

**Economic and Social Council**Distr.: General
11 March 2013

Original: English

Economic Commission for Europe

Conference of European Statisticians

Sixty-first plenary session

Geneva, 10-12 June 2013

Item 4(b) of the provisional agenda

How should national statistical offices respond - moving from risk avoidance to risk management**Risk management for micro-data access: two Canadian examples****Note prepared by Statistics Canada***Summary*

This paper outlines the nature of the process that Canada has adopted to start giving deemed employees access to business micro-data while managing the risks inherent in this process in a small country such as Canada. This process has been adopted since the inauguration of the Canadian Centre for Data Development and Economic Research. This paper will cover the legal issues behind the decision, the systems that have been developed to ensure that the data are not used inappropriately, the procedures that have been adopted to ensure safe premises and the nature of the accreditation process, how disclosure control is being organized and the directions that are being contemplated for future developments. The paper will also explain the process of researcher accreditation for access to social and business data.

I. Introduction

1. Statistics Canada has a mandate to conduct analysis on the data it collects. The potential for analysis of our data holdings vastly exceeds our own capacity to conduct the analysis. To ensure Canadians obtain the full measure of benefit from the information we have collected at their expense we form partnerships with academics and other organizations to expand our analytical capacity. We provide these partners with access to our data via a variety of data access modes that can be understood as a continuum that deals with different levels of risk. Suitable risk management procedures are implemented at each point along the continuum.

2. For example, public use micro-data files undergo careful disclosure analysis prior to wide release and confidential micro-data are protected by close direct control and/or technology such as remote access technology. For direct access to detailed micro-data Statistics Canada uses its legislative powers to recognize approved researchers as deemed employees of Statistics Canada. As deemed employees, researchers carry out scientifically sound analysis that is appropriate for a national statistical office, conducted for statistical purposes only. By deeming these researchers as employees we bring them in scope of the severe penalties under the *Statistics Act* for any breach of respondent confidentiality.

3. Access by deemed employees to social and business micro-data is managed by policies, rules and procedures that limit the access to approved researchers in safe centres, specifically the Research Data Centres and the Centre for Data Development and Economic Research. Although deemed employees are subject to all the same legal penalties of paid employees, Statistics Canada has implemented further risk management strategies to ensure that the confidence and trust of respondents – whether individuals or businesses – is maintained.

4. National Statistical Offices (NSOs) sometimes adopt a strategy that minimizes access so as to virtually eliminate risk when it comes to micro-data access. Until recently, this approach has been taken for business data in Canada where, given the fairly small population, it can be relatively easy to identify dominant businesses. Nevertheless, where there is a system supporting ethically responsible research that informs the policy process and a professional research community that includes expert analysts, complete risk avoidance may not be the best position for National Statistical Offices to take. This paper describes how Canada has moved from risk avoidance to risk management with the intent to increase the analytical output from national datasets to benefit all Canadians. Recent developments in Canada that move us along the continuum are demonstrated by two new activities: streamlining procedures for accredited researchers and access to business micro-data.

II. Researcher accreditation

5. Access by deemed employees to social micro-data through Canada's Research Data Centres (RDC) requires a series of proposals, reviews, approvals, security procedures and paperwork prior to deemed employee researchers commencing work on their project. To identify where risk is low and to reduce the procedural burden for our experienced researchers with a proven track record around confidentiality, Canada has developed a researcher accreditation process that involves two tiers. The objectives for developing the two-tiered researcher accreditation were: Tier 1) to ensure that deemed employee researchers accessing confidential Canadian micro-data were qualified and sufficiently

experienced to carry out the work independently and abide by the procedures in place to protect the confidentiality of Canadians; and Tier 2) to streamline the project proposal application process for highly qualified and experienced researchers by accrediting the researcher for a ten year period and eliminating the requirement for accreditation for each project for these researchers.

6. Accreditation for Tier 1 has been in place in Canada for over ten years and is completed for each project proposal submitted. The applicant's information is reviewed and approved by peers – academic or government as appropriate. This level of accreditation allows access to confidential StatCan data, but requires a full peer review for each project. Applicants must complete an application that demonstrates research experience and technical competence to work with the data requested. In addition, graduate students must provide a letter of support indicating that a senior researcher will advise or assist the student to resolve statistical or analytical questions arising during the student's research. Specific requirements for this level of accreditation are listed in Table 1.

Table 1

Tier 1 accreditation requirements

Most significant research contributions
Graduate student supervision
Membership in professional bodies (e.g., professional associations or societies, grant evaluation committees, etc.)
Funded research projects
Research chairs or other awards recognizing research achievements
Relevant presentations to government or non-governmental groups
Invited lectures
Refereed, publications in the past 5 years that illustrate technical competence

7. Tier 2 accreditation has recently been developed to streamline access to data by experienced and accomplished deemed employee researchers. The researcher must have recognized research achievements and expertise within a field or discipline, must demonstrate that they have the knowledge and experience needed to handle confidential personal information and must provide evidence of technical competence to carry out the kind of research conducted in a Statistics Canada Research Data Centre (RDC). Table 2 provides the list of specific requirements for Tier 2 accreditation. Each application will be reviewed and approved by a panel of academic peers and Statistics Canada senior management. Researchers with Tier 2 accreditation will not be required to go through a full peer review for subsequent projects.

Table 2

Tier 2 accreditation requirements

Previous experience working with confidential information in a Research Data Centre; or equivalent experience; or confirmation showing the applicant has completed an online program on confidentiality and security.
Identify the public good arising from your current or previous research (e.g. government policy or program changes, new knowledge to

understand a current social or economic issue, etc.)
Most significant research contributions
Graduate student supervision
Membership in professional bodies (e.g., professional associations or societies, grant evaluation committees, etc.)
Association with a research institute or organization
Funded research projects
Research chairs or other awards recognizing research achievements
Relevant presentations to government or non-governmental groups
Invited lectures
Reviews for peer reviewed scholarly journals
Refereed, publications in the past 5 years that illustrate technical competence
Teaching a research methods course

8. Tier 1 accreditation has been in place in the RDC program since its inception in 2000. It is now being used in the new Canadian Centre for Data Development and Economic Research (CDER) as will be outlined below. Tier 2 accreditation has been approved and the application process is currently under development. Implementation is anticipated in 2013/2014.

III. Canadian Centre for Data Development and Economic Research

9. With the inauguration of the Canadian Centre for Data Development and Economic Research (CDER) in October 2012, secure access to business micro-data at Statistics Canada's headquarters in Ottawa was made available to a wide community of professional deemed employee researchers working on peer-reviewed, research-oriented projects. Before the creation of CDER, access to business micro-data existed, but was limited. Statistics Canada managed a fellowship program that supported a small number of doctoral and post-doctoral students. Contracts were written with a small number of academics whose work directly supported programs or the development of programs within Statistics Canada. A few analysts from other government departments and agencies, who worked on priority projects, were hosted. With a limited capacity, the strategy stressed risk avoidance.

10. Business data offers challenges in disclosure control—particularly when large numbers of tables are being requested. The *Statistics Act* forbids the disclosure of the 'particulars' of data that have been provided to the agency by a respondent. Given the skewness in business data and the small number of businesses in some industries and regions, it is relatively easy to ascertain particular entities within business data sets. These two characteristics (small size of populations and dominance of totals by a small number of observations) mean that the quality of the data produced depends on ensuring high response rates. And getting good response rates depends on the strength of the long-standing partnership between Statistics Canada and its respondents.

11. This partnership relies on more than the legal clauses contained within the *Statistics Act* that focus on the rights of the agency to demand data and its responsibility to protect them. It also depends on the tradition that the agency and its respondents have developed to work together to create valued statistical information. When new surveys are developed, the agency works hand in hand with respondents to explain the use to which they are to be put. The agency works with respondents to reduce respondent burden. So too, it communicates with this community to assure them that their data are continuously being protected.

12. All this is important to protect the quality of the business data produced. The loss of trust of even a few business respondents—especially the largest—would have a significant and direct impact on the business data produced by the statistical system. This in turn could lead to a general respondent ambivalence to reporting and a decline in the quality of the output of the statistical system.

13. Business data not only offer high costs if data confidentiality are breached, but the likelihood that this may occur is also high because of potential personal incentives for outsiders to try to decipher data bases and identify individual observations. Researchers can gain financially from misusing and breaching the confidentiality of business data. Therefore, the risk of a breach of confidentiality of business data, and the high potential impact of such a breach, meant that historically Statistics Canada adopted a very risk-averse approach when it came to granting access to business micro-data.

14. Statistics Canada has nevertheless recognized that there is a mandate and clear benefits to making use of business micro-data for analytical purposes. It has done so by developing in-house capabilities in economic research in analytical divisions that partner with other researchers from the federal policy community and the academic community; but these partnerships have been limited relative to the access provided by the RDCs that contain social data. Statistics Canada recognized that, should it use its legislative powers to deem employees for access and develop procedures that additionally control the risks of access to micro-data, there would be benefits of providing greater access to business micro-data. By making use of expertise and resources outside the agency in other government departments and in academic and research institutions, it would be better able to serve its mandate of providing information to the Canadian public through the analysis of “statistical information relating to the commercial, industrial, financial, social, economic and general activities and condition of the people”—a function given to it by the *Statistics Act*.

15. Therefore CDER was established to enable highly-trained deemed employee researchers to undertake important research on business that informs policy and improves its development. This benefit, however, had to be weighed against the increased risk of a breach in confidentiality that the greater access entailed. Therefore, before the decision to create CDER was made, policies and procedures were developed to manage the risk—to reduce it to acceptable levels. The following discussion focuses mainly on the measures that have been put in place that are in addition to the ones in place at the RDCs where Statistics Canada provides access to social micro-data. The procedures are meant to ensure that deemed employees work in secured accommodations with safe procedures on projects that fall under the mandate of the Statistical agency.

A. Nature of the approval process

16. As is the case in the RDCs, projects submitted to CDER undergo two types of reviews: a peer review that ensures that the projects meet a certain professional standard, and an institutional review that ensures that the projects fall under Statistics Canada’s mandate and do not interfere with Statistics Canada’s ability to maintain respondent trust or affect its reputation for objectivity or neutrality. Researchers associated with the projects also undergo an accreditation process that assesses the researcher’s ability to carry out the

proposed project. A project must pass both reviews and researchers associated with the project must be accredited for the project to be approved.

17. The peer review and accreditation process can proceed in one of two ways, but in both cases the review and the accreditation process is carried out by researchers in the academic community, thereby allowing Statistics Canada to make use of outside expertise to help it with the accreditation process. If the project has been reviewed and granted funds by an official research funding agency, such as the Social Science and Humanities Research Council, then Statistics Canada takes that as an indication that both the project and associated researchers meet the necessary professional standards. If the project has not been granted funds from an official granting agency, then the professional merit of the project and the qualifications of the researchers are evaluated by two members of a committee of external experts consisting mainly of academics appointed by Statistics Canada.

18. It is in the institutional review phase that CDER differs from the RDCs. The institutional review of projects is conducted by a committee of senior executives in the Agency in addition to the subject matter experts that review RDC proposals. The executive review considers the project proposal so as to assure the agency that the subject of the project meets Statistics Canada's mandate. It is also based upon supplementary material provided by the researchers when they submit their proposals. The aim of the review is to ensure that: the research will be used only for the purposes indicated in their proposal; that the researchers have completely revealed all their sources of funding and support; and that no other person or organization has an 'interest' in the project. To that end, researchers are asked to provide information on:

- (a) The sources of monetary or in-kind support they are receiving to carry out their project;
- (b) Their previous and intended work on the business population that they want to study in their current project;
- (c) Their access to other business micro-data at Statistics Canada or from other countries;
- (d) Their contact with other researchers that are working on similar or related micro-data at Statistics Canada;
- (e) Their affiliations with bodies that have administrative authority over the businesses whose data that they will access; and
- (f) Their affiliation with entities that would gain financially from obtaining information on businesses to which the researchers have access.

19. As argued earlier, the potential financial gains for a researcher to misuse business data represent a risk, so a careful review of the affiliations of the researchers using business micro-data is undertaken to manage that risk. Once the project and researcher have been approved, the researcher goes through the process to become a deemed employee under the *Statistics Act*.

B. Systems developed to prevent identification of individual business data

20. While the *Statistics Act* prohibits the revelation of data obtained from respondents, the implicit contract between the agency and its respondents depends on a broader set of understandings with regards to the information respondents provide. Information provided by a business could be used to harm it in ways other than by Statistics Canada inadvertently releasing and publishing individual numbers. The perusal of micro-data business data bases

by deemed employee researchers who have been approved for one project but who may also ingest that information for their own purposes is one of the risks for which a set of operating procedures have been designed.

21. As in case of social micro-data and the RDCs, all output produced in CDER must be vetted for confidentiality before it is allowed to leave CDER premises. Despite the fact that all output must be vetted and the fact that deemed employee researchers do not have printer privileges, confidential information can leave CDER in a number of ways which would require unacceptably invasive measures to prevent. Moreover, nothing can be done to prevent the removal of data in the memory banks of individuals.

22. In order to address this issue, a two-stage access system was developed for business data. In the first stage, researchers are given a shuffled database to use. The algorithm to produce this database moves data from one firm randomly to another firm (firm A could receive firm B's revenues and firm C's assets), while preserving to some degree the correlations and autocorrelations between the continuous variables, some accounting identities, and the marginal distribution of each variable. With this shuffled database, researchers can investigate the structure of the database and develop their computer programs.

23. Deemed employee researchers are then allowed to use the actual data, but only non-interactively. That is to say, they submit their programs to be run in batch mode. This prevents the viewing of the actual data by researchers, and allows for the researcher's use of the data to be monitored and recorded in a way that is more easily audited. Researchers also agree not to try to view the individual observations. They are also informed that their key strokes are being recorded and that audits will be performed on a random basis to ensure that inappropriate activities are not being conducted. Finally, researchers agree to immediately inform CDER if they have inadvertently accessed individual data points.

C. Institutional signatory on research contracts

24. In both the RDCs and CDER, researchers must take the Oath of office as outlined in Section 6(1) of the *Statistics Act*, which makes them subject to all the same conditions and penalties of the *Statistics Act* as other Statistics Canada employees. They also agree to abide by all security policies as outlined in the security manual of the agency. In addition to the penalties under the *Act*, regular Statistics Canada employees face the possibility of the loss of employment at Statistics Canada with no re-employment in the Public Service in the event of a confidentially breach. In comparison, researchers risk to be barred for life from the RDCs and CDER.

25. To strengthen the potential penalties and sanctions in case of a violation of the *Statistics Act* or another one of Statistics Canada's policies, researchers sign a research contract and for CDER researchers, the contract is also signed by the researcher's employer. While this contract does not oblige the employer of the researcher to discipline the researcher if he/she violates the terms and conditions in their contract, it does make the employer aware of the terms and conditions of data access by their employee and the seriousness of any violation. It places a moral obligation on the employer to deal with aberrant behaviour on the part of their employee should it occur. It thereby offers additional disincentives to engaging in non-approved activities for deemed employee researchers employed by institutions who value the opportunity to send researchers to Statistics Canada to make use of micro-data.

D. Future directions for access to business data

26. Risk management was a major consideration that determined the choices of CDER (its procedures and its location at head office) on the continuum of data access modes, but it was not the only consideration. Other considerations included the preparedness of the business data for research, the cost and complexity of developing a secure system that would enable access to data in locations across the country and the perception that Statistics Canada had to show its success in a limited mode before it opened up wider access. All this led Statistics Canada to situate CDER at head office in Ottawa. The necessity to have experts involved in setting up and supporting databases led to its creation adjacent to, and part of, the primary economic analysis research unit in the Agency—the Economic Analysis Division.

27. As the documentation of business micro-data reaches the standards set by the documentation of the social data and as possible partnerships with academics to fund the costs of expanding secure access develop, it is anticipated that modifications may be made to the location of CDER on the continuum of access. If CDER can demonstrate to Canadians that the policies, procedures and systems that have been put in place are working to manage the risk and maintain the trust of the business community, then moving business data access further up the continuum of access may be possible. A key tool for building trust will be to show that the research coming out of CDER can provide important information to Canadians and therefore should be seen to be valuable to the respondent community.

E. Conclusions

28. At Statistics Canada, risk management procedures have been implemented at different points along the continuum of access. It is clear that risk management strategies must be in place to ensure that the confidence and trust of respondents is maintained. The careful implementation of policies, rules and procedures can help NSOs manage the risk associated with access and allow for incremental expansion of such access to enhance the public good. Two recent developments in Canada show that it is possible for NSOs to move along the continuum from risk avoidance to risk management to make access easier for professional analysts who become deemed employees of Statistics Canada.
