

**Economic and Social Council**Distr.: General
5 May 2017

English only

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on National Accounts**Sixteenth session**

Geneva, 31 May – 2 June 2017

Item 3 of the provisional agenda

Data Exchange and Confrontation**New approach to data suppression in macroeconomic accounts****Prepared by the Statistics Canada***Summary*

This document presents the outcome of a recent review of the data suppression approach for the economic statistics program in Canada. In order to respect the requirements of the Statistics Act, Statistics Canada develops methodologies and approaches that suppress the release of tabulated or aggregated information if the published data can be manipulated to reveal the information provided to Statistics Canada in such a way that the information can be related back to an individual person, business or organization. The agency periodically updates its approach to reflect the latest statistical methods. The latest review in 2015 resulted in the development of a new approach that reflects an updated interpretation of the Statistics Act. It takes into account the increasing diversity of the Canadian business population and the evolution and increasing complexity of Statistics Canada's data products.

I. Background

1. Trust is one of the cornerstones of an effective national statistics office. Users of official statistics trust national statistical organizations to provide timely, relevant and high quality statistics. They also trust national statistical organizations to protect the information that has been provided to them in confidence.
2. The idea of trust is fundamental to the Statistics Act – the legislation that governs Statistics Canada. The Statistics Act provides Statistics Canada with the mandate to produce a wide range of statistics about Canada’s economy, society and environment. At the same time the Statistics Act also lays out a number of provisions against divulging information provided to Statistics Canada under the Statistics Act.
3. Data users of economic statistics are increasingly demanding access to very detailed aggregate economic statistics, whether they be by industry or geography or activity. Faced with increasing user demand for more detail, some of Statistics Canada’s products subjected to existing data suppression methods were no longer relevant or useable. Part of the role of Statistics Canada is to determine ways to maximize the release of its information holdings while respecting the secrecy provisions of the Statistics Act. Recently, Statistics Canada conducted a review of the approach it uses to determine the information that can be placed in the public domain and the information that needs to be suppressed for reasons of confidentiality. This paper presents the results of this review by outlining the approach used at Statistics Canada to determine if a given economic statistics should be suppressed or can be released to the public.

II. Interpreting the Statistics Act

4. The decision to re-examine the agency’s approach to data suppression was motivated by the desire to balance two important risks facing Statistics Canada – risks related to relevance and trust. Data suppression reduces the relevance of Statistics Canada and the loss of relevance is a substantial risk to the agency. Releasing or being perceived to release confidential information breaks the trust relationship between Statistics Canada and the respondent. The basic question this review addressed was whether Statistics Canada could develop a new approach to data suppression that substantially reduces the risk to relevance and has minimal or no impact on the risk to our trust relationship.
5. The review started from the premise that the current confidentiality rules developed by the agency are not written in legislation – they are the agency’s own constructs / best practices implemented to comply to the legislation. In addition, while the confidentiality rules may have been systematized and embedded in methods, at their core they reflect the agency’s judgement and risk tolerance when the methods were developed. Risk tolerance can change over time. For example, many statistical organizations have dominance rules in place that state if a firm contributes more than X% to an aggregate measure the information should be suppressed. The ‘X%’ in this equation is subjective and reflects the risk tolerance of the agency at the time when the method was developed. Statistical processes have become increasingly complex and individual and business notions of privacy and confidentiality have shifted widely in the last decade. These factors combined with the fact that the current confidentiality rules were Statistics Canada’s own constructs, have led the Agency to re-examine its approach in identifying which data should be suppressed for reasons of confidentiality.
6. In developing this new approach, Statistics Canada conducted a review of the confidentiality provisions embedded in section 17(1)(b) of the Statistics Act which states that:

“no person who has been sworn under section 6 shall disclose or knowingly cause to be disclosed, by any means, any information obtained under this Act in such a manner that it is possible from the disclosure to relate the particulars obtained from any individual return to any identifiable individual person, business or organization.”

7. The review focused on the term and interpretation of ‘*particulars obtained*’ and the term ‘*possible from the disclosure to relate*’. The current interpretation led to the development of a series of rules which resulted in the suppression of an aggregate statistics:

- if there is only a single firm that reported results or if there is a dominant firm among a group of firms (even if there are thousands of firms that could have reported the information).
- if the aggregate statistic can be related to a firm even if the data are not the particulars obtained from the firm or if the data represent a statistical construct rather than a reported observation.

8. The review noted that given the increasingly complex nature of our statistical methods and data products, the statistics that the agency releases are often far removed from the “particulars collected” from respondents. In addition, given the increasing complex nature of the business population it is possible that firms engage in multiple activities and those activities change from one period to another. Therefore, just because only one firm engages in an activity at a given point in time it does not mean that it is the only firm likely to engage in the activity and therefore by disclosing the information it is not possible to trace the information back to the firm because there are many firms that are likely to have engaged in the activity.

9. This led the agency to alter its interpretation of the terms “particulars” and “possible from the disclosure to relate” in section 17 (1)(b) of the Statistics Act such that the agency can:

- Release data that could be related to a business provided the statistics are not the exact particulars obtained from any individual return (i.e. the information has been transformed to such an extent that it is no longer statistically close to the data originally provided by the respondent or obtained from administrative information)
- Release the particulars obtained from any individual return provided there is no way, through the release of the information, to relate the particulars to any identifiable business (i.e. we can release the particulars provided by a firm provided there is no way to trace those particulars back to the firm – many other firms could engage in the activity).

10. Based on this updated interpretation of the Statistics Act, the agency had to develop a new approach to identify data cells to be suppressed for reasons of confidentiality. In developing this new approach several core principles had to be respected These include:

- The new approach must respect the provisions of the Statistics Act (using the new interpretation of the term particulars and possible from the disclosure to relate).
- The new approach needs to be consistently applied across all programs and across temporal aspects within a program.
- The new approach needs to provide room to consider a subjective assessment of how the statistic was compiled and that the sensitivity can vary from one statistic to another.
- The new approach needs to consider an assessment of the potential population rather than the reporting population.

11. This change in approach means that data suppression is as much a qualitative (subjective) question as it is a quantitative question. To guide programs at Statistics Canada in the application of this new approach a decision tree was developed. The decision tree provides a clear and consistent set of rules that can be used to determine whether aggregate economic statistics can be released or whether they need to be suppressed for reasons of confidentiality.

12. The decision tree was developed to evoke a binary response, to leave out, as much as possible, any ambiguity in the application of the data suppression rules across the organization. Each branch in the decision tree is represented by a question for which there is a yes or no response. The response determines the next branch in the tree. It is expected that each response be supported by appropriate evidence such as tabulations of population counts or narratives from the program manager.

13. The decision tree is applied to each ‘data cell’ in a table proposed for release. The term “data cell” represents the intersection of dimensions in a data table proposed for publication (such as the intersection of industry, geography and time, or the intersection of an industry, commodity) as illustrated in the following table.

Revenue, 2016

Industry/Geography	Geography 1	Geography 2	Geography 3	Geography 4
Industry 1	(Data Cell)			
Industry 2				
Industry 3				

14. In total, there are eight questions in the decision tree. The response to a given question can either lead to a data cell being published, suppressed or requiring further evaluation. The remainder of this paper outlines the questions, the possible responses and the course of action to be taken for each response along with the type of supporting evidence.

15. The decision tree presented in this paper is aligned with Canada’s Statistics Act. As such the decision tree, can be used to inform the approach in other countries but the exact form of a data suppression decision tree for a given country will need to be based upon the country’s legislation.

Data Release Decision Tree¹

First Question

1. Has the data cell proposed for release been derived using information obtained under the Statistics Act?

No – the data cell can be released as per the terms and conditions under which Statistics Canada obtained the information.

Supporting evidence: legal agreement outlining the terms and conditions associated with the acquisition of the data

Yes – The data cell proposed for release was obtained by means of the Statistics Act. Further assessment is required. Go to question 2.

Second Question

2. Has the data cell proposed for release been authorized to be disclosed by order by the Chief Statistician as per 17(2) of the Statistics Act?

Yes – The data cell can be released as per the order issued by the Chief Statistician go to question 8.

Supporting Evidence: disclosure Order signed by the Chief Statistician of Canada

No – The data cell has not been authorized for disclosure by order by the Chief Statistician. Further assessment is required. Go to question 3.

Third Question

3. Is the data cell proposed for release an index for which the base period relative is not available such that the data cell no longer reflects the particulars obtained from any individual return nor is it possible to construct the particulars obtained from any individual return as outlined in section 17(1)(b) of the Statistics Act.

Yes – the data cell is an index for which the base period relative is not available. The data cell proposed for release no longer represents the particulars obtained from any individual return nor is it possible to construct the particulars obtained from any individual return. The data cell can be released subject to residual and sensitivity disclosure considerations. This form of release is authorized in section 17(2) of the Statistics Act. Go to question 7.

Supporting Evidence: sources and Methods documentation indicating the statistical product being released is an index for which the base year relative is not published.

No – The data cell is not an index or is an index for which the base period relative is available. Further assessment is required. Go to question 4.

¹ Green means the data cell can be released, blue means the data cell requires further evaluation and red means the data cell needs to be suppressed.

Fourth Question

4. Has the data cell proposed for release been derived by combining information from various sources and/or integrating the information into an economic statistics framework such that the data cell no longer reflects the particulars obtained from any individual return as outlined in section 17(1)(b) of the Statistics Act².

Yes – the data cell was derived by combining information from various sources and/or incorporating the information into an economic statistics framework such that the data cell proposed for release no longer represents the particulars obtained from any individual return. The data cell can be released subject to residual and sensitivity disclosure considerations. Go to question 7.

Supporting Evidence: sources and Methods document outlining the data sources and statistical framework used to compile the estimates.

No –The data cell is a direct aggregation of the particulars obtained from an individual return as outlined in section 17(1)(b) and has not been subjected to conceptual or methodological transformations. Further assessment is required. Go to question 5.

Fifth Question

5. Are there five or more businesses or organizations in the population (e.g. business register, individual register) that could engage in the activity represented by the data cell such that it would not be possible by a third party (i.e. a party other than Statistics Canada or the responding individual, business or organization) to relate the particulars obtained to any identifiable individual person, business or organization?

Yes – there are 5 or more businesses or organizations that could engage in the activity represented by the data cell such that it would not be able to relate the information contained in the data cell to any identifiable individual person, business or organization. The data cell can be released subject to residual and sensitivity disclosure considerations go to question 8.

Supporting Evidence: population counts indicating the number of firms within the population that could likely engage in the activity represented by the data cell.

No –There are 4 or fewer businesses or organizations in the population that engage in the activity represented by the data cell and therefore users may be able to relate the data cell proposed for release by Statistics Canada to any identifiable individual person, business or organization. Further assessment is required go to question 6.

Sixth Question

6. Have corporately approved data suppression methods (such as dominance rules) flagged the data cell for suppression?

Yes –The data cell must be suppressed.

No –The data cell can be released subject to residual and sensitivity disclosure considerations go to question 7.

Supporting evidence: reports indicating which data cells have passed corporately approved data suppression methods (such as dominance rules)

² The conceptual framework itself acts to mask of the particulars obtained from the respondent and therefore no further suppression or masking is required.

Seventh Question

7. Does the data cell reveal through residual disclosure (that is by indirect calculation) a data cell that has been suppressed from release as per the secrecy provision of the Statistics Act?

No – The data cell can be released subject to sensitivity disclosure considerations. Go to question 9.

Supporting Evidence: report indicating that the disclosure of the data cell does not reveal by residual disclosure a data cell that has been suppressed from release.

Yes – the information falls under the secrecy provisions of the Statistics Act and cannot be released. The data cell must be suppressed as per the secrecy provision 17(1)(b) of the Statistics Act.

Eighth Question

8. Is there any other factor (sensitivity of the data cell, compromises the future receipt of information or damage to the reputation of Statistics Canada, etc.) that merit further suppression of the data or the removal of suppression identified in the proceeding steps?

No – The data cell can be released.

Supporting Evidence: Program manager attestation that there are no other factors that merit further suppression of the data.

Yes – additional suppression should be applied or removed as justified by the program manager.

III. Conclusion

16. In order to respect the requirements of the Statistics Act, the agency develops methodologies and approaches that suppress the release of tabulated or aggregated information if the published data can be manipulated to reveal the information provided to Statistics Canada in such a way that it can be related back to an individual person, business or organization. The agency periodically updates its approach to reflect the latest statistical methods and balance risk tolerance with relevance. The new approach reflects an updated interpretation of the Statistics Act that considers the increasing diversity of the Canadian business population and the evolution and increasing complexity of Statistics Canada's data products.

17. This new approach does not reduce the necessary level of protection of information, rather it better aligns the approach with the requirements of the Statistics Act, including the protection of information and the agency's mandate to provide relevant and comprehensive data to the public.

18. For certain programs the application of the new approach will result in more data being available in the public domain. The approach takes into account the changing nature of the Canadian business population and the evolution and increasing complexity of Statistics Canada's data products.

19. The new approach continues to respect the confidentiality provisions of the Statistics Act. Statistics Canada will not disclose information obtained under the Statistics Act in

such a manner that it is possible from the disclosure to relate the particulars obtained to any identifiable individual person, business or organization. Statistics Canada will continue, when feasible, to review its data suppression approach and implement changes as required. This strategy furthers the Agency's policy of maintaining the relevancy of its economic statistics while continuing to protect the confidentiality of the information obtained under the Statistics Act.

Appendix

One of the first programs that made us of this new approach was Statistics Canada's foreign direct investment program. The foreign direct investment program publishes annual estimates of Inward and outward foreign direct investment by country and industry. In the past, a substantial number of these cells were suppressed for reasons of confidentiality. For the most part the main reason cells were suppressed was because the business population evaluated for confidentiality was the reporting population rather than the population that could potentially engage in foreign direct investment activity. For example, assume that a single firm in Canada in the mining industry invested in the US in 2016. In the past, this flow would have been suppressed. Assume that there are 100 large firms in Canada, all with the same likelihood of being able to invest in the US. If we consider the potential reporting population, there is no need to suppress the cell because through its disclosure there is nothing that indicates which of the 100 firms engaged in the investment.

The following tables show the level of suppression prior to using this new approach and the level of suppression after using this new approach.

Canadian FDI Abroad by region

	2012	2013	2014	2015
FDI Flow	704,335	778,371	825,303	1,005,227
Number of Data Cells	171	171	171	171
Previously Suppressed Cells	65	71	77	77
Current Suppression	11	13	23	23
New information	54	58	54	54

FDI in Canada by region

	2012	2013	2014	2015
FDI Flow	633,778	688,873	719,574	767,675
Number of Data Cells	171	171	171	171
Previously Suppressed Cells	74	44	44	44
Current Suppression	0	2	4	4
New information	74	42	40	40

Canadian FDI Abroad by Industry and Region

	2012	2013	2014	2015
FDI Flow	704,335	778,371	825,303	1,005,227
Number of Data Cells	241	241	241	241
Previously Suppressed Cells	99	84	90	90
Current Suppression	4	10	6	6
New information	95	74	84	84

FDI in Canada by Industry and Region

	2012	2013	2014	2015
FDI Flow	633,778	688,873	719,574	767,675
Number of Data Cells	241	241	241	241
Previously Suppressed Cells	107	91	99	99
Current Suppression	9	4	11	11
New information	98	87	88	88
