

**STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE**

**CONFERENCE OF EUROPEAN STATISTICIANS**

**WORKING PAPER No. 19**

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

**ENGLISH ONLY**

**REVISING STATISTICAL INFORMATION RELEASED TO THE MEDIA ON THE INTERNET**

Paper submitted by Statistics Canada and U.S. Department of Justice<sup>1</sup>

**1. Introduction**

Despite all the best efforts to disseminate statistical information that is error-free, statistical agencies do have from time to time to publish corrections, errata or revisions to already-released data.

The Internet enables us to release information more rapidly than before; this has increased user expectations with regard to rapidity of access and timeliness. These user expectations can make it tempting to accelerate release, but faster release increases the potential for errors.

This paper presents some of the practices of Statistics Canada and the U.S. Department of Justice's Bureau of Justice Statistics with regard to both handling changes to already-published statistical information destined for the media on the Internet and informing the media of those revisions and corrections. It also examines some of the constraints encountered.

The mission and scope of these two agencies differ considerably. Statistics Canada provides statistical information on virtually every aspect of the nation's society and economy. Much of the data produced by Statistics Canada is frequently updated. Statistics Canada also sells many data products, including access to online databases and downloadable publications. The U.S. Bureau of Justice Statistics focuses on crime and justice data which are usually updated annually. According to the United States model, data cannot be sold. Reimbursement for reproduction such as distribution on CD-ROM is permitted. No charges are made for data made available on the Internet.

**2. The nature of the revisions dictates the strategy**

At **Statistics Canada**, journalists rely mostly on *The Daily / Le Quotidien*, the Agency's official release vehicle, to get their statistical information. Over 1200 releases are published each year in *The Daily*. There are two official release versions<sup>2</sup> of *The Daily*: the print version, delivered at 8:30 a.m. each business day to

---

<sup>1</sup> Paper prepared by Mr. Daniel Scott, Statistics Canada and Marianne Zawitz, U.S. Department of Justice

<sup>2</sup> Other versions produced include PDF, ASCII, large print, FAX and listserv.

200 journalists at the Parliamentary Press Gallery in Ottawa; and the HTML version, released on the Agency's Web site at exactly the same time.<sup>3</sup>

At the **Bureau of Justice Statistics** (BJS), most of the material targeted for media attention is released by news release. Usually, these data are included in a published report although some data is released exclusively through a news release. BJS currently produces about 24 news releases a year, which represents less than half of the total publication production. Because BJS data release is not economically sensitive, the news release and copies of the reports are provided to the media a day or two before release under an embargo. The Internet versions of reports are considered the official version.

In both agencies, the nature of the error will dictate the strategy to adopt.

## 2.1 HTML documents

At **Statistics Canada**, all changes to already-published numbers in releases to the media get corrected, at a minimum on the HTML version. Occasionally, changes to the analysis will also be made.

If the error is of little consequence, only the HTML version will be corrected, sometime the same day or the next day, without any notification to media and other users. However, if the error is considered important, it will be corrected as soon as detected in the HTML version and an erratum will appear in every version of the following edition of *The Daily*. For example, this is the case for a release that would receive wide media coverage or an error that could affect markets.

At the **Bureau of Justice Statistics**, data that appear on HTML pages usually come from published reports. In all cases of changes to numbers, the Internet versions are corrected. If an error is found in a report, corrections to the HTML pages are made at the time the electronic versions of the report are corrected. As these pages frequently change, no special notification is included. All HTML pages include the date of last revision. For original data presented on the Web site including downloadable spreadsheets, the procedures and requirements for initial verification as well as revision are the same as those for data that are included in printed publications.

For text and HTML versions of reports and news releases, a notice of revision including the revision date appears immediately after the title, and may also appear at the end of the file. On some occasions, a notice of revision is also included on the HTML abstract for the report where users find the links to the report.

## 2.2 PDF documents

At **Statistics Canada**, when an error is discovered in a PDF product, it remains available on the Web site, whether the product is free or not, until a revised version is published, provided the revision can be made quickly and the error is of no serious consequence. In the meantime, all purchases of the paying product are tracked down so the revised PDF file can be sent to clients when it becomes available. For both free and priced products, authors are encouraged to release the product again in *The Daily* to notify users of the revision. A notification also appears in the on-line catalogue.

The same procedures apply for HTML products other than *The Daily*, except that clients are directed to the appropriate URL.

The **Bureau of Justice Statistics** produces all publications in both PDF and text formats. Publications that involve news releases are generally not sent to the printer until several weeks after release. During this

---

<sup>3</sup> There are two exceptions: The official release time for the Labour Force Survey and the Consumer Price Index releases is 7:00 a.m.

time, editorial corrections are made to the PDF and text versions that are made available at the time of release, and a revised version is posted on the Internet when the publication goes to the printer. For PDF documents with data changes, a notice of revision including the revision date is included on the front page and on any other pages affected by the change. Spreadsheets related to the publication are also corrected and the date noted in the date field required on all spreadsheets.

If the data are critical to the findings and the publication is not a periodical, an erratum may be issued. This printed enclosure is produced as if it were a regular publication and is inserted in all of the remaining stock. In some instances, errata are also sent to the original dissemination list for the paper product.

Periodic reports that are updated at least annually and include trend data may include the revisions in the next edition. In the relevant tables, a note is included stating that the data from prior years are revisions to earlier releases.

### **2.3 On-line and other databases**

The strategy adopted for revisions to data in **Statistics Canada**'s on-line databases depends on the gravity of the problem. Access to the database can be restricted until the extent of the problem can be ascertained. Secondary distributors and major clients are also immediately notified by e-mail and asked not to use the data until further notice. They are also expected to inform their clients of the situation.

As soon as the data are revised in the database, a decision—based on the extent of the changes—is made as to the timing of their release. They can be sent as files on the same day to distributors and clients or re-released the following morning at official release time, which gives clients access to the revised data. In addition, a letter is sent to users to explain the situation and apologize for any inconvenience.

At Statistics Canada, the creation of publications and other products is undergoing profound change with the implementation of dynamic database publishing from CANSIM<sup>4</sup>. This new publishing method enables the Agency to produce several mediums (such as paper, HTML, PDF, CD-ROM) with minimal effort using only one data source, which can simplify the process associated with the dissemination of revised data. Whenever errors are discovered, they can be corrected in CANSIM, thus CANSIM is always the authoritative source and is always available. This is very important also for scheduled revisions, as in the case of national accounts. However, with this new publishing system, quality control before release is becoming even more important, since errors in CANSIM can be passed on to many channels.

The **Bureau of Justice Statistics** data are archived at the National Archive of Criminal Justice Data at the Inter-university Consortium for Political and Social Research at the University of Michigan. Errors in the data that could affect an analysis, particularly a replication of published results, are corrected. Some errors, such as switching one missing code for another (e.g., "Not applicable" to "Missing") are not. In documentation, errors that could affect the analysis, such as incorrectly listed codes or factual errors in descriptions of sampling plans and the represented universe, are corrected. Clarifications of summary descriptions are not always made. In annual collections, the clarification is made in the subsequent year's documentation.

The Inter-university Consortium for Political and Social Research database about the available data sets includes a field for "collection changes." This summarizes all changes to the data collection between its first release and the current release. Exact dates of revision usually are not given. Typically, it indicates the

---

<sup>4</sup> The Canadian Socio-Economic Information Management System (CANSIM) is a time series database containing more than 650,000 items. CANSIM is being converted to a multi-dimensional model which will become the corporate data warehouse for Statistics Canada.

version number (i.e., second, third...) of the data collection. This information is presented on the citation and title pages.

### 3. Important revisions

In cases when revisions are considered important, **Statistics Canada**'s media relations officers alert major Canadian news agencies and key reporters directly by telephone as soon as the problem is found to limit the use of the erroneous data or analysis. The media are given the correct data as soon as they are available. Media agencies, in turn, are expected to inform their media clients through their wire services.

In addition, a message is sent to *The Daily*'s listserv clients to inform them of the error and give them the correct information.

From time to time, journalists themselves discover errors in major economic releases during media lock-ups, which are held 30 to 45 minutes before the data are officially released. It is at that point too late to change the various versions (in Canada's two official languages) of the *Daily* release and accompanying products, if any.

Statistics Canada has an excellent track record with regard to meeting pre-established release dates for its major economic indicators. Thus, the Agency would not postpone the publication of such releases unless a major flaw were discovered, and this has never occurred. Instead, the various versions are amended as early as possible after their release. If appropriate, an erratum is published the next day.

The Agency is continually exploring ways of further improving the flexibility of its production system to allow for last-minute changes. The HTML version of *The Daily* in both official languages could potentially be changed shortly before the official release time. This would depend on a variety of factors, including the nature of the changes. The delivery of the listserv and the availability of the PDF versions, which are not considered official release versions, could be delayed. However, there would not be enough time to reprint the actual copies of *The Daily* and still have them hand-delivered to reporters at the official release time.

Reporters have noticed errors in **Bureau of Justice Statistics** reports during the embargo period. These errors are generally revised before the actual release. To date, none of the errors encountered has been deemed to be serious enough to merit any additional notification. If such a serious error were noticed, it is likely that those media who received pre-release notification would be notified through the Bureau's media listserv and fax broadcasting services. Other broader notifications might also be considered through a media alert on the wire. If the error were very serious, a notice would probably also be included in JUSTINFO, a listserv that provides a biweekly newsletter on information from the Office of Justice Programs which includes BJS.

Notification of other users is not done. In the United States, considerable attention is being given to Internet privacy. Government agencies are required to have a statement of the data they collect about visitors on their web sites. This summer, the Office of Management and Budget issued a requirement that prohibits the use of cookies on government web sites in order to protect the privacy of visitors to web sites. Therefore, collecting any information from web users which would allow subsequent contact is problematic.

As Statistics Canada is increasingly using the Internet as a means of communicating statistical information to the public, the Agency is legally and institutionally committed to ensuring full protection of Internet and other client information, as required by the *Access to Information Act* and the *Privacy Act*. Individuals who register on the Agency's web site are both informed of the intended use of their client information and given the opportunity to limit the use of this information. Clients can request, for example, that use of their information be limited to product delivery. Client registration can thus be used as a means to alert clients to data revisions.

#### **4. Conclusion**

This paper presented practices of only two statistical agencies. Despite differences in scope and mission, both agencies follow very similar procedures concerning revisions due to data errors. They both try to make the changes as quickly as possible and include notification of data changes wherever possible. Additional notification through contacts with the media, listserves, and client lists can also be employed. The greatest challenge both face is notifying users of the free web products since they are anonymous.

Clearly the speed of delivery and the size of the audience have changed with the Internet. Expectations are that changes will be instantaneous and effortless. The agencies included in this paper are realizing that this is not the case. Last minute changes are not always feasible. Data changes that once resulted only in a paper erratum now affect many files. We no longer know who all of our customers are. Media and others from our own shores or foreign countries may access our free sites and we would never know who they are. Additional techniques need to be developed to ensure proper notification of revisions to data.

It would be useful to study more agencies in order to identify best practices.