5.2 Cross-National Data Center in Luxembourg, LIS[1]

Janet Gornick, Berglind Hólm Ragnarsdóttir, and Sarah Kostecki
City University of New York

5.2.1 Description and importance

LIS – formerly known as the Luxembourg Income Study – is a unique micro-data archive and research center dedicated to cross-national analysis. LIS is located in Luxembourg and also has a satellite office at the Graduate Center of the City University of New York. LIS' mission is to enable, facilitate, promote, and conduct cross-national comparative research on socio-economic outcomes and on the institutional factors that shape those outcomes.

LIS was founded in Luxembourg in 1983 by two American academics, economist Timothy Smeeding and sociologist Lee Rainwater, and a Luxembourgish psychologist, Gaston Schaber. Noticing a considerable gap in available and dependable micro-data for cross-national research, their aim was to construct harmonized cross-national income data in order to enable meaningful comparative research on poverty and inequality. They assembled a cross-national team that provided the micro-data and consulted on the harmonization; the first harmonized datasets were made available to researchers in 1987.

In 2002, LIS and its staff became an independent non-profit institution, incorporated in Luxembourg. In 2005/2006, Smeeding and Rainwater retired from LIS. They were succeeded by Janet Gornick, a political economist and sociologist based in the United States, and Markus Jäntti, an economist based in Sweden. As of 2012, Gornick and Jäntti continue to direct LIS.

LIS has four longstanding goals: (1) to harmonize cross-national micro-datasets that have been collected from participating countries and which include data on income, demography, employment, and wealth; (2) to provide a secure method that allows researchers to access micro-data that would otherwise not be available due to country-specific

1 This paper is a slightly shortened version of a longer paper about LIS that is available on the LIS website at: http://www.lisdatacenter.org/. The full paper – LIS Technical Paper #5 – includes over 140 citations corresponding to many of the research findings summarized in this entry.

privacy restrictions; (3) to create and maintain a remote-access system that allows research requests to be received and that quickly returns results to users at off-site locations; and (4) to promote the use of micro-data in comparative research on social and economic wellbeing on a global level, to conduct research onsite, and to sponsor and host scholars using the LIS data.

LIS is now home to two databases, the Luxembourg Income Study Database (originated in the 1980s and referred to as "LIS") and the Luxembourg Wealth Study Database (originated in 2007 and referred to as "LWS" – pronounced "loose"). The LIS Database, under constant expansion, is the largest available database of harmonized income micro-data collected from multiple countries over a period of decades. The database contains over 220 datasets, mainly organized into seven repeated cross-sections (known as waves); the datasets now span the years 1968 to 2010. The LIS Database, which contains variables at the household- and person-level, focuses on income data, from both private and public sources. The LIS datasets also contain data on household characteristics, labour market engagement, and, in some cases, household expenditures for consumption. The LIS data are mainly used to study poverty, income distribution, and labour market outcomes.

The newer Luxembourg Wealth Study Database (LWS) is the only cross-national wealth micro-database in existence. It currently includes 20 datasets from 12 countries, spanning the years 1994 to 2007. The LWS datasets focus on wealth data, including both assets and debt; they also contain household demographic and labour market characteristics, and, in some cases, behavioral variables. The LWS data, while still new, are gaining a base of users; these data have been used so far to study household wealth portfolios as well as the joint distribution of income and wealth.

Due to confidentiality protections, no micro-data may leave the LIS office. LIS therefore provides access to the LIS and LWS Databases in three ways: via the remote-execution system ("LISSY"), the Web Tabulator (an online table-maker), and the LIS Key Figures (standardized national indicators constructed by the LIS staff). Access through LISSY or the Web Tabulator requires registration. The LIS Key Figures are publicly accessible to all visitors to the LIS website, and do not require registration. LISSY is a remote-execution data access system for the LIS and LWS micro-data. LISSY allows registered users to submit programs using common statistical software packages (SAS, SPSS, or

Stata), while respecting the confidentiality restrictions imposed by many of the participating countries. The Web Tabulator is an online table-making tool that allows registered users to design and generate cross-national descriptive tables without the need for programming. The "Web Tab" includes datasets starting from LIS Wave V (centered around the year 2000), but does not presently include LWS datasets.

In addition, LIS has created two sets of national indicators based on micro-data from the LIS Database. The *Inequality and Poverty Key Figures* include multiple inequality measures, relative poverty rates for various demographic groups, and median and mean disposable household income. The *Employment Key Figures by Gender* display information about economic outcomes by gender as well as gender disparities in poverty and employment.

Extensive documentation for each dataset details technical aspects of the original survey, a record of the harmonization process, and institutional information on tax and transfer programs corresponding to the micro-data variables. The LIS website also houses several complementary country-level databases, including, e.g., a comparative welfare states database, and a family policy database; these contain an array of country-level policy indicators. These country-level databases are widely used by LIS' micro-data researchers, who often seek to link macro-level variables to micro-level outcomes.

Since LIS' inception, the databases have been used by about 3,500 researchers world-wide to analyze diverse outcomes including poverty, income inequality, employment status, wage patterns, gender inequality, family formation, immigration, and (since 2007) wealth and debt accumulation. A large share of LIS- and LWS-based studies focus on the ways in which, and the extent to which, economic and social policies shape these outcomes. Reports based on the LIS data – and recently, on the LWS data – have appeared in books, journal articles, and dissertations, and are often featured in the popular media. Each completed study is published in the LIS or LWS Working Paper series. In addition, LIS conducts annual summer training workshops that introduce researchers to the LIS and LWS Databases and to cross-national research on wages, income, employment, wealth, and social policy. Since 1988, over 500 scholars have attended the workshops. In 2010, LIS established a pre-doctoral and post-doctoral scholars program. LIS also hosts a longstanding Visiting Scholar program.

5.2.2 Structure, decision-making, and funding

The main office of LIS is located in Luxembourg, in a free-standing dwelling in Luxembourg City. The on-site staff includes a Director of Operations (who also serves as LIS' IT Director), an Administrator, a Data Team Coordinator, and the data staff (which usually includes four to six data harmonizers).

The operation of LIS is overseen by LIS' Director Janet Gornick. She is responsible for all aspects of LIS' work, including its administrative, management, development, and data production functions. She is joined by LIS' Research Director Markus Jäntti. His main responsibility is advising the LIS data staff on data harmonization, data quality, data production planning, and research methods. The Luxembourg office of LIS is complemented, and supported, by a satellite office in the United States, based at the Graduate Center (a doctoral granting campus) of the City University of New York (CUNY). The LIS office at CUNY, also directed by Janet Gornick, is a venue for additional administrative, development, and fundraising work, and it is a center for research using the LIS data, staffed largely by PhD students.

The LIS' Directors and staff (both on-site and off-site) are advised and guided by an active international governing board, composed of representatives from many of LIS' data-providing and funding institutions. The board is led by a President (currently, Oxford Professor A.B. Atkinson) and an Executive Committee. Board meetings are held annually – with "virtual" meetings in even-numbered years and "face to face" meetings in Luxembourg in odd-numbered years. Although the board plays a crucial role in advising LIS about its practices and priorities, decision-making about day-to-day operations is the responsibility of the Director, the Research Director, and the Director of Operations.

Decisions about harmonization practices are overseen by the LIS Data Team Coordinator, in conjunction with the LIS Research Director. The LIS data team frequently consults with the original data providers (mainly national statistical offices). They also consider and incorporate on-going feedback from LIS' many data users, and they follow international data harmonization standards. (For example, they follow, as closely as possible, recommendations contained in the Final Report and Recommendations made by the Expert Group on Household Income Statistics – i.e., the Canberra Report.) All that said, final

decisions about data harmonization are entirely the responsibility of the LIS Directors and staff.

LIS is funded by the Luxembourg Government – mainly the Ministry of Higher Education and Research, and the National Research Fund – and by contributions from 16 of the participating countries (http://www.lisdatacenter.org/about-lis/contributors), as well as from four supranational organizations (the Organization for Economic Cooperation and Development, the World Bank, the United Nations Development Programme, and the International Monetary Fund). Supplemental funds, usually for time-limited projects, come from assorted grants, mainly from foundations.

5.2.3 Possible and demonstrated benefits – intellectual impact

In its 29-year history, LIS has provided the means by which researchers can make accurate cross-national comparisons of diverse social and economic outcomes. While LIS' resources have most prominently provided the basis for descriptive results and policy-oriented analyses, they also provide the basis for a substantial amount of methodological and theoretical work. The main benefit of the LIS and LWS data is that they provide researchers around the world with harmonized micro-data that enable high-quality, cross-national, comparative research. Here, we will briefly review research contributions in economics, sociology, political science and related fields, based on the LIS and LWS data. For ease of reading, most of the many citations to specific studies have been omitted, but all citations can be found in the full paper on the LIS website (http://www.lisdatacenter.org/).

LIS' contribution to comparative scholarship

Income inequality – theory, measurement, and empirical analyses. LIS has provided a basis on which scholars can examine income inequality across countries and over time. Researchers have questioned the Kuznets hypothesis that income inequality grows and then declines as countries experience growth. Income inequality has risen, after 1980, not only in the U.S. and the U.K. (two of the more unequal national income distributions among LIS countries), but also in Sweden (one of the most equal) and in several other (but not all) Western countries. Atkinson (2004) suggests that explaining these changes is one of the key uses of LIS and one where it has a large impact.

LIS' ground-breaking study (Atkinson et al., 1995), produced in collaboration with the OECD, led many researchers to adjust the methodology for studying income distributions, towards one that begins with post-tax-and-transfer income and decomposes income distributions "backwards" to market income. Because many households in western countries rely on income sources other than earned income, this allows researchers to identify the elements of policy that shape household income packages. LIS has provided the basis by which researchers can measure the distributive and redistributive effects of labor market institutions, direct taxes, and income transfers, across countries and over time. A new and innovative line of research links labor market institutions to income inequality; another links income inequality with inequality in working hours.

The LIS Database has also provided the basis for the development of theoretical and empirical measures of income inequality, such as Lorenz dominance, Sen indices, and economic polarization. It has also provided a basis for sensitivity testing regarding top- and bottom-coding of income sources, family size adjustments, and other methodological decisions. LIS was also at the forefront of setting international standards for comparing income distributions.

LIS has been used recently to study the relationship between income inequality and economic growth, and between income inequality and both child wellbeing and health outcomes, including an article published in The Lancet in 2001 (Lynch et al., 2001). These are among the leading questions in income-related research and LIS has enabled a cross-national comparative dimension. The contribution of LIS to the economic study of inequality is further evident in its role in the *Handbook of Income Distribution*, edited by Atkinson and Bourguignon (2000), in which two of three empirical chapters were based on LIS. Atkinson and Bourguignon have just begun a follow-up volume (in late 2011); it too will include several chapters based on LIS. The 2009 *Oxford Handbook of Economic Inequality*, edited by Salverda, Nolan, and Smeeding, includes several chapters based on LIS. The widely-publicized 2008 OECD report *Growing Unequal: Income Distribution and Poverty in OECD Countries* also drew heavily on LIS and LWS micro-data, as did the December 2011 follow-up report, *Divided We Stand.* In 2013, Stanford University Press will publish a volume, edited by Gornick and Jäntti, titled *Income Inequality: Economic Disparities and the Middle Class in Affluent Countries.* The book includes 17 commissioned studies,

based on datasets included in (or soon to be included in) the LIS or LWS Databases.

Poverty measurement and analysis. The LIS Database has been used extensively to compare levels of relative and absolute poverty across countries and over time, with many studies focusing on economically vulnerable groups, such as children, the elderly, women (especially single mothers and older women) and immigrants. Some studies have used the LIS data to assess the U.S. specifically in cross-national perspective, focusing on its comparatively high relative poverty rates, in conjunction with variation across the U.S. states. This body of LIS-based poverty research has led to work on purchasing power parities for micro-level comparisons of absolute standards of living, and to a large literature on income packaging. LIS-based research on absolute and relative poverty has formed the basis for analyses of poverty in the 1997 and later editions of the UN *Human Development Report.* Gornick and Jäntti (2009) produced a report on women's poverty, commissioned by the United Nations Research Institute for Social Development (UNRISD).

Gender gaps in employment, earnings, occupations, and income. The LIS Database has allowed researchers to analyze cross-national variation in gender gaps and/or women's status in employment and household economic wellbeing. LIS has allowed researchers to assess the effects of tax and transfer policy, and work-family services, on women's employment rates, work hours, earnings, and family income. LIS has enabled research on the employment/earnings of single mothers, young women, and older women, and on household bargaining. See Gornick (2004) for a summary of 20 years of LIS research (as of 2003) on gender differentials and women's economic status across countries. In recent years, innovative new work assesses the earnings penalty associated with working in a "care" profession and the influence of family on occupation.

Political economy, politics, and policy. The inter-relationships among public policy, inequality, and political outcomes (e.g., political behavior and public opinion) constitute a new frontier in research using LIS. Research on varieties of capitalism and inequality are part of the mix of political science and sociology enabled by LIS.

Wealth. The new LWS Database will enable entirely new lines of research on wealth across countries and, in the future, over time. The first empirical papers using these data are diverse; they assess, e.g., the impact of age adjustments on wealth inequality rankings; homeownership as social insurance and/or retirement income; the relationship between

inheritances and wealth distributions; homeownership patterns and access to credit; consumption effects of various types of wealth; older women's wealth and income packages; and cross-national variation in wealth distributions.

5.2.4 Evolution and direction

Adapting to evolving needs

LIS has, for nearly 30 years, grown and evolved in order to adapt to the needs of researchers throughout the world. Beginning with seven countries, the LIS databases now include data from more than 40 countries – spanning Europe, North America, Latin America, the Middle East and North Africa, and Asia. In 2005, incoming LIS Director Janet Gornick initiated a comprehensive review of the LIS Database's 20-year-old data template and harmonization rules. Working with Research Director Jäntti and the LIS staff, she supervised a major review with multiple goals: to assess the quality of a large number of LIS variables and identify ways to increase comparability; to update the LIS data template due to changes in the last two decades in countries' social policies and survey content; and to increase the number and quality of the labor market variables included in LIS. Following this review, several changes were made and introduced into the public use files (PUFs) in 2007. Among them, the construction rules were refined for a number of LIS variables; in some cases (e.g., the expenditure data) LIS adjusted its template to bring it more closely in line with international data conventions.

Furthermore, the pension and family benefit classification rules were revised to reflect widespread policy changes. The treatment of net versus gross incomes was also revised and additional person-level income variables were added, both enabling more precise analyses. The labor market variables were restructured and substantially expanded. This enables the many researchers who use LIS primarily for employment studies to go further in their comparative analyses. In June 2007, LIS introduced the revised template to its users. These changes dramatically improved both cross-country and over-time comparability.

In 2009, the LIS staff began a second major round of LIS Database template revisions.[2] Two major factors motivated the timing and content of this restructuring: the inclusion of an increasing number of

2 Implementation of this revision – referred to as the 2011 Template – is linked to the release of the Wave VII (centered on 2007) micro-data. Most

datasets from middle-income countries, which necessitated some conceptual adjustments, and changes to the list of harmonized variables. This revision began with the recognition that while the previous template revision (described above) increased the quality of the harmonized LIS data, it did not necessarily increase its user-friendliness. Thus, the main objectives of this second restructuring were to adapt the LIS Database template to maximize its applicability to datasets from both high- and middle-income countries, and to introduce a more user-friendly structure for LIS' data and documentation.

To meet these overarching objectives, this revision was guided by several principles and goals: (1) to restructure the variables, especially the income variables, to achieve a more logical, comparable, and comprehensive list; (2) to standardize most of the variables, which led to the use of fewer country-specific codes; and (3) to introduce easy-to-use dummy or categorical variables to complement the more detailed ones that are still provided. In summary, the goal was to produce a revised template that would increase both over-time and cross-national comparability, and that would require LIS' data users to make fewer assumptions and to do less recoding as they carry out their research.

In the last five years, other advancements have modernized and expanded LIS. Research using the LIS databases was recently facilitated by the launch of a new job submission interface ("JSI"), which offers a range of innovative features that improve the process of composing, sending, and tracking programs sent to LISSY. In addition, LIS recently launched an email-based "User Outreach Campaign" aimed at introducing LIS to thousands of data users in the countries that participate in the LIS and LWS Databases. Finally, in 2011, LIS launched a dramatically upgraded new website; the revised website also introduced online registration for the first time.

Future goals

LIS' core goal over the next five years is to increase and diversify its data holdings. Traditionally, the LIS Database' surveys came primarily from high-income countries – with a concentration in Europe and North America. During most of LIS' existence, a few middle-income

---

components of this revised template have also been applied, retroactively, to all earlier waves of the micro-data.

countries have been included, e.g., Mexico and from a few Eastern European countries.

Starting in 2007, LIS made it a high priority to add more datasets from middle-income countries. This growth plan was launched in 2007 with the addition of datasets from five Latin American countries: Brazil, Colombia, Guatemala, Peru, and Uruguay. Datasets from four more Latin American countries (Chile, Dominican Republic, Panama, Paraguay) will be added soon. In addition, LIS recently added datasets from South Africa, India, and China, and the data staff is now harmonizing data from Russia and Egypt. As its data holdings grow, LIS will enable researchers to study an increasingly diverse set of countries, and to tackle a larger universe of questions. LIS also has plans to expand the number of datasets contained in the LWS Database. As of 2012, the LIS staff and Directors are working with several international organizations, including the OECD and the European Central Bank, to acquire additional wealth datasets and also to further the process of developing universal wealth data and measurement standards.

Between 2012 and 2016, LIS' overarching plan is – first – to continue its core work: acquiring, harmonizing, documenting, and making available (through the three pathways) high quality, cross-national income, wealth, and employment data. LIS also plans to continue its teaching and networking activities, including its User Support services, the international and national training workshops, the LIS and LWS Working Paper series, its annual Research Awards, Visiting Scholar program, and the holding of periodic international research conferences. Furthermore, LIS will continue its on-going data quality improvement efforts, institutionalized since 2004, aimed at improving cross-country and inter-temporal comparability.

Finally, LIS anticipates adding some new products. In the next five years, LIS intends to build and introduce a streamlined, modernized, searchable, storage system for dataset documentation, as well as an entirely new set of "e-learning tools". One result of the recent developments (chronicled in this article) has been a remarkable increase in the use of LIS' micro-data: the number of newly registered LIS users more than doubled between 2010 and 2011, from 236 to well over 500; LIS expects steadily increasing user numbers. Each summer, about 30 researchers learn to use the LIS and LWS micro-data on-site in Luxembourg – at LIS' well-known summer workshops – and some of them return to their home countries to teach others. Other researchers

learn to use the micro-data entirely "virtually", meaning without ever attending a LIS training workshop or learning ("face to face") from past workshop attendees or other LIS users. Currently, about 50 per cent of LIS users learn to use the data entirely virtually. One of LIS' goals is to raise that percentage to approximately 80 per cent. That increase is necessary because LIS' growing number of users will quickly outstrip its workshops' capacity. It is also necessary because a rising share of LIS users are from under-resourced countries, and for these users, international travel is often not feasible.

5.2.5 Barriers

At LIS, building and maintaining the data infrastructure – both LIS and LWS – is at the heart of its responsibilities and accomplishments. LIS relies on its funders, mostly public institutions in the participating countries, to maintain that infrastructure. Although LIS' funding has largely survived the strains of the recent global financial crisis, the current level of funding is clearly inadequate. Financial resource constraints are such that the LIS staff is far too small for the current workload. One of the main consequences of the staffing shortage is that the lag time between dataset acquisition and release is longer than would be ideal.

A second barrier for LIS stems from the diversity of data collection methods across countries. Although there has been some development of standard practices of collecting income data, the data that LIS receives are still very diverse. If income and especially wealth micro-data were more similar across countries, LIS would be able to provide a greater quantity of harmonized micro-data and the harmonized data would be even more cross-nationally comparable; furthermore LIS would be able to provide the harmonized data in a more timely way. Third, a few important "high profile" countries have not yet agreed to participate, usually for legal and/or political reasons. That, unfortunately, limits the overall scope and usefulness of the LIS and LWS Databases.