



OECD Science, Technology and Industry Working Papers
1998/03

Institutional Arrangements
for Access to Confidential
Micro-Level Data in OECD
Countries

Kazuyuki Motohashi

<https://dx.doi.org/10.1787/041104314765>



PARIS

DSTI/DOC(98)3
Unclassified

English text only

Unclassified

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

DSTI/DOC(98)3

OLIS : 09-Jul-1998
Dist. : 13-Jul-1998

English text only

DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INDUSTRY

STI WORKING PAPERS

1998/3

**INSTITUTIONAL ARRANGEMENTS FOR ACCESS TO CONFIDENTIAL
MICRO-LEVEL DATA IN OECD COUNTRIES**

Kazuyuki Motohashi

67463

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

STI Working Paper Series

The Working Papers series of the OECD Directorate for Science, Technology and Industry of the OECD is designed to make available to a wider readership selected studies prepared by staff in the Directorate or by outside consultants working on OECD projects. The papers included in the series are of a technical and analytical nature and deal with issues of data, methodology and empirical analysis in the areas of work of DSTI. The Working Papers are generally available only in their original language – English or French – with a summary in the other.

Comment on the papers is invited, and should be sent to Directorate for Science, Technology and Industry, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

The opinions expressed in these papers are the sole responsibility of the author(s) and do not necessarily reflect those of the OECD or of the governments of its Member countries.

Copyright OECD, 1998

**Applications for permission to reproduce or translate all or part of this material should be made to:
Head of Publication Service, OECD, 2 rue André-Pascal, 75775 Paris, Cedex 16, France.**

INSTITUTIONAL ARRANGEMENTS FOR ACCESS TO CONFIDENTIAL MICRO-LEVEL DATA IN OECD COUNTRIES

by Kazuyuki Motohashi

This document, based on country surveys carried out by the OECD, compares the experiences of OECD countries in dealing with confidentiality constraints in an effort to meet the demand for policy-relevant analysis using micro-data. It shows that countries have made significant efforts to implement confidentiality rules and practices which balance increasing demands for microdata for policy analysis with the protection of confidential information on individual firms or establishments. Country-specific descriptions of data sources for micro datasets, the legal framework, and conditions for access to and use of data are provided for Australia, Canada, the Czech Republic, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Switzerland, the United Kingdom and the United States.

DISPOSITIONS INSTITUTIONNELLES RELATIVES A L'ACCÈS AUX MICRODONNÉES CONFIDENTIELLES DANS LES PAYS MEMBRES DE L'OCDE

par Kazuyuki Motohashi

Ce document présente une étude comparative des dispositions adoptées par les pays Membres pour faire face aux contraintes de confidentialité tout en répondant au besoin de microdonnées pour les analyses de l'action gouvernementale, sur la base d'une enquête par pays de l'OCDE. On y montre que chaque pays s'est efforcé de mettre en place des règles et pratiques de confidentialité qui répondent au besoin croissant de microdonnées pour les analyses de l'action gouvernementale, tout en assurant une protection adéquate des informations confidentielles des différents établissements ou entreprises. On y trouvera également des descriptions pays par pays des sources d'ensembles de microdonnées, du cadre juridique et des modalités de l'accès aux données et de leur utilisation pour l'Allemagne, l'Australie, le Canada, le Danemark, les États-Unis, la Finlande, la France, l'Italie, le Japon, la Norvège, les Pays-Bas, la République tchèque, le Royaume-Uni et la Suisse.

Introduction

1. Large sets of data at the firm, establishment or individual level are an important tool for policy analysis, and the majority of the OECD countries have developed and use such longitudinal micro-level datasets (LMDs) for research, particularly on policy-relevant issues. LMDs contain information that cannot be obtained from publicly available aggregated statistics. At the same time, the use of individual firm, establishment or personal data raises concerns about confidentiality.
2. LMDs based on administrative records, such as business registers, are confidential in nature, in the sense that such administrative data should only be used for a designated public purpose and individual information should not be disclosed. In most OECD countries, the use of establishment- or firm-level data from statistical surveys is strictly regulated. Such regulation is the counterpart to the obligation to respond, in that its absence might reduce response rates (McGuckin and Reznek, 1996).
3. This report compares the experiences of OECD countries in dealing with confidentiality constraints in an effort to meet the increasing demand for policy-relevant analysis using microdata. The difficulty here is the variety of approaches, as these are rooted in the institutions specific to each country, such as the legal framework for statistical surveys, the structure of governmental organisations and the enforcement of laws on privacy. To ascertain the institutional arrangements for access to confidential micro-level data in place in each country, a survey was conducted under the auspices of the OECD's Working Party on Industrial Statistics.
4. The survey revealed substantial cross-country differences in terms of conditions and procedures for access. It also found common factors that may help prepare for further and increasing demand for access to microdata.
5. This report is presented in two parts. The first offers a general discussion of institutional arrangements for access to microdata. The second presents a country-by-country description of data sources of LMDs, the legal framework, and access to and use of LMDs. The countries covered are Australia, Canada, the Czech Republic, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Switzerland, the United Kingdom and the United States.

PART I. DATA ACCESS ISSUES FOR LMDs

6. First, rules on confidentiality depend on the *source of the data* in the LMDs. In most countries, LMDs are based on business registers or on government administrative data. These data are confidential in nature, and data are handled confidentially within government. Statistical survey data are also used to construct LMDs or to enrich business register data, which usually have very few variables. Use of individual data from statistical surveys is often regulated by a law guaranteeing the confidentiality of the data reported. In the case of linked datasets of information from business registers and statistical surveys, the issue is more complicated. The regulations applicable to both datasets normally apply, but, in many cases, they are treated under the relevant law, since survey items also cover the business register variables. For LMDs based on statistical surveys, confidentiality may be negotiated with those surveyed.
7. The data unit, the period(s) covered and the information contained in the dataset may also affect access conditions. The data unit may be the firm, the establishment or even the individual worker. Public

firm-level data are published in annual reports, with the degree of secrecy increasing as the data units become more detailed. On the other hand, data anonymity can be more easily maintained for personal than for business data. Even when the name, address etc. of the firm are deleted from firm-level data, publicly available information on an individual firm, such as the number of employees and sales, often make it possible to identify the firm. For this reason, anonymous personal data from population censuses or labour force surveys have been published in many countries, while this is not the case for firm- or establishment-level data (Als, 1996). The period covered by the data is also a factor to be taken into consideration, as earlier data are less sensitive than more recent data. Finally, from the viewpoint of the firm surveyed, some variables, such as R&D expenditures, are more sensitive than others, such as total sales. Appendix Table 1 provides an overview of available LMDs in OECD countries.

8. The second issue is the *legal framework* for confidentiality. Statistical survey data are usually regulated under the Statistics Law, while public administrative records are generally protected by the relevant law, such as the United Kingdom's 1994 Value Added Tax Act for tax register data. As Appendix Table 1 shows, data from two data sources are combined in some countries, including Canada, Denmark, France and the United Kingdom. In this case, the level of confidentiality for both kinds of data should be the same. In Canada, for example, all data transferred to Statistics Canada are regulated under the Statistics Canada Act and are managed consistently, whether they originate from administrative records or from statistical surveys.

9. General confidentiality rules do not allow potential users free access to either statistical survey data or public administrative records. However, microdata may be used for policy analysis if individual data are not disclosed in the course of the analysis. In most OECD countries, LMDs are used for scientific research, but the majority of users are employees of national statistics offices or relevant governmental agencies, who have the right to access such data. The question of maintaining data confidentiality in cases of access by external users is a more difficult one.

10. Some countries, including Japan and the United Kingdom, have implemented explicit legislation regarding access by external users to confidential microdata. In Japan, the conditions and procedures for external use of statistical survey LMDs are stipulated under the Statistical Law. A government or related public organisation employee may be allowed to use microdata for public policy oriented research. In terms of procedures, a research proposal must be submitted to the government, which may give its approval for a specific research purpose and a limited time period.

11. In other countries, no specific legislation on external access to confidential data exists, and criteria for granting access rights and procedures are set by a data management body, usually a national statistics office. In terms of institutional settings, France has a unique system, the *Comité du secret statistique*, which meets three times a year to judge applications by potential users and grant access rights to confidential data (see Lang, 1996, for details). This method presents several advantages as it both maintains transparency and improves communication between data suppliers and users.

12. Finally, *conditions of access to and use of LMDs* are also important, because the legal framework generally leaves substantive matters to the discretion of the executing body, i.e. the national statistics office. In many countries, decisions on the following questions are made on a case-by-case basis by the national statistics office or other relevant institution.

- Who can use LMDs? Government employees, private academic researchers and/or business people?

- When can LMDs be used? A research programme initiated by the government, research proposed by the researcher?
- How is access to an LMD obtained, in terms of application procedures?
- To what extent is access allowed? To subsets of datasets only? Are there restrictions on access to specific variables or to information that allows identification of individual firms?
- What kinds of restriction are placed on the disclosure of research results?

13. Although there are substantial cross-country differences in policy regarding access to microdata, in all countries users are expected to comply with the following basic rules:

- external users cannot access datasets containing identifying information such as name and address of firm;
- external users should work under the supervision of statistics office employees;
- prior to publication, research results must be checked by the statistics office for compliance with confidentiality rules;
- external users must not disclose any information on confidential data during the course of the research project and must destroy all work files upon completion of the project.

14. For the last of these points, the US Census Bureau applies a special sworn status (SSS) system, whereby external users have the same obligations as Census Bureau employees and are subject to penalties in the case of violation. The intention is not only to maintain data confidentiality but also to promote external use of microdata (McGuckin and Reznick, 1996). The Center for Economic Studies (CES) has been set up in the Census Bureau to serve both as a gatekeeper for microdata through working closely with external researchers to improve data, as well as to check that their results do not violate confidentiality rules. The Central Statistics Office of the Netherlands plans to establish a similar institution, the Centre for Research on Economic Microdata.

15. The issue of whether users of micro-level datasets should work at designated sites or are allowed to work elsewhere is also important. The first solution is preferable for security reasons, but it is costly to maintain computer facilities and may be inconvenient for users. In countries such as Denmark, Italy and the United States, users work on computer terminals at the data management body's site. In contrast, Japan's statistics offices do not have such facilities and users work in a secure environment on their own.

16. Another issue involves the practical mechanisms for restraining direct use of confidential data by external users. Some countries have implemented ways to ease confidentiality constraints.

17. The first is *use of published micro-level datasets*. The sensibility of information on statistical units varies: for example, corporate firm-level data is often publicly available and LMDs simply provide efficient access to it. On the other hand, establishment-based data or data on individuals are more sensitive. Data sensitivity also depends on the type of variable. For example, variables such as R&D expenditures or profits may be considered more confidential than information on employment. Therefore, some parts of confidential micro-level datasets have been published in some countries. For example, Statistics Finland publishes firm-level data on certain variables from the Business Register, such as industry, number of employees and turnover (see Groot and Citteur, 1996, for details).

18. Second, *data by smart aggregation* may suffice for the researcher's purpose. Random aggregation of a small number of firms in each sub-industry or the creation of stratified samples by sector

and size make individual units unidentifiable and yet very little of the richness of a micro-level dataset is lost for analysis purposes. For example, aggregated microdata from the Community Innovation Survey in EU countries are made available by Eurostat.

19. Finally, external users may be provided with *indirect access*. Statistics office officials carry out tabulations or regression analyses on behalf of outside researchers. The statistics office of Lower Saxony in Germany provides this service for certain research projects, while Statistics Denmark maintains an electronic mailbox, which allows users to send their programmes by e-mail and receive results obtained by statistics office employees.

Conclusion

20. There is a non-negligible cost associated with maintaining a system of external access to confidential micro-level datasets. Therefore, whether they are made available to external users or kept exclusively for internal use depends on the perceived benefits of external use of microdata. The United States sees a number of benefits, including data improvement and research beyond what the resources of the statistics office allow (McGuckin and Nguyen, 1990). However, it should be noted that, although its first dataset dates back to the early 1980s, the Longitudinal Research Database (LRD) has only been recognised as a core asset of the US Census Bureau within the last two or three years (McGuckin and Reznek, 1996).

21. Possibilities for external access to microdata vary widely. Germany does not allow direct external access to microdata, while the United States has a separate institution to attract external users. If external demand is very small, indirect access, as in Lower Saxony in Germany, may suffice. However, further demand may be created by changing the management of confidential microdata. The examples of active research projects and policy analysis in the United States are likely to influence other countries' activities, and may lead to further deregulation of access to data.

22. At the same time, the potential risks of failing to maintain confidentiality owing to external access should not be underestimated. If research needs can be met through publicly available data or other aggregated data, access to confidential microdata should not be granted. Innovative cross-tabulations and smart aggregations are also important. Furthermore, international comparisons are obliged to rely on such methods as most countries prohibit cross-border transfer of confidential microdata.

REFERENCES

- ALS, G. (1996), "Statistical Confidentiality in the 15 Member States of the European Union, A Critical Comparison", paper presented at Third International Conference on Statistical Confidentiality, Slovenia.
- GROOT, A. and C.A.W. CITTEUR (1996), "Accessibility of Business Microdata", Research Paper No. 9630, Statistics Netherlands.
- LANG, G. (1996), "The Statistical Confidentiality and the French Committee of Statistical Confidentiality Concerning Enterprises", paper presented at Third International Conference on Statistical Confidentiality, Slovenia.
- McGUCKIN, R.H. and S.V. NGUYEN (1990), "Public Use Microdata: Disclosure and Usefulness", *Journal of Economic and Social Measurement* 16.
- McGUCKIN, R.H. and A.P. REZNEK (1996), "The Development and Use of Longitudinal Microdata: The US Census Bureau's Center for Economic Studies' Experience with Confidential Survey Microdata", mimeo.

PART II. COUNTRY NOTES

AUSTRALIA

Data source

Since 1994/95, the Australian Bureau of Statistics (ABS) has undertaken annual longitudinal surveys of private-sector, non-agricultural firms. The survey is a stratified random sample and includes firms of all sizes except those with no employees. The aim of the survey is to assess the factors that have the greatest impact on firm growth and performance over time.

Every year, the sample consists of about 6 000 firms (the first-year sample was larger). In addition, each year a sample of new firms is added to the panel, thereby making it possible to develop statistical estimates for cohorts of new firms as well as for the entire population of firms. The survey is funded by the Australian Department of Industry, Science and Tourism, and is expected to run for five years.

Legal framework

The 1905 Census and Statistics Act and the Statistics Determination issued under that Act enable, with the written approval of the Australian Statistician, the disclosure of unidentified unit record information where:

- all identifying information, such as names and addresses, has been removed;
- the information is disclosed in such a manner that the relevant person or organisation is not likely to be identifiable;
- the person to whom the information is being disclosed has given written assurance that no attempt will be made to identify particular persons or organisations, that the information will be used only for statistical purposes, that the information will not be disclosed to any other person or organisation, and that any other conditions that may be imposed by the Australian Statistician will be satisfied.

This is an enabling provision only, and the Australian Statistician is not obliged to disclose unidentified unit record information. Identifiable information that is not of a personal or domestic nature may also be released with the written consent of the supplier of that information.

Access to and use of LMDs

As of 1997, only one partial LMD of economic data, relating to a set of specifically selected manufacturing establishments for the period 1977/78 to 1992/93, has been released (with the consent of each of the establishments). Micro-level datasets of samples of persons included in the Population Census and in social surveys are released. Researchers and analysts in other government departments, academia, etc., are able to specify desired tables or analyses; ABS staff will undertake the necessary data extraction (for a fee) and inspect it for any potential confidentiality disclosure before release. LMDs can be constructed for several industries (such as agriculture, mining and manufacturing) if the costs of compilation are recovered.

Reference

“Evaluating the performance of SMEs: New statistical initiatives in Australia”,
DSTI/EAS/IND/WP9(96)16/23.

CANADA

Data source

Statistics Canada actively compiles and uses longitudinal micro-level datasets for analysis purposes. Its activities include:

- The Business Register;
- The Longitudinal Employment Analysis Programme (LEAP);
- The Linked Quinquennial Census of Agriculture;
- Industrial Organisation and Finance Micro Dataset;
- The Statistical Information System (SIS);
- The Census of Manufactures.

From these basic datasets, Statistics Canada creates a body of micro-level datasets. For example, statistical surveys, such as the Census of Manufactures, provides detailed information on establishment-level activities but does not cover small establishments, so that Business Register or LEAP data, which cover all firms, are used to supplement the data from the Census of Manufactures. In addition, SIS provides information that is important for linking firm-level data (from the Business Register and LEAP) with establishment-level data (from the Quinquennial Census of Agriculture and the Census of Manufactures). Furthermore, the Industrial Organisation and Finance Micro Dataset has consolidated basic financial statement information for firm group level activities.

Other regular or *ad hoc* surveys can be linked with these datasets. For example, surveys on advanced technology and innovation have been linked for use by research programmes at Statistics Canada.

Legal framework

The Statistics Canada Act regulates the confidentiality of microdata. The law applies not only to statistical survey data, but also to administrative data, such as business registers, which have been transmitted to Statistics Canada. In order to use its microdata, outside researchers have to be research partners with Statistics Canada, and the publication of results based on these data requires approval of the Chief Statistician of Statistics Canada.

Access to and use of LMDs

Since outside researchers cannot use confidential microdata independently, such studies are carried out inside Statistics Canada. Analytical topics include:

- entry and exit, gross job turnover;
- job creation and the role of SMEs;
- human capital development and innovation for SMEs;
- business strategy in innovative and non-innovative firms;
- advanced technology use, training and firm performance.

References

“The Collection and Use of Longitudinal Data”, DSTI/EAS/IND/WP9(95)24/04.

Analytical Studies Branch Research Paper Series, Statistics Canada.

CZECH REPUBLIC

Data source

The Czech Statistical Office (CzSO) currently has only one source of statistical microdata: statistical surveys of economic and social statistics. The Business Register, which provides mainly qualitative characteristics of enterprises, can also be considered as a source of microdata.

The use of administrative data is under negotiation with the various central administrative bodies. In 1995 data from the Social Security Agency were used to update the Business Register. The main effort has focused on the use of taxation data provided by the Ministry of Finance. The current negotiations should permit use of tax data in 1997 for the first time for statistical and business purposes.

Legal framework

The Act of the State Statistical Service (passed in April 1995) states that microdata entering CzSO or other bodies of the national statistics office via statistical surveys or administrative sources are confidential and may be used for statistical purposes only. Employees of the CzSO (or other bodies of the statistics office) using this data for statistical analysis and production of statistical information are sworn to secrecy.

The Statistical Act also provides for specific cases where microdata may be used for research purposes outside the statistical service on the condition that such data remain anonymous.

Access to and use of LMDs

In accordance with the terms of the Statistical Act, microdata collected and stored in the framework of statistical surveys have been used exclusively by employees of the CzSO. Standard disclosure rules for individual data apply to the publication and dissemination of the statistical data and results.

DENMARK

Data source

Statistics Denmark manages and links data from various administrative sources, including the Central Business Register for firms and establishments and the Salary Information Register for workers. In addition, the Central Population Register can be used for information on individuals not engaged in work.

Statistical survey data can be linked with these administrative data through unique identification numbers, and industrial variables, such as value added, employment, wages and capital, are available since 1984 from *Industrial Accounts Statistics*, an annual survey of firms employing 20 or more workers. Also since 1984, surveys on AMT (advanced manufacturing technology) have been conducted for selected industries; these provide information on firms' use of 16 types of IT-related technologies.

Legal framework

The Danish Public Authorities' Registers Act regulates microdata derived from administrative register data. According to this act, confidential microdata can be used for other statistical purposes, subject to approval by the Data Protection Agency.

Statistics Denmark has stricter rules for business data than for personal data, since business units are more easily identifiable. External access to confidential business data is restricted to supervised use of computer terminals at Statistics Denmark (in Copenhagen and in the regional office in Aarhus). For personal data, an electronic mailbox is available, allowing users to send a programme and receive results from Statistics Denmark. A similar system for business data is currently being planned.

Access to and use of LMDs

Administrative register data is extensively used for labour market studies. For example, the *Integreret Database for Arbejdsmarkedsforskning* (IDA – Integrated Database for Labour Market Research) is used to assess worker mobility and industry dynamics.

In addition, statistical survey data on AMT have been used to evaluate the use of advanced technology and firm-level productivity under a Ministry of Business and Industry research project.

Reference

Spieker, Finn (1996), "Access to Statistical Information at the Micro Level", in the conference volume of the Third International Seminar on Statistical Confidentiality, Slovenia.

FINLAND

Data source

Statistics Finland has compiled and used longitudinal micro-level datasets for various research projects analysing the Finnish economy. Three main datasets exist:

- *ANALTOP (datasets for analysis of manufacturing plants)*: This micro-level dataset at establishment level is based on the annual industrial production census, which covers all manufacturing establishments employing at least five workers. Industrial output and input variables, as well as industry-level deflators and plant-level capital stock estimates, are available. The period covered is 1975-94.
- *Firm-level R&D dataset*: This dataset is constructed from biannual statistical surveys of R&D and covers the period 1985-95. A 1991 innovation survey is linked to this dataset.
- *WEDB (longitudinal worker-establishment database)*: The data source is the administrative records of Statistics Finland. The database covers workers, establishments and enterprises. It covers all workers in the labour force employed by an establishment or an enterprise for at least one year on the basis of information from regional employment statistics. The variables at the establishment and enterprise level are obtained from the Business Register.

Legal framework

Chapter 5 of the Statistics Act regulates the confidentiality and release of data. Under certain conditions, Statistics Finland may release limited data (such as data on a person's age and sex), but most data, particularly for establishments and enterprises, can only be used for scientific research.

Access to and use of LMDs

Due to the strict regulations contained in the Statistics Act, only Statistics Finland employees can work on confidential microdata. Outside researchers wishing to work on these data have to be attached to Statistics Finland or have to obtain special permission. They must comply with the rules set by Statistics Finland and their work is supervised.

In practice, all studies using Finish confidential microdata are conducted by Statistics Finland. The following are examples of research topics:

- dynamics of plant-level productivity, by entry and exit;
- education and the innovative capability of firms;
- R&D and productivity;
- enterprise demography and job flows.

References

The Statistics Act.

S. Laaksonen, ed. (1997), *The Evolution of Firms and Industries*, Research Report 223, Statistics Finland.

FRANCE

Data source

France has two major statistical sources which serve as bases for LMDs: its business register SIRENE and SUSE (*Système unifié de statistiques d'entreprises*), which integrates balancesheet data from the tax authorities (*Direction générale des impôts*) with data from the annual survey of firms (*Enquête annuelle d'entreprises*). Legislation imposes the use of each firm's SIRENE identification code in all administrative acts. In addition, the code is used in various statistical surveys, and thus provides a unique method for linking business registration data, other administrative sources and survey results.

There are five major LMDs: *i*) BRIDGE, an establishment-based dataset designed to respond to the statistical needs of regional planning and policies; *ii*) BALSUS, based on SUSE, is a comprehensive set of enterprise data covering the period 1984-93; *iii*) LIFI (*Liaisons financières*), is a dataset permitting the identification and monitoring of the development of groups, i.e. enterprises with financial links, which currently covers the years 1981-90; *iv*) SINE (*Système d'information sur les nouvelles entreprises*), a dataset under development designed to improve the tools for enterprise demography; and *v*) the MSE sample (*L'échantillon marché et stratégie d'entreprises*), a representative sample of about 13 000 firms for analytical purposes. In addition to these five major LMDs, a number of supplementary LMDs exist, in particular with technology-related variables.

Legal framework

For government employees, access to confidential data is generally permitted within their own administration under a general prohibition on disclosing confidential data. Access by outside researchers is governed by the Comité du secret statistique, whose president is a member of the Conseil d'État (Council of State) and whose 11 members come from the statistics office, other parts of the administration, and the business community. Potential users of LMDs submit to the committee for approval a research proposal detailing the purpose of the research, the type of data needed, and the means of securing confidential data. Between 1984 and 1994, 324 out of 360 requests received a favourable response. However, this high rate of acceptance is partly due to extensive prior filtering of requests by the committee secretariat.

Access to and use of LMDs

In a substantial number of cases, LMDs have been used by government officials, especially from INSEE. In addition, outside researchers have obtained permission from the *Comité du secret statistique* to use micro-level data for research purposes. Eurostat has also submitted four requests for transmission of confidential data. Normal disclosure rules for individual data apply to individual data, and research results can be published.

Reference

“Les bases de données micro-économiques longitudinales sur les entreprises à l'INSEE”, DSTI/EAS/IND/WP9(95)24/08.

GERMANY

Data source

To date, the German federal statistics office has not carried out enterprise-level surveys, and it would appear that no panels are currently planned. German experience with micro-level firm data is limited to an academic research project to develop and exploit LMDs in co-operation with the statistics office of the federal state of Lower Saxony. The statistics office created two panels with data from 1978 to 1994 by linking surveys of manufacturing establishments via identification numbers. One covers all manufacturing establishments active in at least one year since 1978. It was created by matching information on all local manufacturing firms with at least 20 employees (or belonging to an enterprise of that size) and information from the annual census of small manufacturing establishments to cover the rest of the industry. Variables include employment, sales, industry and regional codes. The second covers only local units with at least 20 employees (or belonging to an enterprise of that size), but provides more detailed information, such as value added, exports and investment.

Legal framework

The Federal Statistics Law regulates the overall statistical activities of the government and provides rules for the disclosure of confidential data. The law stipulates that individual personal data provided for federal statistics must not be disclosed by the specially sworn employees who are entrusted with federal statistics. Exceptions must be explicitly authorised by a provision of the Federal Statistics Law or the law initiating the relevant survey.

According to the Federal Statistics Law, individual data rendered anonymous in such a way that it is absolutely impossible to identify the respondent or the person concerned may be released, provided the respondent gives written permission for transmission or publication, or if the data have been combined with individual data of other respondents in such a way that they can no longer be linked to any individual.

Efforts have been made to supply certain data to scientific and other agencies for their own work, provided that the data is fully protected. However, as it has become increasingly easy to retrieve individual identities, this goal is difficult to attain. Therefore, the Federal Statistics Law provides that data may be disclosed to scientific or to other institutions entrusted with independent scientific research in Germany if attributing the data to individual respondents would require an unreasonable amount of time, expense, and effort (*de facto* anonymity).

Access to and use of LMDs

Absolute and *de facto* anonymity are of course very difficult to ensure for industry statistics. Therefore, in co-operation with the statisticians of the statistics office, researchers have developed methods for using LMDs so that the rules of confidentiality are maintained but researchers obtain their results. These methods are used, for example, in the above-mentioned research project in Lower Saxony. Researchers are provided with a random sample of 100 firms from both datasets. The sample contains no information on industry or region so that the firms cannot be identified. The samples are used to test programmes written by researchers. In a second stage, researchers bring their programmes on diskettes to the statistics office where they are copied onto a specially designed local PC. Employees of the statistics office run the programmes and check the output for disclosure of confidential information. Researchers have to provide a detailed description of the use they intend to make of the data. If no violation of data protection rules is

found, the output is delivered to the researchers. Various kinds of analysis have already been conducted with these datasets, with publicly available results.

References

Law on Statistics for Federal Purposes (Federal Statistics Law – FSL) of 22 January 1987, [*Bundesgesetzblatt* (Federal Law Gazette)] 1987, Part I, pp. 462, 565, last modified pursuant to the Law of 17 January 1996, *Bundesgesetzblatt* 1996, Part I, p. 34.

“Development and Use of Establishment-level Longitudinal Datasets from Official Statistics: A Report on the Case of Lower Saxony”, DSTI/EAS/IND/WP9(95)24/01.

Wagner, Joachim (1996), “Firm Size, Firm Age and Job Duration”, *Review of Industrial Organization* 11, pp. 201-210.

ITALY

Data source

ISTAT (*Istituto Nazionale di Statistica*) has compiled annual longitudinal microdata from the structural survey of industry and services from 1989-92, which covered all firms in the manufacturing and service sectors employing 20 or more workers. This dataset covers most fundamental variables for business activities, such as turnover, employment and capital investment. Other survey data can be linked with this core dataset, such as data from the Italian innovation surveys conducted in 1989 and 1991, using methodology and definitions common to other European countries. Other relevant surveys are the annual R&D survey and the firm-level occupation mix survey. Technically, these survey data can be linked at firm level via firm codes, but as of 1996 this had not been done.

Legal framework

There is a general legal obligation to ensure confidentiality of survey data, but the actual management of micro-level data for research purposes is based on ISTAT policy. ISTAT's policy on confidentiality is that if an individual included in anonymous data can be identified through reference to published data, such data cannot be disclosed. In the case of firm-level data, since data are publicly available via financial statements a firm can be identified even from anonymous data. As a result, firm-level microdata is strictly regulated, while anonymous household data is published. However, use of microdata for research is possible through *ad hoc* arrangements with ISTAT. Administrative data are subject to different regulations, but they are not currently used for policy analysis.

Access to and use of microdata

Analysis of microdata is mostly carried out by ISTAT researchers. Contract researchers can only access data at ISTAT's headquarters in Rome and at its Milan office. One of ISTAT's current analytical focuses links panel data on firm performance with innovation survey data to study the relationship between technology and economic performance. Owing to Italy's persistently high unemployment rate, the implications of technology for employment are an important analytical topic.

References

Conti economici delle imprese con 20 addetti ed oltre, anno 1991, 1994, ISTAT.

Politi, Mauro and Piero Taccini (1996), "Concentration, Productivity and Technological Innovation of the Manufacturing Enterprises in Italy", paper presented at the International Conference on Comparative Analysis of Enterprise Data, Helsinki.

JAPAN

Data source

The sole LMD in Japan is maintained by the Ministry of International Trade and Industry (MITI). It is the result of a pilot study of the Manufacturing Census, compiled at establishment level and covering the years 1970-92. The main variables are output, employment, capital and material inputs. A major constraint of this dataset is that it covers only existing establishments with at least 30 employees; no surveys on entry and exit were carried out. More generally, business registers or other administrative records are difficult to use for setting up LMDs in Japan, due both to strict regulation of their use and to difficulties involved in inter-ministerial co-ordination of research activities.

Legal framework

The use of individual census survey data in Japan is regulated by the Statistics Law. Under this law, detailed administrative procedures are defined in a regulation (“Ordinance concerning the use of individual data for outside purposes under the designated survey”). Since the statistical survey is carried out to derive *aggregate* figures, compilation and analysis of LMDs is considered an “outside purpose” and thus requires special approval to comply with confidentiality rules. These rules set out an application procedure for the use of LMDs. For approval, the following main conditions must be met:

- the data user should be a government official or an employee of a public research institute;
- the purpose of the analysis should be public policy oriented and not for private profitable activities;
- approval is granted for a limited period of time and for specified data.

The last condition makes continuous maintenance of general purpose LMDs, such as SUSE in France or LRD in the United States, impossible. Thus, LMDs based on the Manufacturing Census are provisional datasets, and data must be discarded once the expiry date is reached.

Access to and use of LMDs

The range of potential data users is very narrowly defined, and it is almost impossible for persons from private research institutes, including private universities, to gain access to microdata. Also, since the Japanese system only allows researchers to use microdata for a specific research purpose, compilation and use of LMDs occurs concurrently. MITI’s LMD based on the Manufacturing Census is currently being used in an ongoing project to compare the microeconomic structures of France, Japan and the United States. The Japanese part of this study must be conducted by a government employee of MITI, and data coverage, period of use and conditions for the disclosure of research results must be clearly set out in the application form and approved. A few other cases of LMD use exist, but all have involved employees of public organisations, such as local government or academics of a public university.

Reference

Motohashi, Kazuyuki (1997), “Japanese Experience with Longitudinal Datasets of the Manufacturing Census”, in S. Laaksonen (ed.), *The Evolution of Firms and Industries*, Statistics Finland Research Report No. 223, Helsinki.

NETHERLANDS

Data source

Statistics Netherlands has compiled a number of longitudinal micro-level datasets from statistical surveys. The core datasets are:

- **Production Survey (PS):** Firm-level data from annual production surveys are used. All firms with 20 or more employees (10 or more employees until 1987) are covered, and basic output and input variables are available.
- **R&D survey (RD):** Quadrennial R&D surveys were conducted in 1985, 1989 and 1993. A detailed breakdown of R&D expenditure (by type of expense, such as labour cost, input purchases, etc., in-house or outsourced R&D) is available, as are numbers of scientists in full-time equivalent.
- **Survey on Manufacturing Technology (SMT):** Quadrennial survey of use of manufacturing technologies such as CAD/CAM.
- **Wage Survey (WS):** Annual survey on the structure of workers' earnings in firms. In 1979, 1985 and 1989, firms were asked about their workers' level of education.

Legal framework

Data collected by Statistics Netherlands under the Act of 1936 (Statistics Act) must not be published in a way that permits the disclosure of information on individuals, enterprises or institutions. There are no explicit procedures for working with confidential microdata for other than statistical purposes.

Access to and use of LMDs

At present, work using confidential micro-level datasets should be carried out in collaboration with an employee of Statistics Netherlands and on its premises. Statistics Netherlands is working towards establishing a Center for Research on Economic Microdata which will provide on-site facilities for external researchers under stringent rules. The following are some examples of Statistics Netherlands' analytical projects:

- energy demand;
- the relation between firm size and firm growth;
- the dynamics of employment;
- productivity and downsizing;
- productivity and new technology.

Reference

Netherlands Official Statistics, Special Issue, Micro-analysis of Firm Data, Vol. 11, Autumn 1996.

NORWAY

Data source

The main LMD available at Statistics Norway is based on the annual statistics on industrial structure and the annual statistics of accounts. LMDs also exist for the construction and service industries.

Statistics on industrial structures and production are available from 1966. The survey covers establishments with more than ten employees (until 1992, five employees). In addition, information on employment, turnover and industrial sector is available for smaller firms from administrative registers. Values for the other variables are estimated for these establishments. The statistics of accounts cover extraction, quarrying, manufacturing, wholesale and retail sales. The LMD is available from 1977 (from 1967 for certain variables).

Legal framework

The Statistics Act of 16 June 1989 (No. 54) regulates the collection of information for statistical purposes. Further details concerning the role and responsibilities of Statistics Norway are provided in the regulations concerning the implementation and supplementation of the Statistics Act (No. 105), laid down by the Ministry of Finance on 13 February 1990. Statistics Norway carries out its surveys within this framework. The Statistics Act also authorises the collection of data for statistical purposes from nation-wide registers, including information collected in accordance with any prescribed obligation to provide information. An independent governmental body, the Norwegian Data Inspectorate, licenses all data registers containing confidential information, and monitors the protection and use of all such registers and data systems. Statistics Norway has a general concession for registering data for statistical purposes, granted by the Data Inspectorate on 27 April 1990.

Access to and use of LMDs

LMDs are available to researchers at Statistics Norway and other research institutes. For access to LMDs, a request containing an extensive description of the research project must be submitted to the Data Inspectorate. If the request is approved, the approval, together with the project description and an estimated termination date, is submitted to the appropriate division of Statistics Norway. The request is then evaluated by Statistics Norway. Upon approval, data vital to the project will be made available to the researcher. All data is anonymous and confidential data is deleted.

If both the Data Inspectorate and Statistics Norway approve, identifiable and confidential information can be released. This is only possible in special cases where such information is indispensable to the project. Researchers and all personnel in contact with the data are sworn to secrecy and sign a research contract stating the conditions for handling, protecting and destroying the information. At the agreed termination date of the project, destruction of the data must be confirmed in writing to Statistics Norway. Tables included in reports, articles, etc., are reviewed by Statistics Norway before publication to ensure that confidentiality is not violated.

References

Håndbok i datasikkerhet og fysisk sikring, Statistisk sentralbyrås håndbøker, Statistics Norway, 1994.

The Statistics Act of 16 June 1989, No. 54.

SWITZERLAND

Data source

The main data source for firm-level micro datasets at the Swiss Federal Statistics Office (SFSO) is “value-added statistics”, an annual survey of all large firms and of sampled smaller firms (the threshold varies between 20 and 200 employees, depending on the industry). Smaller firms are not included in the sample for more than five years. The first accounting year for which “value-added statistics” are available is 1990. Information on individual firms is maintained as a separate database. Continuity of data for individual firms cannot be guaranteed because the value-added surveys are not compulsory.

Legal framework

The use of individual firm-level data in Office of Federal Statistics is regulated under the Law of Federal Statistics of 9 October 1992, the Decree concerning the Realization of Federal Statistical Surveys of 30 June 1993 and the Law of Data Protection of 19 June 1992.

Access to and use of LMDs

Under the above laws, external access to microdata is possible under the following conditions.

- data are used exclusively for statistical purposes;
- data do not contain any information element allowing identification of individual firms;
- the user does not disclose any information in confidential data and keeps it in a secure place;
- the user returns and destroys data once the analysis is terminated.

In order to access microdata, users submit an application with a detailed description of how the data are to be used. The SFSO then accepts or rejects the proposal; once approval has been granted, the conditions for data access are stipulated in a contract. Access to personal data is managed in a restrictive way.

The following are examples of analytical projects:

- Market dynamics, innovation and performance for industrial activities in Switzerland;
- Financial arrangement of SMEs.

UNITED KINGDOM

Data source

The UK Office of National Statistics (ONS) manages statistical survey data, as well as the business register for all UK businesses, IDBR (Inter-Departmental Business Register). Since every ONS statistical survey uses the same IDBR codes, micro-level data from any survey, such as the annual production survey and the R&D survey, can be linked. Since IDBR is based on information on VAT and PAYE (a firm's employee pension code), the firm is the core unit of analysis. However, it is also possible to identify establishments and firm groups from other data sources. The variables available from IDBR are industrial classification, number of employees and turnover. Availability of other variables depends on the methods used in individual statistical surveys. For example, the annual production census covers input and output production variables in greater detail, but it is impossible to construct panel data for SMEs, since surveys of smaller firms are conducted on the basis of rotating samples.

Legal framework

Various laws [e.g. Statistics of Trade Act 1947, Statistics of Trade and Employment (NI) Order 1988] govern the use of statistical survey data and ensure the confidentiality of individual data. In addition, laws for administrative records, such as the Value Added Tax Act 1994, provide rules for using these data. The Deregulation and Contract Out Law of 1995 opened up the possibility for researchers from outside the statistics office to work with individual data in ONS. Procedures for application to use microdata are still *ad hoc*. ONS may grant approval, based on the purpose and duration of the research project. Data users must not disclose any confidential data, and results will be checked by ONS prior to their publication.

Access to and use of LMDs

External access to LMDs is very rare, and they are little used even within ONS, since IDBR was only constructed recently. One example of analysis based on IDBR is a study of firm demography and the economic presence of SMEs, a project conducted along the lines of Eurostat's project on firm demography, and substantial summary statistics on SMEs have been compiled.

References

The Inter-Departmental Business Register, Office of National Statistics, 109.23-109.29.

Small and Medium-sized Enterprises: How Many and How Important?, Office of National Statistics, 109.17-109.22.

UNITED STATES

Data source

The most widely used LMD in the United States is the Longitudinal Research Database (LRD). An unbalanced panel covering the years 1972-92 at the establishment level, it has been compiled from the Manufacturing Census and other related statistical surveys by the Bureau of the Census. This database covers production variables such as output, employment, capital and material inputs. Though establishment-based, data can be aggregated to firm-level information via a firm identifier and can be linked with other firm-level statistical surveys, such as the R&D survey of the National Science Foundation and the inward foreign direct investment data compiled by the Bureau of Economic Analysis. Links are also possible at the level of establishments, for example through the Survey of Manufacturing Technology (SMT). Most of the datasets around LRD are based on statistical surveys, supplemented by information from administrative records for small plants.

Legal framework

The US Statistics Law generally states that microdata from the US Bureau of the Census survey are confidential and may only be used for statistical purposes at a secure site at the Bureau of the Census and by specially sworn employees of the US Bureau of the Census. However, outside researchers can also gain access to the LRD family under clearly specified circumstances. In particular, a potential outside user of LRD must obtain special sworn status by taking a legal oath not to disclose confidential data. The Center of Economic Studies (CES) at the Bureau of the Census is in charge of selecting projects submitted by potential users. The project has to be beneficial to the data collection work of the Bureau of the Census, e.g. improved survey concepts or questionnaires, better survey processing procedures, etc. For the most part, researchers who obtain access rights from CES pay the cost of their projects in the form of laboratory fees.

Actual use of LMDs

Various projects and analyses based on the LRD family have been carried out. Once a researcher's project proposal is accepted, CES employees often work closely with the researcher to manipulate the data and to point out data issues that may bear on the research objectives. In addition, CES employees protect data confidentiality by performing "disclosure analysis" on any results researchers wish to publish. Researchers can obtain access to all authorised data by CES at the CES terminal in the Washington DC suburbs, at the regional centre in Boston, and at Carnegie-Mellon University in Pittsburgh.

References

- McGuckin, Robert (1995), *Longitudinal Microdata Development and Use: The US Census Bureau's Experience*, Room document presented at the 1995 meeting of the OECD Working Party on Industrial Statistics.
- McGuckin, Robert and Sang V. Nguyen (1990), "Public Use Microdata: Disclosure and Usefulness", *Journal of Economic and Social Measurement*, Vol. 16.

Appendix Table 1. Overview of LMDs of industry and technology in OECD countries

	Data	Coverage	Sample Characteristics	Industry Classification	Sales	Value Added	Employment	Wage	Capital/GFCF	Material	Technology	Others
Canada												
Census of Manufacturing (CM)	Establishment	1970-89, annually	All manufacturing establishments	SIC, 4-digit Level	Yes	Yes, partly	Yes	Yes	Yes	Yes	No	
Survey of Advanced Technology	Establishment	1989, 1993	Sample of 4,000 establishment	-	-	-	-	-	-	-	Use of 22 kinds of technology	
Denmark												
Integrated Database for Arbejdsmarkedsforskning (IDA)	Establishment	1980-90, annually	All establishment	NACE 70	Yes	Yes	Yes	Yes	Yes	Yes	No	
Finland												
Longitudinal Data File of Industrial Establishments (DIE)	Establishment	1974-93, annually	Manufacturing w/ 5+ employees	Finnish SIC (63 groups)	Yes	Yes	Yes	Yes	Yes, Partly	No	No	Export, foreign Ownership
R&D, Financial and Industrial Statistics	Firm	1987-93, annually	Sample of 70 Firms	-	-	-	-	-	-	-	R&D Expenditure	
France												
Systeme Unifie de Statistiques d'Entreprises (SUSE)	Firm	1978-92, annually	All private firms	NAP600, 4-digit level	Yes	Yes	Yes	Yes	Yes	Yes	No	
Recherche	Firm	1974-92, annually	All R&D firms	-	-	-	-	-	-	-	R&D expenditure, # of Scientists	
Innovation Survey	Firm	1991	Sample of firms	-	-	-	-	-	-	-	Innovation counts, etc.	
Germany												
Manufacturing Census in Lower Saxony	Establishment	1978-94, annually	All manufacturing establishments	-	Yes	Yes, Partly	Yes	Yes, Partly	Yes, Partly	No	No	Export
IAB Establishment Panel	Establishment	1993-95, annually	Sample of 4,000 firms	-	Yes	Yes	Yes	Yes	No	No	Technology level in 5 point scale	Compensation plan variables
Stifterverband, R&D Panel	Firm	1977-89, annually	Sample of 800 firms	2-3 digit German System(20groups)	Yes	Yes	Yes	Yes	Yes	Yes	R&D Expenditure	

Appendix Table 1. Overview of LMDs of industry and technology in OECD countries (cont'd.)

	Data Unit	Years of Coverage	Sample Characteristics	Industry Classification	Sales	Value Added	Employment	Wage	Capital/GFCF	Material	Technology	Others
Italy												
Nuova Archivio Imprese (NAI)	Firm	1985-92, annually	Manuf.(20+emp) Service(10+emp)	NAP70, 3-digit code	Yes	Yes	Yes	Yes	Yes	Yes	No	
Innovation Survey	Firm	1992	Sample of 7 000 firms	-	-	-	-	-	-	-	R&D expenditure, innovation	
Japan												
Manufacturing Census of Japan	Establishment	1970-92, annually	Manufacturing. (30+employees)	4-digit, Japanese system	Yes	Yes	Yes	Yes	Yes	Yes	No	
Japanese R&D Panel	Firm	1987,89,91	Manufacturing. (100+employees)	3-digit, Japanese system	Yes	Yes	Yes	Yes	Yes	Yes	R&D expenditure, # of scientists	Export, Foreign Ownership
Netherlands												
Annual Production Survey (APS)	Firm	1979-93, annually	Manufacturing (20+employees)	3-digit, Dutch system	Yes	Yes	Yes	Yes	Yes	Yes	No	
R&D Survey	Firm	1985,89,93	Sample of 300 firms	-	-	-	-	-	-	-	R&D expenditure, # of scientists	
Survey of Manufacturing Technology	Firm	1985, 89	Sample of firms	-	-	-	-	-	-	-	Use of technology	
Norway												
Annual Census of Manufacturing (ACM)	Establishment	1983-92, annually	Manufacturing (2+ employees)	2-digit Norway system	Yes	Yes	Yes	Yes	Yes	Yes	No	
R&D Survey	Line of business of Firm	1983,84,85,87,89,91	Manufacturing (20+ employees)	2-digit Norway system	-	-	-	-	-	-	R&D expenditure	
Sweden												
Central Register of Enterprises and Establishments (CFAR)	Establishment/ Firm	1985-93, annually	All industries	3-digit Swedish system	Yes	Yes	Yes	Yes	Yes	Yes	No	

Appendix Table 1. Overview of LMDs of industry and technology in OECD countries (cont'd.)

	Data Unit	Coverage	Sample Characteristics	Industry Classification	Sales	Value Added	Employment	Wage	Capital/GFCF	Material	Technology	Others
United Kingdom												
VAT Register	Firm	1973-present	All private firms	NACE70, 2-digit code	Yes	No	Yes, partly	No	No	No	No	
United States												
Longitudinal Research Database (LRD)	Establishment	1963,67,72-92 annually	Manufacturing(all for census years)	4-digit SIC code	Yes	Yes	Yes	Yes	Yes	Yes	No	
NSF R&D Survey	Firm	1972-92, annually	Sample of 2 000 firms	-	-	-	-	-	-	-	R&D expenditure, # of scientists	
Survey of Manufacturing Technology (SMT)	Establishment	1988, 93	Sample of 8 000 establishments	-	-	-	-	-	-	-	Use of 17 kinds of technology	
Eurostat												
European Enterprise Panel	Firm	1987-93, annually	Sample, method depends on country	Nace 3-digit code	Yes	Yes	Yes	Yes	No	No	No	Export/import, partly

STI WORKING PAPERS

1996

1. Embodied Technology Diffusion: An Empirical Analysis for 10 OECD Countries
George Papaconstantinou, Norihisa Sakurai and Andrew Wyckoff
2. The Impact of R&D and Technology Diffusion on Productivity Growth: Evidence for 10 OECD Countries in the 1970s and 1980s
Norihisa Sakurai, Evangelos Ioannidis and George Papaconstantinou
3. Short-term Indicators: Using Qualitative Indicators to Update Production Indices
Paul Schreyer and Corinne Emery
4. SMEs and Employment Creation: Overview of Selected Quantitative Studies in OECD Member Countries
Paul Schreyer
5. Globalisation and Competitiveness: Relevant Indicators
Thomas Hatzichronoglou
6. Factors Influencing the Steel Work Force: 1990 to 1995
Donald F. Barnett
7. Measuring R&D in the Services
Alison Young
8. The Evolution of Skills in OECD Countries and the Role of Technology
A. Colecchia and G. Papaconstantinou

1997

1. Bibliometric Indicators and Analysis of Research Systems: Methods and Examples
(disponible en français sous le titre "Indicateurs bibliométriques et analyse des systèmes de recherche : méthodes et exemples")
Yoshiko Okubo
2. Revision of the High-technology Sector and Product Classification
(disponible en français sous le titre "Révision des classifications des secteurs et des produits de haute technologie")
Thomas Hatzichronoglou

.../...

1998

1. Relative Trade-weighted Unit Labour Costs by Industry
Karine Lepron and Paul Schreyer
2. Information and Communication Technology and the Measurement of Real Output, Final Demand and Productivity
Paul Schreyer
3. Institutional Arrangements for Access to Confidential Micro-level Data in OECD Countries
Kazuyuki Motohashi