish government, users of statistics, other statistical offices and SCB employees.

The objective of the review was to further develop the ASPIRE approach at Statistics Sweden (SCB) by addressing identified weaknesses, while maintaining identified strengths of the previous approach. Other important objectives were to strengthen the connection between ASPIRE and the Evaluation of the Quality of Official Statistics (SOS-Q), applicable to all statistical agencies in Sweden, and to learn from the quality review approaches used by other statistical offices.

This paper reports briefly on the conclusions from the *Evidence Gathering Stage* and, based on these, presents the *Final Proposal* for a changed ASPIRE approach for SCB.

The *Evidence Gathering Stage* studied:

1. Information needs and requirements of the stakeholders for the outputs of SCB's quality evaluations
2. International practice of quality reviews used at other statistical offices
3. ASPIRE in the context of SCB's quality management

The *Final Proposal* for the changed ASPIRE approach is based on the conclusions from the *Evidence Gathering Stage* building on the strengths and addressing the weaknesses which were identified at the start of the ASPIRE review.

2 Evidence gathering

2.1 Information needs and requirements

Understanding the needs and requirements of different stakeholders regarding quality evaluations is key to the design of an evaluation approach that can provide relevant input for decisions, analysis or improvement work.

The needs and requirements of the following stakeholders have been investigated:

- Swedish government
- SCB senior management
- the product areas presently selected for ASPIRE

In addition, the review benefitted from the reflections of the expert ASPIRE team based on their recent experience with ASPIRE.

The following notes were made based upon these investigations:

- No stakeholders expressed the need for quantification or metrics regarding the results of ASPIRE, since they were first produced in 2011.
- Top management needs feedback from quality reviews to confirm that quality assurance practices are being applied, for example that quality guidelines set out in the statistical production support system (SPS) are being followed. They also need feedback on issues requiring more corporate attention and solutions.
- Top management would also like to consider what aspects of the SCB quality management approach, including ASPIRE and a risk-based approach, can be broadened to apply to the System of Official Statistics of Sweden, accepting that any such development would require senior management in other agencies to first be brought on board.
- The preparation work undertaken by areas being reviewed in ASPIRE is seen by participants as being useful to those areas, and they benefit from their discussions with the experts, as this yields useful insights regarding the quality of their product. However, the preparation of the detailed scoring is burdensome, and it is important to strike a balance in the cost/benefit trade-off to avoid unnecessary burden on staff and at the same time challenge them to reflect more deeply on quality issues.
- A clearer process is needed at SCB to address product level issues that are more of a cross-cutting nature and where a more corporate approach is needed. More generally the ASPIRE expert team would welcome feedback on progress being made with cross-cutting recommendations.
- Only a few products are selected for ASPIRE, and although these all have strategic significance for the SCB, some of them are quite limited in scope and it would be good to have a more systematic approach to product selection.
- An important objective of ASPIRE in the future is to use it to encourage the quality culture in the organization to a larger extent.
- Given the existence of the Evaluation of the Quality of Official Statistics (SOS-Q), which operates in parallel to ASPIRE, more integration is needed between the two, to avoid duplication.

2.2 International Practice

The information in this study was collected from the statistical offices of eight countries in addition to Sweden: UK, Australia, Canada, Costa Rica, Lithuania, Austria, Malta and Switzerland. The reason for choosing these countries is partly the experts' career background. The statistical offices of the other countries were chosen due to recent interactions between these statistical offices, for example, in the context of the ongoing peer review round in the European Statistical System. Thus, the choice of statistical offices was made on the basis that the sought information was readily available. Therefore, we do not view this group of countries as a representative sample of quality review approaches, nor do we view these as necessarily the most important ones.

The nine countries showed differences in approaches regarding the different features of quality reviews which are categorized in table 1. More detail per country is provided in Appendix 1.

In some instances where countries use more than one approach, double counting of a country has been allowed in the same row. Countries are indicated by the so-called Alpha-2 codes.

Table 1 Alternative approaches regarding quality review features in the nine countries studied

Quality review features	Alternative approaches			
What is reviewed	All products shallow study UK, SE	All products on rotation, deeper dive CR, LT, MT	Significant products UK, CA, SE	Products deemed at risk (recent major changes, errors) AU, AT, CH, UK
What aspects of quality are reviewed	All quality components and production processes UK, AU, LT	All quality components AU, SE, CR, AT	Selection of quality components CA, MT, SE	Production, management and support processes CH
Gradings	5 or more levels UK, SE	3 levels CR	2 levels (compliance or non-compliance, or strengths and weaknesses) LT, MT, CH, CA, AT	No gradings AU
Who does the review	Self-assessment CR, SE	Internal but independent experts CA, UK, AU, LT, CH	External, independent experts in a certain field, chosen specifically for the review AU	External expert panel with broad experience on more of a strategic level AT, MT, SE
Follow up	Internal follow up only, by product area SE, AU	Internal but independent follow up, e.g., thru methodology or Quality Review Team SE, UK, CR, AT, AU, CA	Management follow up AU, LT, CH	Management follow up outcomes published MT
Publication of review, recommendations and corporate response	No publication. Response remains internal to area involved.	Reported within organisation but not externally. UK, CA, CR, LT, CH	Publicly available. SE, MT	Degree of publication varies by review and degree of public interest, e.g., significant reviews and follow up published. AU

Other general aspects of quality reviews were noted and are summarized below.

Evaluation frameworks
Reviews tend to be against a framework such as the GSBPM or the ES CoP comprising the five principles of statistical output i.e., Relevance, Accuracy, Timeliness and punctuality, Accessibility and clarity, Comparability, and coherence (or the five main components of SCB's quality concept for Official Statistics). A combined approach looking at the quality components across a production cycle is not uncommon either.

Self-assessments and documentation
Generally, there is some type of self-assessment using a standard questionnaire which is often but not always facilitated by staff external to the product. The self-assessment can stand alone with improvement actions left up to the product itself or it can give input to a more formal discussion to evaluate the product or process. Some reviews are limited to existing documentation, others require further research.

Internal or external experts
It is common to employ experts who are at least external to the product. Experts may also be external to the statistical office. Reviews can be carried out nearly entirely by the business area, or can draw on methodological support, or outside expertise (including users), or be entirely external.

In many cases, countries involve methodologists in the review process, partly because of their relevant skills and partly to provide a somewhat independent challenge to the thinking of the product areas.

Resources
Reviews can require a lot of resources and the deepest reviews can lead to many recommendations which can be difficult to resource. They can point to a need for resourcing and funding for the area. A number of countries use a risk management approach, which can benefit the organization in considering the allocation of funding and implication on quality.

Sustainability
Quality review systems need to be sustainable into the future, and different countries have developed ways of following up recommendations, trying to build quality into the organization, and having an approach for follow up reviews.

2.3 ASPIRE in the context of SCB's quality management

An aim of the review team was to ensure that a revised version of ASPIRE would integrate well with the existing quality framework of SCB and build on its strengths. The components of the quality framework most relevant to ASPIRE were seen to be:

- The quality culture throughout the organization,
- The current review processes: SOS-Q and ASPIRE,
- The quality support tools: quality reports ("kvalitetsdeklarationer", in Swedish), the Statistics Production Support (SPS), the quality policy,
- The specialized organizational roles: the Quality Committee and the Quality Centre.

A strong quality culture requires everyone in the organisation to understand the SCB quality policy and framework, and their role in it. Specifically, it requires product areas to:

- Understand the quality of their product and how far it meets the needs of users
- Seek to continuously improve
- Evaluate what they do and report the outcomes

A review process can get product areas thinking in this sort of way, and the feedback from those involved in the reviews was that the ASPIRE approach had helped them to set aside time for this sort of thinking and these discussions. A question for the review of ASPIRE was how to broaden exposure to this thinking across more areas, with a strategy being to select more new products in ASPIRE and repeat review fewer products.

In terms of the two review processes SOS-Q and ASPIRE, SOS-Q covers all products with a light touch, covering all quality components, whereas ASPIRE is a deep dive, focusing on Accuracy. Table 2 below give a comparison of the two.

Table 2. Comparison between the processes for SOS-Q and ASPIRE

Evaluation phase:	SOS-Q	ASPIRE
Self-assessment	According to 2 forms with questions for statistical area and statistical product	According to the so-called checklists as preparation for the experts' assessment
Assessment with experts	--	Done by the external expert team with discussions based on self-assessments and other documentation
Purpose 1	Reflection on quality with SAMs (fitness for purpose perspective)	Identify improvement activities (Fitness for purpose perspective)
Purpose 2	Report to govt office on quality changes over time for Official Statistics	A reporting to govt office on quality changes over time for SCB
Coverage of statistical products and frequency	All statistical products which have made an official release of official statistics during the evaluation year	A selection of SCB's "important" products which are evaluated every other year (presently 4+4 products)
Coverage on quality components	All five main quality components	Mostly the main component, Accuracy
Report	SOS-report part 2, delivered annually to govt office by 31 March	An independent report from the experts approx. 3 weeks after the evaluation with product-specific and cross-cutting recommendations. Results reported in SCB's annual report.
Follow-up	--	Product recommendations followed up by the unit and section head's council and by the expert team

The quality report is important to users and a good quality report contributes significantly to the quality component, "Clarity". The quality report is also a starting point for ASPIRE reviewers in understanding the attributes of the product. Past experience with ASPIRE has shown that the quality report is often inadequate and

somewhat out of date. We suggest that in the revised approach to ASPIRE, the full updating of the quality reports by the product areas is encouraged as a first step in providing information to ASPIRE reviewers. This will not only reduce the need for further information requests, but also result in an immediate improvement in the quality of the product in terms of the Clarity component.

The SPS is a powerful repository of standards, best practice and tools for undertaking some forms of statistical processes at SCB, and unless exceptions have been approved, should be implemented across all product areas. For products selected for ASPIRE, a cross check on compliance with SPS would provide an insight into the effectiveness of SPS as well as an indication of potential cross cutting issues.

The Quality committee is a high-level committee newly established to oversee quality in SCB and has the potential to design and drive a very strategic approach to quality management in the organization. A revised ASPIRE should encourage and use this strategic capability.

A revised ASPIRE approach should conform to and support the SCB quality policy and regulations, and in particular the cyclical procedure seen in Figure 1.

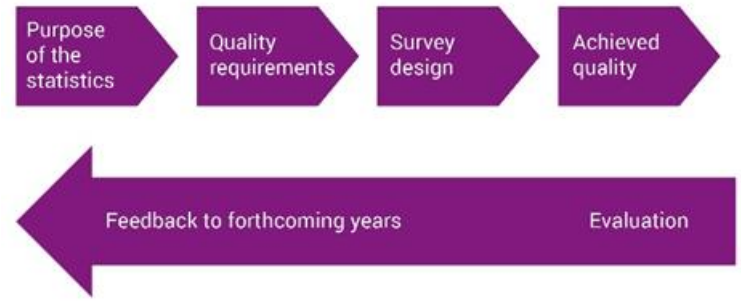

Figure 1. The cyclical procedure of the annual self-evaluation of the quality of official statistics.
Source: Evaluation of the Quality of Official Statistics – a handbook, version 3.3

2.4 Conclusions from the Evidence Gathering Stage

The information collected at the Evidence Gathering stage suggest the following conclusions:

1. Quantifiable evidence of quality levels is no longer needed
2. There is need to simplify the scoring
3. We should build further on SOS-Q and quality reports for a more integrated approach
4. We should consider a risk-based assessment in the selection of products
5. The Quality Committee should be asked to provide a more strategic approach to choosing reviews and progressing significant recommendations
6. There is scope for ASPIRE to contribute to building a stronger quality culture at SCB
7. The SPS exists as a framework for best practice and provides a potential basis for assessing process quality
8. We should consider whether to continue to focus on Accuracy, as other components are covered by good user consultations
9. Some clarification of roles is needed together with clearer ownership for the results and follow-up.
10. More involvement from methodologists would provide a basis for some internal challenge in the quality considerations being input to the reviews and in the responses to recommendations.

3 The Final Proposal

3.1 A Risk Approach to Quality Management

One of the increasingly common features of quality reviews in other countries is the use of a risk approach. This enables top management to focus on quality issues which are of major importance and for the office to determine which products should be subject to a full 'deep dive' review. Evaluation of risk is commonly used in statistical agencies as part of project management for statistical developments but is increasingly also being used as a way to manage ongoing statistical production.

The SCB risk framework

The SCB guidelines[1] on risk management require risk to be managed for operations at all levels, with responsibility for identifying, evaluating and managing significant risks delegated to heads of department or heads of unit. According to the guidance, a risk analysis is carried out to identify and evaluate risks in order to prevent them from occurring and disrupting operations or to limit damage if they occur, but also to provide a basis for planning, continuity plans and decisions. Control measures are put in place to reduce the likelihood of a risk occurring or to minimize the damage if it does occur.

Risks to statistical quality are recognized by SCB as an important component of operational risk and may also contribute to strategic risk. They are described as 'errors in data input, failures, errors in statistics, incorrect survey design, errors in methods, errors in administrative systems, errors in accounting, etc.'

Risks are generally identified at a risk workshop. This is a useful approach for assessing risks to statistical quality as it enables product area staff and methodologists to work together and agree the scores for each risk.

SCB uses a three-level scale for its risk assessment based on likelihood and consequence.

Likelihood

Likelihood is assessed as follows:

Likelihood score			
1	Low likelihood	<5%	Hardly arises
2	Medium likelihood	5-25%	May occur
3	High likelihood	>25%	Likely to occur

Consequence

Consequence is assessed as follows:

1 = negligible, mild

2 = noticeable

3 = serious

[1] This section reflects SCB guidelines 'Risk Management at SCB, 2022, and 'Routine Description of Risk Analysis, 2022'

The SCB note 'Routine description of risk analysis' has the following descriptions for Accuracy for the three levels of Consequence:

Accuracy

Consequence score	Level	Consequence
1.	Negligible, mild	Superficial flaws
2.	Noticeable	Errors with influence on several users, but without the negative effect of 3.
3.	Serious	Errors that have large negative consequences for financial markets and /or planning for society and the economy.

We noted that the above descriptions for the consequences of Accuracy failure in the Risk Guidelines focus solely on flaws/errors and do not consider the risk for excessive levels of uncertainty the latter of which is a central perspective in the descriptions of Accuracy in SCB's quality regulations. We suggest adding "uncertainty" to the risk descriptions for Accuracy in the following way for each of the three levels of Consequence:

Accuracy (amended with underlined text)

Consequence score	Level	Consequence
1.	Negligible, mild	Superficial flaws or minimal levels of uncertainty
2.	Noticeable	Errors, or levels of uncertainty, with influence on several users, but without the negative effect of 3.
3.	Serious	Errors, or levels of uncertainty, that have large negative consequences for financial markets and /or planning for society and the economy.

A risk value is equal to the likelihood value multiplied by the consequence value. Risk value is then assigned a colour linked to the different risk values. Actions or control measures need to be put in place to reduce Likelihood or Consequence, as follows:

	Risk value	Level	Actions
Green	1–2	Low	None, unless the actions are simple and do not involve any real costs.
Yellow	3–4	Medium	monitor and act in regular planning or decision-making process
Red	6–9	High	critical risk that must be addressed immediately

Usually, each risk would have a risk 'owner' with responsibility for ensuring the control measures are in place, and for regularly reviewing the risk score.

The risks can be graphically illustrated in a risk matrix as follows:

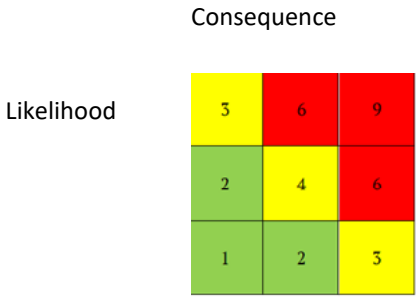

Figure 2. Risk matrix.
Source: Guidelines for Risk Management at SCB, 2022-08-22

Using SOS-Q as a tool to determine statistical risk

SOS-Q is the annual evaluation that SCB publishes on the quality of Swedish official statistics. It covers statistics from all government agencies, including SCB, and is based on a self-evaluation questionnaire.

In 2022, the questionnaire reported on 115 statistical areas and 363 statistical products. For statistical areas, the questionnaire asks about new external impacts, impact of response burden, shortfall of information needs, change in user needs, and whether quality studies had been carried out in the recent year. For statistical products the questions are more detailed and include quality requirements (whether they were established and met), whether achieved quality met the objective, which source of uncertainty had the greatest impact, and what had the greatest impact on quality.

The results of the questionnaire enable SCB to produce a report which summarises whether the quality of official statistics has improved since the previous year and in what ways.

It seems to the ASPIRE team that the SOS-Q approach, by providing an across-the-board assessment of quality, has great potential to assist ASPIRE (it is not currently used by the team), and in particular would help SCB choose candidate products for ASPIRE on the basis of statistical risk, if it were extended and modified. However, we understand that the questionnaire is largely seen by product areas as simply a reporting tool rather than a means of self-reflection and self-evaluation, and there is little challenge to the product level results.

We are proposing a development of SOS-Q that would:

- Include a brief self-evaluation of statistical risk
- Require it to be owned and signed off (with appropriate management challenge) at Section Head level, and by methodology

The detail of the self-evaluation of statistical risk is in the next section. Only three additional questions are proposed. It is important however that this evaluation be authoritative and consistent over products. SOS-Q has great potential as a tool to drive up quality and contribute to the developing quality culture at SCB.

SOS-Q is a self-assessment tool, and this is a very useful approach if the self-assessments are undertaken objectively. There are two complementary ways to ensure this. One is for Sections Heads to take responsibility for the SOS-Q reports in their Section, and to confirm they agree with assessments. The other is for methodologists to be involved in their production and challenge the product areas on the assessments and ensure consistency across the office. There could in effect be a dual 'sign off' between the product area and methodology of the SOS-Q. This is especially important if the SOS-Q is developed to include a risk analysis, as we suggest below.

3.2 Building on SOS-Q

The proposed approach is to use the existing SCB approach to risk – using likelihood and consequence in a matrix formulation – to measure the risk of quality failure. It is suggested that for ASPIRE the most important

factor affecting statistical quality is overall accuracy, taking into account the related sources of uncertainty: sampling, frame coverage, measurement, non-response, data processing and model assumptions.

Product areas would assess the likelihood of accuracy issues leading to a quality failure for the product, and also the consequence of such a failure for SCB. Each product area would be asked to answer the following questions:

1. Looking at overall accuracy, what is the likelihood that the current level of accuracy would lead to a quality failure for the product? (Low, Medium, High) – using the scales described earlier.
2. Where the score is M or H: which sources of uncertainty are mainly contributing to this: sampling, frame coverage, measurement, non-response, data processing and model assumptions?
3. Looking at overall accuracy, what would be the consequence of such a failure for SCB? (Low, Medium, High)

As noted above, these assessments would be jointly signed off by the product area and methodology. The Quality Centre would combine the likelihood and consequence scores to produce an analysis of the total risk for each product, together with information on the accuracy components most contributing to risk. This report would be considered by the Quality Committee which would determine the products that should be included in the coming ASPIRE round. The report would be made available to the ASPIRE expert team for those products included in the ASPIRE round.

3.3 Proposed Clarification and Strengthening of Roles in relation to ASPIRE

Experience with past cycles of ASPIRE has helped identify the need to clarify roles and expectations of different groups and levels within the organisation. This would support more effective reviews, provide a clearer pathway for addressing recommendations, and help develop action plans and implementation strategies.

The roles we are proposing are set out in Appendix 2, but in summary this involves the following proposed changes to roles and responsibilities for the revised ASPIRE.

The Quality Committee**,** as part of its responsibility for the overall quality framework and driving forward the quality culture, becomes responsible for selecting products for review, focusing reviews on areas of greatest perceived risk. The Committee subsequently receives and considers a report from Unit Heads identifying a summary of responses to recommendations, information on any significant quality issues identified, cross cutting issues arising, and approves action plans on cross-cutting and significant issues.

Methodology becomes a joint owner of each review and plays a challenge role, where applicable, to ensure quality issues are openly addressed. Methodology becomes accountable to the organisation for playing this challenge role if required.

Section heads of both the product areas and methodology areas become accountable for appropriate quality risks being raised, and for providing a response to the recommendations, with an appropriate action plan to their unit head.

The Product unit heads become responsible for reporting across their area, on significant issues arising from the reviews, and proposed action plans to address them.

The Methodology and Architecture methodologists take on a role in working with product area methodologists to develop a response and action plan for cross cutting recommendations, and escalating issues for corporate consideration through the head of Methodology and Architecture.

3.4 Quality components to be reviewed in ASPIRE

Previous ASPIRE rounds have focused on the quality component, Accuracy. While users can observe and evaluate other quality components for themselves, statistical expertise is needed to objectively assess the methods and activities used to assure accuracy. Hence, we concluded that the ASPIRE reviews should continue to focus on accuracy. Understanding user requirements is a necessary element of understanding whether accuracy

requirements are being met. Given this, some quality components such as timeliness and relevance are likely to continue to be covered in the reviews, even though they are focused on accuracy.

3.5 Changes to the ASPIRE review process

The ASPIRE review process was developed in 2011 in response to a request from the Swedish government for an annual report with metrics reflecting current quality in statistics and capturing changes in quality over time. The original ASPIRE design included checklists with six so-called evaluation criteria which contribute to quality assurance. For each criterion, a score was assigned reflecting the production team's level of thoroughness in addressing each of several different sources of uncertainty.

After more than a decade of annual ASPIRE reviews, it has been noted that the task of assigning the scores is overly burdensome compared to the value of having the scores, while the value of the discussions sparked by going through the checklists remains high. Currently, the Swedish government is no longer requesting annual metrics. Therefore, it seems logical that the ASIPRE redesign include a simplification of the scoring process.

The evaluation criteria have also been renamed/regrouped to better align similar activities. Also, we have added a frequency distribution chart to facilitate interpretation of comparisons across products or over time.

Scoring levels

The previous scoring had 10 levels which were incremental, so that an increase in quality initiatives would typically result in an increase in the score. Each of the 10 levels was described as a list of activities. Where it became burdensome was trying to match up activities undertaken by production teams with the activities listed in the checklist. If for example a production team had done all the activities at level 6 and a couple of activities at level 8, but they have not done two of the activities listed at level 4, what should the score be?

The new scoring has been simplified to just 4 levels, with level 3 being a fully acceptable level of thoroughness. Level 3 is the target that production teams should aim for. Production teams scoring a level 3 are asked to summarize the key elements they have in place to demonstrate that their knowledge and activities in this area are fully acceptable, or in other words, fit for purpose. A level 2 is used where some good effort is recognized but there remains a gap. Production teams scoring a level 2 are asked to describe what is lacking and what plans are in place to bridge the gap. Level 1 is used to indicate an unacceptable absence of knowledge or quality assurance activities. Production teams scoring a level 1 are asked to describe the obstacles they are facing in this area. Conversely, level 4 indicates that the target level is fully reached and exceeded, for example the production team is setting best practice. Production teams scoring at a level 4 are asked to describe in what way they are exceeding the target.

The new checklists are further simplified by formulating the fully acceptable description of knowledge and activities for each quality assurance factor into one statement and asking production teams to indicate to what extent the statement holds true for their product. An example is shown below of the quality assurance factor, **F2 Compliance with standards and best practices**.

> To what extent is the following statement true?
>
> The production team has good knowledge of standards and best practices (including international best practices) related to this source of uncertainty and the product. There is compliance with all significant standards and best practices.
>
> 1. Not at all
> 2. Partially true
> 3. Fully true
> 4. Fully true and we do more
>
> Description: …….

See Appendix 3 for the statements that apply to the other quality assurance factors, F1-F6.

The quality assurance factors (previously evaluation criteria)

The table below shows the renaming and regrouping of evaluation criteria to better align similar activities.

Table 3. Comparison of the new set of Quality Assurance factors and the previous set of the so-called Evaluation criteria

New set of Quality Assurance factors	Previous set of Evaluation criteria
F1 Available expertise	C1 Available expertise
F2 Compliance with standards and best practices	C2 Compliance with standards and best practices
F3 Communication with users and knowledge of requirements	C3 Knowledge of requirements, achievements and improvement needs
F4 Knowledge of achieved quality and results of improvement activities	C5 Results of improvement activities and findings from other evaluations
F5 Knowledge of improvement needs and plans for improvement activities	C4 Plans for improvement activities
F6 Communication with data suppliers	C6 Communication with users and data suppliers

Graphical representation of results

Harvey Ball charts are graphical representations used for visual communication of qualitative information. Harvey Ball charts are commonly used in comparison tables to indicate the degree to which a particular item meets a particular criterion. In this respect they are well suited to the purpose of demonstrating the results of an ASPIRE review. An example from 2021 is shown below in Figure 3.

Sub and sub-subcomponents of Accuracy	Average Score Previous Round	Average Score Current Round	C1. Available Expertise	C2. Compliance with standards & best practices	C3. Knowledge of requirements, achievements, and improvement needs	C4. Plans for improvement activities	C5. Results of improvement activities and findings from other evaluations	C6. Communication with users and data suppliers	Importance to Overall accuracy (single sources of uncertainty)
Overall accuracy	53	55	○	◒	○	◒	◓	◓	
Sources of uncertainty:	55	56							
-Sampling	55	55	◒	◒	◒	◓	◓	◓	H
-Frame coverage	58	60	◒	◒	○	○	○	○	M
-Measurement	53	58	◒	◒	○	○	◓	◓	H
-Non-response	57	52	◒	○	○	○	◓	◓	L
-Data processing	58	55	○	◒	◒	○	◓	◓	M
-Model assumptions	52	53	◒	◒	○	○	◓	◓	H
Preliminary statistics compared with final statistics	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	

Ratings					Importance to Overall accuracy			
●	◓	○	◒	◎	Not applicable (N/A)	Low (L)	Medium (M)	High (H)
Weak	Fair	Good	Very good	Excellent	Weights			
1, 2	3, 4	5, 6	7, 8	9, 10	0	1	2	3

Figure 3. A Harvey Ball chart for the results of an ASPIRE review of one product.

The table shows the evaluation criteria in columns and the sources of uncertainty in rows, with the average scores (previous and current) in the first 2 columns and Harvey Ball symbols elsewhere. Below the chart is the legend explaining how to interpret the symbols. The previous ASPIRE scoring had 10 levels, which were aggregated to 5 levels in the Harvey Ball chart. In the new proposed scoring, Harvey Ball charts can still be used to show product level results, with the 4 levels being represented by partially or fully filled circles, and a N/A response being represented with a fully open circle.

A summary of scores across all products reviewed has previously been shown as a table of average scores by sources of uncertainty. An example from 2022 is shown below in Figure 4.

Sub and sub-subcomponents of Accuracy	PVI	QEA	Building	Population	Average score	Weighted average score
Overall Accuracy	40	47	35	43	55	N/A
Sources of uncertainty:	40	47	37	43	42	N/A
-Sampling	32	40	N/A	N/A	36	35
-Frame coverage	25	53	35	42	39	26
-Measurement	52	42	38	40	43	33
-Non-response	42	42	35	47	42	30
-Data processing	45	50	40	48	46	27
-Model assumptions	47	48	38	43	44	40
Preliminary statistics compared to final statistics	40	33	35	40	37	N/A

Figure 4. Example of a table of average scores summarizing the results of the 2022 ASPIRE reviews.

An alternative graphical representation is a frequency distribution. This can be used to show the relative frequency of levels for a single product, as shown below in figure 5, or to compare products, as shown in figure 6.

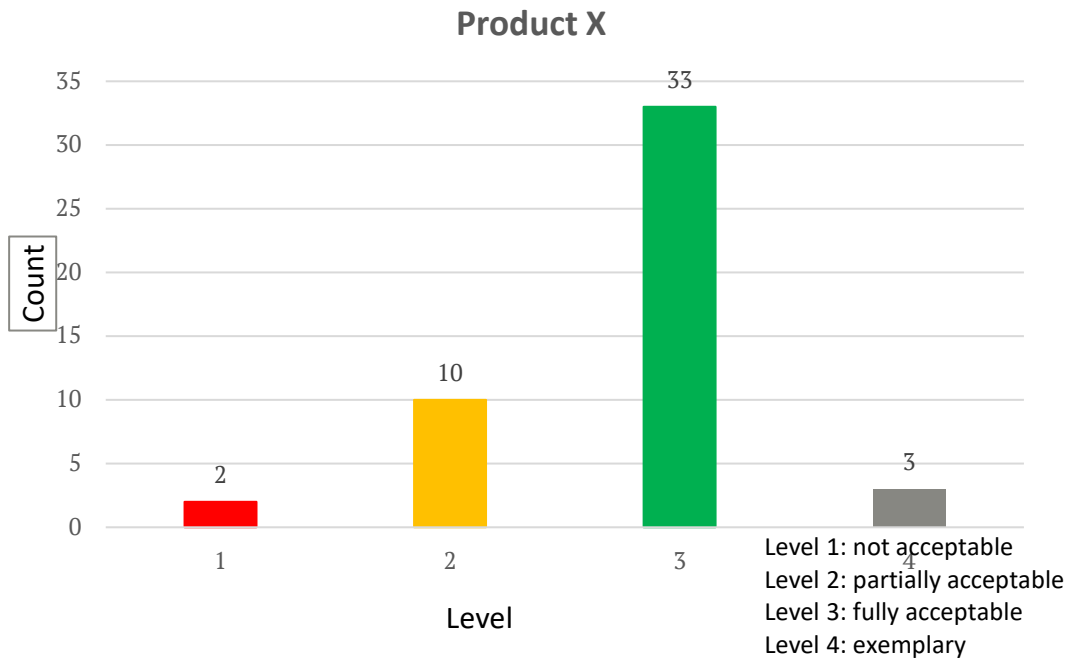

Figure 5. Example of a frequency distribution chart summarizing all the levels counted across all quality assurance factors and all sources of uncertainty for a particular product.

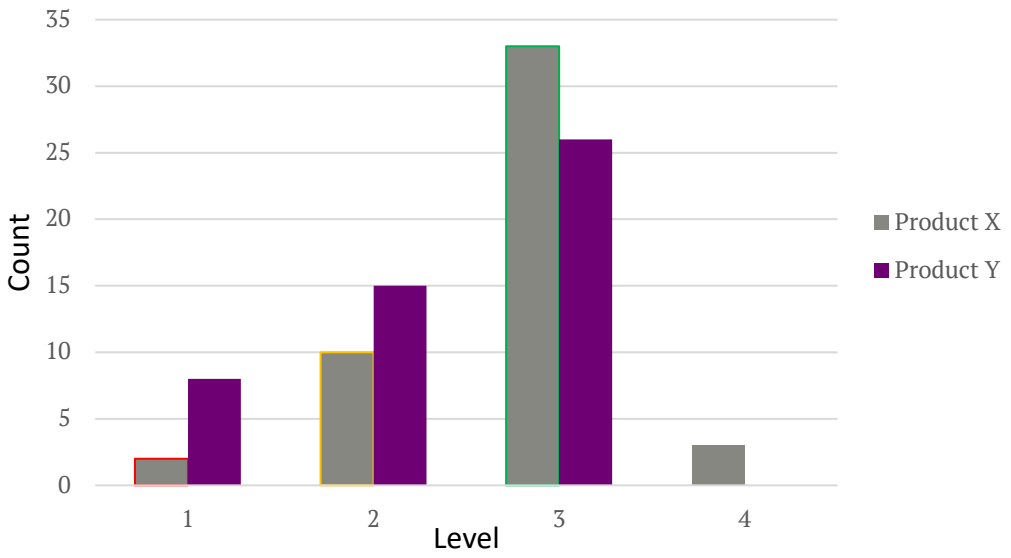

Figure 6. Example of a frequency distribution chart summarizing all the levels counted across all quality assurance factors and all sources of uncertainty, comparing frequency of levels for 2 different products.

Other variations can of course be used, for example comparing the levels across the different sources of uncertainty or across the different quality assurance factors.

3.6 What is not covered by the proposed revision to ASPIRE

The above proposal has been developed to meet as many of the needs and conclusions identified in the evidence gathering phase as possible, while recognising the limitations of any one review team working in a limited time frame to undertake reviews.

The proposal seeks to direct the reviews to areas of greatest risk in the organisation. It integrates with existing components of the quality environment, building on SOS-Q for risk identification. It simplifies the checklists and scoring processes used in self-assessment. Offsetting this it will make greater use of documentation from the quality reports, encouraging the update of these prior to review, and it will build in a cross check on compliance within the product area with the standards and tools provided in SPS. It allows for an external challenge to the product area in undertaking the self-assessment through the role of methodology. It provides a clear accountability for following up recommendations and a pathway for raising issues of significant concern, or of a cross cutting nature, for consideration at a corporate level.

The proposal does not move from focusing on evaluating the accuracy component of quality (while picking up related issues that arise in the course of a review). Quality components such as Relevance and Timeliness are considered very important, but better assessed through involvement of key users than with the ASPIRE model. Nor does it move from focusing on evaluation of products rather than processes. This means, for example, it would not review ‘data collection’ or ‘editing’. Reviews of process are still warranted and should be considered by the Quality Committee as a component of the overall review strategy. One possibility may be to initiate and run these process reviews through the Methodology and Architecture unit.

The proposal does not cover Official Statistics more broadly than SCB. A broadening of aspects of the proposal may be appropriate over time. In particular the use of SOS-Q as a tool for identifying statistical risk may be attractive to senior managers of statistical products outside of SCB.

4 Recommendations

The following are the recommendations we are making in this review, based on the evidence we have gathered and the assessments we have made:

1. Product areas should assess the risk of quality issues for their product arising and leading to a failure. This would be done by adding a small number of questions to SOS-Q

2. The descriptions for the consequences of Accuracy failure in the *SCB Routines for risk analysis* should be updated to align with the descriptions regarding Accuracy in SCB's quality regulation regarding the risk for excessive levels of uncertainty

3. There should be a dual sign off of the SOS-Q, including the assessment of risk, by the product area and methodology to improve its value and objectivity

4. This would be supported by the Quality Centre who would calculate risk scores for each product and report summarised findings to the Quality Committee

5. SOS-Q risk assessments should be made available to the ASPIRE expert team for those products included in ASPIRE

6. Quality reports should be kept up to date and complete by the product areas as a first step in providing information to ASPIRE reviewers. This will reduce the need for further information requests and provide an immediate upgrade in the quality of the product for users, in terms of the clarity component

7. Review of processes should be an important part of quality assurance but should not be included in ASPIRE

8. For products selected for ASPIRE, a cross check on compliance with SPS would provide an insight into the effectiveness of SPS as well as an indication of cross cutting issues

9. Roles and responsibilities of management levels, methodology and the Quality Committee should be clarified and strengthened as set out in section 3.3

10. ASPIRE should continue to focus on Accuracy, addressing other components of quality as required

11. The scoring process should be simplified to reduce the number of levels from 10 to 4 (section 3.5.1)

12. The evaluation criteria should be renamed and regrouped so they are better aligned and easier to use

13. The graphical representation of the results of ASPIRE should include frequency charts

5 Implementation Strategy

We have given some thought to what our proposals might mean for ASPIRE next year. The recommendations on introducing risk assessment into SOS-Q, the clarification of roles and responsibilities and the simpler scoring system will need further work to operationalise into SCB. This work is best undertaken by those who understand the constraints and opportunities that there might be, not by the ASPIRE team.

We can see that it might be possible to run the new version of ASPIRE on a partial basis, not including all of these new elements. That would be for the Quality Committee to decide. Whatever is decided it will be important that an implementation plan is worked through quickly so that all players are aware of what is required.

Appendix 1. International comparisons of quality reviews

Name and scope of the evaluations /approach

Country	Name of the evaluations/or approach	Scope
UK	1. ONS Statistical Quality Maturity Model (SQMM) - a self-assessment by the product team. The first line of defence! 2. Deep Dive Quality Reviews (targeted) 3. Error Review (a shorter deep dive)	Product and divisional level
Australia	ABS Risk Control Framework with three lines of defence: product area quality assurance, independent oversight roles within the ABS (Chief Methodologist, Chief Data Officer and Chief Information Officer), external oversight (Audit and Risk Committee) and ad hoc reviews	All product areas
Canada	Quality reviews 2023	CPI, NA products
Costa Rica	Quality reviews	All quality components of all entities/programs in the NSS are evaluated every 4 years for compliance with the Code of Good Statistical Practices of Costa Rica
Lithuania	1. Quality Management Audits 2. Annual plan to monitor QPIs 3. Self-assessments, based on the simplified DESAP checklist	3-year audit plan. Every year about 25 surveys are assessed with the commitment to perform the exercise once per 5 years for the same survey.
Austria	1. Feedback Talks by Statistical Council, 2. Internal quality and risk management audits	1. All products are to be covered over time according to a schedule. 2. Key products; audit plans taking into account of resource situation
Malta	Reviews of statistical output.	Statistical outputs
Switzerland	Quality reviews	Mandatory quality reviews for each department considering all activities of a statistical organisation, according to the UN GAMSO model.
Sweden	Evaluation of the Quality of Official Statistics (SOS-Q) A System for Product Improvement, Review and Evaluation (ASPIRE)	SOS-Q: All official statistics products which have published during the reference year ASPIRE: selected important products

Objectives and format of the quality reviews/evaluations

Country	Objectives	Format/method
UK	Identify quality improvements needed re user needs and quality assurance, identify topics for Deep Dive Quality Reviews, support learning from errors and avoid recurrence, and provide overall view of quality across ONS	Quality champions in each division take the lead. There are peer challenge sessions, and moderation by the ONS Methodology and Quality Directorate.
Australia	The Risk Control Framework is an ongoing strategy of oversight and support aimed at building quality into products, and identifying and addressing areas of concern	Managers (risk owners) are required to implement and maintain operational controls and demonstrate the controls are effective aided by methodologists. Along with user consultation and feedback, quality gates and measures can indicate quality concerns and trigger a response. A response might include a review.
Canada	A risk-based methodology assessment, all methods, not just statistical methods, implicitly evaluating compliance with quality guidelines and international guidelines where they exist.	Reviews are performed by the Quality unit staff who are internal to StatCan but independent from the programs. They study existing documentation and meet with program staff a couple of times. This is a new process and as of August 2023, the pilot (CPI) is underway but not yet completed.
Costa Rica	Compliance to Code	Self-assessment questionnaire with approximately 200 answers to grade: Full compliance: 2 points; Partial compliance: 1 point; Non-compliance: 0 points
Lithuania	Check compliance to selected requirements based on legal acts, standards, and QMS documents.	QM audits are led by internal working group on Quality management audits, according to the approved 3-year audit plan, following the Description of procedures for Quality management audits, approved by the order of DG. These audits inspect all phases of the statistical production process and other processes in accordance with approved procedures and defined quality requirements. Internal experts collect, generalize and present results, together with the recommendations for improvements.
Austria	1. Feedback talks evaluate implementation of Statistics Austria's standard documentation of quality for statistical outputs given user needs. 2. Quality audits measure compliance with the NSI's quality guidelines	Feedback talks are conducted by the Statistics Council by way of the technical quality committee (comprising key users). Three products are selected every 6 months a review. These are selected based on the outcome of the internal quality audits. Another criterion for selection is if a product is to undergo re-engineering. The selection for quality audits is done in close cooperation with the internal audit unit.
Malta	Assess quality impacts, efficiency, accuracy and reliability of statistics, if methodologies are internationally comparable, dissemination re. user needs, metadata documentation, methodological notes, strengths and weaknesses.	Quality reviews study suitable methodologies to ensure compliance with manuals/regulations. The centralised Methodology and Quality unit checks issues of non-compliance/areas for improvement consulting with the relevant domain units. Improvement action plans are to be followed by experts in the respective domain fields. Presently, this is only done for areas for domain units expressing concern about a process or ask for the assistance from the Methodology and Quality Unit.
Switzerland	- identify best practices established in relation to an activity that could be used / adopted by other organisational units; - identify strengths and opportunities for improvement of an activity; - Anticipate future requirements impacting the activity.	The Quality and Process Unit plans and coordinates the reviews via the quality review manager. To better understand the aim/purpose of quality reviews and to define the scope of them a preparation meeting is held with the participants. The product area does a self-assessment in preparation. During the review, a list with recommendations/improvement actions is made to be approved by the process owner whose superior is to be informed after the review.
Sweden	SOS-Q: provide assurance to govt of the quality of official statistics in support of its Commitment on Confidence, and to stimulate quality improvements ASPIRE: to stimulate Accuracy improvements and report yearly to the govt.	SOS-Q: The central quality team coordinates the annual self-assessment questionnaire for all statistical agencies. ASPIRE: self-assessments are done by the products prior to the external review. These are facilitated by a central quality coordinator.

Focused quality components/aspects and results

Country	Focused quality components/aspects	Results/output
UK	**Product level**: All five quality components as well as the following themes: sources; methods and systems, processes, users, reputation and people. The themes are graded. **Divisional level**: Quality Improvement Plans; Quality Assurance; Quality Champions Network; Errors and Near Misses; Training and guidance; and Relationships with data suppliers. The themes are graded.	Product level: Index of statistical quality based on gradings: 1. unacceptable, 2. improvement required, 3. acceptable, 4. good, 5. exemplar Division level: Index for quality culture based on gradings In general: identified issues feed into Quality Improvement Plans. Results reported to senior mgmt., and annual report on cross-cutting issues is reported to the Quality Committee chaired by DG.
Australia	Holistic and risk approach based on the traditional quality dimensions, aided by user consultation and feedback as well as quality measures/ gates to help identify areas of concern	If quality issues arise, the response can be a Quality Incident Response Plan (QIRP) or a local Quality Investigation. A QIRP will ensure relevant expertise engages to fix immediate issues and to review the nature of and cause for the concern to prevent future repetition. It is formally the subject matter unit who decide if a QIRP is required as an incident response. Where significant issues arise, either from user feedback for a product, or from quality measures, an internal or external review may be undertaken.
Canada	Accuracy, Clarity (interpretability), Timeliness, Accessibility of inputs	A document with recommendations, strengths and weaknesses
Costa Rica	19 principles of the Code including the familiar quality components	Outputs are 2 reports: 1. The individual report: it has the scores of each entity. The score is the sum of the points divided by the total possible, expressed as a percentage. Totals on 3 levels: <40%; 40-70%, >70% 2. The general report: it has the scores of the NSS, some conclusions and recommendations on what to improve and how.
Lithuania	All quality components according to the DESAP questionnaire (Estat), statistical production process in line with procedures and defined quality requirements.	A report is compiled with recommendations which are deployed into action plans, approved by the DG.
Austria	Relevance, Timeliness and punctuality, Comparability and coherence, Accessibility and Clarity. Quality and risk audits cover Accuracy including error sources.	Results of quality audits and feedback talks are available to senior management the latter of which are reported to the Statistical Council.
Malta	Coherence, clarity, quality, consistency	Published reports on breaches of best practice and errors, with advice on how to increase confidence in official statistics.
Switzerland	- Good practices - Strengths and weaknesses for each GSBPM process or for management and support processes	Action plans. Quality reviews are an effective way to get management attention and an overall assessment of how well statistics and management, or support processes perform without a heavy administrative burden for the staff.
Sweden	SOS-Q: all quality components, ASPIRE: Accuracy	SOS-Q: Aggregated results are compiled by the central quality team who formulates improvement actions. Report published and sent to the govt. ASPIRE: independent report from experts published with product-specific and cross-cutting recommendations.

Follow-up routines, possible weaknesses and lessoned learnt from experience

Country	Follow-up	Possible weakness or lessons learnt from experience
UK	Quality Improvement Plans are followed up by senior mgmt. An annual report is compiled on cross-cutting issues by the Methodology and Quality Directorate.	Lessons learnt from the past: National Statistics Quality Reviews (NSQRs) were performed by the UK Statistical Authority 2000-2007 addressing the questions: What are users' needs of the output? Do the outputs meet users' needs and are they of adequate quality? The reviews were quite effective in many cases, but resource intensive and apparently the follow up of recommendations became an issue. The Office for Statistics Regulation assessments (OSRs) which were external started in 2008. Assessments are quite long, include evidence from users, and reviews of evidence with the producer. The final report is submitted to the UKSA Regulation Committee, which decides whether the statistical output should remain as National Statistics. Reports usually include recommendations for improvement and are published. The NSQRs ran again in parallel 2012-2016, only with the ONS, to be carried out every 3-5 years, according to a risk based planned timetable. Methodologists were involved. Quality, Methods and Harmonization Tool (QMHT) – a self-assessment questionnaire, started in 2013 involving annual reviews of all products against the UK CoP and ES CoP (also according to the GSBPM). This approach lacked independent methodological input and led to many recommendations that couldn't always be implemented. Seems that the assessments ceased after a short time. Regular Quality Reviews (RQRs) started in 2014 and expanded to include a 'Divisional Director walkthrough', of quality assurance procedures in place for the outputs. The main challenges were scheduling the reviews; getting buy-in and internal communications; and addressing the recommendations – issues in any review system. On the last one it was noted that there needed to be more time between RQRs for the recommendations to be addressed. The Quality Centre advised on prioritization and picked up cross cutting recommendations.
Australia	Follow up of recommendations for external reviews are done through line management to the Executive Board and a quarterly report to the Audit and Risk Committee.	Follow-up is adapted to the purpose of the review, with tighter and more independent follow-up for more significant reviews.
Canada	The quality unit follows up recommendations after 6-12 months.	The Quality Unit is worried that if recommendations are not considered in resource allocation, then they won't get done. For previous quality reviews (2007-2016) there was no follow-up/accountability for recommendations, no measurable impact for the resources spent doing reviews, some overlap with other audits, similar weaknesses/risks were coming repeatedly requiring corporate level solutions, not on the program level.
Costa Rica	The quality unit follows-up with reviewed programs over the next 5 years to ensure that the improvements are made.	The quality unit at INEC developed, executes and manages the quality reviews, and present the reports to senior management. Given the weaknesses that have been discovered, the quality unit takes the initiative to write guidelines to help the NSS entities to become more compliant with the best quality practices in their Code.

		• Guide for the elaboration of metadata of statistical operations • Guide for the elaboration of calendars for the dissemination of official statistics • Guidelines for the dissemination of statistics and access to microdata of the NSS • Guidelines for sharing confidential data • Protocol to preserve the principle of data confidentiality • Guidelines to incorporate the gender perspective in the production and dissemination of NSS statistics • Technical guidelines to incorporate the ethnic-racial approach in the production of official statistics
Lithuania	The implementation of recommendations is continuously monitored and reported to the Advisory Commission of the DG of Statistics Lithuania	The activity report (the implementation of recommendations) of the internal audit working group is presented to the Advisory Commission of the DG of Statistics Lithuania every February. The implementation of the measure is monitored in the electronic document management system.
Austria	There is regular follow-up of improvement actions	There are sometimes problems to attract and motivate users participate in the feedback talks. The findings in quality audits show that there is a need for further modernization which often leads to problems since additional know-how is needed.
Malta	Improvement action plans are designed by the Methodology and Quality Unit to be followed by the experts in the respective domain fields.	
Switzerland	The Quality and Process Unit monitors to ensure follow-up of the action plans. Once a year, the state of current improvement actions is presented to NSI's top management	Earlier the quality reviews have been voluntary which has resulted in too few being conducted. 2023 it became mandatory for each division (9 in total) to suggest an area for a quality review, annually. Management and support processes included.
Sweden	SOS-Q: No systematic follow up per product ASPIRE: follow-up is made of product-specific recommendations but not cross-cutting recommendations	Between 2004-2006 a type of quality reviews was done by teams who were external to the product but internal to SCB, according to a plan to cover all products over the course of 5 years. The teams who had expertise from subject-matter, methodology and IT, studied the available documentation, discussed with the products, and made recommendations – during one full week. Annual follow-up on improvement actions was done. The program was costly and the expertise in the different teams were unevenly distributed even though it was a valuable experience for staff to participate in a team. The program was discontinued with a re-organisation of the office in 2006 towards more process-orientation and standardisation of methods and tools. Another approach with quality audits started in 2011 as a requirement for the ISO-certification (ISO 20252). Different themes, usually sub-processes of the GSBPM, in combination with products were reviewed against the SCB Quality Assurance Framework, the SCB Process Support System. It was hard to see the impact on quality of the audit results relative to the cost of the programme. Many of the issues and recommendations came up repetitively. The programme was discontinued when the ISO-certification was let go in 2020.

Appendix 2. A more detailed Proposal for Roles and Responsibilities with ASPIRE

The following descriptions set out a suggested strategy and set of roles for undertaking ASPIRE reviews, with the aim of maximizing the value of the reviews to the organisation and integrating the review process with quality management in SCB more generally.

Product Area Section Heads are responsible for achieving 'fit for purpose' quality for their products. Under the revised ASPIRE, Section Heads would be responsible for completing SOS-Q both as a self-assessment of quality, and a light touch risk assessment. In terms of ASPIRE if selected, Section Heads would be responsible for ensuring any risks to product quality are identified, and also responsible for providing a response to recommendations made, and developing associated action plans, and raising particular concerns with the unit head.

The role of **Product Area Unit Heads** in terms of the revised ASPIRE, would be to work with section heads to review the recommendations and develop and implement action plans to address quality concerns. They would use input from Section Heads to identify and report on issues generating significant quality risk to the Quality Committee, together with a proposed action plan to address the risk. As relevant, they would liaise with the Unit Head Methodology and Architecture on cross cutting issues, to bring strategies to the Quality Committee.

The **Methodology Section Head** working with the product area, is responsible for achieving quality fit for purpose and cost efficiency in the methods embedded in products. In terms of the revised version of SOS-Q and ASPIRE, the Methodology Section Head would be responsible for ensuring all significant issues generating quality risk within products are identified, challenging product sections heads if appropriate, and if necessary, raising issues of concern with the product unit head. They would also review recommendations from ASPIRE and work with Section Heads of the product areas to prepare a response and an action plan. For issues remaining a concern, they would be responsible for raising these with the Head of Methodology and Architecture.

The **Section Head of Methodology and Architecture** supports the development and maintenance of fit for purpose common tools and methods solutions used across products, and review of use of these tools across areas, as well as review of effectiveness of these tools. In terms of the revised approach to ASPIRE, they would be responsible for reviewing recommendations with a cross cutting component, and identifying risks and opportunities related to common methods and tools. They would also raise issues of concern or opportunity, to Unit Head Methodology and Architecture, working to develop action plan to address them where appropriate.

The **Head of Methodology and Architecture**, under the revised approach to ASPIRE would work with their section head to develop and implement action plans to address cross cutting quality concerns, liaising with product Unit Heads. They would identify significant issues generating quality risk, or opportunities, to the Quality Committee together with proposed strategy/ action plans to address.

The **Quality Committee** is responsible for the design of the overall quality framework within the organisation (including SOS-Q, ASPIRE, SPS related reviews of cross cutting methods, policy, guidelines, tools and supporting the overall quality culture). In terms of the proposed revised approach to ASPIRE, the Committee would consider reports generated by Unit Heads (eg. annual) of significant risk to quality across products. Risk analysis from SOS-Q would provide one input to such a report. The Quality Committee would select products for review in the next ASPIRE round and would receive and consider a report from Unit Heads, summarizing the response to recommendations, significant issues arising, and proposed actions relating to significant quality concerns and cross cutting issues. The Quality Committee would report to the Leadership Group and DG.

The **Quality Centre** supports the Quality Committee, as well as supporting cross cutting elements of the quality framework. In terms of the revised approach to ASPIRE, the Quality Centre would support the 'light touch' approach to risk assessment across SCB products through SOS-Q and the implementation of the revised ASPIRE approach.

Appendix 3 Statements, regarding level 3, for the Quality Assurance factors, F1-F6.

Qualifying statements in order to score at level 3, fit for purpose. See more details in section 3.5.

Note that the following statements are in draft form and should be subject to further consultation and approval at SCB prior to implementation.

F1 Available expertise

There is a good level of expertise available to the production team (including expertise in methodology and IT) in terms of skills and knowledge to study this source of uncertainty, as well as sufficient resources for interaction with internal and external colleagues and to do the work required. There is also some capacity to contribute improvement ideas.

F2 Compliance with standards and best practices

The production team has good knowledge of standards and best practices (including international best practices) related to this source of uncertainty and the product. There is compliance with all significant standards and best practices.

F3 Communication with users and knowledge of requirements

Communication with key users includes the prioritized needs and uses of the statistics, business priorities, quality requirements that follow from the purpose of the statistics, and any gap between requirements and achieved accuracy.

F4 Knowledge of achieved quality and the results of improvement activities

Key components of the achieved quality are recognised and routinely measured and documented. Results of improvement activities are analyzed and documented.

F5 Knowledge of quality improvement needs, and plans for improvement activities

Potential quality improvement activities are on a list that is prioritised and regularly updated. Resources have been allocated for at least one prioritised activity, and an overall plan exists for the other key activities.

F6 Communication with data suppliers

There is ongoing communication with data suppliers which covers factors of importance in the supply stage of the data for the accuracy regarding the intended uses of the statistics. The results lead to agreements on priorities for the work with data suppliers.