Smeeding, Timothy M.

Working Paper

Procuring Microdata Files for the LIS Project Databank: Progress and Promise

LIS Working Paper Series, No. 250

Provided in Cooperation with:
Luxembourg Income Study (LIS)

Suggested Citation: Smeeding, Timothy M. (2001) : Procuring Microdata Files for the LIS Project Databank: Progress and Promise, LIS Working Paper Series, No. 250, Luxembourg Income Study (LIS), Luxembourg

This Version is available at:
https://hdl.handle.net/10419/160922

Luxembourg Income Study Working Paper No. 250

PROCURING MICRODATA FILES FOR THE LIS PROJECT DATABANK: PROGESS AND PROMISE

Timothy M. Smeeding

February 2001

Procuring Microdata Files for the LIS Project Databank: Progress and Promise

Timothy M. Smeeding
Overall Project Director
Luxembourg Income Study
and
Director
Center for Policy Research
Maxwell School of Citizenship and Public Affairs
Syracuse University
426 Eggers Hall
Syracuse, NY 13244-1020
(315) 443-9042 | fax (315) 443-1081
tmsmeed@maxwell.syr.edu

Prepared for the "Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI)"

February 2001

Key Words: Luxemburg Income Study (LIS); harmonization, comparability, remote access, infrastructure, household income data

*The author wishes to thank the German government and the LIS country sponsors for their support in producing this paper. Esther Gray, Mary Santy, and Kati Foley provided excellent assistance in preparing the manuscript. The paper has also benefited from the comments of David Jesuit and Paul Alkemade. However, the author assumes responsibility for any and all remaining errors of omission and commission.

Abstract

The Luxembourg Income Study (LIS) project is one of the oldest and best known examples of crossnational social science infrastructure. Some 25 nations and 20 sponsors team together to provide internet accessible, privacy-protected, household income microdata to over 400 users in 30 nations. The project is financed by annual contributions by 16 nations' National Science Foundations and/or National Statistical Offices. One of the most crucial pieces of the LIS structure is the source and type of data that it offers to its users. This paper describes these data, how they are obtained, harmonized, and made available. It presents a critical discussion of where the project is today and where and how international data collection efforts can improve upon both the quality of income data and its dissemination to qualified researchers.

I. Introduction

The Luxembourg Income Study (LIS) research and databank project has provided harmonized cross-national household income microdata for social science research for over 15 years. These data provide the basis for cross-national comparative research projects by providing access to household income microdata for all research users who are connected to the internet, who promise to respect the privacy of survey respondents, and who promise to make use of the LIS microdata for research purposes only.

The purpose of this paper is to describe the types of data used by LIS and the issues involved with obtaining, harmonizing, and making the data available to users. We begin with a description of LIS and the types of data it employs. We then turn to a more in-depth discussion of data type, data quality, and cases of "typical" data from a subset of nations. Finally, we discuss additional cases in which microdata have not yet been obtained, and dilemmas regarding privacy protection for data that have been made available to LIS. We close with a brief view of future LIS plans. The objective here is to give the nonuser a brief overview of the data sources used by LIS and the way that they are harmonized, deployed, and accessed in a time-tested privacy-protected manner by over 400 users in 30 nations, 24 hours a day, 365 days a year.

II. The Luxemburg Income Study: A Brief Overview

The Luxembourg Income Study (LIS) project began in 1983 under the joint sponsorship of the government of Luxembourg and the Center for Population, Poverty, and Policy Studies (CEPS) in Luxembourg. The LIS project has five goals:

- to *harmonize* cross-national data (through relieving researchers of this task) and by building an expert staff to accomplish this task and to handle user questions and user services;

- to test the feasibility of *creating* a database consisting of social and economic household survey microdata from different countries:
- to provide a method of allowing researchers to *access* these data under various privacy restrictions required by the countries providing the data;
- to create a system that will allow research requests to be *quickly processed* and the responses returned to users at remote locations; and
- to *promote comparative research* on the economic and social status of populations in different countries, through training and networking activities.

CEPS, which later became CEPS/INSTEAD (International Networks for Studies in Technology, Environment, Alternatives, Development, or INSTEAD) was important because its leader, Gaston Schaber, understood the importance of cross-national data and cross-national comparative research studies for the tiny country of Luxembourg, which otherwise would attract very little international scholarly attention. He saw the production of harmonized (comparative) cross-national data as a key step forward in the social sciences. Thus, he funded the beginning of the LIS project, with the author as project director, with Lee Rainwater at Harvard University as research director, and with John Coder of the U.S. Census Bureau as technical director. Also important to LIS was the Luxembourgish government, whose data protection laws are very strict, thus allowing guaranteed privacy protection for household microdata to national statistical offices who had no experience with making data available outside of their national boundaries.

From the beginning, the LIS project was supported by a key group of academics and social statisticians who were valuable because of their intellectual capital and their ability to make datasets and technical expertise available to LIS.[1] LIS produced its first research outputs in 1985, and became fully available to researchers in 1987. In terms of general access, therefore, it is about 14 to 15 years old. It stands as one of the few truly cross-national and comparable

data infrastructures extant (OECD 2000), and thus provides a useful model for other similar projects.

Financing a Public Good

The project is now funded on a continuing basis by the national science foundations and social science research foundations of its member countries, by the ingenuity of its directors, and by the Luxembourg government through CEPS/INSTEAD, a government agency within which LIS is located. The original grant from the government of Luxembourg in 1983 was followed by a "bridge grant" from the Ford Foundation, allowing LIS the time to find sponsoring partners in most of its member nations. By 2000, 16 nations help support the LIS core via 24 different funders, each of whom pay between $10,000 and $30,000 per year (see Table 1 for a list of these sponsors). The list is a broad one: central statistical offices, national science foundations, social science foundations, and national research institutes. These nations cover, in an average year, over 90 percent of the core costs of the LIS database. Various foundations in the United States (e.g., Ford Foundation, National Institutes of Health, National Science Foundation, Russell Sage Foundation, MacArthur Foundation) and foreign sponsors (European Community) have helped us fill in the funds to meet project demands that exceeded national contributions. For instance, the Ford Foundation covers the cost of maintaining the United States office of LIS at Syracuse University. LIS pays its own expenses except for office rent, which is covered by CEPS/INSTEAD. Because of some previous shortfalls in national contributions, the directors of LIS spend about one-third of their time on the LIS project working on research projects which will ultimately fund the LIS core budget, and which will add new dimensions to the database.

The goals of LIS have in effect made it a "public good." Once the data are harmonized, they are made available to qualified academic users at zero marginal monetary cost. Moreover, LIS holds summer workshops and other seminars aimed explicitly at increasing the base of users,

especially among junior scholars. Over 500 scholars have attended these workshops, many of whom have become long term LIS users. All of these students have been treated to an intellectually rich and academically strenuous opportunity to learn to use the data.

The problem with public goods, however, is that they can be used without helping covering their fixed costs. Within nations, public goods are funded by national bodies and national research funds. But across nations there are few, if any, organizations with the scope or interest to fund a microdata infrastructure (OECD 2000).

As with all real public goods, LIS is therefore, almost continually underfunded. Major international research organizations (e.g., Office of Economic Cooperation and Development (OECD), International Labor Office (ILO), United Nations (UN), World Bank, United Nations Children's Fund (UNICEF), and dozens of others) have followed economic logic and have basically become free riders, making use of LIS data and user services at zero marginal cost. Occasionally, these organizations will hire LIS to provide custom-made analyses which provide some support for LIS. More often, they will hire one or more of the 400 currently active LIS users at marginal cost to carry out these analyses. The result is that LIS has continually been faced with "projectizing" the core…adding "data costs" to research projects so that the core could be maintained, updated, and improved.

Yet, country contributions have been amazingly steady. In one country (France), the sponsoring organizations which send checks to LIS were disbanded and thus we have had to find new support. In some nations there are lapses in sponsorship (e.g., Spain, Austria). In all other nations, support has not only continued but has been increased when asked. Most of the sponsors listed in Table 1 have been LIS contributors for over ten years. However, raising the core country membership price will, at some point, reduce membership. Thus, we have not until very recently been able to totally fund LIS from the core alone.

Should country contributions and core staff costs continue at current levels, we will for the first time be able to fund the core from country revenues alone. In this situation, we can turn our attention directly to improving the core: data quality, user access and user services, substituting these activities for the time intensive work of projecting the core.

Data Harmonization

The most important goal for LIS, we believe, is data harmonization. Data availability is an important issue that is slowly being overcome (see sections III and IV below). But the access and availability of 3, 4, or more national income surveys with no idea of how sources or definitions of "income" are arrived at does not permit comparability. Harmonization of data—reshaping and reclassifying components of income or definitions of household structure into comparable categories—is the real value of LIS. It allows the researcher to address important social issues without having to invest countless hours getting every variable that will be analyzed into a comparable format.

Because of data restrictions and privacy concerns of many governments, LIS must keep the data in one location where it can be accessible yet "protected" against misuse. The LIS micro datasets are therefore accessed globally at zero direct cost to their user using electronic mail. More general release of LIS data to national archives is difficult due to differential national interests in data protection for clients and governments (e.g., Japan, Sweden, Finland, others); sale of national data to recover costs (e.g., Canada, Australia, the United Kingdom, others), and other complicated political prerogatives (e.g., the European Community Household Panel Dataset (ECHP)), all of which are described more fully below in sections III and IV. Despite these issues, national or international statistical bodies which would like to make data available but also protect privacy and confidentiality ought to consider LIS or similar

organizations as a method of providing access to their data at reasonable cost and with no risk of violating the confidentiality and privacy of survey respondents.

Countries Covered and Access

Since its beginning, the LIS experiment has grown into a cooperative research project with a membership that includes countries in Europe, North America, the Far East, and Australia. Our countries are largely covered by the OECD, G-7, and in the European Community broadly defined. The database now contains information for almost 30 countries for one or more years of data. Negotiations are underway to add data from New Zealand, Korea, Japan, South Africa, and other countries. The LIS data bank includes more than 100 datasets covering the period of 1968 to1997. During 2001 additional surveys are being added to more fully represent the period of the middle 1990s for most of the nations, and in 2002 we will begin a new "millennium" round of datasets for 2000. A list of countries and years for which data are available is attached (Table 2).

A new operating system for our remote access network was implemented in 1998, and already improved again in 2000. It is much more flexible and faster than the old system, and it provides additional data access options including SAS, SPSS, and STATA software packages. Extensive documentation concerning technical aspects of the survey data and the social institutions of income provision in member countries is also available to users via electronic mail. In 1999 we began to provide direct web-access to "mesodata" and "metadata" in the form of comparable output on income distribution, poverty, and related issues. Inexperienced users can now download these tabular data in comparable form. In future years LIS will add a new "web tabulator" system that allows inexperienced users the ability to obtain summary data by only entering a few key words into a worldwide web-based system which will generate these tabulations directly.

II. Data Details and Case Studies

As seen in Table 3, there are numerous types of data to which LIS has access. Each nation's data is almost a story in and of itself. The various nations follow very different policies with respect to data access, data quality, and data availability. Types of survey data available are listed in Table 3, where we present data by type (3A) and by one measure of dataset quality (3B).

Survey Types and Data Quality

Perhaps the most important issue of comparability lies with the relative quality and consistency of LIS datasets themselves. The types of survey data used by the LIS are not uniform in nature, purpose, or objective. The lowest common denominator the LIS requires is the existence of a substantial level of detail concerning income sources and income totals. The surveys themselves are quite diverse, as illustrated in Table 3A. Some surveys are designed first and foremost to collect income data; others are derived from income tax records; and still others come from special supplements to labor force surveys. Some LIS datasets are based on income questions taken from expenditure surveys (as in the case of the United Kingdom); others are separate waves of longitudinal household panel data from a scientific university or research center based data collection (e.g., Germany, Russia); and still others are taken, at least in part, directly from government administrative data. In many nations, several different types of data are available, allowing LIS to choose the "best" survey for comparability reasons (see Atkinson, Rainwater, and Smeeding 1995).

The second part of Table 3B presents a reasonable way to envision how these differences are likely to affect the quality of income data. Five conceptual levels of income reporting are suggested and the approximate level at which each LIS country dataset lies. Income reporting in the upper rows is considered more complete than in lower rows. This is not intended to be more than indicative. There are studies of income distribution which lie between rows 1 and 2, such as

those combining administrative data with survey information, as in the case of Sweden and Finland (see below) or the Blue Book series in the United Kingdom, which combines survey and administrative data.

Up the rows from bottom to top, Table 3B begins with the amount of income actually reported by the population, excluding entire non-interviews but leaving partial or "item" non-response intact (row 5), as in the case in the Dutch, German, and Swiss surveys. The Dutch and Swiss make extensive imputations for some types of income (e.g., social security, child allowances). The German Socioeconomic Panel Data, however, leaves item non-response as missing values, allowing the user to make further imputations for non-reporting of income items. This income data is perhaps more congruous with the next level, which is edited income (row 4) whereby all item non-responses are corrected. These adjustments may take many forms, including "hot-deck" imputation (e.g., the United States Census Bureau technique), where there is imputation of a value taken from the most recent (on the tape) respondent with the same characteristics as the non-respondent, or "cold deck" imputation, whereby the imputation is taken from a matrix which computes the average value of all respondents and assigns the average value to the non-respondent. (See Atkinson, Rainwater, and Smeeding 1995, Appendix 4 for more detail on this topic.)

Row 3 refers to the amount of income recorded in data taken from tax records. Norwegian and French data are at this level. Table 3B suggests that incomes for tax purposes are more reliably reported than survey incomes, which may be true for some but not all countries. Tax-based surveys may also suffer from omissions of certain types of non-taxable income or non-taxpayers, in addition to tax evasion and tax avoidance. Row 2 raises gross incomes to the total amount recorded by some administrative intermediary, based on totals drawn from national income accounts or administrative records of government agencies. Swedish data, for example,

are mainly drawn from such records. Differences between the top row, "true income," and the administrative amounts usually arise from amounts of income which in principle are recorded in the national accounts, but are not readily allocated to individual households. This largely includes the underground, informal, or "shadow" economy as well as fiduciary accounts such as pension funds. These differences in data quality can manifest themselves as differences in the amount and type of income data collected, an issue on which we can briefly comment.

Similarities and differences in the quality of reported income amounts are important in survey measurement. What can be learned about the overall quality of income data from comparisons with national accounts and other external sources is an important question for the LIS; but one for which there is no firm answer. Three points should be made before comparing reported income amounts from surveys and administrative sources. First, national income accounts or administrative data may not always be superior to survey data in some countries. National accounts aggregates are themselves estimates whose reliability is the subject of much literature. Self-employment income, for example, is poorly reported and differs according to the accounting convention employed by the data tabulator. In the case of property income, which is derived as a residual in National Accounts, estimates may be very suspect.

Second, administrative data need adjusting to produce estimates for comparable income concepts and populations before comparing it to survey data (or tax data). For example, national accounts may include households together with non-profit organizations. It may be necessary to subtract the interest income received by charities, or income received by households not in the survey population (e.g., non-residents, the deceased, and the institutionalized), or payments to institutions.

As put by the U.S. Bureau of the Census,

> Deriving independent estimates of aggregate income for purposes of evaluating the survey data is difficult. The survey and administrative sources use different

> definitions, cover different universes, and are based on concepts that are not exactly the same. Therefore, adjustments to the administrative sources must be made to help correct for these inconsistencies and arrive at a valid independent estimate that can be used to make fair and accurate estimates of the quality of the survey estimates (…). [In the United States, these adjustments] attempt to 1) remove income that is received by the institutional population, the deceased, and persons not residing in the United States at the time of interview, 2) remove any components of income that are received as "in-kind" payments or benefits, and 3) remove any lump-sum or one-time payments, withdrawals, etc. (U.S. Department of Commerce 1991, p. 215).

Third, it is important when comparing income amounts to bear in mind that differences between income aggregates may arise from different sources: varying non-response to the survey (for example, a low response rate from high income groups may cause understated investment income); item non-response by households taking part; or inaccurate reporting by respondents. If reported wages and salaries are, say 95 percent, of the comparable aggregate, this does not mean that all individuals reported 95 percent of their true wages and salaries. This is an average based on some individuals who have over-reported or under-reported their incomes. Multiplying reported amounts by the reciprocal of the percentage reported is not the appropriate way to make an adjustment for under-reporting. A direct record-for-record comparison is needed for further information here. Under-recording may appear as failure to report in income source, but it may be indistinguishable form genuine zero entries, creating another type of dilemma. Overall ratings of data quality do not therefore provide all of the ingredients necessary to adjust microdata for reporting errors. Simple "grossing up" will therefore not improve the accuracy of income reporting, even if it produces a higher (but not a better) reported income amount.

Most of the datasets in LIS conform to a reported amount that is overall 85 to 90 percent of the comparable aggregate among the dozen nations who have made these calculations (e.g., see Atkinson, Rainwater, and Smeeding 1995, Table 3.7). Wage and salary income tends to be

reported with 95 percent or above accuracy. Self-employment income and income from property (interest rents and dividends) are less well-reported. Income transfers fall somewhere in between. However, until we are able to "exactly" match reported incomes with administrative records for the same persons and units (e.g., Radner 1983) we are unable to thoroughly assess data quality.

The bottom line is that all survey income have some error. The degree of error which is tolerable depends on the purpose to which the data will be used. As reported in Gottschalk and Smeeding (2000), the importance of data quality depends on the ratio of the signal (accurate data) to noise (or spurious data). LIS can improve the ratio of signal to noise by making data more comparable;. it cannot improve the quality of the data themselves. Others, e.g., the Canberra Group (see below), can improve data quality directly and are therefore of great interest to LIS.

LIS Criteria for Data Selection

Several considerations go into deciding which survey is "best" for LIS purposes:

- **data quality**. The overriding criteria for inclusion in LIS is that this is the highest quality and most consistent and reliable national dataset for measuring annual household income and its components.

- **income detail**. The more detail on an income survey, the better the estimate of income. In particular, surveys explicitly designed to measure "income" do a better job.

- **national staff support**. Every LIS dataset has one or more national country coordinators, who help with technical documentation, harmonization of data, and with user support that goes beyond the knowledge of the LIS team.

- **periodicity**. In general we now try to have data for most nations on a four to five year period rotating basis. We cannot include every year's data for every nation due to cost. On the other hand if a nation has only one or two years of "good" data, we will include these years even if they do not closely match to other nations. In general, LIS seeks to "space" datasets first, and second, to find a "given" year, e.g., 1995 or 2000. Even if all datasets were for the same year, different business conditions will produce different cyclical outcomes across datasets.

- **time consistency**. LIS pays a great deal of attention to intraperiod or cross-sectional consistency of data. We seek the best dataset for each period. Time trend analyses of income inequality when datasets change, or when datasets are substituted for one another are not recommended. For instance, while time trends in inequality from LIS normally track those found in any given nation, one should also compare these to the time trend data produced in each country itself (see Atkinson, Brandolini, Smeeding, and van der Laan 2000).

These criteria have been applied to each nation's data supplied to LIS. How they have been balanced in each country differs, but can be inferred from the specific country discussions that follow shortly below.

Costs of Data and Basic "LISification" Procedures

It is important to note that LIS makes it a point **not** to pay for the datasets that it obtains from its users. In fact, LIS has *never* paid for a dataset beyond the marginal cost of shipping the data. Datasets are either freely donated by the nation in question, or they are paid for by the LIS domestic financial sponsor. In return, LIS provides each nation with the value added by the comparability it provides and by the return of a "LISified" data file to the data originator. Once collected, we feel that data should, in fact, be a public good and thus we urge our member countries to treat it in the same way that LIS treats it.

The data harmonization, or "LISification" process involves several steps. First, LIS is usually concerned with a limited set of the total number of variables on a dataset. The basic LIS household variable list and demographic variable list is included in Table 4. From this list are derived basic subaggregations of household income according to the LIS definitions (e.g., see Atkinson, Rainwater, and Smeeding 1995; Smeeding and Weinberg 2001). We also have individual person records as well as household records, but the variable list is basically what is provided in Table 4, section A. The aggregates in section B of Table 4 are derived from adding and subtracting the list of variables in 4A. In addition, LIS collects a set of consistent classifiers

or "demographic" data shown in Table 4, section C. For more on how these are combined to produce the aggregates and for analytic purposes, the reader should consult the publications cited in the references, or the LIS website (<www.lis.ceps.lu>).

Once a dataset has been identified as acceptable, LIS asks the country to send their "full" data file to us, with completed documentation and other information. The LIS staff will then make the LISification itself, standardize the documentation, and return the LIS estimated and harmonized dataset to the originator so that it might be further checked by the data owner, and further adjusted for inconsistencies. Often nations will add income top codes or suppress geographic detail for privacy reasons before allowing LIS to make their data available to researchers. We request permission to keep a copy of the basic unharmonized file so that any errors later uncovered by users can be corrected by LIS staff. If this is not possible, we return the original dataset to the owner. Once the data owner has signed off, and once we have received the required documentation, the dataset is made electronically available to users, using the LIS remote access system (Coder 2000).

Original Data Sources

Beyond these criteria and basic data harmonization issues, almost every LIS nation has its own history and idiosyncrasies. To provide a richer picture of these nations and their data, we have selected a set of nations for additional comment and description. These countries include a set of nations where LIS data comes from a statistical office; data from scientific non-governmental sources; and data from nations which make heavy use of administrative data.

Statistical Office Data. Many government bodies and particularly Central Statistical Offices (CSO's) conduct income surveys which are used by LIS (Table 3). Two countries with ongoing important surveys which are of high quality and which are made available to the LIS are Statistics Canada's, *Survey of Consumer Finance (SCF)*, and the United

States' *Current Population Survey*'s (CPS) annual March Income Supplement, which is collected, cleaned, and edited by the U.S. Bureau of the Census. These surveys have many strengths and a few weaknesses for LIS purposes. The main strength is that both nations have decades of experience with "public use," microdata files. Hence, well documented codebooks, very complete and user friendly documentation, and helpful staff were already in place before LIS. Moreover, each of these nations have already cleaned and edited their data for non-reporting. In the end, they were easy to arrange into LIS income categories. The weaknesses are suppression of some United States geographic identifiers from smaller United States states, and top and bottom coding of income variables in both surveys. These top and bottom codes artificially reduce the highest and lowest incomes to some fixed amount. Such "codes" may independently bias Lorenz curve based income inequality measures and "depth" of poverty measures.[2] Otherwise the files are clean, neat and relatively unrestricted.

Australia and the United Kingdom also have strong traditions of "public use" datafiles. Australia's *Income and Housing Survey* (AHS) is similar to the Canadian SCF. In Britain, the Family Expenditure Survey (FES) includes detailed periodic subannual income measures (monthly and weekly incomes, and some periodic subannual income measures for capital income) and was, until recently, the best source of household income for the United Kingdom. While the AHS is relatively easily "LISified," the United Kingdom FES has proven to be more difficult to work with, simply because of the many ways in which annual income can be constructed from subannual reports.

Both datasets also present additional hurdles for LIS. In Australia, the raw microdata could only be exported under very restrictive conditions. LIS has, of course, satisfied those conditions with the help of the Australian Bureau of Statistics (ABS) and with the support of Peter Saunders, Director of the Social Policy Research Centre in Sydney. And these constraints

are very real. For instance, to date, LIS has been unable to convince the New Zealand Statistical office to allow their data to leave the Central Statistical Office because of similar constraints, and because of different interpretations of national rules by the New Zealanders who remain most earnest in their desire to make these data available. We are working toward a solution which meets both the New Zealand government's needs for privacy assurance, and the needs of LIS, and we are hopeful for a good solution soon.

In the United Kingdom, the mid-1980's brought a change in data distribution such that the FES was available only from the data archive at Essex and then only if United Kingdom users signed a pledge separate from and in addition to, the Standard LIS pledge of confidentiality (Table 5a and 5b). The bureaucratic maneuvering to arrive at this arrangement with the United Kingdom took almost two years to be accomplished, during which time access to the United Kingdom files was severely restricted.

Finally, all four of these datasets show how times and surveys change and hence affect continuity in time series data from LIS. Besides changing income top codes, the United States' CPS underwent a major change in data collection made in the mid-1990's, producing a large jump in summary inequality measures. In 1997, Canada abandoned the SCF and moved their annual income data source to the new Survey of Low Income Dynamics (SLID) which LIS will use from now on. The comparability of the new survey to the old will be questioned, no doubt. In a similar vein, Australia redesigned their AHS in the 1990's and debates about comparability to the old survey are still ongoing. And while England continues with the FES, it also has a new "Income Survey" (The Family Resources Survey) since 1994, which collects improved annual household income information from a random sample of British households. Future LIS data will include both surveys. And so, while the documentation and overall data quality of these surveys are constantly improving, they have also changed dramatically in recent years. For the

most part, use of these datasets has been regularized and works smoothly, though time trend comparisons are somewhat problematic (see Atkinson et al. 2000).

In contrast, many non-anglo-saxon central national statistical offices (CSO's) have not much experience at all with "public use" household income data. In Japan, hardly anyone but the CSO can use the data. In many other nations, e.g., Spain, Israel, and Germany, LIS was the first instance in which any research user outside the country obtained national data. In these cases, creation of both a clean user friendly file and appropriate documentation in English is sometimes problematic. Luckily, a number of national research teams had worked already with this data, and LIS could build on their groundwork.[3] Still the process is more cumbersome and the documentation weaker under such arrangements, regardless of the overall quality of the data.

Science-Based Data. In some nations, there are no CSO data that can be made available for LIS. This can happen because the data cannot leave the country, because the national data are weak, or because there are no national data to begin with. In Germany, during the 1980s, the Bundesamt would not allow data to leave the country, and further wanted to charge a very high price for such data, if it could be exported. In Italy, there is only a periodic CSO Expenditure/Budget Survey with weak household income information, and in Luxembourg there was no income survey at all. In each of these cases, a scientific organization, or a scientific research branch (of the Bank of Italy) is the main source of reliable household income data. In Germany, the Socioeconomic Panel (GSOEP) of the Deutsches Institut für Wirtschaftsforschung (DIW) has been the main database for LIS since 1984, while the Luxembourg Socio-Economic Panel (PSELL) has been the only source of Luxembourgish income data since 1985. In general, because these databases are collected and directed by researchers for research purposes, the data are easier to use. However, sampling issues and population weighting (to national totals) are more problematic than with national statistical office data. Technical documentation is

sometimes less available than in the case of the Anglophone public data use nations. On the other hand, the researchers at each of these institutions are keenly aware of the cross-national research opportunities which LIS brings and are therefore more likely to be helpful in maintaining, advertising, and using the harmonized data created by LIS.

The Bank of Italy Survey is an annual household cross-sectional income survey like all of the CSO sources cited above, and is the best source of Italian income data. In contrast, the GSOEP and PSELL data are taken from household panel surveys which follow the same people over time and which collect subannual income components. In both of these cases, LIS needs to construct a cross-section of annual income from the longitudinal files. Other than some possible biases from sampling attrition in the panel surveys, this process has been regularized and now works smoothly.

Administrative Data. The Scandinavian and Nordic nations that are part of LIS (Denmark, Finland, Norway, and Sweden) have yet another type of data. In these nations, national registers from labor, tax, and social security offices are used to provide administratively matched data which is obtained with the permission of the respondent. Essentially, a very high percentage of households (over 95 percent typically) fill out a short demographic survey and then sign a waiver which allows the CSO to obtain exact source data reports for each person and household income component from various government income registers (Denmark, Finland, Sweden), or from tax forms (Norway). The data thus produced is neither topcoded, nor in any way edited. National CSO's may make some additional top or bottom codes before the data are transferred to LIS, but essentially there is no underreporting or misreporting or other type of reporting error with this data (see Table 3B).

There are slight variations across countries. For instance, the Norway data comes from the tax registers and may exclude small amounts of local income maintenance benefits that are

not taxed. The Swedish data were until recently aggregated only to the level of the tax unit, not the "household," thus making it difficult to identify children aged 18 or over still living with their parents. Recently, in large part on behalf of LIS, the Swedish Income Distribution Survey has become aggregable into households. Finland has a complete set of data, but had to enact a separate law to allow the Income Distribution Survey to leave the country. In Denmark, the law was not changed, but had to be interpreted such that the data could come to LIS. Each of these data files is well documented with the help of the CSO's, and are by some criteria, the highest quality data in LIS.

Summary. Thus, LIS contains several types of data. As can be seen in Table 3, there are other variations to this mix as well. France provides both tax data and budget data, but only periodically. The Swiss national science foundation has sponsored two independent income/ poverty surveys which are the only available files for LIS. Austria has a labor force supplement to its microcensus that contains a small number of income questions for LIS, and Belgium has a scientific group's panel data survey from which income distribution data is derived. However, the main types of data used by LIS are described above. Additional information on any and all of these topics is available from the LIS website (www.lis.ceps.lu).

III. Privacy, Availability, and Quality

Over the past 15 years, LIS has managed to provide safe, privacy-protected remote access by internet to a wide range of harmonized datafiles. This has been accomplished with absolutely no breach of confidentiality or privacy. At this point in time, LIS's attention to detail, willingness to work with users, and our strong support of the "Canberra Group," (a set of major central statistical offices and projects such as LIS, working together to improve the guidelines

that define household income and its components for purposes of international comparability; see below), has given LIS a strong international backing. However, this was not always the case.

Privacy and Access: An Historical Perspective

Early on, the LIS project had to remove a large number of hurdles to obtain data. First of all, the LIS project stands for open and low cost (zero money cost) access to data by researchers who sign the privacy pledge (Table 5). Access to household income microdata by university or "think tank" researchers in a national context was essentially accepted practice in only a handful of nations. To provide flexible access and also to maintain the privacy and confidentiality of respondents was unheard of in the early 1980s. In fact, one of the major reasons that LIS ended up in Luxembourg was because Luxembourg has the strongest data protection and confidentiality laws in all of the OECD nations. Thus nations which provided their data had to be reassured that there would be no direct distribution of data outside of Luxembourg.

The obstacles were many. Suppose that LIS data could be used under restricted access conditions in Luxembourg (with the actual household income data being stored and used on the Luxembourg Central Government computers). This access would be useful only if the data could be harmonized and if the results proved feasible and attractive to researchers. And even then, one would have to travel to Luxembourg to make use of the data; something that researchers are not likely to do on a regular basis. All of these obstacles had to be overcome to make LIS work.

Three particular individuals were instrumental in making LIS work as it does. First, Professor Gaston Schaber, head of a Luxembourg "think tank," CEPS provided the central funding for the creation of the harmonized LIS data, while the author did the harmonization and most of the initial variable definitions and construction, and wrote most of the initial papers. Professor Lee Rainwater (LIS Research Director) and Mr. John Coder (then on leave from the

U.S. Census Bureau) worked out a method by which the internet could be used to provide remote access to the harmonized but restricted data files. Thus, reasonable remote access to harmonized data could be obtained for a very low cost.

The final piece in the technical puzzle was the willingness of the original seven nations to share their data with LIS. For the United States, United Kingdom, and Canada, the task was relatively easy since there were no restrictions on foreign use of "public use" microdata files. In Germany, a small 1981-research center database ("Transferumfrage" or Transfer Survey) was made available, while in Israel, there were no objections to access government collected data, but not much documentation. For Norway and Sweden, however, where administrative data was used, exporting data to LIS was a bold step for their CSO's to take. With the help of Stein Ringer (Norway), Sten Johannson (Sweden) an Robert Erickson (Sweden), the Norwegian and Swedish data were made available to LIS in 1984.

All that remained was the issue of obtaining academic approval of this experiment. The result was the first LIS conference in Luxembourg in August 1985 and the subsequent publication of the first LIS book a few years later (Smeeding; O'Higgins, and Rainwater 1990), with a stirring introduction by Anthony Atkinson heralding the importance and usefulness of LIS.

By 1987 remote access via the SPSS software system was in place and users around the world began to use LIS. New datasets and countries were added; national sources of funding paid for the "public good" that LIS became, and the project moved forward.

Programs and Progress: LIS Perspectives

What was revolutionary in 1983 is by some standards "backward" in 2000. Now LIS is pressured to release its own public use microdata files to users around the world. However, the privacy restrictions and restrictions on added use by the majority of LIS countries have made it

impossible to do so. LIS has added several software packages (SAS and STATA as well as SPSS), several service-oriented staff, documentation of institutional data for national transfer programs and summary statistics. Yet it still cannot provide household income microdata offsite.

In many nations, for the World Bank, and for other data producers, household income microdata files are easily obtained in non-harmonized form by researchers who usually apply for such permission and pay a marginal cost for accessing these data. In many ways, then, the world of data access has moved beyond LIS. Still, LIS offers a product that few others can match: a set of harmonized datasets which are as comparable as can be made possible using the resources of the LIS database team.[4] Other data sources are neither harmonized and often not comparable; but still they are widely used and treated as if they were comparable (e.g., see Atkinson et al. 2000; Smeeding 2000).

In contrast, some central statistical offices have not even come up to the LIS level of access. For a series of complicated reasons, the European Community Household Panel (ECHP) datasets collected from 1995 through 1999 for 15 European Community nations have not been made available to LIS or to independent scientific researchers more generally. The European Statistical Office, Eurostat, has set up a complicated process of access which is very expensive and very restrictive, almost bordering on the need for explicit permission from Eurostat to publish research results used in this data. As a result, scientific publications and research use of these data have been restricted and even minimized. For many of the less rich nations in Europe, e.g., Greece, Portugal, (until recently) Ireland, and Spain, these are the only recent income survey data available. Five years of negotiation with Eurostat by LIS have been totally unproductive in gaining access to these data. And, in effect, the lack of access has reduced both the demand for these data and their usefulness to academic and policy researchers in Europe. In

so doing, it has also likely reduced academic and public support for the ECHP itself since so few scientific results have been made public.

Improving Data Quality Directly: The Canberra Group

The best way to improve national survey data on income is to begin with improving the data itself. And just such a movement has recently begun. In 1996, the initiative to organize an International Expert Group on Household Income Statistics was taken by the Australian Bureau of Statistics in order to work on the development of statistics on household economic well-being and particularly on household income. The initiative reacted to a growing awareness that, in advancing the quality of their own household income statistics, National Statistical Institutes and CSO's shared many problems. In particular the comparative OECD study on income distribution (Atkinson, Rainwater, and Smeeding 1995) triggered a renewed discussion on the underlying quality and comparability of income data. Expectations were that combining forces would help solve conceptual and methodological problems, result in more relevant and reliable national statistics, and provide better data to be used for international comparisons on income distribution.

The primary objective of the Canberra Group is enhancing national household income statistics by developing standards on practical and conceptual issues which are related to the production of income distribution statistics. Its work was in support of a revision of international guidelines on income distribution statistics provided in draft form in 1977 by the United Nations. The Group collectively addressed the common conceptual, definitional, and practical problems faced by national and international statistical agencies in this subject area and has acted as a forum for expert opinions on conceptual and methodological issues and for obtaining endorsement for guidelines. This combined approach to solving these conceptual and

methodological problems will hopefully result in improved national statistics, and also in improved data for international comparisons on household income distribution.

The International Expert Group met for the first time in Canberra, Australia in 1996 and, taking its name from the venue of the First Meeting, is known as the "Canberra Group." It follows a now well-established phenomenon of City-named Expert Groups set up under the auspices of the United Nations Statistical Commission. From the beginning, the Canberra Group was designed to be a flexible working group of experts in household income statistics from both national and international organizations. Members of the Group included representatives from national statistical agencies, government departments and research agencies from Europe, North and South America, Asia, Australia and New Zealand, as well as from a number of international organizations and research agencies. The final report of the Canberra Group will be published early in 2001. The previous reports are available on the Canberra Group website, at LIS, or in hard copy (International Expert Group on Household Income Statistics 1997, 1998, 1999, 2000).

LIS has remained a primary sponsoring party to the Canberra Group from its beginning, also hosting the fourth and final meeting of the Group in May 2000. Yet the importance of the Canberra Group report for LIS is yet to be determined. When the final report is published in early 2001, its usefulness will depend upon the extent to which its recommendations and guidelines are used by national CSO's and other data producers. To the extent that the comparability of these "source" data are improved, the LIS comparability of their harmonized data will also improve.

IV. The Future: Summary and Conclusion

The LIS project is now stronger than ever, with adequate funding, and good scientific reputation and excellent staff. LIS is expanding its horizons to add Mexico, South Africa and a

second wave of Central and Eastern European nations. We are developing new "web access" tools to substitute for complicated software so that non-programmers can have basic, but still restricted, access to LIS files. Response time for over 95 percent of remotely submitted jobs is now 15 minutes or less and less than 5 minutes for 60 percent of all jobs (Coder 2000). Moreover, several CSO's have been in touch with the LIS technical team to assess the feasibility of making their own data available via remote access. The final report of the Canberra Group (see LIS website at www.lis.ceps.lu/canberra) will be published in February 2001 and then we hope that better source data will come to LIS; something that will make the harmonization process easier to beginning with. Thus, the future is bright for LIS and its process of restricted data in a safe, user friendly environment. We can only hope that the statistical offices, which have been so restrictive in their access to data, come to see the net benefits for users, providers, and governments more generally from participating in the LIS and in other similar projects.

Table 1. Luxembourg Income Study

Country	National Sponsor(s)
Australia	Australian Bureau of Statistics Social Policy Research Center, U. of New South Wales
Austria	Ministry of Social Security and Generations
Belgium	National Office of Science and Technology
Canada	Statistics Canada; Social Sciences and Humanities Research Council
Denmark	Social Research Institute
Finland	Abo Akademi University
Germany	Zentrum für Umfragen, Methoden und Analysen (Center for Survey Research and Methodology)
Israel	National Insurance Institute
Italy	Banca d'Italia; Consiglio Nazionale della Ricerche/Instituto di Ricerche sulla Popolazione (National Research Council/Institute for Population Research)
Luxembourg	Center for Population, Poverty and Public Policy Studies/ International Networks for Studies in Technology, Environment, Alternatives, Development (CEPS/INSTEAD)
Netherlands	Ministry of Social Affairs and Employment
Norway	Institute for Social Research
Sweden	Swedish Institute for Social Research and Swedish Government
Switzerland	Swiss National Science Foundation
United Kingdom	Economic and Social Research Council (ESRC); HMS Office of National Statistics; HMS Department of Social Security; HMS Department of the Treasury
United States	U.S. National Science Foundation Ford Foundation MacArthur Foundation

Source: Luxembourg Income Study.

Table 2. LIS Database List: Country and Year [a]

Country [b]	Historical Databases		Wave I	Wave II	Wave III	Wave IV
Australia			AS81	AS85	AS89	AS94
Austria				OS87		OS95
Belgium				BE85	BE88/BE92	BE96
Canada	CN71	CN75	CN81	CN87	CN91	CN94/CN97
Czech Republic					CZ92	CZ96*
Denmark				DK87	DK92	DK95
Finland				FI87	FI91	FI95
France [c]			FR79/FR81	FR84A/FR84B	FR89	FR94
Germany [d]	GE73	GE78	GE81	GE83/GE84	GE89	GE94
Hungary					HU91	HU94
Ireland				IR87		IR95*
Israel			IS79	IS86	IS92	IS97
Italy				IT86	IT91	IT95
Luxembourg				LX85	LX91	LX94
Mexico					MX90*	MX94*
Netherlands			NL83	NL86/NL87	NL91	NL94
New Zealand			NZ81*	NZ86*	NZ90*	NZ94*/NZ98*
Norway			NW79	NW86	NW91	NW95
Poland				PL86	PL92	PL95
ROC-Taiwan			RC81	RC86	RC91	RC95
Russia					RL92	RL95
Slovak Republic					SV92	SV96*
Spain			SP80		SP90	SP95*
Sweden	SW67	SW75	SW81	SW87	SW92	SW95
Switzerland			CH82		CH92	
United Kingdom	UK69	UK74	UK79	UK86	UK91	UK95
United States	US69	US74	US79	US86	US91	US94/97
(U.S. State File) [e]						(US199567)

[a] Year given is reference year, not necessarily the year that the data were collected. Codes within the cells are the LIS database country/year abbreviations.

[b] We are also in negotiation with Greece (1995), Korea (1993), South Africa (1993), Portugal (1990, 1995) and Japan (1993).

[c] France has an income survey (1979, 1984) and a budget survey (1984, 1989, 1994).

[d] Germany has three different databases: an income and expenditure survey (1973, 1978, 1983); a transfer income survey (1981); and three cross-sections from the Socio-Economic Panel Study (GSOEP) (1984, 1989, 1994).

[e] U.S. State file is a merged set of three annual CPS databases which provides the capability of comparisons within the United States.

*Anticipated that this will be available during 2001.

Source: Luxembourg Income Study.

Table 3. Types of Survey Data and Quality

3A. Data Types

Row	Income Concept	
1.	Income or Living Standard Survey [a]	Netherlands, Australia, Canada, Israel Republic of China, Spain, New Zealand, Mexico, Czech Republic, Slovak Republic; Poland, Hungary, Ireland, Italy, Switzerland; United Kingdom[e], Germany[e]
	Combination of survey and administrative records	Denmark, Finland, Sweden
2.	Income Tax Records [b]	France [e], Norway
3.	Panel study from scientific group	Belgium, Germany [e], Luxembourg, Russia, Switzerland
4.	Labor Force Survey Supplement [c]	United States, Austria
5.	Expenditure Survey [d]	United Kingdom[a, e], Germany [e], France[e]

[a]Survey primarily aimed at necessary living standards or income. Secondary aims may include other items such as wealth, expenditure, earnings, home ownership, finances, etc. All but Italy came from government statistical office.

[b]Survey basis is from income tax records. Additional imputations are made for non-taxed income sources and related issues. In Finland, additional information is obtained from interviews.

[c]Primary survey objective is labor force participation, employment, unemployment, etc., special supplement provides income data.

[d]Primary purpose of survey is expenditure data, but monthly/weekly income information is also gathered.

[e]The United Kingdom, France and Germany have both income data from expenditure surveys and form income surveys. Germany and the United Kingdom also have privately and publicly financed data sources available from "scientific" sources. Only for Germany does LIS use all three sources.

Source: Luxembourg Income Study.

Table 3. Continued

3B. Differential Income Data Quality: A Conceptual Breakdown

Row	Income Concept	Difference	
1.	"True Income"	Black Economy [a]	
2.	Administrative Record Income	Tax Evasion and Avoidance [b]	Sweden, Finland, Denmark
3.	Tax Reported Income	Reporting Error [c]	Norway, France
4.	Edited Survey Income [d]	Item Non-response [e]	Australia, United States, United Kingdom, Germany, Luxembourg, Canada, Belgium, Italy, Ireland, Israel, Republic of China, Spain, New Zealand, Mexico, Czech Republic, Slovak Republic, Poland, Hungary, Russia, Austria
5.	Reported Survey Income		Netherlands, Switzerland, Germany

[a] *Black economy* consists of net income from illegal activities.

[b] *Tax evasion* refers to legal sources of income which are not reported to income tax authorities, while *tax avoidance* refers to use of legal means of reducing tax liabilities.

[c] *Reporting error* refers to the difference between the amount of income reported on a survey and the amount actually received.

[d] *Edited survey income* refers to survey income that has been adjusted for item non-response.

[e] *Item non-response* refers to the failure of a respondent to report the amount of income received from a specific income source.

Source: Luxembourg Income Study.

Table 4. LIS Income, Income Aggregates and Demographic Variables*

A.	**Income Variables**	
	Gross wages and salaries	Other social insurance
	Mandatory employer contribution	Means-tested cash benefits
	Nonmandatory employer contribution	All near cash benefits
	Farm self-employment income	Food benefits
	Self employment income	Housing benefits
	In-kind earnings	Medical benefits
	Mandatory contribution for self-employment	Heating benefits
	Cash property income	Education benefits
	Noncash property income	Private pensions
	Market value: residence (homeowners)	Public sector pensions
	Income taxes	Alimony or child support
	Property or wealth taxes	Other regular private income
	Mandatory employee contribution	Other cash income
	Other direct taxes	Realized lump sum income
	Indirect taxes	Gross wage/salary head
	Sick pay	Net wage/salary head
	Accident pay	Hourly wage rate head
	Disability pay	Gross wage/salary spouse
	Social retirement benefits	Net wage/salary spouse
	Child or family allowances	Hourly wage/salary spouse
	Unemployment Compensation	Alternate Non-cash income
	Maternity allowances	Near cash housing benefits
	Military/vet/war benefits	Near cash except housing
B.	**LIS Income Aggregates (combined from variables above)**	
	Total self employment income	Total social insurance transfer
	Total earnings	Total social transfers
	Total factor income	Total private transfers
	Total occupational pensions	Total transfer income
	Total market income	Total gross income
	Total means-tested income	Total mandatory payroll taxes
	Total social insurance	Net disposable income
C.	**Demographic Variables**	
	Married couple indicator	Marital status head
	Age of head	Marital status spouse
	Age of spouse	Tenure (owned/rented housing)
	Sex of head	Disability status head
	Number of persons in household	Disability status spouse
	Family (unit) structure	Number of children under age 18
	Number of earners in household	Age of the youngest child
	Geographic location indicator	Number of persons aged 65 to 74
	Ethnicity/Nationality of head	Number of persons aged 75 or more
	Ethnicity/Nationality of spouse	Labor force status head
	Education level of head	Labor force status spouse
	Education level of spouse	Weeks worked full time head
	Occupational training of head	Weeks worked full time spouse
	Occupational training of spouse	Weeks worked part time head
	Occupation of head	Weeks worked part time spouse
	Occupation of spouse	Weeks unemployed head
	Industry of head	Weeks unemployed spouse
	Industry of spouse	Hours worked per week head
	Type (status) of worker head	Hours worked per week spouse
	Type (status) of worker spouse	

Source: Luxembourg Income Study

Table 5. LIS/LES Pledge of Confidentiality and Data Security

5A. Basic Pledge for All Nations

I, _____ , have completed the LIS/LES Project Information Form and hereby promise to use these materials only for purposes of academic research or teaching as specified in the attached application.

I further promise to act at all times so as to preserve the confidentiality of individuals and institutions whose information is recorded in these materials. In particular I undertake not to use or attempt to use these materials alone or in combination with any other data to derive information relating specifically to an identified individual or institution nor to claim to have done so.

I understand that attempts to make copies of the data, in whole or in part, stored in the LIS/LES database, or any violation of the above clauses, may be subject to censure, fine or imprisonment. I understand that it is my responsibility that any research papers written or assisted by me and based on LIS/LES must be entered into the LIS/LES working paper series before they are published elsewhere.

The U.K. Office for National Statistics and the ESRC Data Archive as their agent have put additional restrictions on the use of the U.K.'s (1986, 1991, 1994, ...) FES and LFS data by LIS/LES subscribers, hence all LIS/LES users must choose either a. or b. below before we can grant you access to the these data.

a. ___ I intend to use the U.K. (1986, 1991, 1994, ...) FES and/or LFS data and will therefore sign and return the additional pledge.

b. ___ I will not use the U.K.'s FES and/or LFS data for these years and therefore do not need to sign and return the additional pledge.

If in the future I decide to use the U.K.'s FES and/or LFS data provided by LIS/LES, I will sign the additional pledge before attempting to access these data.

Signature: _____ Date: ___ / ___ / _____

As soon as this form is processed by the LIS office in Luxembourg, you will be given a LIS "ID" and LIS password. After you have received a message sent by our system that you have been added to the database you are allowed to submit your first runs.

Source: Luxembourg Income Study

5b. United Kingdom Special Pledge

LIS/LES Undertaking of Conditions of Use
of the U.K.'s FES and LFS Data (1986, 1991, 1994, ...)
and Explanatory Documentation Supplied by the Luxembourg Income Study

I, _____ , have completed the LIS/LES Project Information Form and have signed the LIS/LES PLEDGE OF CONFIDENTIALITY AND DATA SECURITY. In order to use the U.K.'s (1986, 1991, 1994, ...) FES and LFS data, included in the LIS/LES database, I further promise to use these materials only for purposes of academic research or teaching as specified in the attached application and to act at all times so as to preserve the confidentiality of individuals and institutions whose information is recorded in these materials.

I promise to acknowledge in any publication, whether printed, electronic or broadcast, based wholly or in part on such materials, both the original depositors and the ESRC Data Archive. The form of wording of the citation for individual datasets is to be found in the documentation distributed by LIS/LES and the ESRC Data Archive. To declare in any such work that those who carried out the original collection and analysis of the data bear no responsibility for their further analysis or interpretation. The acknowledgement is "Material from _____ is Crown Copyright; has been made available by the Office for National Statistics through the ESRC Data Archive; and has been used by permission. Neither the Office for National Statistics nor the ESRC Data Archive bear any responsibility for the analysis or the interpretation of the data reported here."

To acknowledge Crown Copyright within the text in the following form:

"Crown Copyright 20__ . Source: Office for National Statistics."

I promise not to distribute copies of the materials to others, nor to make copies of them except as necessary to carry out the purpose specified in the LIS/LES application form.

I promise to meet any charges for which I am liable and accept that the ESRC Data Archive and the depositor of the materials supplied bear no legal responsibility for their accuracy or comprehensiveness.

Signature: _____ Date: ___ / ___ / _____

Name (Block Capitals): _____

This form has been drafted on the request of the U.K.'s Office for National Statistics and the ESRC Data Archive. After you have received an acknowledgement that this pledge, together with the LIS/LES Project Information Form and LIS/LES Pledge of Confidentiality and Data Security, have been processed by the LIS Office in Luxembourg, you will be allowed to submit jobs using the UK data stored in the LIS/LES Database.

Source: Luxembourg Income Study

References

Atkinson, A.B., L. Rainwater, and T.M Smeeding. 1995. *Income Distribution in OECD Countries: The Evidence from the Luxembourg Income Study (LIS),* Social Policy Studies No.18. OECD, Paris, October.

Atkinson, A.B., A. Brandolini, T.M. Smeeding, and P. van der Laan. 2000. "Producing Time Series Data for Income Distribution: Sources, Methods, and Techniques," working paper, Center for Policy Research, The Maxwell School. Syracuse, NY: Syracuse University, September

Burkhauser, R.V., F. Behringer, and G. Wagner. 1993. "The English Public Use File of the German Socioeconomic Panel," *Journal of Human Resources,* 24(2): 429-433.

Coder, John. 2000. "User Services and Data Management in the Luxembourg Income Study," mimeo, Luxembourg Income Study. Syracuse, NY: Syracuse University, October.

Gottschalk, P. and T.M. Smeeding. 2000. "Empirical Evidence on Income Inequality in Industrialized Countries." In A. B. Atkinson and F. Bourgignon (eds.), *Handbook of Income Distribution*. New York: Elsevier-North Holland Publishers, Vol. 1: 263-308.

International Expert Group on Household Income Statistics ('Canberra Group'). 1997. *Papers and Final Report of the First Meeting on Household Income Statistics, Canberra, Australia, 2-4 December 1996*. Belconnen, ACT: Australian Bureau of Statistics, February.

International Expert Group on Household Income Statistics ('Canberra Group'). 1998. *Papers and Final Report of the Second Meeting on Household Income Statistics, Voorburg, The Netherlands, 9-11 March 1998* Voorburg and Heerlen: Statistics Netherlands, May.

International Expert Group on Household Income Statistics ('Canberra Group'). 1999 *Papers and Final Report of the Third Meeting on Household Income Statistics, Ottawa, Ontario, Canada, 7-9 June 1999*. Ottawa, Ontario: Statistics Canada, November.

International Expert Group on Household Income Statistics ('Canberra Group'). 2000. *Papers and Final Report of the Fourth Meeting on Household Income Statistics, Differdange, Luxembourg, 15-17 May 2000*. Differdange, Luxembourg: Luxembourg Income Study at CEPS/INSTEAD. <http:www.lis.ceps.lu/canberra>.

Organization for Economic Co-operation and Development. 2000. *Social Sciences for a Digital World: Building Infrastructure and Databases for the Future*. Paris: OECD.

Radner, D. 1983. "Adjusted Estimates of the Size Distribution of Money Income," *Journal of Business and Economic Statistics*, 1: 135-146.

Smeeding, T. 2000. "Changing Income Inequality in OECD Countries: Updated Results from the Luxembourg Income Study (LIS)." In R. Hauser and I. Becker (eds.), *The Personal*

Distribution of Income in an International Perspective. Berlin, Germany: Springer-Verlag.

Smeeding, T.M., M. O'Higgins, and L. Rainwater. 1990. *Poverty, Inequality and the Distribution of Income in an International Context: Initial Research from the Luxembourg Income Study (LIS) Project London*: Wheatsheaf Books, and Washington, DC: Urban Institute Press.

Smeeding, T.M. and D. Weinberg. 2001. "Toward a Uniform Definition of Household Income," *Review of Income and Wealth*, in press.

U.S. Department of Commerce. 1991. *Poverty in the United States, 1990*. Current Population Reports, Series P-60, No. 175. Washington, DC: Government Printing Office.

Endnotes

1. These included Professors Lee Rainwater and Martin Rein (United States); Prof. Dr. Richard Hauser (Germany); Prof. Robert Erikson (Sweden); Dr. Stein Ringen (Norway); Dr. Michael O'Higgins (United Kingdom); and Ms. Lea Achdut (Israel).

2. For more on this topic, please see Atkinson, Rainwater, and Smeeding (1995), and Smeeding (2000).

3. For example, the work of Prof. Dr. Richard Hauser, Sonderforschungs Bereich III, at the University of Frankfurt-Main for Germany; and Javier Ruiz-Castillo at Juan Carlos University, Madrid, for Spain.

4. See Burkhauser, Behringer, and Wagner (1993) for an important exception: the German-United States Panel Data Comparability Project.