

**UNITED NATIONS STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE  
CONFERENCE OF EUROPEAN STATISTICIANS**

**UNECE Work Session on Statistical Dissemination and Communication**  
(12-14 September 2006, Washington D.C., United States of America)

Topic (iv) Managing revisions and version control to maintain credibility

**FIXING MISTAKES ON THE INTERNET:  
A CASE HISTORY AT STATISTICS CANADA**

**Invited Paper**

Submitted by Statistics Canada<sup>1</sup>

**I. INTRODUCTION**

1. Over the course of a year, statistical agencies release a phenomenal amount of data. At Statistics Canada, our official release bulletin, *The Daily*, contains on average more than 1,300 data releases a year, not to mention electronic links to myriad tables and charts in the scores of modules elsewhere on our website.
2. For the most part, we get things right. If a user were to do an electronic search of *The Daily* for “erratum”, he would get only 18 hits between January 1, 2000 and July 21, 2006. This doesn’t count mistakes in other Agency products such as publications or studies, but it does send a signal to our clients, and the public, that the Agency’s goal is accuracy.
3. This is crucial. One of the foremost challenges for a statistical agency is public confidence in the integrity of its data. Few users can validate directly the data released by a statistical agency. Instead, they rely on the source of the information. Consequently, the value of statistical information hinges directly on the credibility of the statistical agency, and this credibility is extremely fragile. A single headline can cripple an agency’s integrity in the eyes of the public. This, in turn, may make collection more difficult, if the organization loses public respect.
4. Despite the very best of efforts, mistakes happen, and when they do, the Internet confronts statistical agencies with some unique problems. For example, given the complexities of the Internet, how can agencies best change or revise information that they have already posted? How can they make sure that the new information gets communicated to all its clients? How do they handle “minor” errors as opposed to “substantive” errors, and who makes the decision on the severity of the mistake anyway?
5. The challenge is to handle each situation quickly and effectively. Since January 1, 2005, Statistics Canada has encountered a number of difficult situations concerning data errors that were published in *The Daily*. These circumstances resulted in adverse media coverage that called into question the integrity of the Agency’s economic data.
6. This paper will examine the nature of some of these errors on the Internet, and how Statistics Canada handled the situation.

**II. THE DAILY: A SHIFT IN DYNAMICS WITH OUR READERS**

7. *The Daily* is Statistics Canada’s first line of communication with the media and the public. It is Agency policy that every new product or dataset must be announced to the public through *The Daily* in

---

<sup>1</sup> Prepared by John Flanders, Senior Media Advisor, john.flanders@statcan.ca.

one form or another, whether through an analytical news release or simply a listing for a new product. Statistics Canada has published *The Daily* for 73 years. The most fundamental change in the publication occurred in 1995 when it made the transition from a paper product to an Internet product. This transition changed the very dynamics of our relationship with our clients.

8. In retrospect, the days of the paper product were relatively simple and straightforward. Copies of *The Daily* were distributed each day to journalists, either by fax or through the office of Parliamentary Press Gallery. To obtain favourable media coverage, Statistics Canada developed a working relationship with journalists, who were very much the “gatekeepers” of access and meaning between the Agency and the general public. (Although the target audience of *The Daily* was journalists, copies were also distributed to other clients, such as financial analysts, banks and economists.)

9. The point is, the circulation of *The Daily* wasn’t extensive, and Statistics Canada knew to a large extent who its readers were. When an error occurred, we could notify our clients right away. For purposes of the publication itself, subject matter analysts prepared an “erratum”, which was published as soon as possible after the offending release.

10. The Internet changed the dynamics of dealing with *The Daily*’s clients. As an online product, *The Daily* is accessible not just by the media, but by anyone with a computer and access to the World Wide Web. Within minutes of release time at 8.30 a.m. Eastern Standard Time, thousands of readers around the world may have accessed a data release. As a result, the speed and capacity of the Internet have eroded the “gatekeeping” role of the media. Although they still play a key role, newspapers and wire services are no longer the only vehicle for distributing factual information from the Agency.

11. This, in turn, made life far more complicated for the Agency. This is because it now has less of a relationship with its readers than ever before. There are many thousand visitors to our website who remain anonymous. The Agency is further removed from users as individuals than it has ever been.

12. As a result, the logistics of correcting an error become daunting. We can correct an error online. But we have no means of communicating directly with everyone who may have read an erroneous release, and we have only limited hope that everyone who had read it will see any eventual correction. It’s no wonder that the Internet seems to magnify even seemingly small errors.

13. Corporate policy requires Statistics Canada to “inform users of the concepts and methodology used in collecting, processing and analyzing its data, of the accuracy of these data, and of any other features that affect their quality of ‘fitness for use’.” However, there is no consistent procedure for following this policy. (One is under consideration and will be explained later in this paper).

14. In the following sections, we’ll examine situations that arose last year at Statistics Canada, and how they were handled.

## II. INTERNATIONAL TRADE DATA: “THE PERFECT STORM”

15. On January 12, 2005, *The Daily* published a routine monthly news release on international merchandise trade data for the reference month of November 2004. As it turned out, the release was not so routine. It kicked off a chain of events that was to culminate in national media coverage calling into question the integrity of much of Statistics Canada’s economic data.

16. According to the release, Canada’s merchandise imports, particularly from the United States, tumbled in November, resulting in a near record high trade surplus. The plunge in imports was attributed to two factors: the soaring value of the Canadian dollar compared with its US counterpart, and a drop in the volume of goods.

17. Unfortunately, the main culprit wasn’t the dollar, it was technology. Canada Border Services Agency, which supplies Statistics Canada with customs records on imported merchandise, had shut down their system for one day to introduce some modifications. Import transactions recorded on that particular day accounted for 8% to 10% of all transactions for the entire month, four times the volume on a regular day. Because these data were missing, the November trade surplus was low by \$2 billion.

18. Officials in Statistics Canada’s International Trade Division said the import data raised eyebrows right from the beginning. They said the division bent over backwards to confirm the data. It discussed them with trade officials in the United States, as well as with large Canadian importing companies and approached the Border Services Agency several times.

19. Ten days after the release, the division discovered that one day's data were in fact missing. The division concluded immediately that it had to fix the problem, and fast. On January 31, 2005, it published a seven-paragraph release in *The Daily*, issuing "revised data" for merchandise imports for November 2004. It said: "A technical problem at Canada Border Services Agency led to the transmission of incomplete customs records to Statistics Canada for November." Normally, the November data would have been revised with the release of the December data, on February 10<sup>th</sup>, but, given the size of the problem, an exceptional earlier release was distributed.

20. The mistake had immense repercussions, and not just for Statistics Canada's credibility. In a statement, the Bureau of Economic Analysis at the US Commerce Department said it uses Statistics Canada's estimates of imports of goods from the United States to prepare estimates for American trade in goods and services, and for the US gross domestic product. Statistics Canada's correction would have a US\$ 1 billion impact on calculations of trade in American goods and services for November, and on calculations of US GDP for the fourth quarter of 2004.

21. "Disappointing GDP? Blame Canada" said a headline in the *Toronto Star* of February 1, 2005. The journalist wrote: "A major blunder in a Statistics Canada report sent the world's financial markets scrambling to revise upward what was originally disappointing U.S. economic growth numbers."

22. As a Statscan trade analyst put it later: "It was the Perfect Storm."

### III. CULTURE DATA: 'NO CRIME' IN MAKING A MISTAKE

23. In June 2005, international trade statistics were again front and centre in another difficult, and complex, situation over errors. This time, the releases concerned data for international trade in culture goods for the reference years 2003 and 2004. They appeared in *The Daily* of March 29, 2005 and May 19, 2005, respectively.

24. Almost immediately, questions were raised about the data, both from subject matter people at Statistics Canada and the government department that sponsored the project. *The Daily* of May 24, 2005 contained a short erratum stating that errors were made in the original release. However, questions about the data were still being raised.

25. As it was explained later, virtually every data point in the original two releases was incorrect. The problem was traced to a processing error committed by a relatively new, inexperienced employee.

26. Between June 3 and June 13, 2005, a series of e-mails circulated among staff members at Statistics Canada and their policy department sponsor about what action should be taken.

27. At one point, the sponsoring department wrote: "It is our view that there is no 'crime' in making a mistake. The offence is the wilful promulgation of that mistake, knowing that people are downloading incorrect data, and continuing to do so for some days after the Bureau (Statistics Canada) discovered the errors." He said communications branches in both departments had their work cut out ahead "to restore full confidence in cultural statistics".

28. On June 14, the two offending news releases, along with the May 24 erratum, were all removed from Statistics Canada's website. They were replaced by a notice that the releases had been recalled due to a technical problem in the data extraction process. This had resulted in errors in many values for the import and export for culture goods for the years 1996 to 2004.

29. "We will replace the news releases once new data and analysis have been produced and validated thoroughly," the notice said. "We deeply regret any inconvenience this error may have caused our project partners...."

30. Up to this point, the situation had not attracted any media attention. But someone was obviously watching.

31. On June 27, 2005, Access to Information officials at Statistics Canada received this request from a journalist: "All records of investigation into errors in data on the international trade of culture goods, released in the Daily of March 29, 2005, May 19, 2005 and May 24, 2005, including e-mails, memos, correspondence with Canadian Heritage, briefing material for the Chief Statistician and minister, including draft or interim versions or final versions not yet complete." A package was delivered to the journalist on July 25, 2005.

32. Six days later, one of Canada's two domestic news agencies put the story on its news wire. The lead paragraph said: "For the second time this year, Statistics Canada has released erroneous trade numbers on its website, further undermining the credibility of federal statistics. And the agency knew about the latest batch of bad numbers more than 10 days before it alerted the public and media to the problem...."

33. The article also referred to the Agency's problems over the trade data for November 2004. The headline in the *Winnipeg Sun* on August 2, 2005: "Statscan goofed again."

#### IV. REPERCUSSIONS: A SCEPTICAL MEDIA

34. Incidents such as these have had major repercussions on Statistics Canada's image in the nation's media. All of a sudden, the Agency's data were no longer being taken at face value by journalists.

35. This became evident in April 2006 after the Agency removed a news release from its website that it had published the previous month. The release concerned estimates of labour productivity growth in the business sector between 2004 and 2005.

36. The text was lifted after analysts discovered they had not properly counted the number of working days in 2005. As a result, labour productivity in Canada's business sector increased much faster than initially reported.

37. The director of the responsible division explained to journalists that revisions are usually made in June. However, the magnitude of this was so big that the agency decided to make the change immediately. Instead of calling it a revision, it was being treated as a correction.

38. This incident returned to haunt the Agency in June when it issued labour force data showing that Canada created 97,600 jobs in May alone, more than the United States, which has a population 10 times larger than Canada's.

39. Financial analysts expressed strong reservations about the growth, but a spokesman for the Agency's Labour Force Survey said it had thoroughly checked the figures.

40. Even so, a journalist for Reuter's news agency phoned Statistics Canada's Media Hotline demanding to speak with someone in authority who could comment on "high-profile stumbles" at the Agency, such as the labour productivity data.

41. An assistant chief statistician told the journalist that these problems had nothing to do with the agency's data collection methods.

#### V. LESSONS: WHAT CAN WE LEARN?

42. No statistical agency is perfect. But when it makes a mistake, there are some appropriate responses it might make. At Statistics Canada, these include:

1. **Build trust through transparency:** A statistical agency gains credibility by admitting its mistakes in an open and timely manner. That is what International Trade Division did in the case of the faulty import data. The division took its lumps, but as an official said: "We were very transparent and open. We reacted quickly. I think we did a very good job once the error was identified."
2. **Determine the severity of the error:** Statistics Canada now has certain procedures in place with respect to erroneous texts in *The Daily*. The first is to determine the severity of the error, which is done in consultation between Communications and Library Services Division and subject matter analysts. Both should apply their own criteria to spell out the differences between minor and substantive errors. These differences might well depend on the context. In some contexts, a change of "0.1" can be material. On the other hand, a 25% error in a price index that, because of low weight, only moves the all-items consumer Price Index by 0.01 points may not be material. Some errors may move markets, others may not.

3. **Prepare a response plan:** The generally accepted plan at Statistics Canada is to remove a news release from *The Daily* if the error is considered to affect findings in a material fashion. (Releases in *The Daily* are archived back to 1995.) The news release is pulled and replaced as soon as possible. In instances in which an error is deemed minor, it could be handled by an erratum in a subsequent edition of *The Daily*.
4. **Notify the media and stakeholders:** Despite their reduced role as “gatekeepers,” the media still play an important role in conveying information to the public. In the event of a substantive error, the Agency’s Media Relations officials notify the media immediately, in particular the two domestic news agencies, and if necessary, journalists with financial wire services, such as Reuters or Dow Jones. This could be the case even for a single erroneous data point, if it is felt that the number is key to a story line. It is the responsibility of subject matter people to keep their stakeholders, such as survey sponsors, informed of developments.
5. **Question unusual data:** Subject matter people and *Daily* editorial staff are paying much more attention to unusual data. If employment data are unusually high, or exhibit behaviour contrary to a longstanding trend, questions are immediately raised. The motto should be check, check and check again.

43. Errors aren’t just the domain of *The Daily*. Every year, author divisions authorize the loading of hundreds of electronic publications onto the Internet. However, there is no corporate process at the moment for correcting errors in these electronic publications. Accordingly, the Agency has established a project to develop such a procedure.

44. The goal is to ensure our external and internal users are made aware of important corrections to the data and analysis; to ensure author divisions are aware of their responsibilities with respect to corrections; to ensure consistency across the organization; and to comply with the Agency’s policy on informing users of data quality and methodology.

45. Data errors provide similar challenges for other Agency products: for example CANSIM, Statistics Canada’s huge databank, which does have a battle plan in place. When the CANSIM database unit is notified of a substantive error, it freezes access to the relevant tables on the Agency’s website. CANSIM has a mailing list of prominent clients, and in the event of an error, they receive notification via an e-mail that access to certain tables has been withheld. They also receive notification when the data have been corrected and the tables replaced.

46. Similarly, in the event of errors, Statistics Canada can temporarily withhold access to products such as “smart pubs”, and to tables on other modules on its website, such as “Canadian Statistics”, that are generated from CANSIM.

## VI. CONCLUSION

47. This paper has summarized certain events during 2005 involving errors in key data releases that led to adverse media coverage. These news reports called into question the integrity of Statistics Canada’s economic data.

48. Events such as these can have a lingering negative impact on the credibility of a national statistical agency. Without a doubt, a single headline can indeed cripple an agency’s integrity in the eyes of the public. We cannot afford to take the public’s confidence in our data for granted.

49. At Statistics Canada, the steps taken last year to correct erroneous releases, albeit painful at times, have nevertheless led to a stronger understanding at the Agency about the nature of the Internet, and its impact on our relationship with our clientele. It is accepted that a statistical agency should be prepared to respond appropriately to errors, without undue delay.

50. And finally, there is a keener understanding about a fundamental responsibility of a statistical agency. Its reputation is only as good as its data – so get it right in the first place!