

Directive on Microdata Linkage

1. Effective date

1.1 This directive takes effect on February 8, 2017.

1.2 It replaces the Directive on Record Linkage, dated August 31, 2011.

2. Application

2.1 This directive applies to all employees who undertake studies that involve microdata linkage, that is to say, the combining of two or more micro-records to form a composite record containing information about the same entity.

2.2 There are different types of microdata linkages, each with its own set of similar procedures.

- Type A – where the linkage falls under the scope of an omnibus microdata linkage authority
- Type B – where the linkage is an amendment of an existing linked file
- Type C – where the linkage is new; i.e., proposal for a new linked file (not falling within types A or B)

Detailed procedures for each of these situations are described in the Operational Steps for Obtaining Approval for a Microdata Linkage, including a more comprehensive definition of the types of linkages included in each type.

3. Legal Context

3.1 Section 3 of the *Statistics Act* gives Statistics Canada the mandate to collect, compile, analyse, abstract and publish statistical information relating to the commercial, industrial, financial, social, economic and general activities and condition of the people.

To fulfill this mandate, Statistics Canada undertakes microdata linkages, an important technique in the development, production, evaluation and analysis of statistical data to shed light on important societal issues. Although by its nature, microdata linkage is privacy intrusive, Statistics Canada undertakes microdata linkages in cases where the public good is clearly evident and outweighs the privacy intrusion. The privacy intrusion results from the fact that information for a particular individual is being put together in a manner that is generally unknown to the individual. As is the case with any microdata, in some situations, this could disadvantage the individual. The mitigating measure is that, by definition, statistical analysis is not focused on an individual, and only non-confidential statistical aggregates are produced.

4. Definitions

4.1 Definitions to be used in the interpretation of this directive are attached in Appendix A.

5. Directive statement

5.1 Objective

5.1.1 Ensure the effective management of microdata linkage activity conducted within Statistics Canada so that the operational and analytical benefits of microdata linkage support the mandate of Statistics Canada while, at the same time addressing and mitigating the inherent privacy intrusive nature of the activity.

5.2 Expected results

5.2.1 Information obtained from microdata linkage is used to produce statistical information that facilitates the better understanding of the Canadian society, economy and environment, and subsequent benefits are clearly in the public interest;

5.2.2 Confidentiality of information relating to individual persons, businesses or organizations (public or private, including public institutions and non-government organizations (NGOs) used in microdata linkages is strictly maintained and the results of the microdata linkage will not be used for purposes that can be detrimental to the persons, businesses or organizations whose information is involved;

5.2.3 Outputs of the microdata linkage will be released only in accordance with the confidentiality provisions of the *Statistics Act* and with any applicable requirements of the *Privacy Act*;

5.2.4 The microdata linkage offers demonstrable cost or respondent burden savings over other alternatives, or is the only feasible option to meet the project objectives;

5.2.5 The microdata linkage is judged not to jeopardize the future conduct of Statistics Canada's programs.

6. Requirements

6.1 Director, Information Management Division (IMD)

The Director, Information Management Division (IMD) has primary responsibility for the implementation of this directive. This will be accomplished by directing statistical and legal analysts in IMD, who are responsible for:

- 6.1.1** Advising program areas planning microdata linkage projects, as required, and providing assistance in all phases of the approval process;
- 6.1.2** Maintaining and managing an inventory of all approved microdata linkages (apart from those approved through an omnibus linkage authority);
- 6.1.3** Ensuring that summaries of new linkages (Type C) are posted on the Statistics Canada website and are removed when a project is completed and linked files have been destroyed. The website also contains general text to describe the types of microdata linkages conducted under an omnibus authority;
- 6.1.4** Including in the Annual Report to Parliament on the *Privacy Act* summaries of microdata linkages involving personal information, except for omnibus and amendments (types A and B linkages) and those linkages involving economic statistics;
- 6.1.5** Undertaking a review of this directive regularly, and revising it as required.

In addition,

- 6.1.6** The Director of IMD approves all Type B linkages.

6.2 Directors of statistical program areas

In directing managers of statistical programs in their division on projects involving microdata linkage, a director is responsible for:

- 6.2.1** Ensuring that no microdata linkage is undertaken without obtaining approval as required by this directive;
- 6.2.2** Ensuring that approval is obtained to use all files required in the linkage held by other divisions;
- 6.2.3** Determining the required approval process for a microdata linkage as described in the Operational Steps for Obtaining Approval for a Microdata Linkage, and adhering to all requirements;
- 6.2.4** If required, requesting the approval of the Chief Statistician to seek consent from respondents for a microdata linkage;
- 6.2.5** Ensuring that linked files, linkage keys and direct identifiers are retained in accordance to the *Directive on the Management of Statistical Microdata Files*, or retained only for the duration of the project.

6.2.6 Approving the use of an existing linked file or other files under their care and control, provided that certain conditions are met (see the Operational Steps for Obtaining Approval for a Microdata Linkage).

6.2.7 Removing, prior to providing the file to the Sponsoring Division, information about survey respondents that opted-out of record linkage.

6.2.8 Tracking the approvals, through corporate tracking systems or other methods, for the use of a file or an existing linked file under their care and control.

6.2.9 Participating in the review of requests by other Sponsoring Division directors for microdata linkages related to his/her divisional mandate.

6.2.10 Assuming the responsibilities of the various divisional roles (i.e., Contributing Division, Custodian Division, Linking Division, Sponsoring Division) as required. (Refer to Appendix A for definitions).

6.3 Director, Microdata Access Division and Director, Economic Analysis Division

The Directors of Microdata Access Division and Economic Analysis Division have responsibility for directing managers of the Research Data Centres (RDCs), the Federal Research Data Centres (FRDCs) and the Canadian Centre for Data Development and Economic Research (CDER), respectively to:

6.3.1 Assume the responsibilities of the Sponsoring Division for all projects that create or use linked files in the RDCs, FRDCs, and CDER.

6.3.2 Ensure that all microdata linkage projects and projects involving the use of already linked microdata files in the RDCs, the FRDCs and the CDER receive the required approval prior to the start of the project.

6.4 Assistant Chief Statisticians

As an advisor to the Chief Statistician, an Assistant Chief Statistician (ACS) is responsible for:

6.4.1 Approving the recommendation for new linkages (Type C) and;

6.4.2 Submitting the microdata linkage requests to the Chief Statistician and/or Executive Management Board for review and approval, when required, based on suggestion from IMD.

6.5 Chief Audit Executive

As responsible for the internal audit functions in Statistics Canada, the Chief Audit Executive:

6.5.1 Assures, on a regular and ongoing basis, compliance with this directive through risk-based compliance audits.

6.6 Chief Statistician

6.6.1 The Chief Statistician has the ultimate authority to approve or deny requests for microdata linkages, based on recommendations from the ACS responsible for the sponsoring division. Approval authority for omnibus and amendments (Types A and B) has been delegated as described in the Operational Steps for Obtaining Approval for a Microdata Linkage.

6.6.2 The Chief Statistician may permit other processes or uses.

7. Consequences

7.1 Statistics Canada has directives, including this one, and procedures which reflect basic principles found in the *Privacy Act*. Non-compliance would be counter to the letter and the spirit of the *Privacy Act* and to Statistics Canada's commitment to the protection and appropriate use of the personal information under its control.

8. References

8.1 Relevant legislation and regulations for this directive are as follows:

Statistics Act

Privacy Act

8.2 Related policy instruments and publications are as follows:

Policy on Privacy Protection (Treasury Board Secretariat (TBS))

Directive on Privacy Impact Assessment (TBS)

Directive on Privacy Practices (TBS)

Directive on Social Insurance Number (TBS)

Principles and Guidelines on Confidentiality Aspects of Data Integration Undertaken for Statistical or Related Research Purposes (United Nations)

Policy on Privacy and Confidentiality (Statistics Canada)

Policy on Information Management (Statistics Canada)

Policy on Official Release (Statistics Canada)

Directive on the Management of Statistical Microdata Files (Statistics Canada)

Directive on Obtaining Administrative Data under the Statistics Act (Statistics Canada)

Directive on Informing Survey Respondents (Statistics Canada)

Directive on the Security of Sensitive Statistical Information (Statistics Canada)

Generic Privacy Impact Assessment (Statistics Canada)

Operational Steps for Obtaining Approval for a Microdata Linkage

9. Enquiries

Please direct enquiries about this directive to the Director, Information Management Division.

Appendix A - Definitions

Definitions for terms used in this directive are included in this appendix.

A **micro-record** is defined as information about an identifiable entity.

An **entity** refers to an individual respondent or unit of observation, such as a person, family, household, dwelling, farm, company, business, establishment, institution, etc.

Microdata linkage is defined as the combining of two or more micro-records to form a composite record containing information about the same entity. The output of a microdata linkage must contain information that originated from more than one data file that were inputs to the microdata linkage activity.

Situations that are considered to be a microdata linkage:

- Matching of information on the same entity from two or more files. The input files may be survey files, census files or administrative files or a combination of these. The input files may relate to different time periods.

Situations that are not considered to be a microdata linkage:

- Subsetting of master files based on an external list. For example, a client wishes to analyze all persons who were diagnosed with a particular health condition in the past year. The list of all in-scope persons would be used only to identify all persons in a health survey who would be the subject of the analytical project. This is not a microdata linkage since the output file would not contain information from more than one input file.
- Imputation methods that identify a "similar" entity to use for imputation, but no attempt is made to actually match the same individuals.
- Linkages of person-level survey or administrative data, with files containing aggregate information at a geographical level. An example is when information about an identifiable person is matched to census summary information for that person's neighbourhood.
- Information contained on statistical registers, such as geography or industry code may be included on survey production files when they are to be used for analytical purposes. Subject to the requirements on direct identifiers, survey managers have the authority to include variables from survey frames without requesting approval for a microdata linkage.
- If the use of the already linked file falls within the approved purpose of the original request. For example, a linked data set approved and created to address a research question could be used for other research questions of similar nature or topic. It does require approval from the Custodian Division

and the Contributing Division for use of the file as per the Directive on Security of Sensitive Statistical Information.

Direct Identifiers are the variables used for linking operations to combine files, but which are not to be used for analytical purposes. They permit the immediate identification of an individual (person, business or organization), such as name and address as well as identifying numbers such as the Social Insurance Number, Provincial Health Number and Business Number.

Amendments of linked files is defined as situations where a currently-existing linked file is used for an already approved purpose but involves the addition of files to the original linkage.

Linked file environments are situations where the purpose of the linkage is to provide an analytical database for future analytical projects. A linked environment may be a linked file or it may simply be the creation of linkage keys that can be used to readily link files for individual projects in an efficient direct deterministic fashion.

Divisional Roles: There are different roles played within the context of a project involving microdata linkage. These roles are described below. It is possible for a single division to play more than one role within a specific project, and in fact this is often the case.

- **Contributing Division:** This is a division which has custody of one or more files that are to be used for a microdata linkage project. In some cases, the information comes from an external organization, in which case, a Statistics Canada division must be identified as the Contributing Division.
- **Custodian Division:** This is the division which has operational responsibility for the linked file, including: use, access, storage, retention and disposition. The director must ensure that all requirements of the Directive on the Security of Sensitive Statistical Information are met.
- **Linking Division:** This is the division which performs the microdata linkage and produces the linked file (or linkage keys).
- **Sponsoring Division:** This is the division which initiates the microdata linkage project. In addition to the requirements of this directive, the Director of a Sponsoring Division must ensure that all the requirements of the Policy on Peer and Institutional Review and the Directive on the Security of Sensitive Statistical Information are met.

Appendix B: Omnibus authority for linkages carried out within the Economic Statistics program

This appendix describes an omnibus microdata linkage authority for the Economic Statistics program (with certain exceptions as described below), and includes the procedures to confirm authorization for each specific linkage.

Microdata linkages within the Economic Statistics program can be for three purposes:

- Data production: Use already collected data for a statistical program rather than re-collect. It also includes frame creation, preparation of contact material, imputation for invalid or non-response.
- Analysis to support production: Analysis for purposes of data certification and data quality evaluation, such as evaluating trends in one dataset by examining the reports of the same businesses in another dataset.
- Analysis to provide information: Making use of combined datasets to support analysis that is not possible by a single existing dataset. In most cases, the outputs from these projects are officially released.

For a particular program, linkages can be one or more of the following:

- Survey data to administrative data for the same time period.
- Survey data for a specific time period to the same survey data from a different time period.
- Survey data for a specific time period to data from another survey for the same time period.
- Administrative file to another administrative file.

In this manner, both cross-sectional and longitudinal linkages are possible.

The following Economic Statistics programs are included (unless specifically excluded below):

- All programs in Statistics Canada's Economic Statistics Field
- All economic statistics and research programs in the Analytical Studies Branch, that do not include household or a person's information.
- Programs that provide supporting infrastructure to these programs: specifically Statistical Registers and Geography Division and Administrative Data Division
- Selected programs outside the Economic Statistics Field: specifically, the "Survey of Employment, Payrolls and Hours" and the "Pension Plans in Canada" program.
- All linkage activities of Agriculture Division.

All microdata files held within the programs listed above are in scope for the purposes of this omnibus linkage.

This omnibus authority includes all linkages that are specified as part of the survey collection process, provided that the intent of the linkage is to aid in the production of a survey master file. This would include, for example, using another data file to impute for item non-response. Linkages to other survey data files, to administrative files, or both would be possible (again provided that they are specified in the survey collection instruments), and would also include longitudinal linkages.

Internal testing, feasibility studies and data validation whose purpose is internal in nature only are also included under this omnibus authority.

The overriding considerations behind this authority are:

- No linkage should damage our relationship with the business respondents.
- No privacy-invasive linkages shall be carried out without a demonstrated public good.
- If the linkage involves an externally-specified group of businesses, no files are to be linked if the results might harm the interests of that group.

To operationalize these considerations, the following are excluded from the umbrella of this omnibus microdata linkage authority:

- Linkages that involve client-supplied lists of businesses;
- Linkages that present a significant risk of residual disclosure.

Linkages excluded under the above bullets must be brought forth for the Chief Statistician's consideration, on a case-by-case basis (see the Operational Steps for Obtaining Approval for a Microdata Linkage for procedures). The Director, IMD, may be consulted on specific projects to help determine whether the omnibus authority applies.

All linkages that are in scope are covered by this omnibus linkage authority, and are considered to be Type A linkages as described in the Operational Steps for Obtaining Approval for a Microdata Linkage.

Retention of linked files, keys and identifiers

As stated in the *Directive on the Management of Statistical Microdata Files*, the linked files are retained until no longer required. The linkage keys and direct identifiers are to be retained until the possibility of re-use for linkage purposes is considered low.

Reporting of microdata linkages covered by the omnibus authority

Linkages conducted under the umbrella of this omnibus authority are considered routine and critical to permit Statistics Canada to meet its mandate. They are also numerous and detailed. Therefore, no reporting or inventory of specific microdata linkages is required.

The Statistics Canada website has a notice that describes the linkage activities conducted in accordance with this directive.

Other considerations

Special care must be taken with respect to data sharing under Sections 11 and 12 of the *Statistics Act* and data disclosure under Section 17(2) of the *Statistics Act*, when it applies to administrative information. Legal requirements must be met. The Director, IMD can advise on specific circumstances.

Appendix C: Omnibus authority for linkages carried out within the Population and Household Statistics programs, and for privacy-sensitive Economic Statistics programs

This appendix describes an omnibus microdata linkage authority related to the Population and Household statistics programs, and includes the procedures to confirm authorization for each specific linkage. Generally, such linkage activities fall into two broad categories.

Category A: Linkages whose primary purpose is for internal use and is not to contribute directly to statistical outputs that are disseminated outside Statistics Canada

The purpose of linkages covered in this part is to do one of the following:

- Obtain information that benefits a survey, such as for stratification in survey design, but that does not directly contribute to estimates;
 - Providing supporting infrastructure to these programs: specifically frame creation with Statistical Registers and Geography Division and Administrative Data Division.
- Study and assess survey data quality, for example, by comparing survey data to data from other sources; and
- Aid in data collection, such as to provide addresses to mail introductory letters or to provide telephone numbers to reduce collection costs by permitting data collection through a telephone interview.
- Internal testing, feasibility studies and data validation whose purpose is internal in nature only.

The linkages can be from survey file to survey file, from survey file to administrative file, or from administrative file to administrative file and may be carried out, for example, to:

- Creation of survey frames or registers
- improve procedures for non-response adjustment, resulting in a reduction of estimation bias. (Census or administrative data are used to determine the characteristics of non-respondents and then adjust the weights of respondents accordingly);
- evaluate the coverage of a survey frame (to assess what types of units are not covered well);
- detect overlap in possible survey samples in an effort to minimize burden on respondents;
- obtain, augment or verify frame information;
- obtain contact information for sample units;
- compare administrative sources;
- evaluate the quality of an administrative file as a source of survey variables;

- assess disclosure risk.

Such linkage activities increase survey quality and in most cases, reduce survey costs and/or response burden. All linkages that are in scope for Category A are covered by this omnibus linkage authority, and are considered to be Type A linkages as described in Operational Steps for Obtaining Approval for a Microdata Linkage.

Category B: Linkages whose primary purpose is to contribute directly to statistical outputs that are disseminated outside Statistics Canada

The following types of linkages are approved:

- Linkage of supplementary surveys to the main survey: This is a situation where a second household survey is designed to use all or part of the sample of another household survey, and to reduce response burden by using content collected in the original survey. This second survey may be collected at the same time as the initial survey or at a later time, but the planned linkage for the second survey must be clearly indicated on the survey questionnaire. Examples of surveys with supplements are the Labour Force Survey (LFS) and the Canadian Community Health Survey (CCHS). Another example is the linkage of the Census of Population data to that for a postcensal survey.
- Longitudinal surveys: The linkage of survey information for the same individual but collected at different times. This applies to surveys that are designed to be longitudinal.
- Linkage of administrative data to survey data. This applies to situations where essential survey content is not collected as part of the direct data collection, but is planned to be taken from administrative files. An example is household surveys who do not ask questions on sources and amounts of personal income, but plan to link to income tax files to obtain the required income information.
- Longitudinal linkage of personal administrative information: This applies to the longitudinal linkage of the same administrative file over time. An example is the Longitudinal Administrative Database (LAD), which links personal income tax files over time. Linkages of different administrative files, either longitudinally or cross-sectionally, are not included.
- All linkages that the intent of the linkage is to aid in the production of a survey master file. This would include, for example, using another data file to impute for item non-response. Linkages to other survey data files, to administrative files, or both would be possible (again, provided that they are specified in the survey collection instruments), and would also include longitudinal linkages.

All linkages that are in scope for Category B are covered by this omnibus linkage authority, and are considered to be Type A linkages as described in the Operational Steps for Obtaining Approval for a Microdata Linkage. Out-of-scope linkages would be defined as another linkage type, and the procedures related to each type are described

in the Operational Steps for Obtaining Approval for a Microdata Linkage. The Director, IMD, may be consulted on specific projects to help determine whether the omnibus authority applies.

Information applicable to both Category A and Category B Linkages

Retention of linked files

As stated in the *Directive on the Management of Statistical Microdata Files*, the linked files are retained until no longer required. The linkage keys and direct identifiers are to be retained until the possibility of re-use for linkage purposes is considered low.

Reporting of microdata linkages covered by the omnibus authority

Linkages conducted under the umbrella of this omnibus authority are considered routine and critical to permit Statistics Canada to meet its mandate. They are also numerous and detailed. Therefore, no reporting or inventory of specific microdata linkages is required.

The Statistics Canada website has a notice that describes the linkage activities conducted in accordance with this directive.

Other considerations

Special care must be taken with respect to data sharing under Sections 11 and 12 of the *Statistics Act* and data disclosure under Section 17(2) of the *Statistics Act*, when it applies to administrative information. Legal requirements must be met. The Director, IMD can advise on specific circumstances.