



Session 4: Peer review methodologies

PEER REVIEWS - A DRIVING FORCE TOWARDS EXCELLENCE IN OFFICIAL STATISTICS?

Hilkka Vihavainen¹, Statistics Finland

Paper to be presented at the Conference on Data Quality for International Organisations, Newport (UK), 27-28 April 2006

1. Introduction

The history of overall assessments and benchmarking of national statistical systems is fairly short in comparison with some other public domain areas where they have been common for quite some time already. There have been assessments of the implementation of different policies in the member countries within the OECD in particular. Peer reviews have often been the used methodology. One reason for the late appearance of comparative assessments or benchmarks in statistics has been the lack of systematic frameworks or standards for entire statistical systems. They have been developed and put into use only in the past few years. Likewise, indicators describing the development of statistical systems are still scarce.

Adoption of the Fundamental Principles of Official Statistics in 1992 was one of the pillars that provided the framework and basic criteria according to which national statistical systems could be built and, if necessary, compared with each other. The substance of the fundamental principles was not new as such, but putting all the issues together in a single document was. The fundamental principles are, naturally, just principles that do not determine the scale and scope of their implementation.

The Data Quality Assessment Framework (DQAF) developed by the IMF was originally devised for the purposes of macro-economic statistics, but it can also be used to assess the quality of statistics in general. The DQAF provides an integrated and flexible framework in which data quality is assessed using a six-part structure that covers institutional environments, statistical processes, and characteristics of statistical products. The approach - in the form of Reports on the Observance of Standards and Codes (ROSC) - has been widely applied as the framework for more than 60 assessments of the quality of national statistics, mainly for macro-economic statistics.

The interest for assessing and comparing national statistical systems did not arise only from international initiatives, but also from the fact that new management frameworks were introduced in the public sector agencies of many countries at the beginning of the 1990s. National Statistical Offices were not an exception. Systems of management by results began to be implemented in various

¹ The opinions expressed in the paper are those of the author and do not necessarily reflect those of her institution.

government institutions. Quite automatically they introduced the notion of service and customer orientation to statistical activities, which lead to customer satisfaction surveys, user evaluation forms, quality declarations, etc. Total Quality Management (TQM) was adopted gradually in some statistical agencies as a further application of their management by results. The need to compare the performances of institutions and the quality of their products and processes became even more evident. There have been a few proposals for indicators that could be used in comparing the performances of national statistical offices (e.g. De Vries 1998), and some attempts to use them as well. The initiators of mutual benchmarking have been the National Statistical Offices themselves.

In their narrowest sense of definition as “an examination of one state’s performance or practices in a particular area by other states’ peers”, quite a few peer reviews have been made in statistics. The best-known examples are the ones conducted by Ivan Fellegi and Jacob Ryten, starting with a peer review of the Swiss Statistical System, followed by a similar review of the Hungarian Statistical System and, finally, a review of the Portuguese Statistical System. In these cases, the peer reviews were also requested by the chief statisticians of the reviewed organisations. The focus was on assessing the adaptability, effectiveness and credibility of the statistical system of a particular country, and on benchmarking it implicitly or explicitly with an advanced statistical system. Due to the background of the peers, frequent references came from Canada.

During the accession phase of new Member States into the European Union, Global Assessments of statistical systems were introduced as a kind of a peer review exercise. At the request of the European Commission, all the candidate countries were assessed during 1999 – 2002 with regard to the overall capacity of their statistical systems to meet the statistical requirements imposed by the European Union on its Member States. The Global Assessments turned out to be such a good tool that the Commission extended their use to other countries approaching the European Union and trying to achieve compliance with the requirements of the pertinent *acquis communautaire* with their statistics.

An extensive round of peer reviews will be launched this year, concentrating on monitoring the implementation of the European Statistics Code of Practice (CoP), adopted by the statistical authorities of the EU Member States and Eurostat in February 2005. The Code comprises 15 principles applying to the institutional environment (such as professional independence, legal mandate, resources, quality commitment, etc.), statistical processes (methods, procedures, cost-effectiveness considerations incl. burden on the respondents) and the quality of outputs, in other words the broad scale of functions of a single statistical office contributing to the European Statistics as well as the whole system in its entirety. Several indicators have been developed for each principle. A special reporting system (including peer reviews) on the implementation of the CoP has been approved and it will closely follow the basic ideas of the EFQM (European Foundation for Quality Management) model.

Intensive development has clearly taken place in recent years with regard to the frameworks and tools for peer reviews. This can be seen as a consequence from the emergence of quality issues in general. This paper discusses the performing of peer reviews of statistical systems. Examples are drawn from the European context and the focus is that of the reviewer.² Although peer reviews do not improve direct the quality of statistical data, they may produce recommendations which stimulate actions in the strive for excellence in official statistics, both nationally and internationally.

² The author has experience of participation in the conducting of three Global Assessments.

2. Contents and methodology of peer reviews of statistical systems

As a method, a peer review means discussion among equals. In the world of official statistics it means discussion between producers of official statistics, who share the same values. The main aim of a peer review is to assess the overall quality, or some quality aspect, of statistical governance, rather than the quality of statistical data. The target in general terms is to help the country under review to improve its overall performance and comply with established standards and principles.

The process adopted for performing peer reviews is quite established. They are usually conducted by a team comprised of 2-3 experts. The process entails a preparatory phase, consulting or discussion phase and reporting phase. The following two examples fit this model well:

Global Assessments

The basic objective of the Global Assessments at the turn of the millennium was to benchmark the sustainability of the national statistical system of a future Member State with respect to her administrative capacity in the field of statistics, degree of the country's compliance with the "acquis communautaire" in statistics, and the technical side of the national statistical system.

The questions underlying the whole review were the following: Is the statistical system of the reviewed country sustainable to meet the ever-increasing statistical requirements of the European Union? Does it have adequate managerial, human, legal, technical and financial resources for doing so? All the accession countries had a very strong development phase in statistics during the past decade. They all had received extensive technical assistance from the EU and from the Member States. Their national statistical systems had gone through fundamental changes. The results were very convincing, but at the same time there were underlying worries of whether the countries would manage without the support that would gradually stop after their accession to the EU.

The process of the assessment took the following forms:

1. Thorough studying of advance documentation in order to get acquainted with the status quo in the country (history, analysis of the economy and social life, development of the statistical system, etc.). Most of the issues arose at this stage. The peers were free to formulate the questions themselves on the basis of the material and propose the agenda for the interviews and discussions. The topics were planned to cover

- Legal framework (legislation on statistics, Statistical Council, consultation and adoption of a programme of statistical surveys, data protection),
- Specific features of the national statistical system (principles, structure, providers of official statistics, liaison with research institutes and universities, dissemination policy, relations with the users and with different levels of government, etc.),
- National Statistical Institute (mission's internal organisation, internal monitoring of quality and performance, finance and budgeting, external accountability, staff recruitment and training, IT, organisation of surveys),
- Statistical domains
 - Statistical infrastructure (classifications, basic registers, etc.)
 - Demographic and social statistics
 - Macro-economic statistics

- Business statistics
- Monetary, financial, trade and balance of payments statistics
- Agricultural statistics
- Multi-domain statistics.

2. Discussions and clarifications of the issues with experts of the statistical office, relevant ministries and the central bank, researchers, and representatives of the mass media. The experts of the NSO were generally well prepared and willing to clarify any point raised. The length of the mission in conducting the Global Assessment gave the reviewers the possibility to digest their remarks and conclusions of the discussions. It also gave the possibility to transfer knowledge of the good practices of their own countries. Usually the process included two visits to the country in question.

3. Dialogue with management. This part was very important for understanding all the other received information. The management of the NSO was willing to answer all the questions, which were sometimes quite awkward and difficult. It also gave the management an opportunity to clarify its own policy regarding the situation in the organisation or the status of the NSO with regard to other administration.

4. Mutual discussions among the reviewers were essential for forming a joint opinion on the country's statistical system. The reviewers came from different kinds of countries as regards size, administrative culture, length of membership in the EU and development of the statistical system. This variety of experiences was an important pre-condition to the formulation of the opinion.

5. The first and the final draft of the report were commented upon by two actors, namely the reviewed NSO and Eurostat. The format of the final report was given. The decision to publish the whole report rested solely on the reviewed NSO. According to the knowledge of the author, most of the NSOs published the final report on their website. In addition, parts of the Global Assessments were published in a special compendium covering the 10 Phare countries in 2003.

Although assessment of institutional infrastructure was the most important part, it was not the only one. The assessors were also asked to review the technical details related to country practises in some statistical domains. However, the target was not to make an assessment of compliance with different EU legal acts, but rather to consider whether or not the underlying pre-conditions of legislation and quality criteria on certain statistics were being met. A formal assessment of compliance with the *acquis communautaire* in statistics had been done by the NSO itself and by some other experts.

Peer reviews on implementation of the CoP

Peer reviews related to the monitoring of implementation of the European Code of Practice (CoP) are currently being planned. The methodology of the peer reviews will be tested in two countries before the final decision of the conduct will be done. So far no experiences can be drawn of the functionality of the eventual method.

The CoP was adopted in response to the invitation by the ECOFIN Council to develop minimum European standards on the independence, integrity and accountability of national and Community statistical authorities. It has the dual purpose of, on the one hand, improving trust and confidence in statistical authorities by proposing certain institutional and organisational arrangements, and, on the other, reinforcing the quality of the statistics they produce and disseminate by promoting coherent

application of best international statistical principles, methods and practices by all producers of official statistics in Europe. The monitoring, including the round of peer reviews, of implementation of the CoP had to serve these purposes.

The peer review process is planned to include the following phases:

- Advance briefing of the reviewers
- Acquaintance with the advance material including self-assessments, ROSC reports, Global Assessments, quality reports etc.
- Short mission (two days), interviewing of the management, key stakeholder groups, experts
- Drafting of the report and recommendations
- Comments and finalising of the report
- Publication of the report.

On the basis of the already available descriptions the review process is to become very effective and transparent. The proposed guidelines and tools seem to be well prepared. The quality awareness of all involved actors will be far higher than it was at the end of last decade and in conduct of the Global Assessments. At the same time, political pressure to demonstrate that the principles of the code have been complied will be strong. One can hope that this pressure, in conjunction with the tight schedule will not prevent open discussion and identification of problems and eventual quality gaps in the statistical systems.

3. Pre-conditions for successful implementation

Bearing in mind that the basic aim of a peer review is to help the country under review to improve its performance, confidence in the process and commitment from all parties involved are important. Several preconditions must be met for successful implementation of a peer review in statistics:

Mutual trust and confidentiality

The starting points are mutual trust and assurance of confidentiality. The management of the organisation under review has to be confident that the peers will act as objectively as possible and, consequently, the peers must be able rely on the information that will be given to them. The manner in which discussions take place is important. They must be talk from one peer to another, and not like auditors' talk, nor like hearings by a superior body that will give a binding judgement or punishment. There should be enough time in the discussions to refer to analogous situations in the peer's own organisation. The creation of a confidential atmosphere depends very much on the peers conducting the review.

Clear and jointly approved targets

Clear understanding of the targets of peer reviews is important. Why and for what purposes are peer reviews conducted? Fulfilment of the targets can give signals or assurance either to the national

authorities or to the international partners. In the Global Assessment exercise the target was to assure that the statistical offices of the new Member States had capacities for meeting emerging challenges and to indicate where further improvement was needed.

In the case of reviewing implementation of the European Code of Practice the target is highly ambitious, namely to assure the users and especially the political users of the credibility of the whole European Statistical System.

Frameworks, procedures

The concepts of the peer review frameworks and the planned procedures had to be described in advance so that both the reviewers and reviewees understood what they meant. Results from the reviews are often expected to be comparable with each other, although the comparisons seldom are the main target of the review. The idea of having a standard list of questions to be discussed in the review process is good and partly helps to solve the problem. However, as with all surveys, you can never be completely sure that the interviewers, let alone the interviewees, understand questions identically. For this reason special training and briefing of reviewers are necessary. Short training was also given to most of those who participated in the conducting of the Global Assessment.

Peer reviews will form just one element in the monitoring of implementation of the CoP. In the first stage, by the end of 2005, the Member States had to fill in an extensive questionnaire covering specific indicators. In the next stage, peer reviews will commence starting from the second half of 2006. They will be carried out “in situ” (in the Member States) by teams consisting of 3-4 experts. An especially positive aspect is that the method will be tested in two countries before extending it to all the 25 Member States. Only after the testing the final procedure will be decided upon by the Heads of the National Statistical Offices of the EU Member States and Eurostat.

Profile of the reviewers

The NSO under review has to be able to assume that the experts coming to conduct a review are professional and very well prepared for their task. Since the peer reviews cover issues related to the governance of the entire statistical system, the experts should have knowledge of statistical systems in the broad sense, and not only of a specific area of statistics but also of management issues and several statistical topics. Such experts are available at national statistical offices. Having the peers coming from at least two different background organisations is also an advantage.

Besides being professionally competent the peers have to be good listeners and capable of communicating properly. For deeper understanding of the situation it would be good if the experts got acquainted with the administrative systems of the reviewed country as well, since the administrative framework has a strong impact on how statistics are compiled (centralised, decentralised systems, legal and co-ordination power of the NSO, government policies regarding public sector reforms, etc.). Even the countries of the European Union have differing administrative arrangements that influence the conditions under which they compile their statistics.

Commitment of management

An overall assessment of a statistical system or statistical organisation concerns strongly the management of the assessed NSO and the way it functions. For this reason it is of utmost importance

that the conductors of the peer review have the full support and understanding of the management. If the management is sceptical about the usefulness of the exercise to further development of the organisation and regard the peer review as only important for the purposes of the international organisation, the results obtained with the review may end up being insignificant. The peer reviews regarding the CoP have a good chance of being very successful, since there is joint approval and strong commitment to their implementation from the Heads of the National Statistical Offices within the EU.

Communicating with staff members

The staff of the organisation under review should be fully aware of the process of the review. In most cases this also applies. It is important that the peers can talk direct with the experts and not only with the managers of the organisation. Ordinary staff members often give a somewhat different picture than the managers of the strengths and weaknesses of the organisation. Sharing information about the results of the peer review is also important for the implementation of further development actions within the organisation.

Facts and documentation

It is impossible to conduct a peer review without a massive amount of facts and documentation. Although interviews and discussions are important, facts are the most convincing. The availability of different performance indicators, organisational information on the Internet, plans and monitoring reports, results from customer, user and personnel surveys, media reports, financial indicators, etc., is essential.

One of the visible benefits from the Global Assessments exercise was that an extensive amount of varying supporting material was prepared and provided. All the material was not available in advance, but compiled in the course of the assessments. For example, some countries had no systematic follow-up reporting for response rates in business and household surveys, or no information about the users of data, etc., prior to the Global Assessments but such documentation was prepared at the request of the peers.

Mutual learning process

At its best the conducting of a peer review can be a mutual learning process, where both parties learn from each other. Learning by discussion takes some time. If the discussions have to be conducted within a very short time frame, it could be that there is no opportunity for sharing good experiences, which is one of the ideas behind peer reviews. The peer review process should serve the purposes of mutual learning and give possibility to both parties to present their positive practices. The advantages from sharing best practices can have long-term effects in an organisation. As the performing of a peer review takes considerable resources, it should be viewed as an investment by both the organisation under review and the organisations sending the peers to conduct it.

4. Challenges in conducting peer reviews

To a certain extent, the performing of a peer review is a sensitive issue, since it lays open the management resources and approaches of the reviewed organisation. If the quality culture and quality awareness of the organisation are high, the peer review can probably only confirm correctness of the chosen direction of the organisation. It may not bring much value-added to the development of the

statistical system, since self-assessments and continuous improvements of processes are already taking place without any external pressure or initiatives. However, if this is not the case, or if the organisation is working under extreme resource constraints for one reason or another, how could the discussions be focused on the problems and issues that need improvement instead of on self-evident matters? From the point of the reviewed organisation the whole exercise may seem like an additional task that has to be performed but that will not bring any improvement to the prevailing situation.

The following challenges are met in conducting peer reviews:

Peer pressure

The success of a peer review relies on the influence and persuasion exercised by the peers during the reviewing process. This is known as “peer pressure”. The peer review process can put pressure through 1) a mix of formal recommendations and informal dialogue; 2) public comparisons and even ranking among countries and, finally 3) publicity of the review. Peer pressure at its best can lead to positive development in the country. The author has no information of negative impacts of peer reviews in statistics but some reviews may not have contributed direct to the development of the statistical system under review.

Peer pressure can also influence by way of excessive solidarity between the peers if they wish to protect the overall credibility of the statistical system and do not want to say critical or negative things about their partner organisation’s performance.

No matter how qualified the peers are, they are not kings or queens who know everything and can tell the absolute truth. Sometimes the peers may have very strong, straightforward opinions about how things should be and how they should be changed in the country. It may also happen that they pay too much attention to too small issues, which would probably have a marginal impact on the quality of the statistical system, or which might only show their limited knowledge of the country in question.

Variety of agendas

Regardless of the agreed guidelines and common tools it may happen that there are quite opposite expectations related to the outcome of the review. The reviewers have one agenda (interpretation of the terms of reference), the management of the reviewed organisation has its own agenda and, finally, the organisation asking for the peer review has its own agenda. Should there even be one agenda for all parties? Maybe not, since the outcomes from the peer review process should be such that the different actors can utilise them for their own purposes. According to experiences gained in the conducting of the Global Assessments the reviewed organisation should “benefit” from the peer review. It has to receive some proposals, recommendations, etc., which can be used to justify changes to legislation or increases to financial or human resources, and as models in setting up new internal procedures and tools for co-operation with different authorities, etc.

Quantitative benchmarks

The ultimate question behind peer reviews in statistics is more or less the following: How effective and qualified is the statistical system of the given country in regard of its resources and in comparison with others? Quality and efficiency are relative concepts. The variety of national statistical systems is enormous even in Europe, let alone globally. In the European Union, there are organisations that have

100-150 experts but have to meet the same requirements as systems with thousands of experts. The ratio of costs between the statistical system and GDP can be indicative, but it does not tell the whole story. Traditions, basic infrastructure, etc., vary considerably from country to country. Nevertheless, some good quantitative indicators would be of great use. Unfortunately, few are available that are comparable with each other.

The peers should be able to give a balanced view of the statistical system while taking into account resources and possibilities.

Involvement of national users

The peer review process involves national users of statistics. There can be interviews with relevant ministries, mass media, research community, etc. The users are normally invited by the local statistical agency, not the reviewers. It can be difficult for them to understand why, since most of them are interested in data and their availability, and not in how well they are described, or whether metadata are available or the used methods are in line with international recommendations, etc. Motivation of the users for giving important feedback deserves consideration.

Technical constrains

At the moment, there is a huge amount of statistical information available in the English language from different countries, and the situation is improving all the time in this respect. Problems arise with supporting information, that is, precisely the information that could be relevant to the assessing of the statistical system. It is difficult to check the quality of plans and monitoring reports, metadata, user/customer survey results, etc., if they are only available in the national language. Such papers should be translated into English, if only for the purposes of the peer review.

Lack of knowledge of local circumstances can also be a problem irrespective of the competence of the reviewers. A certain amount of local knowledge is needed. The situation is in fact improving continuously since a lot of information today is available on the Internet about national legislation and policies in different areas, enabling the reviewers to get acquainted with them.

5. Conclusions

Peer reviews in statistics are successful when they help National Statistical Offices to understand eventual gaps in their performance and take necessary steps to fill them. They can also serve the managements of the organisations in confirming the appropriateness of their direction. All in all they should be understood as an element of the quality management of the organisation and as a tool of sharing of good practices and improving of statistical capacity.

In the long term, peer reviews reinforce confidence and trust in statistics; even when the reports may include critical remarks about certain elements in the national statistical system.

Recommendations and proposals should not remain in the peer review report only, but should be communicated and put in practice widely in the organisation. Follow-up mechanism has to be developed both at the reviewed organisation and internationally. In this respect, monitoring of the implementation of the European Code of Practice seems promising.

Better performance (even financial) indicators of statistical systems should be developed. It would give a good picture of the international community of official statisticians if we were able to develop indicators for comparing our own activities with similar activities in another country. We are able to do that while compiling comparative economic statistics on thousands of different kinds of activities in various countries and assuring that they are comparable.

The strength of international statistics comes from the comparability and coverage of national statistics and from the quality of processes within international organisations. While focusing on the comparability and quality of national statistics and statistical systems, peer reviews contribute to the quality of international statistics as well, and can thus become a driving force towards excellence in official statistics widely.

Sources

Fellegi, I. & Ryten, J. (2005): The effectiveness of a supranational statistical office: pluses, minuses and challenges viewed from the outside. *Journal of Official Statistics*, 21, 2, pp. 145-169.

Fellegi, I. & Ryten, J. (2000): A Peer Review of the Swiss Statistical System, Swiss Federal Statistical Office, Neuchatel

Fellegi, I. & Ryten, J. (2004): Strengths, weaknesses, and matters for immediate attention (regarding the Portuguese Statistical System) Ottawa, September 2004

Grunewald W. & Körner T. (2005): Quality on Its Way to Maturity: Results of the European Conference on Quality and Methodology in Official Statistics (Q2004), *Journal of Official Statistics*, 21, 4, pp. 747-759

Pagani, F. (2002): Peer Review: A Tool for co-operation and change, OECD, (SG/LEG(2002)1

Strode, M. (2005): Peer Reviews for Statistical Systems, The Paris 21 Secretariat

De Vries W.: Are we measuring Up? (1998) Questions on the Performance of National Statistical Systems, *International Statistical Review*

Enlarging the EU Statistical Network (Phare Candidate Countries), Eurostat 2003 (Compendium)

Implementation of the European Statistics Code of Practice/Code of Practice peer reviews, Doc. ESTAT/DDG-02/Quality/2006/3d1, Working paper for the working group Quality in Statistics, 8. meeting, 2006

Implementation of the European Statistics Code of Practice: Peer review methodology (room document), SPC 58th meeting, 2006, CPS 2006/58/8b/EN

IMF web site: <http://dsbb.imf.org/Applications/web/dqrs/dqrshome/>

Global assessments of the EU accession countries

Peer review: a tool for co-operation and change, OECD Observer, December 2003