

**STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

WORKING PAPER No. 13

Work Session on Statistical Output for
Dissemination to Information Media
(Voorburg, Netherlands, 25-27 September 2000)

ENGLISH ONLY

**“For concepts and methodology, click here” –
How Statistics Canada packages metadata in its Internet-based media releases**

Paper submitted by Statistics Canada¹

1. Introduction

National statistical agencies, as professional agencies in charge of producing official statistics, have the responsibility to inform users of the concepts and methodology used in collecting, processing and analysing our data, of the accuracy of these data, and of any other features that affect their quality or "fitness for use".

All data users—generalists or experts—are entitled to information about the data with which they are provided so that they get exactly what they are looking for.

Technical data users are certainly more likely to consult such metadata. General users, including journalists, tend just to want data, any data, to meet a particular need. Usually, they are not interested in, or perhaps cannot make sense of, methodology or data quality considerations, or they simply do not have the time to digest such information. Methodology is a primary concern in about one per cent of calls received by Media Relations at Statistics Canada, and a secondary concern in many other calls.

General users assume that the data provided are quality data because they come from a professional statistical agency. So they assume these data can be used over and over again, in different contexts, without much questioning.

While general users may not normally be interested in methodology, when survey results are seen as surprising or controversial, the media and the public will take an interest in metadata and data quality considerations. When statistical releases report that one city or province has a crime rate higher than others, the media will be interested. When a new method of calculation is used that results in fewer people being designated low-income, the media will be interested. And when statistical and other agencies reach different conclusions on key issues because of differing methods of calculation, the media will become interested.

This paper presents Statistics Canada's current practices and planned measures for providing metadata in media releases. It also looks at the type of basic documentation on data quality and methodology

¹ Paper prepared by Mr. Daniel Scott, Statistics Canada.

provided. Since most Canadian media get their statistical information from The Daily, Statistics Canada's official release bulletin, this paper will concentrate specifically on practices in The Daily.

2. Knowing your media clients

Some journalists will take extra care in making sure the figures they are using correspond to the topic of their story, and that the concepts and methodology used to produce the data fit their intended use.

But the vast majority of Canadian reporters are not statistical or economic experts. They are general news journalists. Moreover, journalists usually do not have much time to research and write their stories. Thus, it is important to ensure the material destined for their use is written as clearly and concisely as possible and is readily understandable.

It is not always possible, however, to write a statistical release without using precise terms and concepts that many may consider to be jargon. Furthermore, being concise when writing releases sometimes means omitting some important contextual information, which could lead to misinterpretation of the intended message. Evaluating how much contextual information to include in a release is a challenge that requires judgement, knowledge of the subject matter and understanding of media needs.

3. Finding the right balance

Subject matter authors at Statistics Canada, as well as The Daily's editors, attempt to produce releases that are just long enough to tell the story but are understandable for a general audience. They evaluate the relevance of virtually every word and sentence in releases passing across their desks.

At the same time, they must bear in mind that The Daily is now widely accessible and consulted. Since it was introduced as an electronic service on the Internet in 1995, it has spun off HTML, PDF, listserv and other versions offering users a wide choice of formats.

3.1 The print format

The Daily is still distributed as a print product to journalists who prefer the print format. Because there is no linking in the print or PDF versions, care is taken not to load the texts with too much metadata. Doing so would just make it more difficult for journalists to find the story lines. In The Daily, there is no such thing as a backgrounder that can be attached to a release. As a result, the print format has definite limitations.

The Daily does inform users of the availability of information on data quality and methodology. But because of its target audience—general news journalists—and its intended use, it does not always present measures or descriptions of data quality and documentation on methodology. Journalists interested in additional metadata must either telephone Statistics Canada with their requests, in which case someone will provide them with the information, or search electronic metadata available on Statistics Canada's Web site.

Metadata are included in Daily releases when they are essential to understanding certain concepts. They can consist of a "Note to readers", which appears as a box and is usually located near the beginning of the release. This box may contain definitions and other information that explain the significance of the release or the concepts and terminology used, or provide background information about the survey or statistical program. Notes to readers also inform users of changes affecting comparability of the data (such as historical revisions or the adoption of new standards), and may also describe any significant external influences.

Subject-matter authors sometimes incorporate definitions into the text if this can be done without interrupting the flow of the text, especially in cases of long definitions or descriptions of methodology.

Each release published in The Daily also includes a statement to advise journalists and other readers that more information is available about the quality of the data released and the concepts and methodology associated with these data. A contact name is provided with a telephone number and e-mail address.

These measures provide journalists and other users an opportunity to verify that the concepts and definitions used in collecting and processing the data released in The Daily meet their particular needs. Users can also obtain help in assessing, or they can assess themselves, the degree to which the accuracy of the data and other quality factors are consistent with their intended use or interpretation.

3.2 The Web-based product

With a print release, statistical agencies are limited with respect to the amount of information—translated into the number of pages—that they can expect reporters to digest. Daily releases are written in journalistic style, that is, the inverted pyramid. The most important messages, the story lines, appear in the first few paragraphs. The remainder follows in descending order of importance, and is in many cases most probably ignored by journalists. Devices such as the Notes to readers in The Daily help attract reporters' attention, but there is no guarantee that they work.

The limitations associated with the print version disappear with the electronic format. On the Internet, we are no longer limited to a static paper medium. Space is still at premium to avoid as much as possible the need for users to scroll horizontally or vertically. However, a lot more information can be provided to users, and it can be organised and presented in such a way that the needs of most users can be met simultaneously.

The two-dimensional journalistic inverted pyramid associated with traditional print press releases has become a three-dimensional concept. For example, through the use of hyperlinks anywhere in the release—from the lead paragraph to the contact name at the end—technical or general users can now be given access to more information or detailed documentation about the data and access to previous releases without having to telephone.

And because the use of this additional information can be monitored, it can be reorganised or relocated to give it more or less prominence, depending on its importance.

4. What metadata to provide

As more and more journalists use the electronic versions of The Daily, particularly the HTML, Statistics Canada is progressively taking advantage of the many possibilities the Internet has to offer.

We are currently testing the provision of additional basic metadata in our Daily releases. When fully implemented, the system will offer reporters and other users, by way of hyperlinks, access to more metadata about the statistical information being released. This metadata will include:

- ◆ General information about the survey program, including program clients, target population, statistical units, frequency, type of survey, survey start date and reference period;
- ◆ An overview of data collection and questions asked, including geographic coverage of collection, design and procedures and organisation responsible for collection of administrative data (where applicable);
- ◆ An overview of data processing;
- ◆ Themes and subjects covered;

- ◆ An overview of data quality, including estimation procedures;
- ◆ Products and services derived from survey program, including geographic units disseminated, release time and reference documents;
- ◆ Contacts for more information, for example the survey manager's name, telephone and fax numbers, and e-mail address.

This information comes from a general meta-database, which has been internally available for nearly 20 years but is now being enlarged and made available on the Internet. The Daily needs just a hyperlink to the specific entry in the meta-database.

Another planned improvement is the linking of terms to a glossary. From Daily releases and other parts of Statistics Canada's Web site, users would have on-line access to a glossary containing definitions of technical terms or terminology specific to a subject. This would be another means for users to ensure they understand the definitions.

Clients who prefer the print or PDF version of The Daily will still be offered an analyst's number to call if they have any questions about the concepts or methods used, or about the quality of the data. Notes to readers will also continue to be included in releases.

5. Conclusion

National statistical agencies are continually looking for ways to improve the clarity of the information they release to the media and other clients, whether over the Internet or through other means. We all want to ensure that our reports are read and interpreted correctly.

Statistics Canada has invested a great deal over the past decade to improve the quality of releases in its official release bulletin, The Daily. We like to think that these efforts have helped journalists do their job better. We hope our additional efforts to inform reporters of the concepts and methodology used in collecting, processing and analysing our data, as well as of the accuracy of these data, will further improve the quality and accuracy of media coverage. At the same time, we hope these initiatives will help improve the statistical literacy of reporters and the Canadian public.

The Internet is becoming a richer and more functional medium, and journalists and other readers are migrating to it. This offers us a unique opportunity to disseminate our message—both data and metadata.