Managing Quality in a Statistical Agency –A Rocky Road

Lilli Japec[1]

Abstract

Statistics Sweden has, like many other National Statistical Institutes (NSIs), a long history of working with quality. More recently, the agency decided to start using a number of frameworks to address organizational, process and product quality. It is important to consider all three levels, since we know that the way we do things, e.g., when asking questions, affects product quality and therefore process quality is an important part of the quality concept. Further, organizational quality, i.e., systematically managing aspects such as training of staff and leadership, is fundamental for achieving process quality.

Statistics Sweden uses EFQM (European Foundation for Quality Management) as a framework for organizational quality and ISO 20252 for market, opinion and social research as a standard for process quality. In April 2014, as the first National Statistical Institute, Statistics Sweden was certified according to the ISO 20252.

One challenge that Statistics Sweden faced in 2011 was to systematically measure and monitor changes in product quality and to clearly present them to stakeholders. Together with external consultants, Paul Biemer and Dennis Trewin, Statistics Sweden developed a tool for this called ASPIRE (A System for Product Improvement, Review and Evaluation).

To assure that quality is maintained and improved, Statistics Sweden has also built an organization for quality comprising a quality manager, quality coaches, and internal and external quality auditors.

In this paper I will present the components of Statistics Sweden's quality management system and some of the challenges we have faced.

Key Words: ISO 20252, ASPIRE, Quality framework, EFQM.

1. Introduction

1.1 Background

Statistics Sweden has a long tradition working on quality. The intensity has however varied over the years. Our first attempt started in 1993 with Total Quality Management. At that time a large number of improvement projects were carried out. The work and training was aided by Westat. Our current journey started in 2008. The Director General at that time made a number of decisions that laid the foundation of today's quality management system. The present Director General has implemented and encouraged further development of the system.

There were a number of reasons why Statistics Sweden chose the road we did. (i) One reason was competition. Since 1996 we have a decentralized system for official statistics in Sweden. There are currently 28 different government agencies responsible for official statistics. The agencies can choose to produce the statistics themselves or have a private firm or Statistics Sweden to produce the statistics. This means that Statistics Sweden has to compete with private firms. Approximately half of our budget comes from work on commission. (ii) Another reason was communication of quality information to customers. Statistics Sweden is perceived to be expensive and back in 2008 we had problems communicating quality to our customers, the quality that they were paying for. (iii) A third reason was to reduce cost and variation. Before 2008 we had a decentralized organization at Statistics Sweden, where each product had its own methodologist and IT-person. It was up to each product manager to decide what standards to use. We had no minimum standard that all products should follow. Of course this made it very difficult

---

[1]Lilli Japec, Statistics Sweden, Box 24300, Sweden, 104 51 Stockholm, (lilli.japec@scb.se)

for Statistics Sweden to communicate how we work on quality on a general level since it varied depending on the methodologist and product manager involved. The cost for maintaining and developing product specific IT-systems was also very high.

In 2008 and 2010 Statistics Sweden experienced some major errors in the Consumer Price Index (CPI) and the National Accounts (NA). Statistics Sweden received extra funding from the government to work on quality improvements and to report these improvements to the Ministry of Finance on a regular basis.

1.2 The Quality Management System at Statistics Sweden

In 2008 the Director General launched an internal investigation with the aim to choose a quality management system for Statistics Sweden. The team that worked on this investigation benchmarked ten different organizations and looked at eight different quality management frameworks. The decision made was to use the European Foundation for Quality Management Model (EFQM) for developing the organizational quality (see Section 2). This is the framework recommended by Eurostat, the European Statistical Agency, to European National Statistical Institutes (NSI). The framework is similar to the Malcolm Baldrige Award Criteria used in the U.S. The second decision was that Statistics Sweden should use the ISO 20252 standard for market, opinion and social research as a minimum standard for process quality (see Section 3). The third decision was that Statistics Sweden would use parts of Six Sigma for improvement work (see Section 5). There are a number of tools that can be used in improvement work. Six Sigma is a toolbox and also a standardized process for improvement work that focuses on results.

Later on due to problems with CPI and NA, the Ministry of Finance wanted to see improvements in our products. We were faced with the challenges of how to measure improvement in product quality and how to present the results of such a measure for someone not familiar with the details of statistics production. We made a first in-house attempt at Statistics Sweden to meet these challenges but soon we realized that this was too big of a challenge to handle on our own. We turned to Paul Biemer, Distinguished Fellow at RTI International and Dennis Trewin, Former Australian Statistician, to help us. ASPIRE (A System for Product Improvement, Review and Evaluation) is a result of this work (see Section 4).

The quality improvement work conducted since 2008 at Statistics Sweden has mainly focused on implementing the ISO standard which is the minimum standard that all our surveys should adhere to. For our ten most important products the quality requirements are higher. The focus since 2012 for those products has also been to make key improvements to product quality according to ASPIRE recommendations.

2. The European Foundation for Quality Management Model —EFQM

In business excellence frameworks, such as EFQM, the focus is on *organizational quality* i.e., components like leadership, strategy and competence. In traditional quality frameworks adopted by statistical organizations such as the European Statistics Code of Practice (Eurostat 2011) the statistical output quality dimensions[2] describe the *product quality*. In the last 20 years, official statistics has also focused on the way we do things, i.e., *process quality*. Managed separately, these three levels of quality are not sufficient, but together, they address all important aspects in an organization (Lyberg et al. 1998).

The EFQM Model is built on three integrated parts: the eight fundamental concepts, the model itself with five enabler and four results criteria[3], and the RADAR[4] assessment tool for evaluation. The principles of an excellence model are generic and go back to Deming's Plan-Do-Check-Act procedure (Deming 1986). RADAR is used to assess the results and the way things are done in the organization. In order to receive high scores the organization

---

[2] The product quality dimensions are relevance, accuracy and reliability, timeliness and punctuality, coherence and comparability, and accessibility and clarity. (Eurostat 2011)

[3] The model criteria are: Leadership, Strategy, People, Partnership & Resources, Processes, Products & Services, Customer results, People results, Society results, and Key results. (EFQM 2013)

[4] The RADAR elements are: Results, Approach, Deployment, and Assessment & Refinement. (EFQM 2013)

should have clear goals and a supporting strategy. Furthermore the organization should (i) develop approaches to fulfill the strategy, (ii) systematically use the approaches in the entire organization, and (iii) assess and improve the way it does things.

At Statistics Sweden we have described our quality management system according to the EFQM model. The main objectives are to give a comprehensive picture of Statistic Sweden and its management system, to create a document that serves as a pedagogical material for e.g., new employees and a base for external assessments and organizational development. The document is also used in the internal training program for managers.

3. The ISO 20252 for Market, Opinion and Social Research

3.1 Requirements

ISO 20252 contains about 450 requirements mainly on the statistics production process (International Standardization Organization 2012). The standard has a client focus and transparency and traceability in methods are important requirements in the standard. The use of checklists and templates are also crucial in order to reduce unnecessary variation in the organization. Validation of results is an important requirement for subprocesses that have a large impact on data quality or cost. Examples of validation requirements include monitoring of interviewers and coding control.

3.2 Implementing the Requirements

In March 2014 Statistics Sweden was certified according to ISO 20252. The process owners at Statistics Sweden developed approaches, checklists and templates that fulfilled the ISO requirements. These approaches were made available to all employees on the intranet in the so called process support system. In order to help the products to implement the standard, Statistics Sweden set up a network of quality coaches. Each department has one quality coach assigned who is specially trained in the ISO requirements and the process support system. All product managers received a couple of hours training on the requirements and all products at Statistics Sweden (about 200 products) were screened in order to make sure that they fulfilled the requirements. In order to make the screening process as easy as possible for the products we set up a helpdesk so that product managers could get support whenever they needed. The helpdesk consisted of ISO-trained quality coaches and the central quality team at the R&D Department. The screening process took about three months. The Director General regularly followed up on the progress. Table 3.2-1 shows the implementation progress, as percent of products that fulfilled the ISO requirements.

Table 3.2-1
Percent of products that fulfilled the ISO requirement

Year	Percent
2009	5%
2010	46%
2011	84%
2012-13	final steps and implementation
2014	certification

The journey to get certified took six years mainly due to some important gaps between the ISO requirements and our production process that were identified in the screening process. Five major development projects were initiated in order to address these long term deficiencies in our production process. The projects addressed deficiencies in disclosure control, coding control, interview monitoring, variance estimation and questionnaire testing.

Compliance monitoring is another feature of the standard. Statistics Sweden set up an organization with 10 internal quality auditors that work part time auditing and part time on their regular work. The internal quality auditors are carefully selected and trained by an external expert. Every year about 25-30 audits are carried out. The internal quality audits are an important tool for changing the culture in an organization. The basic philosophy is that each

employee at Statistics Sweden should be in some kind of contact or hear about these audits either as a result of their product being audited or that someone that they know has been audited or is working as a quality auditor.

Statistics Sweden's approach to fulfill the ISO requirements is described in a quality manual. The quality manual is presented to the external evaluator that certifies the organization.

3.4 Experiences Gained

One of the main benefits of the ISO work is that some important long term deficiencies in the production process such as lack of monitoring of interviewers and coding control have been addressed. Another main benefit is the educational boost in the organization. The staff involvement substantially increased the awareness of the need for quality assurance and quality control at Statistics Sweden. Another important experience is that it was very helpful to have an international standard to lean on. Without the standard Statistics Sweden would have been forced to develop a standard and that would have taken much longer. The communication with external customers would have been more difficult with an in-house developed standard. Most people are familiar with ISO and ISO certification is viewed as a quality mark.

The most important success factor has been the Director General's and top management's commitment throughout the whole process.

4. ASPIRE –A System for Product Improvement, Review and Evaluation

As a result of the errors experienced in the critical products CPI and NA in 2011, the Ministry of Finance wanted to see improvements in Statistics Sweden's products. We needed quantitative and objective measures of product quality. We decided to focus on the accuracy component and prioritize the ten most important statistical products. Paul Biemer, Distinguished Fellow at RTI International and Dennis Trewin, former Australian Statistician, helped Statistics Sweden to develop ASPIRE, a management tool with two main goals (Biemer et al 2014). One goal is to evaluate our products and the second goal is to inspire staff to make important quality improvements in their products.

4.1 Error Sources and Quality Criteria

The ten products that are part of the ASPIRE effort are shown in Table 4.1.-1. The products that are evaluated are surveys, registers and compilations. The error sources are slightly different for these different types of products and the framework has been adjusted accordingly.

The quality criteria that we use for all products are:
• Knowledge (of the producers of statistics) of the risks affecting data quality for each error source,
• Communication of these risks to the users and suppliers of data and information,
• Available expertise to deal with these risks (in areas such as methodology, measurement or IT),
• Compliance with appropriate standards and best practices relevant to the given error source, and,
• Plans and achievements for mitigating the risks.

Table 4.1-1
Error source considered by product

Product	Error Sources
Survey Products Foreign Trade of Goods Survey (FTG) Labour Force Survey (LFS) Annual Municipal Accounts (RS) Structural Business Survey (SBS) Consumer Price Index (CPI) Living Conditions Survey (ULF/SILC)	Specification error Frame error Nonresponse error Measurement error Data processing error Sampling error Model/estimation error Revision error

Registers
Business Register (BR)
Total Population Register (TPR)
	Specification error
	Frame: Overcoverage
	      Undercoverage
	      Duplication
	Missing Data
	Content Error

Compilations
Quarterly Gross Domestic Product (GDP)
Annual GDP
	Input data error
	Compilation error
	   Data processing error
	   Modelling error
	Balancing error
	Revision error

4.3 The Review Process

Having external evaluators is an important feature of ASPIRE. The main reasons are that we want to achieve objectivity, factual and perceived. We also believe that an external influence, by highly competent and respected evaluators, is a great source of inspiration for improvement work. This would be much harder to achieve with a self-assessment approach with internal evaluators.

We have developed guidelines and checklists for the review process to make it as transparent as possible and to minimize the variation between judgments made by evaluators. The production team starts by making a self-assessment. The assessment and relevant documentation are sent to the evaluators. The next step is the main interview focusing on discussions of changes from previous year, review of the quality declarations, progress made on previous recommendations and then preliminary ratings are assigned. In the main interview recommendations on improvements are also discussed. There is also a control step where the production team receives a draft of the report. This is an opportunity to provide feedback to evaluators and for the production team to discuss items if they disagree with the evaluators. The scores are then finalized. This process is repeated annually. (When ASPIRE is implemented for the first time there are of course no previous recommendations to discuss and evaluate.)

4.4 Results from ASPIRE

In the final report the evaluators provide examples of types of studies or improvements each product should make. The results for each product are presented in a summary table. In table 4.4-1 the results for the Labour Force Survey are shown. In the rows we find the different error sources and in the columns the quality criteria are displayed. The scale that is being used ranges from poor to excellent. In our example we can, for instance, see that the available expertise on measurement error in the LFS is very good. We have one red spot and that is for frame error and compliance with standards. This is due to the fact that the frame covers the registered population and the International Labour Organization's recommendation is to cover the resident population.

Another feature of ASPIRE is that we assign a risk to each error source. The risk will vary between products. For instance, in the LFS the nonresponse and measurement errors are deemed to be high risk areas. The risk score is used to calculate the total score for each product. High risk areas have a higher impact on the total score. This is to help the product to focus on important error sources and to set priorities. In our example we can see that the total score is 64.3. Compared to the previous year this is an improvement. The shaded green and pink cells indicate the change from the previous year. In the LFS example work has been done on measurement errors and studies have been carried out to estimate them. This is considered to be an improvement compared to the previous year.

Table 4.4-1
Results for the Labour Force Survey, Round 3 in 2013

	Error Source	Average score round 2	Average score round 3	Knowledge of Risks	Communication	Available Expertise	Compliance with standards & best practices	Plans or Achievement towards mitigation of risks	Risk to data quality
Accuracy (control for error sources)	Specification error	70	70	▼	▼	▼	▼	▼	L
	Frame error	58	58	▼	▼	▼	◓	○	L
	Non-response error	52	52	○	○	○	○	○	H
	Measurement error	56	68	▼	▼	▼	○	▼	H
	Data processing error	62	62	○	○	▼	▼	▼	M
	Sampling error	78	80	▼	O	▼	O	▼	M
	Model/estimation error	60	64	○	○	▼	▼	▼	M
	Revision error	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Total score	60,9	64,3						

Scores					Levels of Risk			Changes from round 2	
🔴	◓	○	▼	O	H	M	L		
Poor	Fair	Good	Very good	Excellent	High	Medium	Low	Improvements	Deteriorations

We have seen concrete improvements as a result of ASPIRE. Here are some examples. In the first round of ASPIRE we found that all of the evaluated products were weak on measurement errors. A project that looked specifically at methods to study measurement errors was initiated. Methodologists were trained in this area and measurement error studies have been carried out. The quality of the quality declarations have improved as well. A special effort was made in this area with hands-on workshops with the specific goal to improve the information and readability of the quality declarations. We have also seen an increased activity in the area of planning for studies and improvement projects. We have also redesigned the Survey of Living Conditions with substantial improvements as a result.

4.5 Strengths and Weaknesses with ASPIRE

The strengths with ASPIRE is that it is a comprehensive tool covering error sources and containing criteria that identify risks to product quality. The evaluator checklists are effective for assigning reliable ratings. The fact that we distinguished between error sources in terms of their impact on the total error is an important feature since we have limited resources and we like to make sure that we use our resources in the best possible way. It is very inspiring for our staff to have the possibility to discuss their products and improvements to their products with very competent evaluators. It is a systematic approach to drive improvements and it is relatively simple and easily understood by managers.

One possible weakness is that ASPIRE does not measure the true accuracy of a statistical product. It also relies on the skills and experience of external evaluators, and on the information provided by the product staff, i.e., there is a certain amount of subjectivity in the approach.

5. Next Steps

Statistics Sweden will continue to use the EFQM framework for identifying general improvements to our quality management system. Another goal is to maintain the current ISO-certification and to improve product quality in our most important products, using ASPIRE. It is crucial, though, to expand the pool of distinguished and respected evaluators. We will also start using the Six Sigma tool box for our improvement projects.

References

Biemer, P., Trewin, D., Bergdahl, H. and Japec, L. (2014). "A System of Managing the Quality in Official Statstics", *Journal of Official Statistics*, 3, pp. 381-573.

Deming, E. (1986). Out of the Crisis. Massachusetts Institute of Technology.

Lyberg, L., Japec, L., and Biemer, P. (1998). Quality Improvement in Surveys – A Process Perspective. Statistics Sweden.

International Standardization Organization (2012). Market-, opinion- and social research – Vocabulary and service requirements, Swedish Standards Institute.

Eurostat (2011). European Statistics Code of Practice, adopted by the European Statistical System Committee 28th September 2011.

EFQM (2013). EFQM Excellence Model. European Foundation for Quality Management. http://www.efqm.org/