UNITED NATIONS

**ECONOMIC COMMISSION
FOR LATIN AMERICA
AND THE CARIBBEAN - ECLAC**

Second meeting of the Statistical Conference of the Americas of the
Economic Commission for Latin America and the Caribbean

Santiago, Chile, 18-20 June 2003

MANAGING DATA QUALITY IN A STATISTICAL AGENCY

03-4-230

CONTENTS

ABSTRACT

Confidence in the quality of the information it produces is a survival issue for a statistical agency. If its information becomes suspect, the credibility of the agency is called into question and its reputation as an independent, objective source of trustworthy information is undermined. Therefore attention to quality is a central preoccupation for the management of a National Statistical Office. But quality is not an easily defined concept, and has become an over-used term in recent years. Quality is defined here to embrace those aspects of statistical outputs that reflect their fitness for use by clients. We identify six dimensions of quality: relevance, accuracy, timeliness, accessibility, interpretability, and coherence. For each dimension of quality, we consider what processes must be in place to manage it and how performance can be assessed. Finally, we try to integrate conclusions across the six dimensions of quality to identify the corporate systems necessary to provide a comprehensive approach to managing quality in a National Statistical Office.

KEY WORDS: **Quality; Official Statistics; Relevance; Accuracy; Timeliness.**

1. Introduction

Confidence in the quality of the information it produces is a survival issue for a statistical agency. If its information becomes suspect, the credibility of the agency is called into question and its reputation as an independent, objective source of trustworthy information is undermined. With this comes the risk that public policy debates become arguments about who has the right set of numbers rather than discussions of the pros and cons of alternative policy options.

Therefore attention to quality is a central preoccupation for the management of a National Statistical Office (we will use the abbreviation NSO to refer to a generic government statistical agency that may go under different names in different countries). Current recognition of the importance of quality to NSO management is reflected in several recent events in the realm of official statistics. For example, *Quality Work and Quality Assurance within Statistics* was chosen as the theme for the May 1998 meeting of the heads of NSO's in the European Community (EUROSTAT, 1998); several principles stressing the importance of relevance, professionalism and openness were included among the ten *Fundamental Principles of Official Statistics* approved by the U.N. (United Nations, 1994); Performance Indicators (which includes quality as a critical dimension of performance) was chosen as the subject for substantive discussion at the 1999 Conference of European Statisticians (UNECE, 1999). This journal through its 25 year history has carried many articles addressing quality issues and it is appropriate that in this anniversary issue we address the topic of quality management in statistical agencies.

But *quality* is not an easily defined concept, so the first issue is what do we mean exactly by quality in this context. Quality has become an over-used term during the past two decades. The Total Quality Management (TQM) movement and other management frameworks have broadened the concept of quality beyond the traditional statistician's concepts of data quality as defined, for example, by the mean square error of an estimator. So our first challenge is to circumscribe the concept of quality as it relates to the work of a NSO. That is the object of section 2 of this paper in which we will suggest six dimensions of quality about which NSO's need to be concerned. In the subsequent six sections we address each of these dimensions in turn, and consider for each: what exactly needs to be managed, what approaches might be used for managing it, and how might we measure performance in managing it.

In section 9 we will attempt to integrate some of the conclusions across the six dimensions of quality, and to identify the agency-wide systems necessary to provide a corporate approach to the management of quality. In the final section we suggest some areas requiring further attention in order to manage quality more effectively.

2. Definition of data quality

The difficulty for statisticians in defining quality as it applies to statistical information is that they thought that was something they had already done. Their whole training is concerned with optimizing the quality of statistical estimates, the fit of statistical models, or the quality of decision-making in the face of uncertainty. Using concepts such as standard error, bias, goodness of fit, and error in hypothesis testing, they have built up methodology for estimation and analysis in which the quality of data, as defined in a certain precise sense, plays a central role.

But the term quality has come to take on a broader meaning in the management of organizations. The TQM movement and other management philosophies have focused on the fitness of final products

and services for users, have emphasized the need to build quality into the production and delivery processes of the organization, and have stressed the importance of employee involvement in process redesign and commitment to improvement of the final product or service. Statistical methods play an important role in these management approaches, but they are part of a larger picture. A question to consider is how this broader notion of quality applies to an organization engaged in the production and delivery of statistical information. The definition and management of quality in government statistics were discussed in several papers presented at the 1995 International Conference on Survey Measurement and Process Quality (Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz and Trewin, 1997; de Leeuw and Collins, 1997; Dippo, 1997; Morganstein and Marker, 1997; Colledge and March, 1997) and more recently in Collins and Sykes (1999). For an earlier approach see Hansen, Hurwitz and Pritzker (1967).

If we accept that the needs of clients or users should be the primary factor in defining the activities, and assessing the success, of a NSO, we can define the concept of quality as embracing those aspects of the statistical outputs of a NSO that reflect their fitness for use by clients. But, since a NSO has many and varied clients, and each may make a variety of uses of statistical information, this does not provide an operational definition. However, it does allow a more systematic consideration of the most important dimensions of this broader concept of quality, a concept which clearly extends beyond the statistician's traditional preoccupation with accuracy, the aspect of quality which most easily lends itself to rigorous mathematical development.

The first aspect is whether the NSO is producing information on the right topics, and utilizing the appropriate concepts for measurement within these topics. Does it have information relevant to topical policy issues or is it still counting buggy whips? Does it utilize a definition of family that is pertinent to today's society? Does its classification of occupations reflect the current labour market? These are examples of questions about the **relevance** of statistical information.

Given that the NSO is measuring relevant topics using appropriate concepts, is it measuring them with sufficient accuracy? Exact measurement is often prohibitively expensive, and sometimes impossible, so the issue is whether an acceptable "margin of error" has been achieved. This is the traditional domain of statisticians with their concepts of standard error, bias, confidence intervals, and so on. We will refer to this dimension of quality as **accuracy**.

The next two dimensions of quality relate to when and how statistical information is made available to clients. Accurate information on relevant topics won't be useful to clients if it arrives after they have to make their decisions. So the **timeliness** of statistical information is another important dimension of its fitness for use. Timeliness to the day may be crucial for key monthly economic series, but less important for measures of slowly changing phenomena.

For statistical information to be useful, clients have to be able to determine what is available and how they could obtain it. It then has to be available to potential clients in a form that they can use and afford. Both searching facilities and statistical products themselves have to use technology that is available to potential clients. This collection of considerations will be referred to as **accessibility**.

To make appropriate use of statistical information from the NSO clients have to know what they have and to understand the properties of the information. That requires the NSO to provide descriptions of the underlying concepts, variables and classifications that have been used, the methods of collection, processing and estimation used in producing the information, and its own assessment of the accuracy of the information. We will refer to this property of statistical information as its **interpretability**.

Finally, as an extension of interpretability, clients are sometimes faced with utilizing different sets of statistical information derived from different sources and at different times. Appropriate use is facilitated if information can be validly compared with other related data sets. This facility is achieved through the use of common, or at least comparable, concepts and methodologies, across products and across occasions. The degree to which statistical information fits into broad frameworks and uses standard concepts, variables, classifications and methods will be referred to as its **coherence**.

These six dimensions are summarized in table 1. Clearly they are not independent of each other. For example, all of the other five have an impact on relevance. Accuracy and timeliness often have to be traded off against each other. Coherence and relevance can sometimes be in conflict as the needs of current relevance and historical consistency compete. Information provided to ensure information is interpretable will also serve to define its coherence. Despite these interactions, these six dimensions provide a useful basis for examining how quality in this broad sense should be managed within a NSO.

It is worth noting that most of the important properties of statistical information are not apparent to users without the provision of supplementary information (or metadata) by the NSO. The accuracy of information cannot be deduced just by looking at the numbers alone ─some comparisons to other sources may shed light, but the NSO, which alone has access to the underlying microdata and first-hand knowledge of the methodology used, has to provide measures of accuracy. The relevance of information may not be apparent without information on the underlying concepts, classifications and methods used. Only timeliness and accessibility are directly observable by users.

It is also worth noting that relevance, accessibility, and coherence usually have to be considered across a whole set of outputs of a NSO, rather than for each output individually. The relevance of statistical information depends on what else is available and therefore needs assessment across a whole program. By definition, the same is true of coherence. Most statistical products are delivered through a common dissemination system for the whole NSO so that questions of accessibility are largely corporate too. On the other hand, accuracy, timeliness, and interpretability can be considered as properties of each statistical output, even though, here too, each output may make use of tools or approaches that are common across programs.

We will next consider the management of quality within each of these dimensions.

Table 1
THE SIX DIMENSIONS OF DATA QUALITY

Relevance

The **relevance** of statistical information reflects the degree to which it meets the real needs of clients. It is concerned with whether the available information sheds light on the issues of most importance to users. Assessing relevance is a subjective matter dependent upon the varying needs of users. The NSO's challenge is to weigh and balance the conflicting needs of different users to produce a program that goes as far as possible in satisfying the most important needs and users within given resource constraints.

Accuracy

The **accuracy** of statistical information is the degree to which the information correctly describes the phenomena it was designed to measure. It is usually characterized in terms of error in statistical estimates and is traditionally decomposed into bias (systematic error) and variance (random error) components. It may also be described in terms of the major sources of error that potentially cause inaccuracy (e.g., coverage, sampling, non-response, response).

Timeliness

The **timeliness** of statistical information refers to the delay between the reference point (or the end of the reference period) to which the information pertains, and the date on which the information becomes available. It is typically involved in a trade-off against accuracy. The timeliness of information will influence its relevance.

Accessibility

The **accessibility** of statistical information refers to the ease with which it can be obtained from the NSO. This includes the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which the information can be accessed. The cost of the information may also be an aspect of accessibility for some users.

Interpretability

The **interpretability** of statistical information reflects the availability of the supplementary information and metadata necessary to interpret and utilize it appropriately. This information normally covers the underlying concepts, variables and classifications used, the methodology of collection, and indications of the accuracy of the statistical information.

Coherence

The **coherence** of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time. The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodology across surveys. Coherence does not necessarily imply full numerical consistency.

---

3. Relevance

Maintaining relevance requires keeping in touch with the full array of current and potential information users, not only to monitor their current needs but also to anticipate their future needs. Information needs are rarely formulated clearly in statistical terms. A major challenge is to translate expressions of interest in particular topics into likely information needs in the future. The relevance of a data set depends on what other data sets are available in related areas of interest. Relevance is therefore more meaningfully managed and assessed at the level of a "statistical program" rather than for an individual data set.

To assure relevance three primary processes need to be in place: client liaison; program review; and priority determination. These are described in the next three sections, followed in section 3.4 by a brief discussion of how performance in the domain of relevance might be assessed.

(a)   Monitoring client needs

The NSO requires a set of mechanisms whereby it stays abreast of the current and future information needs of its main user communities. These mechanisms need to include an array of consultative and intelligence-gathering processes to keep the NSO tuned in to the issues and challenges being faced by major users and which could lead to new or revised information needs on their part. Examples of possible mechanisms are given by the following selection of mechanisms used at Statistics Canada (Fellegi, 1996):

- A National Statistics Council to provide advice on policy and priorities for statistical programs;

- Professional advisory committees in major subject areas;

- Special bilateral liaison arrangements with key federal government ministries;

- Participation of the Chief Statistician in policy and program discussions among Deputy Ministers, including access to proposals to Ministers so that the statistical data needs implicit in proposed decisions or new programs can be identified;

- A Federal-Provincial Consultative Council on Statistical Policy, and subsidiary committees on specific subject-matters, for maintaining awareness of provincial and territorial governments' statistical needs;

- Special Federal-Provincial arrangements in the areas of education, health and justice to manage statistical development in these areas of largely provincial jurisdiction;

- Meetings with major industry and small business associations;

- Feedback through individual users and user enquiries.

These mechanisms are designed to identify gaps in the statistical system ─information required by users that is not currently available or good enough for the desired purposes.

(b)   Program review

The client liaison mechanisms described above will generate user feedback on current programs in addition to information about new and future needs. But periodically some form of explicit program review is required to assess whether existing programs are satisfying user needs, not only in terms of the topics addressed, but also in terms of the accuracy and timeliness of information being produced. Such reviews would utilize information generated by the regular client liaison mechanisms, might also assemble additional data, and would certainly integrate and assess this information to provide a comprehensive picture of how well the program is satisfying client needs.

There are several approaches to such an assessment. An independent expert may be commissioned to consult the user community and make recommendations on program changes. The program area itself may be required to periodically gather and assess the feedback information it is receiving, and prepare a report identifying possible changes to the program. Programs may be required to identify their lowest priority sub-programs so that these can be compared in importance with potential new investments in the same program or elsewhere.

Centrally, the NSO may conduct user satisfaction surveys covering various components of the statistical program, and monitor sales or usage of statistical products. It may also, as a result of its own integrating analytic work, identify gaps or deficiencies in the NSO's products.

All of these approaches have the common feature that, periodically, they call into question, at least on the margins, the continued existence of current programs. They help to identify investment options, both disinvestment from programs no longer relevant, and reinvestment to fill gaps in programs not keeping up with client needs.

(c)   Priority determination

The final leg of the stool is the process for considering, and acting upon, the information gleaned from user consultations and program review. Since demands will always outstrip the availability of funds, this is a process that requires the exercise of judgement in weighing the diverse needs of different user constituencies. An additional dimension of this process involves recognizing and pursuing opportunities for obtaining new financing to meet high priority information needs, thus reducing the pressure on existing programs to yield resources for reinvestment elsewhere.

At Statistics Canada, the regular annual planning cycle is the core of this process. In this process decisions may be made to invest in feasibility studies in preparation for filling recognized data gaps, to provide seed money to demonstrate how information could be produced with a larger investment, or to invest in improvements to the accuracy, timeliness or efficiency of existing programs. The launching of major new data collection initiatives usually requires resources beyond the means of internal reallocation, so the planning cycle is supplemented by periodic exercises to obtain support and funding from key federal data users for addressing major data gaps (Statistics Canada, 1998b). In determining priorities a balance has to be struck between the need for change and improvement, and the need to satisfy the important ongoing requirements served by the core program. In practice, changes from one year to the next are marginal compared to the overall program.

(d)   Monitoring performance

Measures of performance in the domain of relevance are of two main types. Firstly, evidence that the processes described above are in place is provided by descriptions of the particular mechanisms used supported by examples, if not measures, of their impact. For example, the coverage of consultative mechanisms may be assessed by systematically considering each of the major client or stakeholder groups and identifying the means of obtaining information on their statistical needs. The operation of such mechanisms can be evidenced by reviewing records of their deliberations or consultations. From the program perspective, evidence of periodic evaluation of the current relevance of each program can be provided and the impact of the results of these evaluations can be assessed.

Secondly, direct evidence of relevance may be provided by measures of usage, by client satisfaction results, and by high-profile examples of statistical information influencing or shedding light on important policy issues. Sales of information products and services provide a direct and convincing indicator of relevance. Usage of free products and services, including Internet hits for example, also reflects levels of interest, though the impact of price on usage can be complex and sometimes misleading. Pointing out and publicizing new analytic findings based on NSO data that shed light on important public policy issues can be especially convincing in demonstrating relevance. More generally, regular publication of analytic results in a readable form provides a continuing illustration of the relevance of a NSO's output, especially when republished broadly in the daily press.

Finally, the real changes that the NSO makes in its programs from year to year are a visible reflection of the working of its client liaison and priority-setting processes.

4. Accuracy

Processes described under relevance determine which programs are going to be carried out, their broad objectives, and the resource parameters within which they must operate. Within those "program parameters" the management of accuracy requires attention during the three key stages of a survey process: design, implementation, and assessment.

(a) Design

The broad program parameters will not usually specify accuracy targets. They will often indicate the key quantities to be estimated, and the level of detail (e.g., geographical, industrial) at which accurate estimates are needed, but the definition of "accurate" will at best be vague. Nor will they deal at all with tolerable levels of non-sampling error. Indeed, given the multiplicity of estimates and analyses, planned and unplanned, that come from any survey program, it would not be feasible or even useful to try to specify, before design begins, target accuracy levels. The objective of survey design is to find an optimum balance between various dimensions of accuracy and timeliness within constraints imposed by budgets and respondent burden considerations. In this process options that result in different levels of accuracy at different costs, within the broad program parameters, may be considered. The output of the design stage is a survey methodology within which some accuracy targets or assumptions, at least for key estimates and key dimensions of accuracy, will often be embedded. For example, a sample survey may aim to achieve a sampling coefficient of variation for its key estimate below a given threshold at the provincial level, and assume a response rate not less than a defined level. A census design may aim at a specified overall coverage rate, with no key sub-group's coverage falling below some lower specified rate.

The purpose here is not to describe the techniques of survey design that assist in finding optimum designs —that is the subject of the survey methodology literature (amply illustrated by the contents of this journal over its first 25 years!). Here we seek to identify some key management questions that need to be asked to ensure that accuracy considerations have received due attention during the design. We suggest eight primary aspects of design to which attention should be evident.

(i)   Explicit consideration of overall trade-offs between accuracy, cost, timeliness and respondent burden during the design stage. The extent and sophistication of these considerations will depend on the size of the program, and the scope for options in light of the program parameters. But evidence that proper consideration was given to these trade-offs should be visible.

(ii)   Explicit consideration of alternative sources of data, including the availability of existing data or administrative records, to minimize new data collection. This issue focuses on the minimization of respondent burden and the avoidance of unnecessary collection.

(iii)   Adequate justification for each question asked, and appropriate pre-testing of questions and questionnaires, while also assuring that the set of questions asked is sufficient to achieve the descriptive and analytical aims of the survey.

(iv)   Assessment of the coverage of the target population by the proposed survey frames.

(v)   Within overall trade-offs, proper consideration of sampling and estimation options and their impact on accuracy, timeliness, cost, response burden and comparisons of data over time.

(vi)   Adequate measures in place for encouraging response, following up non-response, and dealing with missing data.

(vii)   Proper consideration of the need for quality assurance processes for all stages of collection and processing.

(viii)   Appropriate internal and external consistency checking of data with corresponding correction or adjustment strategies.

While these eight areas do not cover all aspects of survey design, and consideration of issues does not necessarily result in the "optimum" decision, evidence that these aspects have been seriously considered will be strongly suggestive of sound survey design. In the end, the strength of the survey methodology will depend on the judgements of survey design teams. However, this list of issues provides a framework to guide those judgements and ensure that key factors are considered. Smith (1995) and Linacre and Trewin (1993) illustrate the balancing of these considerations in theory and practice.

Not included in the above list is a ninth area for attention: built in assessments of accuracy. This will be covered in section 4.3 below.

(b)   Implementation

But a good design can be negated in implementation. While a very good design will contain built-in protection against implementation errors (through quality assurance processes, for example), things can always go wrong. From the management perspective, two types of information are needed at the implementation stage. The first is information to monitor and correct, in real time, any problems arising during implementation. This requires a timely information system that provides managers with the information they need to adjust or correct problems while the survey is in progress. The second need is for information to assess, after the event, whether the design was carried out as planned, whether some aspects of the design were problematic in operation, and what lessons were learned from the operational standpoint to aid design in the future. This too requires information to be recorded during implementation (though not necessarily with the same fast feedback as for the first need), but it can also include information gleaned from post-implementation studies and debriefings of staff involved in implementation.

Of course, information pertaining directly to accuracy itself may only be a small subset of the information required by operational managers. But information related to costs and timing of operations is equally important to the consideration of accuracy for future designs.

(c)   Accuracy assessment

The third key stage of the survey process is the assessment of accuracy ─what level of accuracy have we actually achieved given our attention to accuracy during design and implementation? Though we describe it last, it needs to be a consideration at the design stage since the measurement of accuracy often requires information to be recorded as the survey is taking place.

As indicated earlier, accuracy is multidimensional and choices have to be made as to what are the most important indicators for each individual survey. Also each survey produces thousands of different estimates, so either generic methods of indicating the accuracy of large numbers of estimates have to be developed, or the indicators are restricted to certain key estimates.

As with design, the extent and sophistication of accuracy assessment measures will depend on the size of the program, and on the significance of the uses of the estimates. Here we propose four primary areas of accuracy assessment that should be considered in all surveys (Statistics Canada, 1992). Other, or more detailed, assessments may be warranted in larger or more important surveys to improve the interpretability of estimates as discussed later.

(i)   Assessment of the coverage of the survey in comparison to a target population, for the population as a whole and for significant sub-populations. This may mean assessing the coverage of a list frame (e.g., a business register by industry), the coverage of a census that seeks to create a list of a population (e.g., the coverage of a census of population by province or by age and sex), or the coverage of an area sample survey in comparison to independent estimates of the target population (e.g., the difference between sample based population estimates from a household survey and official population estimates).

(ii)   Assessment of sampling error where sampling was used. Standard errors, or coefficients of variation, should be provided for key estimates. Methods of deriving approximate standard errors should be indicated for estimates not provided with explicit standard errors.

(iii)   Non-response rates, or percentages of estimates imputed. The objective is to indicate the extent to which estimates are composed of "manufactured" data. For skew populations (such as most business populations), non-response or imputation rates weighted by a measure of size are usually more informative than unweighted ones.

(iv)   Any other serious accuracy or consistency problems with the survey results. This heading allows for the possibility that problems were experienced with a particular aspect of a survey causing a need for caution in using results. For example, a widely misunderstood question might lead to misleading estimates for a particular variable. It also allows any serious inconsistencies between the results and other comparable series to be flagged.

The choice of how much effort to invest in measuring accuracy is a management decision that has to be made in the context of the usual trade-offs in survey design. But requiring that, at a minimum, information on these four aspects of accuracy be available for all programs ensures that attention is paid

to accuracy assessment across the NSO. It also provides a basis for monitoring some key accuracy indicators corporately. For example, tracking trends in response rates across surveys of a similar type can provide valuable management information on a changing respondent climate, or on difficulties in particular surveys. Regular measures of the coverage of major survey frames such as a business register or an address register also provide information that is important both to individual programs using these frames, and to NSO management. More will be said about the provision of information on accuracy to users under interpretability in section 7.

5. Timeliness

Timeliness of information refers to the length of time between the reference point, or the end of the reference period, to which the information relates, and its availability to users. As we have seen, the desired timeliness of information derives from considerations of relevance —for what period does the information remain useful for its main purposes? The answer to this question varies with the rate of change of the phenomena being measured, with the frequency of measurement, and with the immediacy of response that users might make to the latest data. As we have also seen, planned timeliness is a design decision often based on trade-offs with accuracy —are later but more accurate data preferable to earlier less accurate data— and cost. Improved timeliness is not, therefore, an unconditional objective. But timeliness is an important characteristic that should be monitored over time to warn of deterioration, and across programs to recognize extremes of tardiness. User expectations of timeliness are likely to heighten as they become accustomed to immediacy in all forms of service delivery thanks to the pervasive impact of technology. Unlike accuracy, timeliness can be directly observed by users who, one can be sure, will be monitoring it whether or not the NSO does.

As indicated under accuracy, the explicit consideration of design trade-offs is a crucial component of the management of timeliness in a NSO. Equally, measures described earlier under implementation (see section 4 (b)) are important in ensuring that planned timeliness objectives are actually achieved. But there are further measures that can be pursued for managing timeliness.

Major information releases should have release dates announced well in advance. This not only helps users plan, but it also provides internal discipline and, importantly, undermines any potential effort by interested parties to influence or delay any particular release for their benefit. Achievement of planned release dates should be monitored as a timeliness performance measure. Changes in planned release dates over longer periods should also be monitored.

For some programs, the release of preliminary data followed by revised and final figures is used as a strategy for making data more timely. In such cases, the tracking of the size and direction of revisions can serve to assess the appropriateness of the chosen timeliness-accuracy trade-off. It also provides a basis for recognizing any persistent or predictable biases in preliminary data that could be removed through estimation.

For ad hoc surveys and new surveys another possible indicator of timeliness is the elapsed time between the commitment to undertake the survey and the release date. This measure reflects the responsiveness of the Agency in planning and setting up a survey as well as its execution after the reference date. But its interpretation must take account of other factors that help to determine how quickly a new survey should be in place —faster is not necessarily better.

For programs that offer customized data retrieval services, the appropriate timeliness measure is the elapsed time between the receipt of a clear request and the delivery of the information to the client. Service standards should be in place for such services, and achievement of them monitored.

6. Accessibility

Statistical information that users don't know about, can't locate, or, having located, can't access or afford, is not of great value to them. Accessibility of information refers to the ease with which users can learn of its existence, locate it, and import it into their own working environment. Most aspects of accessibility are determined by corporate-wide dissemination policies and delivery systems. At the program level the main responsibility is to choose appropriate delivery systems and ensure that statistical products are properly included within corporate catalogue systems.

So the management of accessibility needs to address four principal aspects of accessibility. Firstly, there is the need to have in place well-indexed corporate "catalogue" systems that allow users to find out what information is available and assist them in locating it. Secondly, there is the need for corporate "delivery" systems that provide access to information through distribution channels, and in formats, that suit users. Thirdly, the coverage of statistical information from individual programs in corporate catalogue systems and the use of appropriate delivery systems (corporate or in some cases program-specific) by each statistical program has to be managed. Finally, there have to be means of obtaining and acting upon usage and user satisfaction measures for the catalogue and delivery systems.

Given the current rate of technology change, the nature of both catalogue and delivery systems is evolving fast. The traditional printed catalogue that was almost always out of date has given way to on-line catalogues of statistical products, whether printed or electronic, linked to metadata bases in which characteristics of the information can be found. A thesaurus that helps users search for information without necessarily knowing the precise terms used by the NSO is also a crucial component of a catalogue system. Access to the catalogue system can be through the Internet, and users who find what they want can immediately place an order to request the desired information. It is also essential that the NSO's catalogue inter-operate with external bibliographic systems so that users searching outside the NSO are directed to it.

In addition to the structured and exhaustive approach of the catalogue, there are at least two other potential entry points for discovering what data are available. The NSO's official release mechanism in which all newly available data are announced, The Daily in the case of Statistics Canada, can provide links to catalogue entries for related products and to sources of more detailed information and metadata. The NSO's main public statistical presentation on its Internet site, known as Canadian Statistics in the case of Statistics Canada, can also include similar links to related information and metadata. While these components are not yet fully operational and integrated in many NSOs, this outlines the nature of catalogue systems for the near future.

The Internet is changing the face of delivery systems and promises to become the hub and entry point of such systems for the coming period. But the traditional delivery system of printed publications is still valued by many users, while electronic products on diskette or CD-ROM meet some needs. On-line databases continue to be a central component of a NSO's information delivery systems, whether accessible via the Internet or directly. Among all this hi-tech turmoil, the NSO has to make sure that the public good information needs of the general public continue to be met whether through the media,

through public libraries, or through the Internet. The special needs of analysts who require access to microdata present an important set of delivery challenges which are being addressed in several NSOs (see SSHRC and Statistics Canada, 1998 for example) but which we will not deal with here.

Increasingly, organizations outside the NSO, both public and private, are playing important roles in improving the accessibility of information produced by the NSO. These organizations may act simply as distributors of data, or may add context or value to NSO data by integrating them with other information or using them in ways that go beyond those that would be appropriate for a NSO. To maximize accessibility, the NSO must be open to opportunities for partnership with such organizations, but must also ensure that its identity as the source of data remains visible and, where appropriate, encourage linkages back to the original, and usually more detailed, data sources held by the NSO.

An important aspect of the accessibility of information is the pricing policy that governs its dissemination. However well-endowed the NSO, resources are limited and the option of providing unrestricted free access to all potential information is not viable. Nor is it desirable because it would destroy a most valuable source of user feedback: measures of real demand for products. A pricing policy needs to balance the desire to make certain basic information freely accessible in the public domain, while recovering the costs of providing specific products, more detailed information, and special requests. Such a policy can promote accessibility, provide a valuable source of information on relevance, and ensure that the resources of the NSO are properly balanced between collecting and processing new data on the one hand, and servicing demands for information from existing data on the other.

Finally, in the process of moving information from statistical programs into the hands of users we have to guard against the introduction of error. At this last hurdle in the process, the wrong information can get loaded into electronic databases; the wrong version of tables can find their way into publications; and enquirers can be given the wrong information over the telephone. Since the potential for these errors occurs at the delivery stage, we include them under accessibility rather than accuracy. Quality assurance systems that minimize the possibility of such errors are a necessary component of these systems.

Since users are the main judge of accessibility, systematic user feedback on catalogue and delivery systems is crucial. This feedback may be derived from (a) automated usage statistics for the various components of these systems, (b) surveys of user satisfaction with particular products, services, or delivery systems, and (c) voluntary user feedback in the form of comments, suggestions, complaints, or plaudits.

Descriptions of cataloguing and delivery systems used by some NSOs can be found in Podehl (1999), Boyko (1999) and by visiting the websites of particular NSOs.

7. Interpretability

Statistical information that users cannot understand, or can easily misunderstand, has no value and may have negative value. Providing sufficient information to allow users to properly interpret statistical information is therefore a responsibility of the NSO. Information about information has come to be known as metainformation or metadata. Managing interpretability is primarily concerned with the provision of metadata.

The information needed to understand statistical data falls under three broad headings: (a) the concepts and classifications that underlie the data; (b) the methodology used to collect and compile the data; and (c) measures of accuracy of the data. Essentially these three headings cover respectively: what has been measured; how it was measured; and how well it was measured. Users clearly need to know what has been measured (to assess its relevance to their needs), how it was measured (to allow appropriate analytic methods to be used), and how well it was measured (to have confidence in the results). Since we can rarely provide a profile of all dimensions of accuracy, the description of methodology also serves as a surrogate indicator of accuracy —it allows the user to assess, if they wish, whether the methods used were scientific, objective and carefully implemented. Under each of these headings, more detailed lists of topics can be formulated (Statistics Canada, 1992).

There are close relationships between these three headings and other dimensions of quality. The underlying concepts and classifications used are also a prime determinant of coherence (see next section) and the degree to which they conform with national or international standards should be apparent from the metadata. They are also important for the systems that allow users to find out what information is available as described under accessibility (section 6). The description of methodology will reflect the kind of design decisions described under accuracy (section 4 (a)) and the use of common tools and methods will be relevant to coherence (section 8). The measures of accuracy should reflect the considerations outlined in section 4 (c).

That information needed to understand statistical data must be comprehensible is a tautology worth stating. The NSO has to make a particular effort to ensure that the information provided under these headings is written in the users' language and not in its own internal jargon. Otherwise it fails on interpretability twice over.

To manage the interpretability dimension of quality, we suggest three elements need to be in place. The first is a policy on informing users of the basic information they need to interpret data. This policy would prescribe what information should be provided with every release of data, and in what form it might be provided. The second element is an integrated base of metadata that contains the information needed to describe each of the NSO's data holdings. Typically, this metadata base would contain more than the minimum required by the policy. Thirdly, there is a need for direct interpretation of the data by the NSO. With each major release, there should be some commentary that focuses on the primary messages that the new information contains. Directed particularly at the media, such commentary increases the odds that at least the first level of interpretation to the public will be correct. Conversely, the NSO should answer or refute serious misinterpretation of its data.

Interpretability is perhaps the one dimension of quality where the NSO should aim to do more than the user is asking. There is an element of user education in the provision of metadata. Spreading the message that all data should be used carefully, and providing the information needed to use data with care, is a responsibility of the NSO that goes beyond simply providing what users seek.

The assessment of success in the area of interpretability requires measuring compliance with the policy proposed above, and seeking user feedback on the usefulness and adequacy of the metadata and analysis provided.

8. Coherence

Coherence of statistical data includes coherence between different data items pertaining to the same point in time, coherence between the same data items for different points in time, and international coherence. The tools for managing coherence within a NSO fall under three broad headings.

The first element is the development and use of standard frameworks, concepts, variables and classifications for all the subject-matter topics that the NSO measures. This aims to ensure that the target of measurement is consistent across programs, that consistent terminology is used across programs (so that, for example, "educational level" means the same thing whether measured in a Census of population or from school records), and that the quantities being estimated bear known relationships to each other. The realization of this element is normally through the adoption and use of frameworks such as the System of National Accounts and standard classification systems for all major variables. The issue of international comparability is addressed by considering the adherence of the standards adopted to international standards where these exist. Policies are required to define program responsibilities for ensuring that data are produced according to the standards adopted.

The second element aims to ensure that the process of measurement does not introduce inconsistency between data sources even when the quantities being measured are defined in a consistent way. The development and use of common frames, methodologies and systems for data collection and processing contribute to this aim. For example, the use of a common business register across all business surveys ensures that differences in frame coverage do not introduce inconsistencies in data (there are other reasons for using a common business register too); the use of commonly formulated questions when the same variables are being collected in different surveys serves to minimize differences due to response error; the use of common methodology and systems for the various processing steps of a survey, especially edit and imputation, helps to ensure that these operations do not introduce spurious differences in data. All of these arguments apply across occasions of a particular survey, as well as across surveys.

With the first two elements we attempt to ensure that we do not build into the design or implementation of statistical programs any unjustified inconsistency. The third element deals with the results of this attempt and focuses on the comparison and integration of data from different sources. Some integration activities are regular and routine, *e.g.*, the integration of data in the national accounts, benchmarking or calibration of estimates to more reliable control totals, seasonal adjustment of data to facilitate temporal comparisons. Other activities are more exploratory and ad hoc. The confrontation of data from different sources, and their subsequent reconciliation or explanation of differences, is an activity that is often needed as part of pre-release review or certification of data to be published. Feedback from external users and analysts of data that point out coherence problems with current data is also an important component of coherence analysis. Some incoherence issues only become apparent with the passage of time and may lead to historical revisions of data.

To assess success in achieving coherence one can identify three broad sets of measures corresponding to the three elements described above. The existence and degree of use of standard frameworks, variables and classification systems; the existence and degree of use of common tools and methodologies for survey design and implementation; and the incidence and size of inconsistencies in published data. Within this latter category, one might include, for example, monitoring the residual error of the national accounts, the closure error in population estimation, or the size of benchmarking adjustments in major surveys.

9. Overall mechanisms

In reviewing each dimension of quality we have identified mechanisms which we believe to be important for the management of quality within a NSO. Some of these mechanisms lead to measures that have to be taken or followed by each individual statistical program within the NSO. Others lead to corporate-wide systems which all programs use, or to which they contribute information. In this section we extract what we consider to be the five major components or subsystems of a quality management system within a NSO.

The user liaison subsystem consists of the series of mechanisms that serve to keep the NSO in touch with its primary user groups. It provides information about current and anticipated information needs, adequacy of current products, and advice on priorities. It plays a key role in assuring the relevance of the NSO's output.

The corporate planning subsystem takes the information coming in from the user liaison system, together with assessments and internal knowledge of program strengths and weaknesses, to identify where program reductions or investments should be made. It sets the program parameters for all programs, and therefore has a direct impact on the relevance, accuracy and timeliness achievable by statistical programs. Through its funding decisions on infrastructure programs, it also influences directly the accessibility, interpretability and coherence of statistical outputs. It must be overseen by the NSO's senior management committee. Funding decisions depend on a robust cost reporting system that accurately captures the component costs of statistical programs.

The methods and standards subsystem establishes the policies and guidelines that govern the design and implementation of statistical programs, including both content and documentation standards, and standards for the methodology and systems used. It is key to achieving coherence and interpretability across statistical outputs, and to the optimization of accuracy and timeliness within programs. Its management must involve senior representation from across the NSO through a management committee.

The dissemination subsystem establishes the policies and guidelines, and puts in place the corporate systems, for delivering information to users. This includes the management and delivery of the metadata needed by users to search and access the NSO's data holdings. It is the key determinant of the accessibility and interpretability of the NSO's data. Its management too must involve senior representation from across the NSO.

Last but not least is the program reporting subsystem. Whatever the level of corporate emphasis on quality, it is within the individual statistical programs that quality is built in to the products. Within the constraints and guidance provided by corporate policies and guidelines, individual programs have to make informed trade-offs and decisions that will influence quality in all its dimensions. Within programs, evaluation and analysis of data provides a first assessment of the accuracy and coherence achieved. It is individual programs that have to defend their accuracy and timeliness records to users. A system for regular reporting by programs to management on their achievements in the different domains of quality provides an essential management input, not only for current monitoring, but more importantly as an input to the corporate planning subsystem where decisions on future investments are made.

Diagram 1 provides a simplified sketch of the relationships between these five subsystems, or key functions, necessary to the management of quality in a NSO. The subsystems are not organizational units. Indeed, the nature of most of them is that they must involve a cross-section of staff from across the NSO in order to build a corporate consensus on the appropriate policies and standards to be followed.

Diagram 1

MANAGING QUALITY IN A NSO
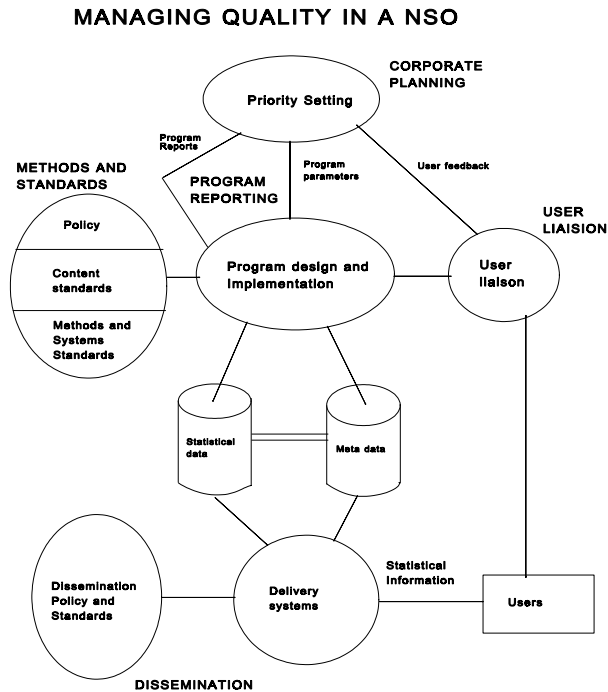
Speaking of staff, the diagram omits the crucial role of staff in all of these subsystems. A NSO is heavily dependent on a strong cadre of "knowledge-workers" covering a wide range of disciplines. As we have seen, professional expertise and judgement are required in many aspects of the design, analysis and evaluation of statistical programs. Competent staff are required to execute all phases of statistical programs with attention to the assurance of quality. Surrounding the subsystems described, we should envisage a human resources subsystem that aims to ensure that the NSO has at all times a well-trained, motivated and versatile staff capable of meeting the challenges facing the NSO. In particular, they need to have an appreciation of the importance of satisfying client needs through the management of all dimensions of quality. For one approach to a human resources subsystem see Statistics Canada (1997).

10. Conclusion

One object of this paper has been to put the statistician's traditional concern for accuracy into a broader context. Accuracy is important, but without attention to other dimensions of quality, accuracy alone will not satisfy users. Nor for many users is it the most important consideration. Trying to look at quality from a user perspective may help in solving the inevitable trade-offs between accuracy and other dimensions of quality.

This broader view of quality also helps to link together several key activities within a NSO as contributors to the management of quality. Training activities within the NSO can take advantage of this linkage to develop a broader understanding among employees of how different activities within the NSO fit together or complement each other ─and particularly of why their own work is so important.

This broader view also helps to reinforce the importance of analysis within a NSO. Analysis has been mentioned as a means of demonstrating relevance, as a means of checking accuracy, as a means of improving interpretability, and as a means of testing coherence. And that list excludes the basic role of analysis in adding to the information content of statistical outputs.

For the future, more can be done to refine the concept of quality in a NSO and to improve quality management. Within the narrower domain of accuracy, there is still more room for work on the control and measurement of non-sampling errors. With the increasing reliance on administrative data, more systematic study of the attributes of data from these sources will be required. The growing interest in longitudinal surveys, sometimes linked with administrative data, raises the need to manage accuracy issues arising in such surveys. Finally, the growing emphasis on the combination of data using integrating frameworks calls for more attention to the quality attributes of data resulting from such manipulations.

BIBLIOGRAPHY

Boyko, E. (1999), Statistical meta-data in context: an overview of statistical meta-data and related meta-data systems. Paper prepared for the 1999 Conference of European Statisticians, UN/ECE Work Session on Statistical Metadata, Working Paper No. 17, Geneva, September, 1999.

Colledge, M., and March, M. (1997), Quality policies, standards, guidelines, and recommended practices at national statistical agencies. *Survey Measurement and Process Quality*, 501-522, New York: John Wiley.

Collins, M., and Sykes, W. (1999), Extending the definition of survey quality. *Journal of Official Statistics,* 15, 1, 57-66.

de Leeuw, E., and Collins, M. (1997), Data collection methods and survey quality: an overview. *Survey Measurement and Process Quality*, 199-220, New York: John Wiley.

Dippo, C. S. (1997), Survey measurement and process improvement: concepts and integration. *Survey Measurement and Process Quality*, 457-474, New York: John Wiley.

EUROSTAT (1998), Papers prepared for the 1998 DGINS meeting are available from EUROSTAT, Luxembourg. See also SIGMA, The Bulletin of European Statistics, 03/1998, Quality in Statistics, published by EUROSTAT.

Fellegi, I.P. (1996), Characteristics of an effective statistical system. *International Statistical Review,* 64, 2.

Hansen, M.H., Hurwitz, W.N., and Pritzker, L. (1967), Standardization of procedures for the evaluation of data: measurement errors and statistical standards in the Bureau of the Census. *Bulletin of the International Statistical Institute, Proceedings of the 36th Session,* 49-66.

Linacre, S., and Trewin, D. (1993), Total survey design - application to a collection of the construction industry. *Journal of Official Statistics*, 9, 3, 611-621.

Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D. (Eds.) (1997), *Survey Measurement and Process Quality*, New York: John Wiley.

Podehl, W.M. (1999), Data base publishing on the internet. *Statistical Journal of the United Nations Economic Commission for Europe*, 16, 145-153.

Smith, T.M.F. (1995), Problems of Resource Allocation. *Proceedings: Symposium 95, From Data to Information - Methods and Systems*, Statistics Canada, 107-114.

SSHRC and Statistics Canada (1998), Final Report of the Joint Working Group of the Social Sciences and Humanities Research Council and Statistics Canada on the Advancement of Research using Social Statistics, December, 1998.

Statistics Canada (1992), Policy on Informing Users of Data Quality and Methodology, April, 1992. Policy Manual 2.3

_____ (1997), Human Resources Development at Statistics Canada, November, 1997. Internal document.

_____ (1998a), *Quality Guidelines.* Third Edition, catalogue No. 12-539-X1E, Statistics Canada.

_____ (1998b), Statistics Canada's Corporate Planning and Program Monitoring System, October, 1998. Internal document.

United Nations Economic Commission for Europe (1999), Papers prepared for the 1999 Plenary discussion. Statistical Division, UNECE, Geneva (www.unece.org/stats/documents/1999.06.ces.htm).

United Nations Economic and Social Council (1994), Report of the Special Session of the Statistical Commission, New York, 11-15 April 1994, E/1994/20.