Smeeding, Timothy M.; Jesuit, David K.; Alkemade, Paul

Article

The LIS /LES Project Databank: Introduction and Overview

Schmollers Jahrbuch – Zeitschrift für Wirtschafts- und Sozialwissenschaften. Journal of Applied Social Science Studies

Provided in Cooperation with:
Duncker & Humblot, Berlin

Suggested Citation: Smeeding, Timothy M.; Jesuit, David K.; Alkemade, Paul (2002) : The LIS /LES Project Databank: Introduction and Overview, Schmollers Jahrbuch – Zeitschrift für Wirtschafts- und Sozialwissenschaften. Journal of Applied Social Science Studies, ISSN 1865-5742, Duncker & Humblot, Berlin, Vol. 122, Iss. 3, pp. 497-517, https://doi.org/10.3790/schm.122.3.497

This Version is available at:
https://hdl.handle.net/10419/292025

Schmollers Jahrbuch 122 (2002), 497 – 517
Duncker & Humblot, Berlin

European Data Watch

This section will offer descriptions as well as discussions of data sources that may be of interest to social scientists engaged in empirical research or teaching courses that include empirical investigations performed by students. The purpose is to describe the information in the data source, to give examples of questions tackled with the data and to tell how to access the data for research and teaching. We will start with data from German speaking countries that allow international comparative research. While most of the data will be at the micro level (individuals, households, or firms), more aggregate data and meta data (for regions, industries, or nations) will be included, too. Suggestions for data sources to be described in future columns (or comments on past columns) should be send to: Joachim Wagner, University of Lueneburg, Institute of Economics, Campus 4.210, 21332 Lueneburg, Germany, or e-mailed to ⟨wagner@uni-lueneburg.de⟩.

The LIS / LES Project Databank: Introduction and Overview

By Timothy M. Smeeding, David K. Jesuit and Paul Alkemade

Abstract

The Luxembourg Income Study (LIS) project, founded in 1983, is one of the oldest and best known examples of cross-national 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.

1. Introduction

The Luxembourg Income Study (LIS) research and databank project has provided harmonized cross-national household income microdata for social science research for over 19 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 article 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 and data quality. 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 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.

2. 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. From the beginning, the LIS project was supported by groups of academics and social statisticians who contributed their expertise to make datasets and technical expertise available to LIS. Today LIS stands as one of the few truly cross-national and comparable data infrastructures extant (OECD 2000).

The LIS project has five goals:

- to *harmonize* cross-national data (thus relieving researchers of this task) 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.

LIS: 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. 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 training seminars aimed explicitly at increasing the base of users, especially among junior scholars.

The problem with public goods, however, is that they can be used without users helping to cover their fixed costs. Within nations, national bodies and national research institutes fund public goods. But across nations there are few, if any, organizations with the scope or interest to fund a microdata infrastructure (OECD 2000). Thus, LIS is fortunate to have a solid base of financial support by the 20 nations who recognize that, without their funding, LIS would close.

Data Harmonization

The most important goal for LIS is data harmonization. International data availability is an important obstacle that is slowly being overcome (see sections 3 and 4 below). But the access and availability of three, four, or more national income surveys with no explanation 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 3 and 4.

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. The countries are largely covered by the OECD, G-8, 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 to 1997. As of 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 have begun a new "millennium" round of datasets for 2000. A list of countries and years for which data are available is attached (Table 1).

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. 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 maintain the privacy and confidentiality of respondents was unheard of in the early 1980s. In fact, one of the major reasons that LIS began in Luxembourg was because Luxembourg has the strongest data protection and confidentiality laws in all of the OECD nations. Thus, nations that 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 researchers are not likely to do on a regular basis. All of these obstacles had to be overcome to make LIS work.

An operating system for our remote access network was implemented in 1987, and researchers around the world began to use LIS. Since that time,

the functionality and flexibility of the remote access system (termed "Lissy") has steadily improved. It performs user requests flexibly and quickly, allowing data access by use of the major statistical software packages SAS, SPSS, and STATA. Moreover, 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 the LIS web site. 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. Finally, in future years, LIS will add a new "web tabulator" system that allows inexperienced users the ability to obtain summary data by simply entering a few key words into a worldwide web-based system which will generate these tabulations directly (see Coder 2000).

3. The Luxembourg Employment Study (LES)

In the early 1990s, labor markets in the developing world were rapidly changing. In order to understand these dynamics from a comparative perspective, the Luxembourg Employment Study (LES) was initiated in 1994. These surveys provide detailed information on areas like job search, employment characteristics, comparable occupations, investment in education, migration, etc.

The basic idea was again to provide users with harmonized data on labor market characteristics in different countries to enable comparative research. As such, LES shares the same principles as LIS but has been enriched by the long-term experience that had developed within the LIS project. Therefore, in this section we will not repeat the similarities, but rather point out some important differences.

The availability of Labour Force Surveys (LFS) to LES appeared much more restricted than the income surveys were to LIS. Due to the large sample sizes and the available detail of labor characteristics of the individual respondents, in a number of cases LFS-data are not allowed to leave the country of origin. At present, the LES database contains 16 countries, compared to 26 in LIS. The list of datasets included in the LES is reported in Table 1A. Also, since the LES project is of more recent date, each country is not yet represented by a whole series of datasets, but rather by only one point in time in most cases.

In terms of comparability (to be discussed in detail shortly), the LES files go one step further than LIS. This means that many LES-variables are not just harmonized, but also fully standardized. The content of a harmonized variable captures the same concept, but the coding of the different categories may vary between countries. In standardized variables however, each category has exactly the same meaning irrespective of the dataset chosen.

Table 1A: **LIS Database List: Country and Year**

Country	Code	Historical Databases	Wave I around 1980	Wave II around 1985	Wave III around 1990	Wave IV around 1995	Wave V around 2000
Australia	AS	.	1981	1985	1989	1994	.
Austria	AT	.	.	1987	.	1995	.
Belgium	BE	.	.	1985	1988 / 1992	1997	.
Canada	CN	1971 / 1975	1981	1987	1991	1994 / 1997 / 1998	.
Czech Republic	CZ	.	.	.	1992	1996	.
Denmark	DK	.	.	1987	1992	1995° / 1997°	.
Estonia	EE	.	.	.	.	.	2000**
Finland	FI	.	.	1987	1991	1995°	2000
France	FR	.	1979 / 1981	1984A / 1984B	1989	1994	.
Germany	GE	1973 / 1978	1981	1983 / 1984	1989	1994	.
Hungary	HU	..	.	.	1991	1994	1999*
Ireland	IR	.	.	1987	.	1995** / 1997*	.
Israel	IS	.	1979	1986	1992	1997	.
Italy	IT	.	.	1986	1991	1995	2000*

Luxembourg	LX	·	·	1985	1991	1994	·
Mexico	MX	·	·	1984	1989/1992	1994/1996/1998	2000*
Netherlands	NL	·	·	1983/1986*/1987	1991	1994	·
Norway	NW	·	1979	1986	1991	1995	·
Poland	PL	·	·	1986	1992	1995	1999
R.O.C Taiwan	RC	·	1981	1986	1991	1995	2000*
Russia	RL	·	·	·	1992	1995	·
Slovak Republic	SV	·	·	·	1992	1996*	·
Slovenia	SI	·	·	·	·	1997	1999
Spain	SP	·	1980	·	1990	1995?	·
Sweden	SW	1967/1975	1981	1987	1992	1995	·
Switzerland	CH	·	1982	·	1992	·	2000*
United Kingdom	UK	1969/1974	1979	1986	1991	1995/1994	1999
United States	US	1969/1974	1979	1986	1991	1994/1997/ State file: 199567	2000**

** = Lissification in process
° = Currently being reviewed
* = received; waiting to be lissified
? = under negotiation

Source: Luxembourg Income study.

Thanks to the larger sample sizes, the LES files offer the possibility of detailed studies on labor market differences and – like the LIS database – are unique. Another known example is Eurostat's attempt to bring different labor force surveys in line with each other by setting up a series of recommendations and definitions, but this of course is limited to member countries within the European Union.

As the two databases are built on different sources, the results from LIS and LES cannot be linked on micro data level. One could construct indicators at the macro level only after aggregating the microdata, which can help the users to get a better understanding of interactions between labor market characteristics and individual well being.

Table 1B

LES Database List: Country and Year

Country	Code	Wave III around 1990	Wave IV around 1995	Wave V around 2000
Austria	AT	1991	·	·
Canada	CN	·	1997	·
Czech Republic	CZ	·	1994	·
Finland	FI	1990	·	·
France	FR	·	1997	·
Hungary	HU	· ·	1993	·
Luxembourg	LX	1992	·	·
Norway	NW	1990	·	·
Poland	PL	·	1994	1999
Slovak Republic	SV	·	1995	·
Slovenia	SI	·	1994	1999
Spain	SP	·	1993	·
Sweden	SW	1990	·	·
Switzerland	CH	·	1997	·
United Kingdom	UK	1989	1997	1999
United States	US	1990	1997	2000**

** = Lissification in process
° = Currently being reviewed
* = received; waiting to be lissified
? = under negociation

Source: Luxembourg Income Study.

At present there are plans being developed to further integrate the two projects. We hope that these plans will facilitate users of both studies and give a further boost to comparative research. More detailed information on the countries and variables available in the Luxembourg Employment Study can be found on the LIS web site.

4. Data Details

As seen in Table 2, there are numerous types of data to which LIS has access. 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 2, where we present data by type (2A) and by one measure of overall dataset quality (2B).

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 LIS are not uniform in nature, purpose, or objective. The lowest common denominator 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 2A. 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 (e.g. for the United Kingdom, France); 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).

Table 2B 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, assuming income reporting in the upper rows to be more complete than in lower rows. In the same table, an attempt has been made to link the LIS country dataset to each of the levels.

Up the rows from bottom to top, Table 2B 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. Item non-response is treated differ-

Table 2

Types of Survey Data and Quality

2A. Data Types		
Row	**Income Concept**	**Country**
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.

ently in the various countries, from leaving the non-response as missing values (allowing the user to make further imputations for non-reporting of income items), up to full imputation whereby all item non-responses are corrected, also called edited income (row 4) (see Atkinson, Rainwater, and Smeeding 1995, Appendix 4 for more detail on this topic).

Next, row 3 refers to the amount of income recorded in data taken from tax records. Norwegian and French data are at this level. Table 2B 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

2B. Differential Income Data Quality: A Conceptual Breakdown			
Row	**Income Concept**	**Difference**	**Country**
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)]

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.
c9 *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.

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 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.

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 from 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 (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 far more problematic to capture. 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 has some error. The degree of error that 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 (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, for example, the Canberra Group (2001 and 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 other nations. In general, LIS seeks to "space" datasets first, and second, to find a "given" year, for example, 1995 or 2000. Even if all datasets were for the same year, different business conditions would produce different cyclical outcomes across datasets.
- **Time consistency.** LIS pays a great deal of attention to intra-period 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 the original survey is substituted by another one is 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 differs but can be inferred from the specific country discussions that follow shortly below.

Basic "LISification" Procedures

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 variable list is included in Table 3, and the LES variable list is reported in Table 4. Included in section B of this list

are derived basic sub-aggregations of household income according to the LIS definitions (see Atkinson, Rainwater, and Smeeding 1995; Smeeding and Weinberg 2001). Besides household records, we also have individual person records. Most of the demographic data shown in Table 3, section C, comes from the individual records within a household. 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.

Table 3

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 4

LES Variables

A. Demographic background	
RELATIONSHIP TO REFERENCE PERSON IN THE HOUSEHOLD SEX	URBAN / RURAL INDICATOR
AGE	HOUSEHOLD TYPE
MARITAL STATUS	FAMILY TYPE
	NUMBER OF PERSONS IN HOUSEHOLD
NATIONALITY	NUMBER OF CHILDREN IN HOUSEHOLD
YEARS OF RESIDENCE IN THIS COUNTRY	NUMBER OF EMPLOYED IN HOUSEHOLD
COUNTRY OF BIRTH	NUMBER OF PENSIONERS IN HOUSEHOLD
ETHNICITY	USUAL / MAIN ECONOMIC STATUS
REGION	
B. Work status	
WORK STATUS DURING REFERENCE WEEK	REASON FOR NOT HAVING WORKED AT ALL THOUGH HAVING A JOB
C. Employment characteristics of the main job	
COUNTRY OF PLACE OF WORK	DURATION OF TEMPORARY JOB OR JOB CONTRACT OF LIMITED DURATION
PROFESSIONAL STATUS / CLASS OF WORKER	NUMBER OF HOURS PER WEEK USUALLY WORKED
ECONOMIC ACTIVITY OF ESTABLISHMENT / INDUSTRY	NUMBER OF HOURS PER WEEK ACTUALLY WORKED
OCCUPATION	MAIN REASON FOR HOURS ACTUALLY WORKED BEING DIFFERENT FROM PERSON'S USUAL HOURS
SECTOR OF ESTABLISHMENT	SHIFT WORK

Table 4 – Continued

NUMBER OF PERSONS WORKING AT THE LOCAL UNIT OF ESTABLISHMENT	EVENING WORK
REGION OF PLACE OF WORK	NIGHT WORK
DURATION OF CURRENT EMPLOYMENT	SATURDAY WORK
FULL-TIME / PART-TIME DISTINCTION	SUNDAY WORK
PERMANENCY OF JOB CONTRACT	WORKING AT HOME
	LOOKING FOR ANOTHER JOB AND REASONS FOR DOING SO
D. Information about second job	
EXISTENCE OF MORE THAN ONE JOB	SECTOR OF ESTABLISHMENT, 2ND JOB
PROFESSIONAL STATUS / CLASS OF WORKER, 2ND JOB	NUMBER OF HOURS ACTUALLY WORKED
ECONOMIC ACTIVITY OF ESTABLISHMENT / INDUSTRY, 2ND JOB	REGULARITY
OCCUPATION, 2ND JOB	
E. Previous work experience of person not in employment	
EXPERIENCE OF EMPLOYMENT	OCCUPATION IN LAST JOB
PROFESSIONAL STATUS / CLASS OF WORKER IN LAST JOB	TIME PASSED SINCE PERSON LAST WORKED
ECONOMIC ACTIVITY OF ESTABLISHMENT / INDUSTRY IN WHICH PERSON LAST WORKED	MAIN REASON FOR LEAVING LAST JOB
F. Search for employment	
SEEKING EMPLOYMENT FOR PERSON WITHOUT EMPLOYMENT DURING THE REFERENCE WEEK	WILLINGNESS TO WORK FOR PERSON NOT SEEKING EMPLOYMENT
TYPE OF EMPLOYMENT SOUGHT	AVAILABILITY TO START WORKING WITHIN TWO WEEKS
DURATION OF SEARCH FOR JOB	SITUATION IMMEDIATELY BEFORE PERSON STARTED TO SEEK EMPLOYMENT (OR WAS WAITING FOR NEW JOB TO START
MAIN METHOD USED DURING PREVIOUS FOUR WEEKS TO FIND A JOB	REGISTRATION AT A PUBLIC EMPLOYMENT OFFICE
SECOND METHOD USED DURING PREVIOUS FOUR WEEKS TO FIND A JOB	TYPE OF BENEFIT THE INDIVIDUAL RECEIVES
THIRD METHOD USED DURING PREVIOUS FOUR WEEKS TO FIND A JOB	REASON FOR LOOKING FOR WORK
G. Situation of inactive persons	
SITUATION OF PERSONWHO NEITHER HAS A JOB NOR IS LOOKING FOR ONE	

H. Education and training	
HIGHEST COMPLETED LEVEL OF GENERAL EDUCATION	TOTAL LENGTH OF TRAINING
HIGHEST COMPLETED LEVEL OF FURTHER EDUCATION OR VOCATIONAL TRAINING	USUAL NUMBER OF HOURS TRAINING PER WEEK
EDUCATION AND TRAINING RECEIVED DURING PREVIOUS FOUR WEEKS	AGE WHEN OBTAINED HIGHEST LEVEL OF EDUCATION
PURPOSE OF THE TRAINING RECEIVED DURING PREVIOUS FOUR WEEKS	
I. Situation one year before survey	
SITUATION WITH REGARD TO ACTIVITY 1YEAR AGO	OCCUPATION 1 YEAR AGO
PROFESSIONAL STATUS / CLASS OF WORKER 1 YEAR AGO	COUNTRY OF RESIDENCE 1 YEAR AGO
ECONOMIC ACTIVITY OF ESTABLISHMENT / INDUSTRY 1 YEAR AGO	REGION OF RESIDENCE 1 YEAR AGO
J. Labour force status	
LABOUR FORCE STATUS	UNEMPLOYMENT STATUS
EMPLOYMENT STATUS	INACTIVITY STATUS
K. Earnings and income	
WAGES / EARNINGS PER HOUR	TOTAL PERSON INCOME
TOTAL PERSON EARNINGS	TOTAL FAMILY (HOUSEHOLD) INCOME
L. Technical items	
SERIAL NUMBER OF HOUSEHOLD	WEIGHTING FACTOR
SERIAL NUMBER OF FAMILY	DATE OF INTERVIEW
SERIAL NUMBER OF PERSON	COUNTRY IDENTIFIER

Source: Luxembourg Income Study.

Once a dataset has been identified as acceptable, LIS asks the country to send their "full" data file with completed documentation and other information. The LIS staff will then make the LISification itself, standar-dize the documentation, and return the LIS estimated and harmonized da-taset 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 LIS staff can correct any errors later uncovered by users. 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 through the LIS remote access system (Coder 2000).

Programs and Progress: LIS Perspectives

What was revolutionary in 1983 is, by some standards, "backward" in 2002. 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 that are as comparable as can be made possible using the resources of the LIS database team.[1] Other data sources are neither harmonized nor comparable, but still they are widely used and treated as if they were comparable (see Atkinson, Brandolini, Smeeding, and van der Laan 2000; Smeeding 2000).

In contrast, some central statistical offices have not even risen 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, 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 wealthy nations in Europe, for example, 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 Eur-

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

ope. 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, particularly household income. The initiative reacted to a growing awareness that, in advancing the quality of their own household income statistics, National Statistical Institutes and CSOs 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 to enhance 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. It has also 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, gov-

ernment 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 was published in early 2001 (Canberra Group 2001). Now, 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 the data are improved, the LIS comparability of their harmonized data will also improve. More information on the Canberra Group can be found on the LIS website.

5. The Future: Summary and Conclusion

The LIS project is now stronger than ever, with adequate funding, a good scientific reputation and excellent staff. LIS is expanding its horizons by adding Mexico and South Africa, and a second wave of Central and Eastern European nations will be included in the future. 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 10 minutes or less and less than five 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 will hopefully make the harmonization process easier. 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 by participating in LIS and in other similar projects.

References

Atkinson, A.B. / *Rainwater,* L. / *Smeeding,* T.M. (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. / *Brandolini,* A. / *Smeeding,* T.M. / *van der Laan,* P. (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. / *Behringer,* F. / *Wagner,* G. (1993): "The English Public Use File of the German Socioeconomic Panel," Journal of Human Resources, 24(2): 429–433.

Canberra Group on Household Income Statistics (2001): "Expert Group on Household Income Statistics: Final Report and Recommendations," Ottawa, Canada.

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

Gottschalk, P. / *Smeeding,* T.M. (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.

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. / *O'Higgins,* M. / *Rainwater,* L. (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. / *Weinberg,* D. (2001): "Toward a Uniform Definition of Household Income," Review of Income and Wealth, 47(1) (March): 1 – 24.

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.