An overview of the LIS databases

Teresa Munzi
Luxembourg Income Study

Fifth Winter School on Inequality and Social Welfare Theory (IT5)
University of Verona
January 11-14 2010, Alba di Canazei

OUTLINE

1. Introduction
2. The LIS databases
3. Working with the data
4. An example of analysis using LIS data

OUTLINE

1. **Introduction**
2. **The LIS databases**
3. **Working with the data**
4. **An example of analysis using LIS data**

History

- **Luxembourg Income Study (LIS):**
  - ✓ **funded in 1983 by a team of multi-disciplinary researchers in Europe and the US**
  - ✓ **aim: to serve a global community of researchers, educators, and policy makers**
  - ✓ **since 2002: independent non-profit institution**

- **Financial support:**
  - ✓ **Luxembourg government**
  - ✓ **national science foundations and other funders in member countries**

- **Who's who**
  - ✓ **7 staff members in LIS offices in Luxembourg**
  - ✓ **Janet Gornick, LIS Director, and her assistant as well as the Research Director Markus Jäntti are located abroad**

Use of LIS data

- **Since the founding of LIS, over 2000 researchers used the data**
  - ✓ **to analyze variation in socio-economic outcomes within and across countries**
  - ✓ **to study the effects of economic and social policies on outcomes**
  - ✓ **outcomes include poverty, income inequality, employment status, wage patterns, gender inequality…**

- **LIS has contributed to four major fields of study**
  - ✓ **refinement of the income concept**
  - ✓ **study of income distributions across the richest countries**
  - ✓ **conceptualization and measurement of income inequality and poverty, and proper identification of international rankings and trends in that regard**
  - ✓ **women's economic status and/or economic gender inequality**

LIS

Today LIS is a large archive of cross-national harmonised microdata on income and (more recently) wealth

- **What is microdata?**
- **What is data harmonisation?**

Microdata

What is microdata?

- **Microdata is original data that contains every individual record (e.g. person, household, company, etc.) in the research samples**

Household level file			
Household identifier	Number of individulas	Region	Total Income
1	4	Piemonte	100
2	1	Sicilia	50
...	...	...	
N	3	Lazio	100

Person level file				
Person identifier	Household identifier	Relationship to head	Sex	Age
1	1	head	male	45
2	1	spouse	female	40
3	1	child	female	15
4	1	child	male	12
5	2	head	female	65
6	2	spouse	male	67
...		...	...	...
n-2	N	head	female	50
n-1	N	sibling	female	56
n	N	child	male	25

Microdata

What is microdata?

- **The same database can consist of several levels nested into each-other**

Household level file			
Household identifier	Number of individulas	Region	Total Income
1	4	Piemonte	100
2	1	Sicilia	50
...	...	...	
N	3	Lazio	100

Person level file				
Person identifier	Household identifier	Relationship to head	Sex	Age
1	1	head	male	45
2	1	spouse	female	40
3	1	child	female	15
4	1	child	male	12
5	2	head	female	65
6	2	spouse	male	67
...		...	...	...
n-2	N	head	female	50
n-1	N	sibling	female	56
n	N	child	male	25

Data harmonisation

Why harmonise?

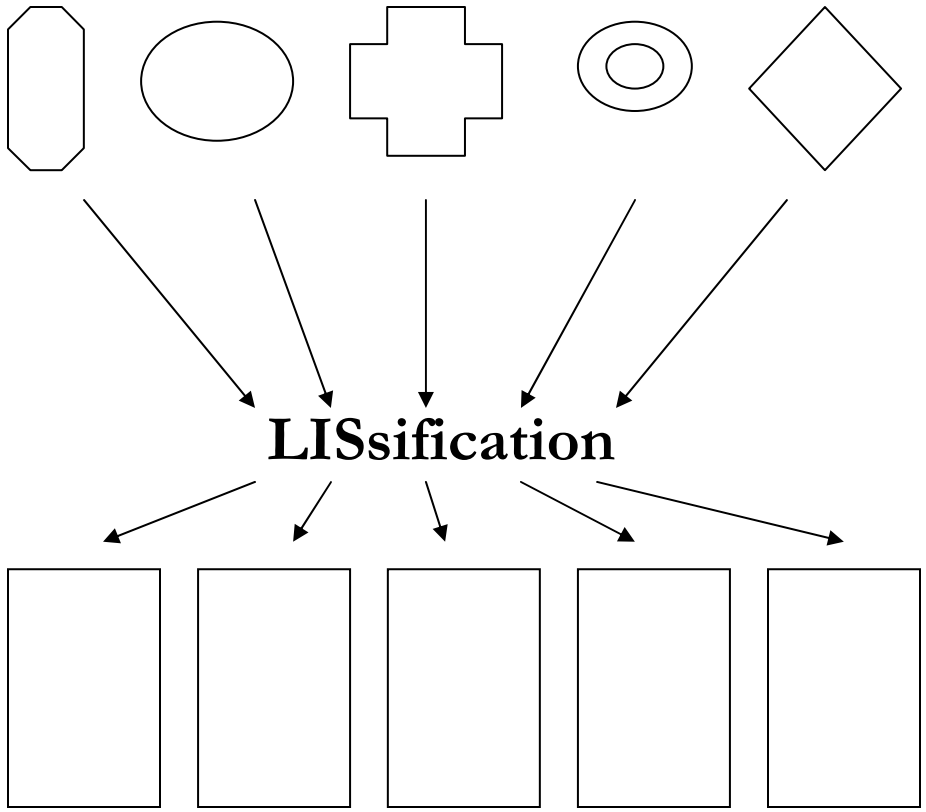
- **Users' perspective**

  Necessity of obtaining comparable micro-datasets for the purpose of carrying out cross-national research
  - ✓ **technically**
  - ✓ **conceptually**

- **LIS' perspective**

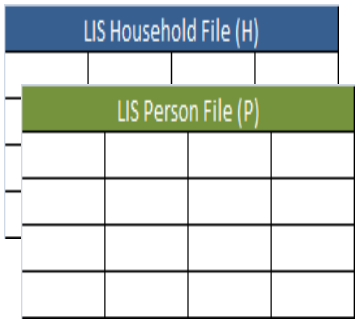
  LIS does not organize surveys but collects data from existing data sources that have:
  - ✓ **microdata (household and person level)**
  - ✓ **representative of the whole population**
  - ✓ **good quality income/wealth information**
  - ✓ **main demographics and (possibly) labour market information**

Harmonization in a nutshell

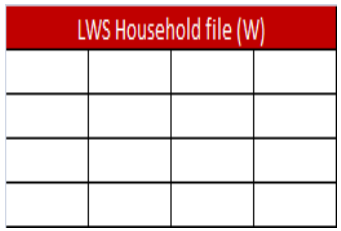


- **Final output: LIS/LWS Datasets**

LIS files



LWS files



- **Final output: LIS/LWS Variables**
  - **same names and labels**
  - **same content - definition as comparable as possible**

Harmonization allows LIS users to eliminate many of the potential sources of non-comparability

Harmonization Process

The 4 steps of LISsification:

- Get the original data
- Collect the documentation (translation!)
- Create uniform file structure
  - ✓ adapt input from panel-surveys for cross-sectional analyses
  - ✓ aggregate data to person or household level
- Harmonize / standardize the variables

Harmonization Challenge

Make comparable original data that are:

- **from various countries**
  → different institutional / societal setups
- **over time**
  → changes in institutions and original surveys
- **household / individual level data**
  → confidentiality issues
- **from various existing datasets**
  → output (or *ex-post*) harmonisation

Output Harmonization

- Different types of original collection instrument
  - ✓ Survey versus administrative data (coverage and contents)
  - ✓ Cross-sections versus panels (sample selection)
- The concepts used in the original data collection are different
  - ✓ Different definitions (employment concept, total household disposable income concept, gross vs net incomes, assets vs net worth, etc.)
  - ✓ Different universes and reference periods
  - ✓ Country-specific categories (especially in education, social security benefits)
- The level of detail of information collected differs
  - ✓ Labor market (e.g.: LFS type of survey)
  - ✓ Incomes /wealth (detailed breakdown vs. overall questions)
- Different statistical techniques
  - ✓ Different sampling procedures (e.g. oversampling of the rich)
  - ✓ Weighting procedures (self-weighted, sampling weights, etc.)
  - ✓ Treatment of missing values, imputation methods

Labour Market Variables

Very difficult to create comparable variables

- **Many international guidelines and recommendations, but those are usually only applicable to data from LFS**
  → different definitions
- **Rigid routing of the labour market questions**
  → different universes
- **Information refers to a variety of time points or ranges**
  → different reference periods
- **Country-specific codes and situations**
  → different categories

Income Variables

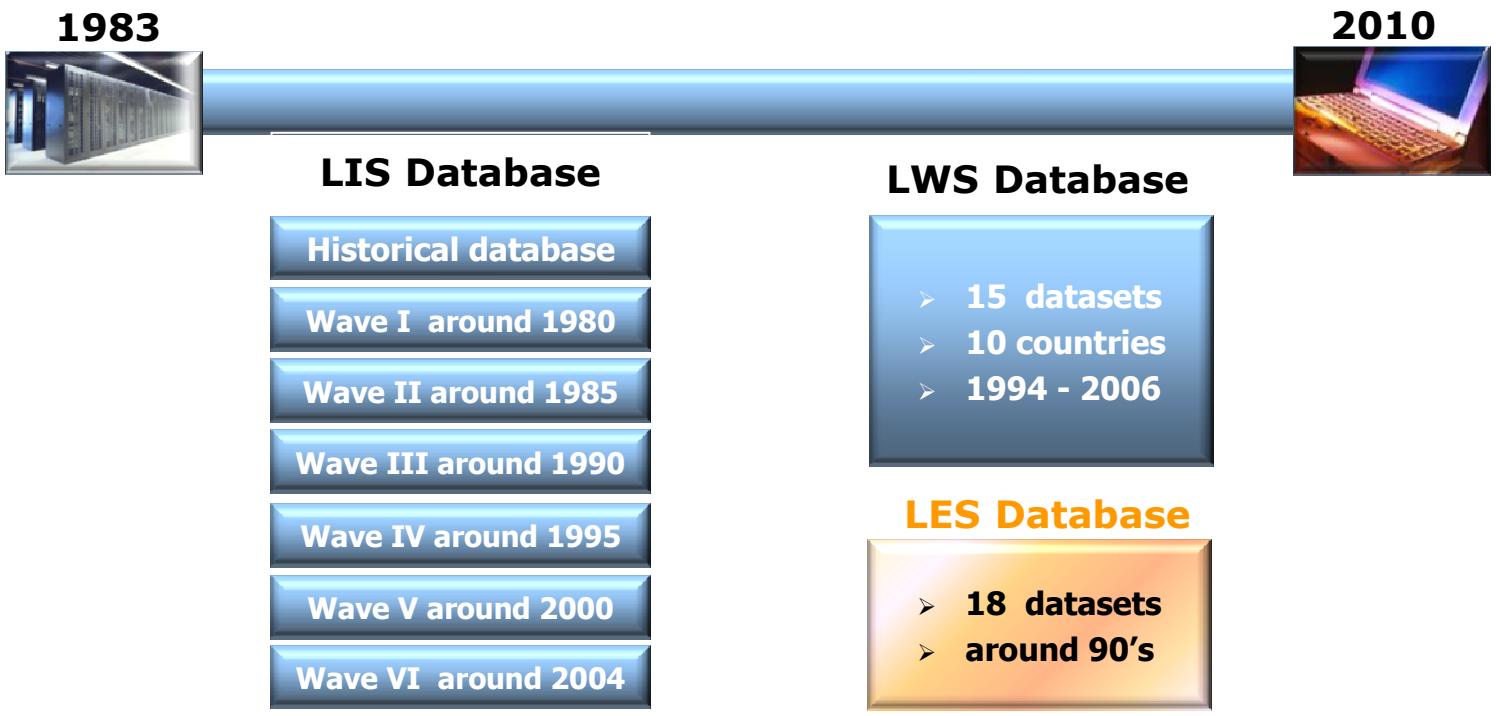
- **Income sources included in total household disposable income (irregular payments, non cash incomes, imputed rents, non-taxable incomes, "informal" incomes)**
- **Current versus annual**
- **Net versus gross (or in between…)**
- **Top- and bottom-coding**
- **Level of detail (e.g. total pensions) and different aggregation (e.g. pensions by detailed fund but not by function)**
- **Classification of incomes:**
  - ✓ **Public versus private**
  - ✓ **Social insurance versus social assistance**

OUTLINE

1. Introduction
2. The LIS databases
   - The LIS database
   - The LWS database
3. Working with the data
4. An example of analysis using LIS data

LIS Micro-Databases

- **LIS now includes micro-data from 37 countries, mostly in Europe and North America, but also including Middle East, Asia-Pacific and Latin America.**



LIS Database

LIS Database

- Historical database
- Wave I around 1980
- Wave II around 1985
- Wave III around 1990
- Wave IV around 1995
- Wave V around 2000
- Wave VI around 2004

Content of the LIS datasets:

- Household composition and characteristics (number of persons, tenure, etc…)
- Socio-demographic characteristics of household members (age, marital status, education, immigration)
- An extensive set of Labor Market data
- Household expenditures (COICOP Level 1)
- Detailed breakdown of Household and Individual Income Data

LIS database

Labor Market Variables

- Labor Force Status (current versus usual, ILO versus main activity)
- Characteristics of the Individuals
  - Information about overall employment (hours and weeks worked, work experience)
  - Other (search for job and care-giving responsibilities)
- Characteristics of the Primary Job
  - Status in employment, occupation, industry, tenure, etc.

Semi-standardization

Coding across countries is similar enough to allow quick comparisons but Country-level detail is retained

	PSEARCH	ES00	SI99
100	LOOKING FOR JOB		X
	101 looking for job replacing main job	X	
	102 looking for job in addition to main job	X	
	103 looking for job in addition or replacing current job	X	
	104 looking for job for nonworkers	X	
200	NOT LOOKING FOR JOB	X	X
	201 not searching for employment		
900	INDISTINGUISHABLE		
	901 will start work soon		X

LIS database

- **Income Variables**
  - **Total income breakdown only at the household level**
  - **Gross of income taxes and social contributions**
  - **Annual**
  - **Units of national currency**

LIS database

Income Variables

- **Private income**
  - ✓ **earnings from labor**
  - ✓ **capital and property income**
  - ✓ **income from private pensions**
  - ✓ **other private transfers**
- **Public Transfers**
  - ✓ **Social insurance transfers**
  - ✓ **Social assistance transfers (cash, near cash and non-cash)**
- **Taxes and contributions**
  - ✓ **income taxes, property taxes, other direct taxes**
  - ✓ **social contributions**

INCOMES

		cash		non-cash	
regular	Labour income	Income from dependent employment	V1	In-kind benefits	V6
				Employers' social security contributions	V2
		Income from self-employment	V4 + V5	Value of goods and services produced and consumed by the household	ALTNCASH
				Value of goods received in barter	
	Capital income	Interest and dividend Rental income Private savings plans	V8 (- V8S3)		
	Private transfers	Alimony received	V34		
		Transfers from other households	V35	Gifts from other households	
	Private pensions	Individual pensions (3rd pillar)	V8S3		
		Occupational pensions	V32 + V33		
	Public income	Public pensions	V17S2 + V18S1 + V19 + V23		
		Temporary replacement income	V16 + V17S1 + V21 + V22		
		Universal benefits	V18S2 + V20 + V24	Health, education, transport (STIK)	
		Social assistance benefits	V25 + V26	Social assistance goods and services	V27 to V31A
irregular	Labour income	Severance pay	V37SR		
	Capital income	Capital gains	V37S1		
		Compensation from private insurance	V37SR		
	Private transfers	Monetary gifts Inheritances Lottery gains	V37SR	Gifts from other households	
	Private pensions	Compensation from private pensions	V37SR		
	Public income	Compensations from public insurance	V37SR		
		One-off universal grants	V37SR	Health, education, transport (STIK)	
		One-off assistance grants	V37SR	Social assistance goods and services	

EXPENDITURES

consumption	Food and non-alcoholic beverages		FOODEXP
	Alcoholic beverages, tobacco and		ALCOEXP
	Clothing and footwear		APPEXP
	Housing, water, electricity, gas and other fuels		HOUSEXP
	Furnishings, household equipment and routine household maintenance		EQUIPEXP
	Health		MEDEXP
	Transport		TRANEXP
	Communication		COMMEXP
	Recreation and culture		CULTEXP
	Education		EDUCEXP
	Restaurants and hotels		RESTOEXP
	Miscellaneous goods and services		MISCEXP
non consumption	mandatory payments	Income taxes	V11
		Other taxes	V12 + V14 + V15
		Social security contributions	V7 + V13
	non mandatory payments	Other insurance contributions	
		Private transfers	V34X + V35X
		Interest	V8X
		Other	

LIS database

Summary Income Variables

Ready-made LIS subtotals to facilitate cross-national comparability

Market Income (MI = EARNING (V1 + SELFI) + V8 + PENSIOI)

+ Transfers (TRANSI = SOCTRANS (SOCI + MEANSI) + PRIVATI)

+ Other Cash Income (V36)

- Taxes and Contributions (V11 + PAYROLL)

= Disposable Household Income (DPI)

LWS Database

LWS Database

- Austria 2004
- Canada 1999
- Cyprus 2002
- Finland 1994 - 1998
- Germany 2001
- Italy 2002 - 2004
- Norway 2002
- Sweden 2002
- United Kingdom 2000
- USA (PSID) 2000
- USA (SCF) 00 – 02 -04

Content of the LWS datasets

- Household composition and characteristics (number of persons, tenure, etc…)
- Socio-demographic and employment characteristics of the household head and spouse
- Household expenditures (COICOP Level 1)
- Household income data
- A detailed set of wealth variables (including some summary wealth variables)
- An extensive set of behavioural variables (risk attitude, health, expectations, reasons for saving, etc.)
- Some dataset-specific vairables

LWS database

Wealth Variables

- **Financial Assets**
  - ✓ **Deposit Accounts: Transaction, Savings and CDs (DA)**
  - ✓ **Total Bonds: Savings and Other Bonds (TB)**
  - ✓ **Stocks (ST)**
  - ✓ **Mutual Funds and other investment funds (TM)**
  - ✓ **Life Insurance (LI)**
  - ✓ **Other investmentments / financial assets (non-pension) (OFA)**
  - ✓ **Pension Assets (PA)**
- **Non Financial Assets**
  - ✓ **Principal Residence (PR)**
  - ✓ **Investment real estate (IR)**
  - ✓ **Business Equity (BE)**
  - ✓ **Vehicles (VH)**
  - ✓ **Durables/Collectibles (DR/ CL)**
  - ✓ **Other non-financial assets (ONF)**

LWS database

- **Wealth Variables (cont)**
  - **Liabilities**
    - **Total Liabilities (TD)**
      - **Total Home secured debt (HSD)**
        - **Principal residence mortgage (MG)**
        - **Other property mortgage (OMG)**
        - **Other home secured debt (including line of credit) (OHSD)**
      - **Vehicle loans (VL)**
      - **Total Installment debt (inc. credit card bal) (IL)**
      - **Educational loans (EL)**
      - **Other loans from financial institutions (OL)**
      - **Informal debt (ID)**

LWS database

Net Worth

3 different concepts (ranging from the most comparable and least complete, to the most comprehensive and least comparable):

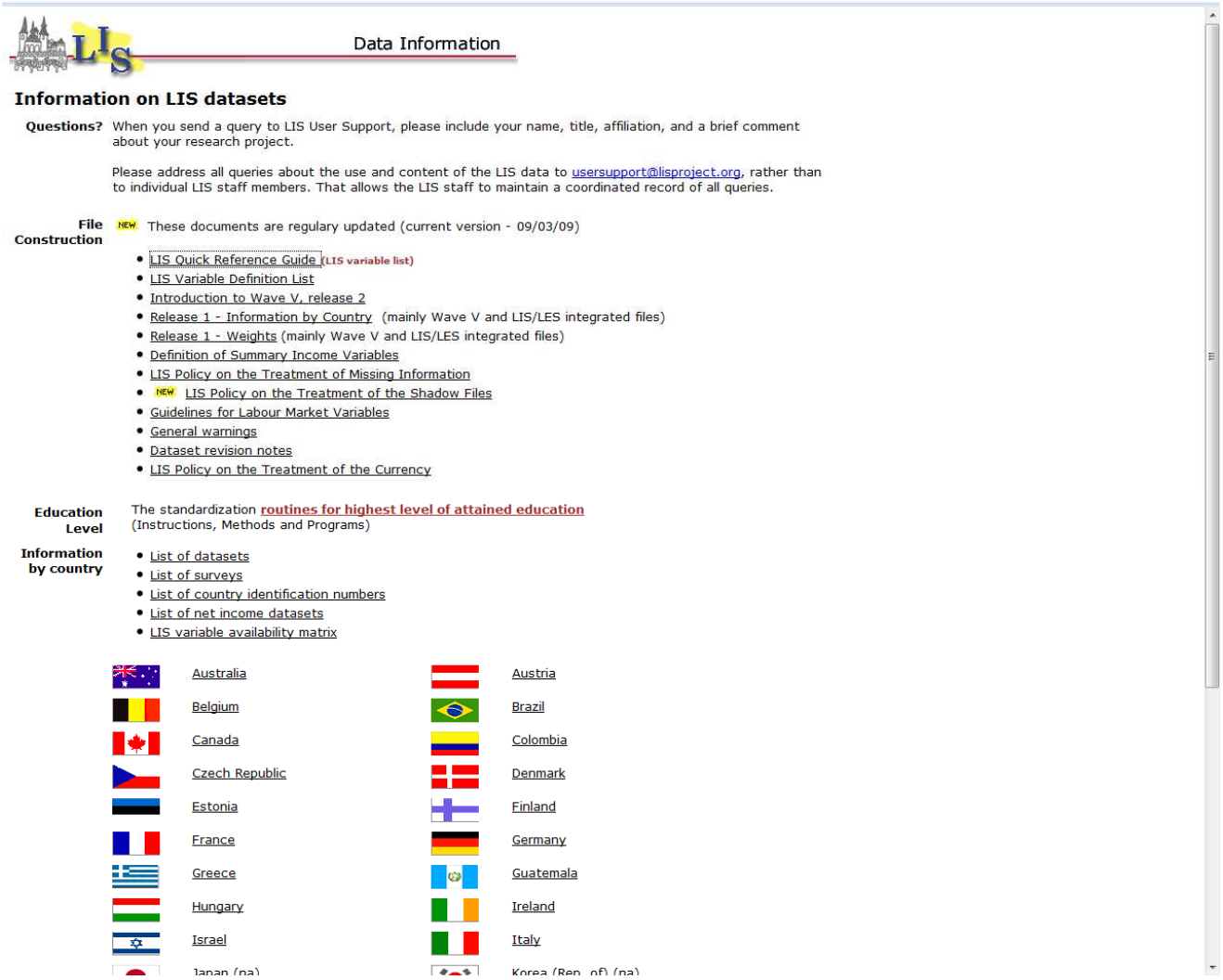
- **NW1 = Total Financial Assets (TFA1 = DA+ST+TB+TM) + Total Non-Financial Assets 1 (TNF1 = PR+IR) – Total Debt (TD)**
- **NW2 = Total Financial Assets (TFA1 = DA+ST+TB+TM) + Total Non-Financial Assets 2 (TNF2 = PR+IR+BA) – Total Debt (TD)**
- **NW = (sum of all assets) - (sum of all debts)**

OUTLINE

1. Introduction
2. The LIS databases
3. **Working with the data**
   - **Documentation**
   - **Data access**
4. An example of analysis using LIS data

LIS Documentation

Documentation available for public access on-line



Data Information

Information on LIS datasets

Questions? When you send a query to LIS User Support, please include your name, title, affiliation, and a brief comment about your research project.

Please address all queries about the use and content of the LIS data to usersupport@lisproject.org, rather than to individual LIS staff members. That allows the LIS staff to maintain a coordinated record of all queries.

File Construction NEW These documents are regulary updated (current version - 09/03/09)

- LIS Quick Reference Guide (LIS variable list)
- LIS Variable Definition List
- Introduction to Wave V, release 2
- Release 1 - Information by Country (mainly Wave V and LIS/LES integrated files)
- Release 1 - Weights (mainly Wave V and LIS/LES integrated files)
- Definition of Summary Income Variables
- LIS Policy on the Treatment of Missing Information
- NEW LIS Policy on the Treatment of the Shadow Files
- Guidelines for Labour Market Variables
- General warnings
- Dataset revision notes
- LIS Policy on the Treatment of the Currency

Education Level The standardization routines for highest level of attained education (Instructions, Methods and Programs)

Information by country

- List of datasets
- List of surveys
- List of country identification numbers
- List of net income datasets
- LIS variable availability matrix

Australia	Austria
Belgium	Brazil
Canada	Colombia
Czech Republic	Denmark
Estonia	Finland
France	Germany
Greece	Guatemala
Hungary	Ireland
Israel	Italy
Japan (na)	Korea (Rep. of) (na)

LWS Documentation

Documentation available for public access on-line

Information on LWS datasets

Introduction
ANNOUNCEMENT - *December 6, 2007*

Development of the project (2002-2007)

File Construction

- Data Availability
- Quick reference guide(including variable list)
- Name and label of extra variables by country
- Missing values policy
- Behavioral variable mapping
- Variable list and variable definitions
- LWS Aggregate Variables Construction
- Technical Report on LWS Income Variables

Information by country

- List of datasets

Questions? Contact our User Support

Austria	Canada
Cyprus	Finland
Germany	Italy
Norway	Sweden
United Kingdom	United States

Generic Documentation

- **Variables standard definitions**
  - For each variable: contents (incl. any comments or warnings), the ideal universe and reference period, the formula if LIS-generated, the standardized values and labels, changes over time and any recommendations for use
  - Specific and detailed guidelines and construction rules for some variables blocks (LIS / LWS summary variables, LIS Labor market variables, LWS behavioral variables)

- **Generic Policies**
  - Policy on the treatment of missing information
  - Policy on the treatment of currencies
  - Technical Report on LWS Income Variables

Dataset overview Documentation

- List of datasets
- List of surveys
- Gross / Net datasets table
- Education standardisation methodology
- Variable availability matrix

Dataset-specific documentation

Data Information

Italy

- Data Provider
- Lissification viewer (Excel format) (wave II to IV)
- Weighting procedures

Warnings:
- IT91, IT95 : dataset revised
- IT86: Person file contains information for a maximum of 5 adults
- All datasets: net income variables only
- Income amounts are in thousands of national currency units.

Wave	Survey Information	Lissification Tables	Institutional Information	Unweighted Basic Descriptives	Labour Market mapping
II	1986	1986	1986(na)	1986H - 1986P	(na)
	1987	1987	1987	1987H - 1987P	(na)
III	1989	1989	1989	1989H - 1989P	(na)
	1991	1991	1991	1991H - 1991P	(na)
IV	1993	1993	1993	1993H - 1993P	(na)
	1995	1995	1995	1995H - 1995P	(na)
V release 2	1998	1998	1998	1998H - 1998P	1998
	2000	2000	2000	2000H - 2000P	2000
VI	2004	2004	2004	2004H - 2004P	2004

Dataset-specific documentation

Survey Information

Necessary information to understand how the data were collected

- **Technical information on the original survey (e.g.: how the data were collected - reference period, sample, response rates, etc.)**

"Lis/Lwssification" Tables, Descriptive and Mapping Tables

Understand the harmonization process and correctly interpret results based on LIS/LWS variables

- **Precise definition and contents of each LIS/LWS variable and its mapping**
- **Main unweighted descriptive statistics of LIS/LWS variables**

Dataset-specific documentation

- **Institutional Information**

  Understand the context within the LIS/LWS data have been collected

  - **Exhaustive information on the tax and transfer programs corresponding to micro-data LIS variables (including legislation, coverage, qualifying conditions, benefits level, accumulation with other income, adjustment and financing)**
  - **Country-level information on policies, practices, and demographics related to the LWS datasets**

- **LIS also houses *external Institutional documentation sources***

  - **Comparative Welfare States and Family Policy Databases that contain arrays of country-level policy indicators to allow to link policy variables to micro-level outcomes**
  - **The Fiscal Redistribution dataset that offers a number of measures of fiscal redistribution based on calculations made with LIS data**

http://www.lisproject.org/publications.htm

Additional source of information

LIS/LWS Working Papers series

- **Each completed study is published in either the LIS or the LWS Working Paper series**
- **The LIS WP series currently numbers 523 papers, while the newly started LWS WP series numbers 9 papers**

Publications

How to look up for LIS working papers

1. Check one or several checkbox(es) to choose the criteria you want to select
2. Build-up your request by selecting the appropriate information in each selected checkbox (ex: less than or equal to 2005)
3. Click on "Send" button to get the result or on "Reset" button to clear your current selection
4. **Notes**
   - You can make a multiple selection (CTRL+*"your selection"*) while making a request based on either keyword or country
   - Selection by countries is not available for the first 140 working papers
   - **TO DISPLAY FULL LIST OF PAPERS, simply press "Send" button (without checking any box)**

Build-up your request

Author

Year: less than or equal to

Keyword(s): poverty, poor, low income; absolute poverty, real income, real standards of living; inequality, income distribution, income spread, wage distribution, wage spread; social exclusion, social inclusion; social policy, social welfare policy

Countr(ies): Australia, Austria, Belgium, Canada, Chinese Taipei/Taiwan

Send Reset

LIS Working Papers series

522 working paper(s) match(es) your request **(selected on 13-Oct-2009, 13:55:34)**

No. 522 - Policy Effects on Class-Gender Employment Intersections
by Lynn Cooke **Sep-2009**

This project explored how the sociopolitical context maps current class-gender intersections in relative employment equality in Australia, East and West Germany, Spain, the United Kingdom, and the United States. The countries were selected based on their diverse policy equality logics codified in initial welfare state provisions. Pooled and individual-country analyses of wave 5.2 of the Luxembourg Income Study revealed gender differences in the impact of individual factors on work hours and wages, as well as national differences controlling for individual characteristics. Two findings bear particular note. First, the differences in relative gender earnings inequality across the class distribution in Australia and West Germany underline that class equality policies do not ensure greater class equality for all social groups. Second, the UK and US results indicate that liberal market forces do not ensure women's greater investment in education and work hours will achieve economic equality with men. As women's 'human capital' increases, men's returns to their own increase such that gender employment equality becomes a moving target.

P: Forthcoming: "The Current State of Employment Equality," Chapter 6 in Equality for Some, New York: Routledge

No. 521 - Poverty and Inequality among the Elderly. A Comparative Study of the Pension Systems in Belgium and the Netherlands
by Ingrid Keupers **Aug-2009**

Text in Dutch. English Abstract only: Pension systems in European countries are under review due to demographic changes. As a solution to the graying population, second and third pillar pensions are advocated. However, it is important not to lose sight of the social consequences of encouraging these private pensions. In this master thesis the paradox of redistribution from Korpi & Palme (1998) is tested on two cases namely Belgium and the Netherlands. Based on micro data on household income as found in the LIS database, pensioners are compared on their poverty rates and inequality. The redistribution paradox states that poverty will be the highest in pension systems that are targeted and will be the lowest in universal systems. In Belgium (targeted minimum pension) poverty rates are much higher than in the Netherlands (universal minimum pension). Poverty rates are higher for elderly women than for elderly men and this gender difference is greater in the Netherlands than in Belgium. The paradox of redistribution further states that inequality will be higher when the first pillar provides flat-rate pensions and will be the lowest for wage coupled pensions that out compete more unequal private pension provisions. In the Netherlands (flat-rate pension), inequality is higher than in Belgium (wage coupled pension) when calculated on the basis of pension income. When disposable income is taken as reference than the inequality is slightly larger in Belgium than in the Netherlands. There is however a much larger reduction in inequality compared to pre-pensioned individuals in Belgium than in the Netherlands. The inequality in income is higher for men than for women both in Belgium and in the Netherlands. In order to minimize poverty and inequality, it is thus advised to further strengthen the pensions systems in both Belgium and the Netherlands not to reduce them.

No. 520 - Wie hat sich die intragenerationale Umverteilung in der staatlichen Säule des Rentensystems verändert? Ein internationaler Vergleich auf Basis von LIS-Daten
by Tim Krieger, Stefan Traub **Jun-2009**

Paper in German Language - Title and abstract in English are provided: "Has intragenerational redistribution become less important in pension systems' public pillar? An international comparison based on LIS microdata. " We empirically investigate whether the significance of intragenerational redistribution in the public pillar of pension systems in 20 OECD countries has changed systematically since the 1980s and whether international convergence of the degree of intragenerational redistribution in terms of the Bismarckian factor can be observed. Based

Data Access

- **Use of data is limited to Social Science Researchers working for an academic institution (incl. Ph.D. & graduate students) or for a government or non-profit research department**
- **No download or direct access is allowed**
- **Three pathways to get access to LIS databases**
  - **The primary means of access is a Remote Execution System called LISSY**
  - **An online table-making service: the Web Tabulator**

    These both paths are for registered users only
  - **Two distinct sets of Key Figures available online**
    - ✓ **LIS Inequality and Poverty Key Figures**
    - ✓ **LIS Gender Key Figures**

LIS Key Figures

- **Two distinct sets of LIS Key Figures**
  - **LIS Inequality and Poverty Key Figures comprise national-level inequality and poverty indicators such as Gini Coefficients, Atkinson coefficients, etc… (all waves)**
  - **LIS Gender Key Figures that include national-level indicators highlighting women's economic outcomes and gender inequality in poverty and employment (Wave V and VI)**

LIS Web Tabulator

- **The Web Tabulator System is an online table-making service that enables to design and generate cross-national descriptive tables based on the underlying LIS datasets**
  - **Without the need for programming**
  - **Prior registration is also required**
  - **Standardized indicators, including multiple measures of real household income, poverty, and income distribution, as well as demographic and labor market variables**
  - **Possibility to export aggregated results in EXCEL or in ASCII formats**

The LISSY System

- The primary means of access is a fully automated software running 24/d 7/week designed specifically for LIS
- SPSS, SAS or Stata batch programs submitted via a Job Submission Interface (JSI) or an email software
- LISSY automatically processes the jobs, generates results and returns them to users on average within two minutes
- To protect the confidentiality, security, and integrity of the micro-data:
  - ✓ Prior to use LISSY, researchers must register
  - ✓ A few exceptions to the user's programming style are needed
  - ✓ Checking for illegal commands or sequences of commands that would end up breaching the rules on data confidentiality are systematically filtered by LISSY before sending the result back

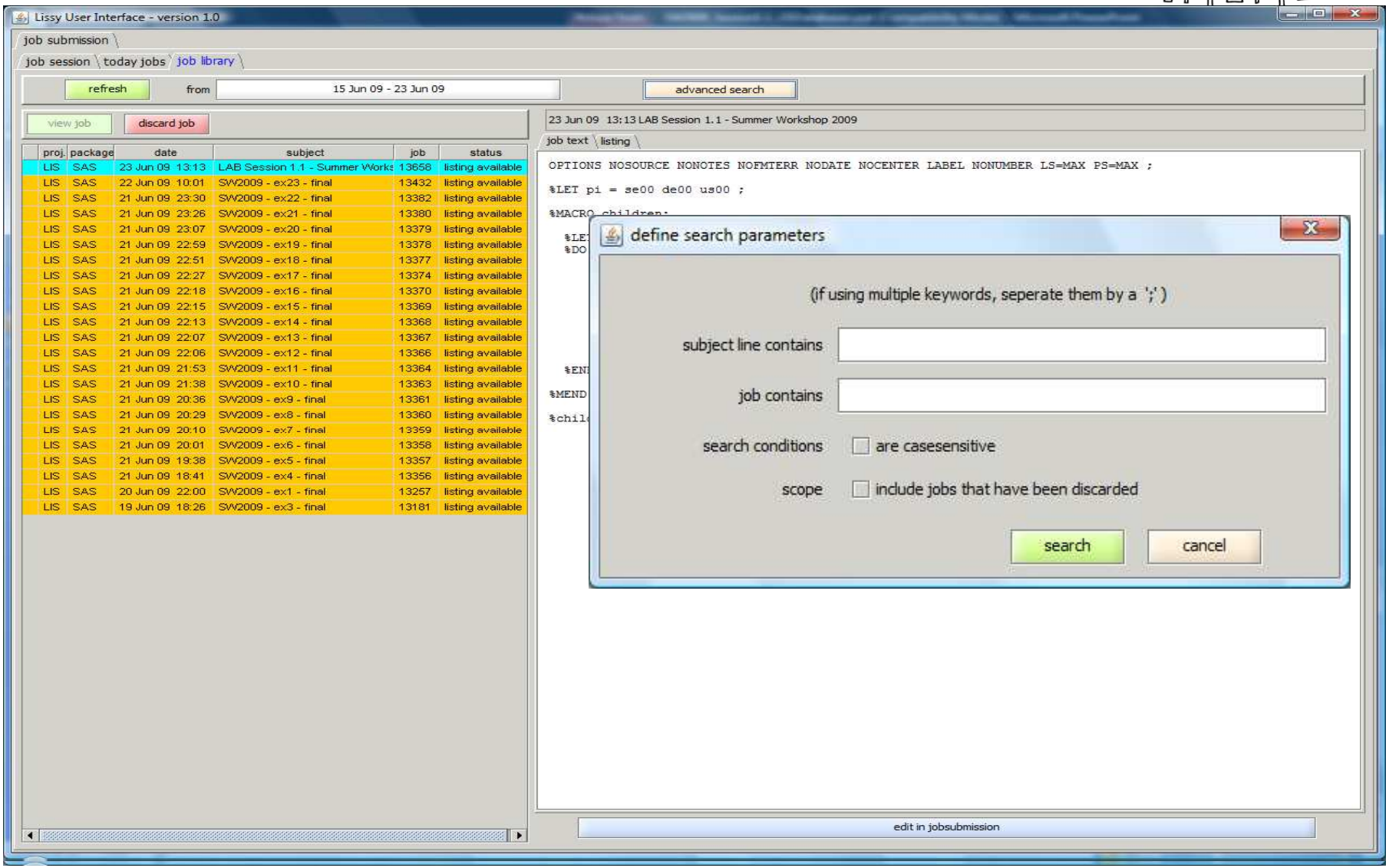
The LISSY JSI

- **Job Submission Interface**
  - **Write, submit and view requests**
  - **Track status of job requests**
  - **Access and manage history of all jobs you ever submitted**

The LISSY JSI

- **The JSI is launched from the LIS Homepage by clicking on the following link - Access LISSY system (version 8)**
  - **Java Runtime (JRE 1.6) must be installed**
  - **Click on Run to accept the LISSY Userinterface Certificate**
- **Connect to LISSY using the userid and password received during the registration process**
  - **the password is case-sensitive**

How JSI Works



Structure of a job

In order for LISSY to properly process requests, a few exceptions to the usual program syntax are required

- **LISSY relies on a three-stage built-in alias to access datasets**
  - **The alias is composed by**
    - ✓ **a statistical package specific heading (&,$ or `*blank*')**
    - ✓ **a two-digit ISO country**
    - ✓ **an abbreviation used to identify the specific type of dataset (h for household, p for individual etc.)**
  - **IT04H to access the LIS 2004 household Italian dataset (STATA)**
- **LISSY does not accept any program commands that allow users to print or read individual records**
  - **Filtering of jobs that would end up breaching the rules on data confidentiality. LISSY automatically puts the job in a security review area to be manually reviewed by the staff**
- **When syntax errors are detected, the language-specific error message are displayed within the listing**

Submission via Email

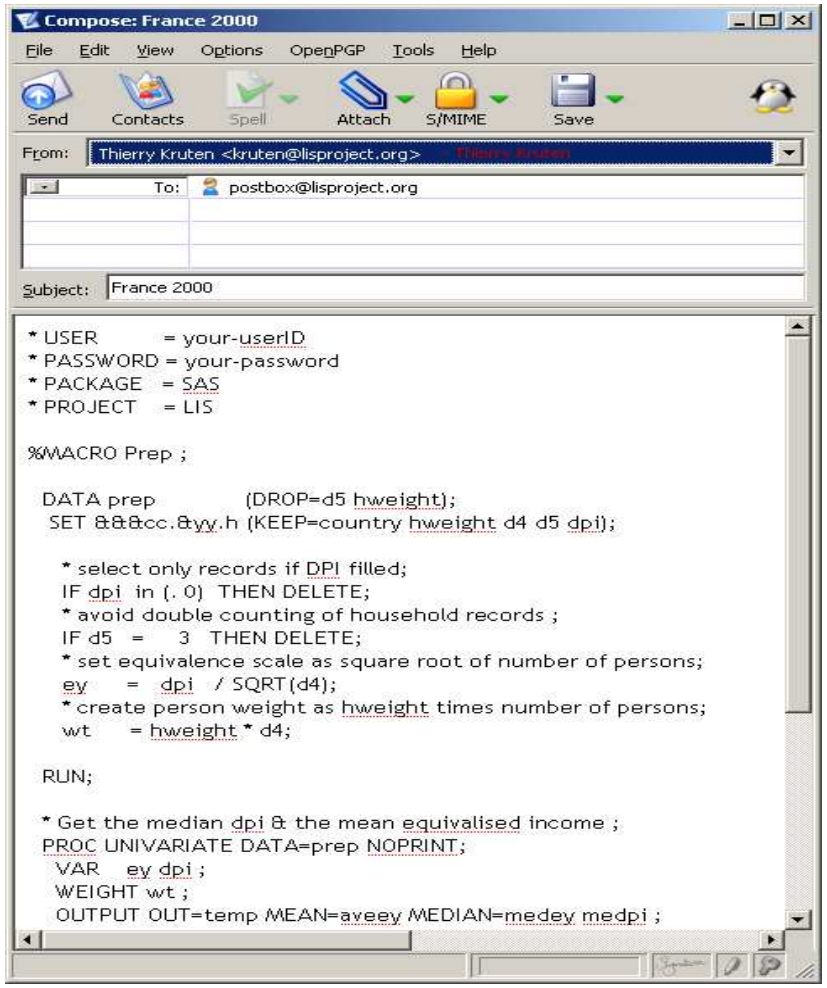
- **Users submit job requests via outlook, thunderbird etc. to postbox@lisproject.org**

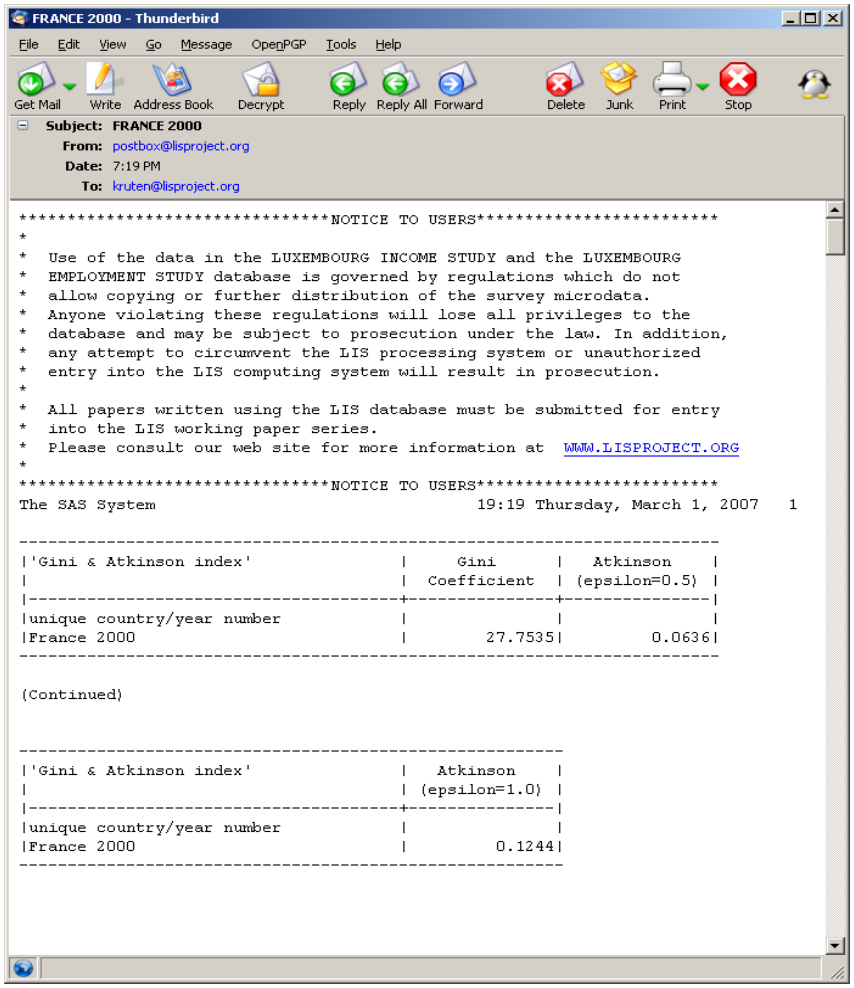
- **Three requirements must be met in order for LISSY to properly process jobs**
  - **All emails must be sent in ascii/plain text format. Users must ensure that this option is enabled in the chosen email package.**
  - **All job instructions must be written inside the body of the email and not as an attachment.**
  - **Each job must start with a specific four-line header at the very beginning of the email body**

```
* user      = your username
* password  = your password
* package   = your statistical package
* project   = LIS or LWS
```

- **Same exceptions to the user's programming style are required**

Submission via Email





Registration Process

- **Access to LISSY under the two following conditions**
  - **Researcher working for an academic, government or non-profit organization**
  - **Use of the micro-data is restricted to Social Science research purposes only. No private or commercial use is permitted**
- **Registration process**
  - **Download, complete, sign and send back a pledge from the LIS website, which states the rules governing the use of the micro-data**
  - **LIS user receives a userid and a password that are strictly personal and must not be shared with anyone**
  - **While registering, researchers must provide an email address. LISSY will only return output resulting from the user requests to this registered email**

OUTLINE

1. Introduction
2. The LIS databases
3. Working with the data
4. **An example of analysis using LIS data**

Job Example

Goal

- We are going to create indicators to help to identify the proportion of poor households (or individuals) and to measure the level of poverty by calculating a main indicator of poverty incidence
- Definition
  - ✓ The Head Count Ratio (HCR) is the percentage of poor individuals in the total population

Measurement issues

- Relative vs absolute poverty: poverty relative to the median of the population (50%)
- Income-resource definition: disposable income
- Persons (Counting Unit) in Households (Sharing Unit)
- Equivalence Scale (LIS equivalence scale: square root of household size)

⇨ **Final measure: Equivalised disposable household income: DPI / square root of household size**

Job Example

Activity

- **Using the 2004 Italian data**
  - ✓ **Run data cleaning procedures and create the equivalence scale**
  - ✓ **Define the poverty line as 50% of the median equivalised household income**
  - ✓ **Calculate the following indicators:**

Median equivalised income	
Poverty line	
How many poor households are in the sample?	
How many poor individuals are there in the total population	
Head Count Ratio	

⇨ **Do not forget to use the weight!!**

Job Example

SAS Program to submit

```
OPTIONS NOSOURCE NONOTES NOFMTERR NODATE NOCENTER LABEL NONUMBER LS=200 PS=MAX ;

DATA pov    (DROP=hweight d4 dpi);
 SET &fi00h (KEEP=hweight d4 dpi);
    IF dpi in (. 0) THEN DELETE;
    ey = dpi /SQRT(d4);
    wt = hweight * d4 ;
RUN;
PROC MEANS DATA=pov NOPRINT;
   VAR ey;
   WEIGHT wt ;
   OUTPUT OUT=temp MEDIAN=medey;
RUN ;
DATA _NULL_;
 SET temp;
   CALL SYMPUT("m",medey);
RUN;
DATA pov (KEEP=eymed povlin gap poor wt);
 SET pov;
    poor = 0 ;
    eymed = &m ;
    povlin = &m * 0.5 ;
    IF ey < povlin THEN
       DO;
         gap = povlin-ey;
         poor = 1 ;
       END;
 RUN ;
 PROC MEANS DATA=pov N MEAN SUMWGT;
   VAR poor eymed povlin gap;
    WEIGHT wt;
 RUN ;
```

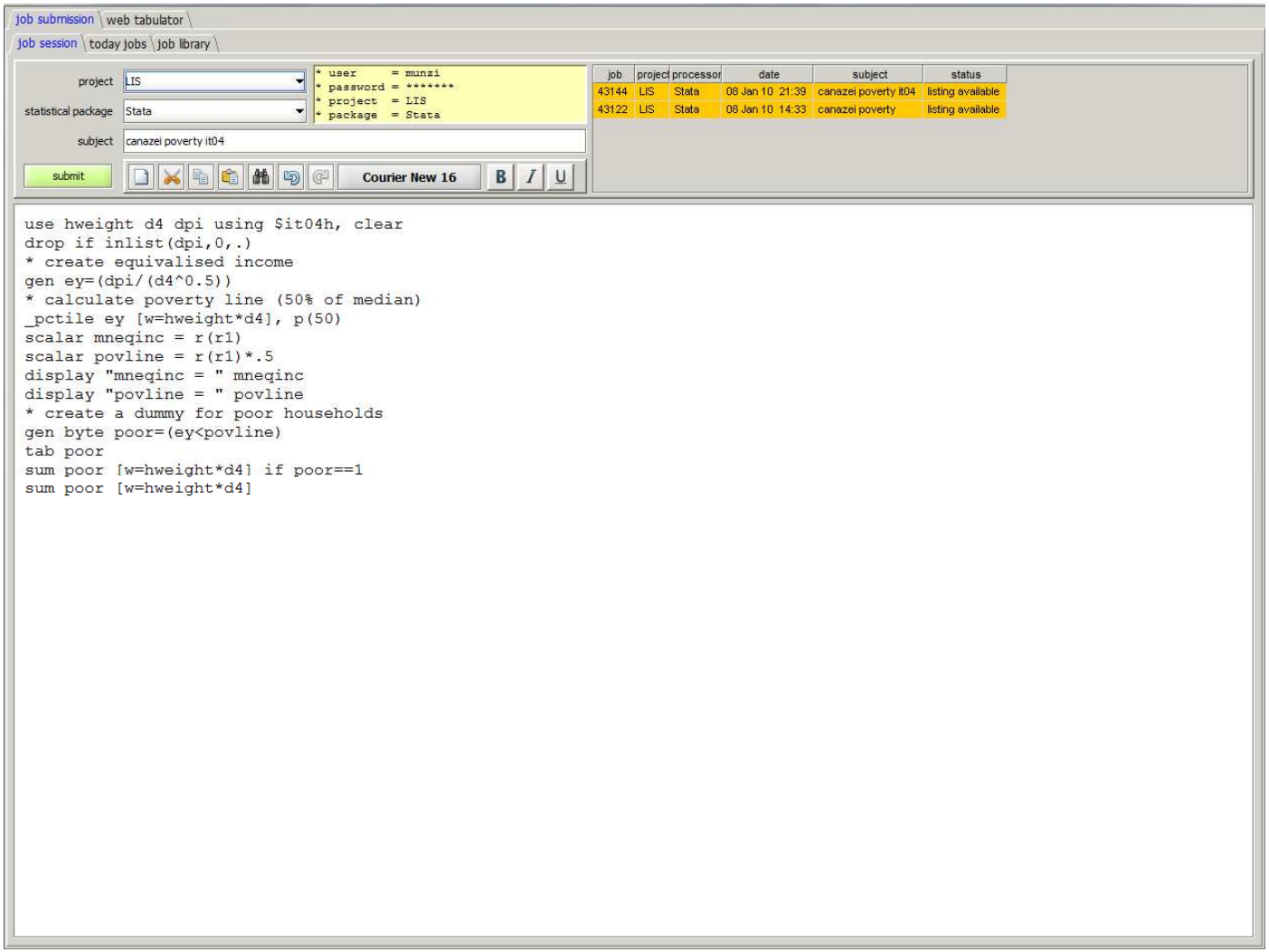
Job Example

- **STATA Program to submit**

```
use hweight d4 dpi using $it04h, clear
drop if inlist(dpi,0,.)
* create equivalised income
gen ey=(dpi/(d4^0.5))
* calculate poverty line (50% of median)
_pctile ey [w=hweight*d4], p(50)
scalar mneqinc = r(r1)
scalar povline = r(r1)*.5
display "mneqinc = " mneqinc
display "povline = " povline
* create a dummy for poor households
gen byte poor=(ey<povline)
tab poor
sum poor [w=hweight*d4] if poor==1
sum poor [w=hweight*d4]
```

Job Example

Job submission



Job Example

Results

Median equivalised income	13,367
Poverty line	13,367 / 2 = 6,684
How many poor households are in the sample?	876
How many poor individuals are there in the total population	6,948,695
Head Count Ratio	12.085%

job submission | web tabulator

job session | today jobs | job library

jobs in process

job	project	processor	date	status	subject

jobs returned view job

job	project	processor	date	subject	status
43144	LIS	Stata	08 Jan 10 21:39	canazei poverty it04	listing available
43122	LIS	Stata	08 Jan 10 14:33	canazei poverty	listing available

```
#  the LIS working paper series.                                              #
#  Please consult our web site for more information at WWW.LISPROJECT.ORG     #
#                                                                             #
#------------------------------ NOTICE TO USERS -----------------------------#

. use hweight d4 dpi using $it04h, clear

. drop if inlist(dpi,0,.)
(16 observations deleted)

. * create equivalised income
. gen ey=(dpi/(d4^0.5))

. * calculate poverty line (50% of median)
. _pctile ey [w=hweight*d4], p(50)
(analytic weights assumed)

. scalar mneqinc = r(r1)

. scalar povline = r(r1)*.5

. display "mneqinc = " mneqinc
mneqinc = 13367.351

. display "povline = " povline
povline = 6683.6753

. * create a dummy for poor households
. gen byte poor=(ey<povline)

. tab poor

      poor |      Freq.     Percent        Cum.
-----------+-----------------------------------
         0 |      7,120       89.04       89.04
         1 |        876       10.96      100.00
-----------+-----------------------------------
     Total |      7,996      100.00

. sum poor [w=hweight*d4] if poor==1
(analytic weights assumed)

    Variable |     Obs      Weight        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------------------
```

edit in jobsubmission

Additional Help

- **If you have any problems contact the LIS user support**

  usersupport@lisproject.org

- **For users who may be unfamiliar with batch coding**
  - **US00 and IT00 (sub-sample) files are downloadable (http://www.lisproject.org/self-teaching.htm)**
  - **Key figures programs (http://www.lisproject.org/key-figures/key-figures-programs.htm)**

Conclusion

Make the access to micro-data easy

- ✓ **Administratively**

  Collection, permission and confidentiality issues dealt at LIS level

- ✓ **Technically**

  Lissy system offers the possibility to perform advanced statistical analyses at your own place of work fast and without any infrastructure

- ✓ **Financially**

  Free for most users, and zero marginal cost for us

- ✓ **Conceptually**

  Eliminate many of the potential sources of non-comparability (core of our business)

Conclusion

Future challenges

- **Extension of geographical area (LMICS project)**
- **Beyond the concept of DPI (inclusion of non-cash and / or irregular incomes?)**
- **Inclusion of microsimulation results (to gross-up net income data)**
- **Further extension of (semi-)standardisation of categories (more semi-standardised variables and/or more standardising include files)**
- **Inclusion of flags for imputed values?**

BUT DILEMMA:

expansion versus quality improvement: where to stop?

Thank you for your attention
Any questions are welcome !

"When you can measure what you are speaking about and express it in numbers you know something about it. But when you cannot measure it or express it in numbers, you knowledge is of a meager and unsatisfactory kind."

Kelvin sir William Thomson
British Mathematician et physician (Belfast, 1824 - Netherhall, 1907)