

Underpinning the reputation of the CSO by maximising the value and utility of data

Central Statistics Office

May 2005

1. Introduction

The CSO is the principal compiler and publisher of official statistics in Ireland. It operates under the Statistics Act, 1993 and collects considerable amounts of information directly from people, households and businesses in both statutory and voluntary surveys. The Statistics Act, 1993 gives the CSO the authority to assess the statistical potential of the records maintained by other public authorities and to ensure that this potential is realised; therefore, the CSO has indirect access on a statutory basis to data on individuals and businesses collected for administrative purposes. The credibility and reputation of the CSO depends on maximising the value and utility of the data available to it, while at the same time protecting the confidentiality of the individual persons and businesses supplying the data.

2. Confidentiality and data protection

The information available (whether directly or indirectly) to the CSO is used by the Office for statistical purposes only. The manner in which the information should be protected is dealt with in the Statistics Act, 1993. Because the issue of confidentiality is of paramount importance to the work of the Office, a *CSO Code of Practice on Statistical Confidentiality* has been published¹. The Office is also fully aware of the need to respect and observe the Data Protection legislation.

3. Response burden

Collecting information places a response burden on both people and businesses. While a certain amount of official form-filling is part of the responsibility of good citizenship, the Office is conscious that it does involve costs for the business sector. It is important, therefore, that the CSO follows a data collection strategy: that minimises the burden and costs for business; and that maximises the utility for statistical purposes of any information collected by the CSO or any other public authority.

4. Maximising the value of data - Statistics Act, 1993 provisions

In the normal course of its business, the CSO publishes those tabulations and analyses which it considers most useful for the very broad user community. However, the CSO has neither the expertise nor the resources to meet directly all of the very detailed and useful research demands on statistics. There are two particular mechanisms under the Statistics Act, 1993 that the Office uses to help satisfy this demand.

Anonymised microdata files

The Statistics Act, 1993² allows the CSO to make available to researchers, for statistical purposes only, information obtained in any way under the Act *in such form that it cannot be directly or indirectly related to an identifiable person or business* subject to such conditions and restrictions as the Director General may determine. This provision is particularly suitable to support social research. It has recently been the practice, under this provision, to make anonymised microdata files available to

¹ See Annex 2.

² Section 34.

researchers through the Irish Social Science Data Archive (ISSDA). The anonymised files lodged with the ISSDA relate to the Quarterly National Household Survey, the Household Budget Survey, a 5% sample of the Census of Population records, and a 15% sample from the 2002 Census of Population of those at work to support analyses of *travel to work* patterns. Researchers have to abide by demanding protocols when working with these data.

Officer of Statistics

Under the Statistics Act, 1993³, the Director General may authorise a researcher to be an Officer of Statistics for a specified period to perform particular statistical analysis on microdata collected under the Act. This is the approach adopted for statistical research on anonymised business microdata since, even when working with anonymised data files, the characteristics of certain businesses make them susceptible to indirect identification. In these circumstances, the researcher is subject to the full rigour and penalties of the Statistics Act, 1993. In these limited instances, the research work has to be carried out within the CSO since the Office will not allow the anonymised microdata out of its premises, and proposed publications are checked by the CSO to ensure that confidentiality is not being breached.

5. Maximising the value of data - linking and integrating data holdings

The increasingly widespread use of personal and other unique identification numbers in the administration of government schemes, coupled with vastly improved computing power, has opened up the potential for enhanced statistical analysis through linking and integrating data holdings from various sources. Such data matching has the potential to reduce the response burden for both persons and business. However since data matching could be undertaken for other non-statistical reasons, some of which could infringe on the privacy of individuals, it is a practice that is closely monitored and regulated in the context of data protection regimes.

The use, including matching, of data from administrative sources for bona fide statistical purposes is acceptable under data protection legislation provided there are adequate safeguards in place to prevent the unlawful disclosure of information on individuals or enterprises. The statistical confidentiality provisions of the Statistics Act, 1993 and corresponding provisions under EU statistical law, provide a comprehensive legislative framework under which such safeguards are applied within the CSO. Since the Act also grants access for statistical purposes (subject to some specific exceptions) by the CSO to administrative data held by public authorities, the Office is uniquely placed to integrate data holdings from a wide range of sources.

The 2003 National Statistics Board report *Developing Irish Social and Equality Statistics to meet Policy Needs* made a compelling case for the more widespread use, with a particular focus on linking and integration, of data from administrative sources. In view of the sensitivity of the issue, the Government in accepting the report's recommendations asked the CSO to set out formally how the integration of administrative data holdings, and the new statistics generated by data integration, can be safely employed without data protection problems. The CSO was also requested to refer its approach to the Data Protection Commissioner for confirmation that the data protection rights of individuals were not undermined.

³ Section 20c.

As a response to the request from Government, the CSO publicly commits itself to the Protocol (or Code of Practice) set out in Annex 1. This Protocol has been discussed and agreed with the National Statistics Board, and approved by the Data Protection Commissioner.

Annex 1: CSO Protocol for linking and integrating data holdings for Statistical Purposes

1. Scope of the Protocol

This Protocol primarily covers any work undertaken within the CSO to match the individual records contained in two or more data holdings, at least one of which originates outside the Office (including data holdings that are a by-product of the administrative system). It also covers any assistance the CSO may give to other public authorities to enable them to link data holdings under their control for statistical purposes.

The linking and integration of data holdings covers both direct matching, i.e. where the linking is based on unique common identifiers (e.g. the PPS Number), and statistical matching that is based on common characteristics covered in the individual data holdings. In all cases the purpose of such linking is to produce a consolidated data file that can serve as the basis for more extensive statistical analysis.

The linking of data holdings derived solely from surveys undertaken directly by the CSO is not covered by this Protocol as their use for statistical purposes is made clear to the respondent at the time the data are collected and the protection of the confidentiality of the individual data is fully covered by the provisions of the Statistics Act, 1993.

2. Undertaking by the CSO

A. Data matching within the CSO

Within the scope of this protocol, the CSO undertakes that it shall engage in data matching through the linking and integration of data holdings held by the Office provided all of the following conditions are met:

1. All data matching activity complies in full with data protection legislation;
2. The data matching is undertaken solely for statistical purposes in line with the legal mandate of the Office as defined in the Statistics Act, 1993 namely: the collection, compilation, extraction and dissemination for statistical purposes of information relating to economic, social and general activities and conditions in the State;
3. The implementation of any proposal for matching has received the prior written approval of the Director General of the CSO before it is implemented; and
4. The provisions of the Statistics Act, 1993 regarding the protection of individual information are strictly adhered to at all stages of the data matching process and are applied in particular to protecting the individual data contained in the linked databases. The CSO has a general Code of Practice covering its procedures for protecting the confidentiality of all individual data under its control. This Code is reproduced in Annex 2 as an integral part of this Protocol insofar as relevant parts will apply to any data matching undertaken in the Office.

B. Data matching assistance the CSO may give to other public authorities

Within the scope of this protocol, the CSO will assist other public authorities in matching data holdings available to them provided all of the following conditions are met:

1. All data matching activity complies in full with data protection legislation;
2. The Director General of the CSO is satisfied that the matching is for statistical purposes only;
3. The public authority has adopted a protocol, which is acceptable to the Data Protection Commissioner, governing its treatment of individual data available to it for statistical purposes (including the linking and integrating of data holdings for statistical purposes); and
4. The Director General of the CSO is satisfied that sufficient safeguards are in place in the authority to ensure that individual information can be adequately protected from disclosure at all stages of the matching process and that the linked file is used for statistical purposes only.

3. Publicly accessible register

The CSO will maintain a register, which will be publicly accessible through its website (www.cso.ie), of all data matching activity undertaken in compliance with Sections 2.A and 2.B of this annex. The register will contain, inter alia, details on: the datasets that have been matched; the reasons why the matching was undertaken; and the statistical outputs that have been obtained.

Annex 2: CSO Code of Practice on Statistical Confidentiality

1. This Code of Practice applies to all information collected by the CSO whether directly in compulsory (statutory) and voluntary statistical inquiries or indirectly from the administrative records of other public authorities.
2. The information collected by the CSO is used only for statistical purposes.
3. The confidentiality of information relating to identifiable persons and undertakings is protected at all stages of statistical operations: collection, storage, processing, and dissemination.
4. The computer files that contain the results of surveys do not, in general, contain the names or addresses of respondents. Limited exceptions are made, however, when there are strong statistical reasons for doing so.
5. In the case of external trade statistics the CSO will, on the request of any importer or exporter, suppress the publication of information that reveals details of its business (the *passive* confidentiality approach).
6. Stringent precautions are actively taken in disseminating all other statistical results to ensure that particulars relating to identifiable persons or undertakings are neither directly nor inadvertently disclosed, and that any disclosure of individual information is made only with the written permission of the data provider concerned (the *active* confidentiality approach). For example, in the case of business statistics, published information is based on a minimum number of three respondents and then only when one or two enterprises do not have too dominant a share of the total.
7. Office and field staff appointed to the CSO sign a Declaration of Secrecy specified in the Statistics Act, 1993 and are fully instructed on their obligations to protect the confidentiality of any identifiable information to which they have access.
8. High security is maintained on the CSO computer network. Authorised access within the Office to computer files containing confidential data is strictly limited and controlled by a system of personal/section passwords that are regularly updated.
9. Information relating to identifiable persons or undertakings may only be provided to a non-CSO person or body in the limited cases specified in the Statistics Act, 1993.
10. Confidential information is transferred in anonymous form to the Statistical Office of the EU (Eurostat) for the compilation of aggregate Community statistics under the protective provisions of Council Regulations (EC) No. 1588/90 and No. 322/97. Under further European⁴ and Irish⁵ legislation, non-anonymised Balance of Payments related statistical microdata may be transmitted by the CSO to the European Central Bank.

⁴ Council Regulation (EC) No. 2533/98 and ECB Guideline (ECB 2004/15).

⁵ European Communities (Collection of Statistics) Regulations, 1999 (S.I. 177 of 1999).

11. The Director General requires Heads of Division to confirm annually the steps taken within their Divisions to ensure ongoing compliance with this Code of Practice.