



**Economic and Social
Council**

Distr.
GENERAL

CES/1999/9
23 March 1999

ORIGINAL : ENGLISH

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Forty-seventh plenary session
(Neuchâtel, Switzerland, 14-16 June 1999)

ANALYTIC ACTIVITIES AT STATISTICS CANADA

Submitted by Statistics Canada¹

I will very briefly² outline Statistics Canada's position on six issues related to analysis:

- benefits of analytic activities
- analytic activities to avoid
- partnerships
- access issues to microdata
- organization of analytic activities
- some safeguards: reviews of analytic products

¹ Prepared by Ivan Fellegi.

² We have a great deal of additional information on these issues that we would be pleased to share with interested persons.

I. Analytic activities and their benefits

1. Every statistical agency is, or should be, engaged in analytic activities. I will outline the types of analytic activities which I believe are appropriate -- even essential -- for statistical agencies. My first two categories are clearly not controversial; the others are perhaps increasingly controversial.

- **Quality assurance.** An important tool of quality assurance is analyzing statistical data from a database to assess "internal consistencies," as well as consistencies with other "related" data. This involves looking for unexpected deviations from known or assumed relationships and is an intrinsically analytical activity. A deviation from the expected relationship *might* signal errors in the data. It can also signal some significant change in the underlying phenomena. The former requires correction, while the latter might be a highly important finding which needs to be publicized. While assessing consistencies is analysis, the differentiation between the two possible causes of *in*consistencies involves an even more sophisticated analysis -- one with huge potential impacts on the statistical office.

- **Statistical outputs which are analytic constructs.** The best known example is the System of National Accounts. Indeed, the Gross Domestic Product (GDP) itself is not an observable entity, but rather an analytic construct. The underlying system of statistical outputs is constructed through a mixture of direct measurement and analytic tools: imputation, extrapolation, interpolation, and adjustments. Other examples include the Consumer Price Index.

2. While the visible output of what we call GDP or CPI is quantitative information, this is not the case with every analytic product. For example Statistics Canada produces several so-called microsimulation products. They are based on integrated anonymized microdata from several household surveys, but an essential feature is a modelling capacity that enables external researchers to assess both the macro and the distributional impacts of alternative policy scenarios, for example: what would be the impact of a particular change in personal income or consumption taxes on the incomes of different types of family; or what would be the impact of such policy changes on the fiscal situation of the federal or provincial governments?

- **Analysis as a means to ensure continued relevance.** It is much more difficult to maintain the longer term relevance of our product lines without a strong analytic capacity within the statistical office. There are several related reasons for this.

- The **discovery of evolving gaps** within the statistical product line is an intrinsically analytic activity. If we want to be proactive in promoting the development of important missing statistical information, we first of all must have a good understanding of *what* is missing, but even more important, *why* it matters to national policy. This requires analytic

activity of a very high order because it calls for attempts to understand not only the current structure of our society and economy, but also their dynamics: the forces that shape their evolution and which might be beneficially impacted through policy. ***I want to emphasize that we cannot wait for others (policy departments) to gain the needed understandings and then tell us what data to collect.*** The needed development is an iterative process of conceptual work, data development, program implementation, program evaluation, followed by conceptual work, and so on. Such work requires the active intellectual partnership of the statistical office.

So the analytic activity by the statistical office can and should highlight issues which new information could help to illuminate.

- **Building partnerships** with the communities of policy analysts and academics. A relationship of peers is a prerequisite for productive interactions and mutual support.
- **Analytic activity as leadership development**, particularly in subject matter areas. It is a very desirable characteristic for leaders of subject matter divisions to have an excellent ability to understand how the statistical information is used that their organization produces -- and therefore what is the real need of its clients. This kind of understanding is helped by analytic work at some time during one's career.
- **Highlighting important findings**. During a television interview I was asked how can I keep on top of the millions of pages of printed material that we produce annually. I explained that it is not my job (nor even of Statistics Canada collectively) to keep on top of billions of data points. Rather, I try to keep on top of the main findings -- the significant information we provide to society about itself.

3. How can society intelligently participate in the setting of national priorities if the population does not understand where we are, and where we are heading on current trends. Yet a large part of society is not used to analyzing statistical data. As a consequence we are really failing to serve them well if data are all of what we produce: most people will not sift through the millions of printed pages we produce in order to find the occasional golden nuggets.

4. If we do not do the necessary analytic work, who will do it *systematically*? And if it is not done, are we not wasting the enormous opportunity we have to make a fundamental contribution to the functioning of democratic societies?

A side benefit of such regular reporting of analytic highlights that the media uses our release texts directly. This not only helps in ensuring that our findings are not misrepresented, but it also raises the prestige of Statistics

Canada which, in turn, helps us in getting the cooperation of Canadians during data collection.

- **Contribution to maintaining political independence.** In my Morris Hansen lecture [1] I outlined in more detail the argument that a steady flow of relevant and objective analytic output makes an enormous contribution to raising the intellectual profile of the statistical agency, while at the same time differentiating this image from that of "the government." While government departments publish material that is mostly designed to justify the programs and approaches of the government, insightful non-partisan analyses by the statistical agency underscore its independence from the political direction of the current government. And the less this independence is understood, the more difficult it is to defend it should the need arise: if society, broadly speaking, assumes that there is no such independence, then it can hardly be expected to believe the leadership of the statistical office should the latter be forced to take a stand in opposing political interference.

II. Self-imposed constraints on analytic activities in Statistics Canada

5. Two types of analytic activities are considered inappropriate for Statistics Canada staff:

- policy advocacy or criticism; and
- forecasting.

6. Both policy advocacy and policy criticism are clearly inappropriate because they are incompatible with the requirement for scrupulous political objectivity. However, this policy requires careful interpretation. On the one hand, we definitely want to highlight objectively whatever trends and associations can rigorously be supported by available data -- whether or not some readers might gain support from our analysis for their own criticism or endorsement of a particular policy.

7. This stance is not easy to maintain. It depends, on the one hand, on a high level of scientific and analytic rigour in highlighting all the assumptions underlying the analysis, as well as the limitations of both the data and the methodology.

8. Typically, we differentiate between highlighting *relationships*, as opposed to *causalities*. Rigorous data analysis often leads to the discovery of strong relationships, but it seldom supports the identification of causalities.

9. We also try very hard to ensure that good judgement is exercised about how far it is appropriate for the statistical agency to go in drawing final conclusions from the analysis. The last section of this note describes the institutional means we use to ensure compliance with these principles.

10. Our policy on projections versus forecasting is more straightforward:

Statistics Canada does not carry out *forecasting* as such, but it is definitely engaged in *projections*. Forecasting pretends to foretell the future, while the projection is an analytic tool which -- within the constraints of a tightly specified model -- enables the analyst to consider the implications of *alternative scenarios*. The issue of scenarios is important: in order to underline the analytic character of projections, it is our policy to publish always a set of possible projections, each corresponding to a well defined analytic scenario.

III. Partnerships

11. Subject to our well established rules governing our neutrality and non-advocacy of policies, we are engaged in a broad analytic program which can be funded by and/or carried out jointly with another government department. There are numerous analytic projects which are carried out jointly with academic staff (and with staff of not-for-profit organizations, so-called "think tanks").

12. We find that partnerships are exceedingly useful in creating bonds with both the policy research and the academic communities. They contribute to securing support for our programs, they help us in maintaining our relevance, and most importantly they contribute to the virtuous circle of conceptual work leading to data development, testing of existing hypotheses and formulation of new ones, and refinement of conceptual frameworks whose validation leads to further data development [2].

IV. Overcoming the confidentiality constraints in making data available for external analysis

13. The statistical office has a tremendous relative advantage in carrying out analytic work: the unrestricted access of its staff to data that, as a result of confidentiality constraints, are inaccessible to external staff. However, this gives rise to the danger of monopoly. While I strongly believe that it is essential for some of our staff to be engaged in analytic work, it is essential to foster analytic activities in both academia and in policy oriented departments.

14. Confidentiality does not impose a constraint on research or analysis based on non-confidential aggregate data. However, an important proportion of analysis requires access to microdata. In order to discuss our approach to dealing with this problem, I have to differentiate (i) business survey data; (ii) cross-sectional household (or person) survey data; and (iii) longitudinal household (or person) survey data.

(i) **Business survey** data typically involve highly skewed observations whose release, in the form of microdata, might be impossible to achieve without

compromising confidentiality. As a result, we generally do *not* release business microdata. Incidentally, there is very little demand for this type of information.

(ii) **Cross-sectional household** (and person) survey data are subjected to an internal process to eliminate excessive geographic detail, rare code combinations or extreme observations. Following this process, and its verification by an internal committee, we release a microdata file as a matter of routine. All cross-sectional household surveys that command a broad interest are released in this manner.

(iii) The most complex problems arise in connection with longitudinal **household (and person based) surveys**. Because of their extremely rich data content, we have not found a way to render these files reasonably anonymous while maintaining their usefulness. Our solution of the confidentiality constraint, therefore, proceeds on several tracks.

(a) First, we are in the negotiation with funding agencies to establish several data centers, located in different parts of the country, where qualified research personnel could come and spend time as research fellows, working under our security control on confidential files. In effect, this will involve a dual "control": the fellowship applications will be reviewed for scientific merit by peers through a process that we do not control; and the resulting analytic product (which will be published as a joint academic and Statistics Canada document) will be reviewed by our staff to ensure that it adheres to our standards of political neutrality. Fellows can publish papers based on their research without our review after the completion of their fellowship,

(b) Second, we have established a mechanism (on a pilot basis) whereby researchers can formulate tabulation requests for confidential data without access to them. This involves our producing a "dummy" version of the confidential file which external researchers can use to formulate and test their tabulation requests. They then submit these requests to us (electronically), we run the requested tabulations against the data, subject the output to a reasonable review of confidentiality, then return the results to the researchers (also electronically). While not as direct an access as under the first alternative, this has proven satisfactory for those researchers whose access needs are occasional and not too intensive.

(c) Third, of course, research using non-confidential aggregate data is unhindered, and submission of retrieval requests from microdata bases in the traditional way is still available.

V. Organization of analytic activities

15. In Statistics Canada we have consciously adopted a mixed mode of organization for analytic activities. On the one hand we have a critical mass of analysts in a central organization. They are typically engaged in longer term analytic projects, projects that span several subject matter areas, and/or the development of broad analytic tools (e.g. our microsimulation model and corresponding database). We also encourage the development of smaller clusters of analysts in subject matter areas throughout the organization. In practice, there are well functioning networks of analysts sharing common interests³.

VI. Some safeguards: review of analytic products

16. The widespread nature of analytic activities, and the intrinsic risks that are associated with it, required that we develop formal policies and assign clear accountabilities for the review processes that we consider as being essential. We therefore have a written policy which mandates that all analytic products must be subjected to a dual review: by peers and by supervisors. The director of any area which is engaged in analytic activities is responsible for managing the execution of the policy of dual review and to monitor its effectiveness.

17. Peer review must cover the following:

- assessment of the author's interpretations and conclusions;
- avoidance of statements that are not statistically valid or justifiable;
- ensuring that interpretive statements can be linked to reported measurements or to established methods;
- assessment of methods used;
- validation of the correspondence between numbers or data patterns cited in the text and any supporting tables or charts;
- ensuring that all analytic assumptions are explicitly articulated;
- ensuring that data quality limitations have been flagged where appropriate.

18. As far as the selection of peer reviewers is concerned, it is explicitly emphasized that while reviewers with sufficient credentials to carry out the review may well be found within Statistics Canada, directors must ensure that their relationship to the author will not constrain their criticism. We require that external reviewers be used in cases where the product is judged to be of a potentially controversial nature.

19. As far as institutional review is concerned, it is to ensure that information products disseminated to the public are free of material which could compromise the Agency's reputation for non-partisanship, objectivity and

³ A report on the organization of analytic activities was commissioned from an experienced external academic person. It is available on request.

neutrality. Institutional review is, therefore, a line management responsibility. In carrying out their institutional review, line managers are encouraged to seek advice inside or outside Statistics Canada concerning statements that may be subject to damaging misuse or misunderstanding.

References

1. **Fellegi, I.P.:** Characteristics of an effective statistical system. *International Statistical Review*, August 1996, pp. 5-34.
2. **Fellegi, I.P. and Wolfson, M.:** Towards a system of social statistics. Invited paper presented at the 1997 meeting of the International Statistical Institute. To be published in *Journal of Official Statistics* (forthcoming).