Luxembourg Income/Wealth Study Database: an overview

Main gaps in research infrastructures in Central-Eastern Europe
Bratislava, 7-8 March 2019

Andrej Cupak

Data Expert and Research Associate, LIS

Outline of the presentation

I. **LIS Data Center**

  – *Who we are*

  – *What we do*

  – *Research with LIS*

II. **Regional gaps**

  – *Country coverage*

  – *Data usage*

  – *Accessibility / funding*

Part I:
LIS Data Center, the institution

LIS history

- LIS was founded in 1983 by two US academics (prof. Tim Smeeding and Lee Rainwater) and a team of multi-disciplinary researchers in Europe. It began as a "study", which later grew and was institutionalized as "LIS"

- For nearly 20 years, LIS was part of a Luxembourg-based research institute known as CEPS (recently renamed LISER). In 2002, LIS became an independent non-profit institution (an ASBL)

- LIS is supported by the Luxembourg government (25%), by the national science foundations and other funders in many of the participating countries (50%), and by several supranational organizations – supplemented by project-specific grants and some private philanthropy (25%)

LIS structure

LIS - Cross-National Data Center **François Bourguignon, President** **LIS Executive Committee** **Daniele Checchi, Secretary General**	
MAIN OFFICE: Luxembourg Office of LIS Daniele Checchi, Director	**SATELLITE OFFICE:** US Office of LIS Janet Gornick, Director
• parent organization • located in Luxembourg, together with the University and LISER • independent, chartered non-profit organization • cross-national, participatory governance • acquires, harmonizes, and disseminates data for research • venue for research, visiting scholars, conferences, and user training	• satellite office • located at the Graduate Center of the City University of New York • one pillar of *Stone Center on Socio-Economic Inequality* • administrative, managerial, development support to parent office • collaborative public programs (lectures, conversations) • venue for research, teaching, and graduate student supervision

L I S

LIS structure (cont'd)

LIS - Cross-National Data Center
François Bourguignon, President
LIS Executive Committee
Daniele Checchi, Secretary General

MAIN OFFICE:	SATELLITE OFFICE:
Luxembourg Office of LIS	US Office of LIS
Daniele Checchi, Director	**Janet Gornick, Director**

Thierry Kruten, Director of Operations and IT	Caroline Batzdorf, Assistant Director
Lucie Scapoli, Administrator Officer	Mei-Ling Israel, Financial Manager
Benjamin Gérard, System and Network Administrator	Laurie Maldonado, Senior Administrator
Data team: Teresa Munzi, Data Team Manager	
Paul Alkemade Heba Omar	
Andrej Cupak Piotr Paradowski	
Jörg Neugschwender Carmen Petrovici	

LIS Senior Scholars	
Prof. Louis Chauvel	Prof. Janet Gornick
Prof. Daniele Checchi	Prof. Paul Krugman
Prof. Conchita D'Ambrosio	Prof. Leslie McCall
Prof. Markus Jäntti	Prof. Branko Milanovic
Prof. Frank Cowell	

What we do at LIS

Core activity: data work

Step 1. Data acquisition

We identify appropriate datasets (reliable, and high-quality data)
We negotiate with each data provider

Step 2. Data harmonization

Common cross-national template
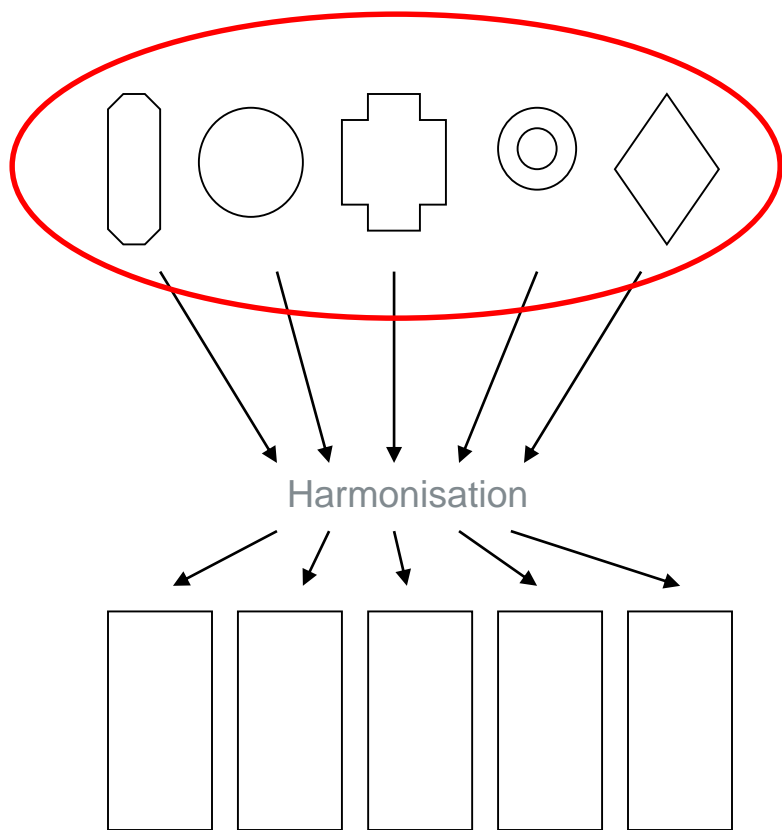Comprehensive documentation

Step 3. Data dissemination

We create national-level indicators (LIS Key Figures)
We provide harmonized microdata to researchers via remote execution (LISSY)

Other activities: research promotion

Conferences, work on methodological issues, collaborations with networks/users/journalists, newsletter, annual report, individual research, user support, research visits
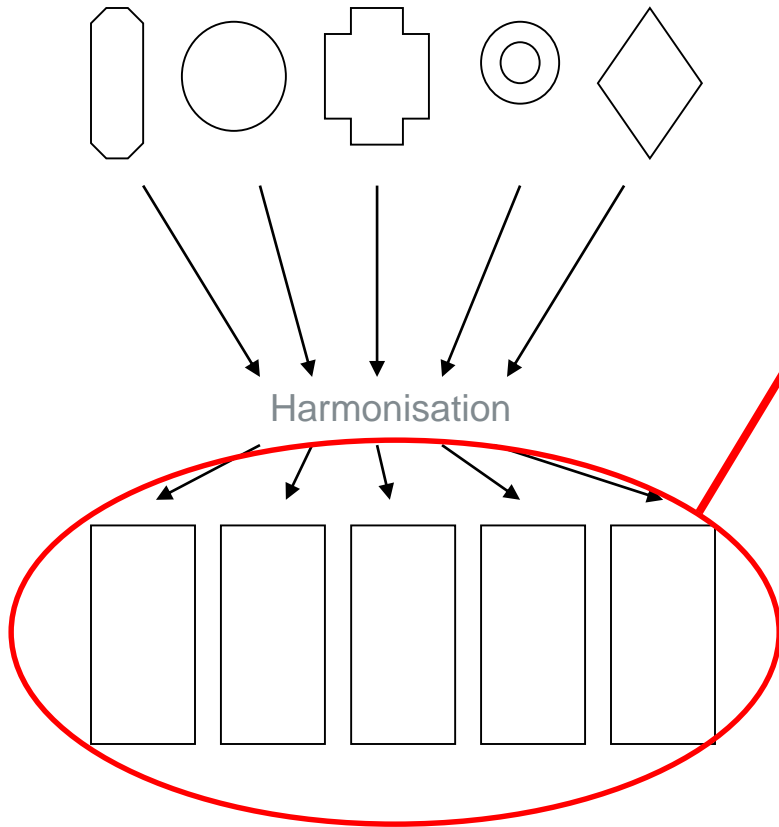
Ex-post harmonisation

The origins of the LIS data

❑ LIS does not organize surveys but collects data from existing data sources:
➢ *Survey data: income, household budget, living conditions, multipurpose, human development*
➢ *Administrative records: tax records, employers records, social security records*
➢ *Any mix of the above*

❑ Common denominator:
➢ *microdata (household- and individual-level)*
➢ *representative of the whole population*
➢ *good quality income/wealth data*
➢ *main demographic and (possibly) labour market information*

Harmonisation

Ex-post harmonisation (cont'd)

Harmonisation

Final output:
the LIS/LWS datasets

❑ Technical harmonization: same file structure, same variables

LIS files

LIS Household File (H)			
HID	...	...	...

	LIS Person File (P)		
HID	PID	...	...

LWS files

LWS Household File (H)			
HID	INUM	...	...

	LWS Person File (P)		
HID	PID	INUM	...

	LWS Replicate Weights File (R)		
HID	...	...	...

❑ Conceptual harmonization
➢ Based on the same definitions
➢ Comparable concepts

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

Luxembourg Income Study (LIS) Database

- <u>First and largest</u> available database of harmonized income data, available at the household and person levels

- In existence since 1983

- Data mostly start in 1980, some go back to the 1960s (recollected every 3-5 years)

- Started with six countries; now over 50 countries

- 300+ datasets (repeated cross sections)

- Used to study: poverty, income inequality, labor market outcomes & earning gaps, social mobility & inequality of opportunity, policy effects (e.g. impact of redistribution)
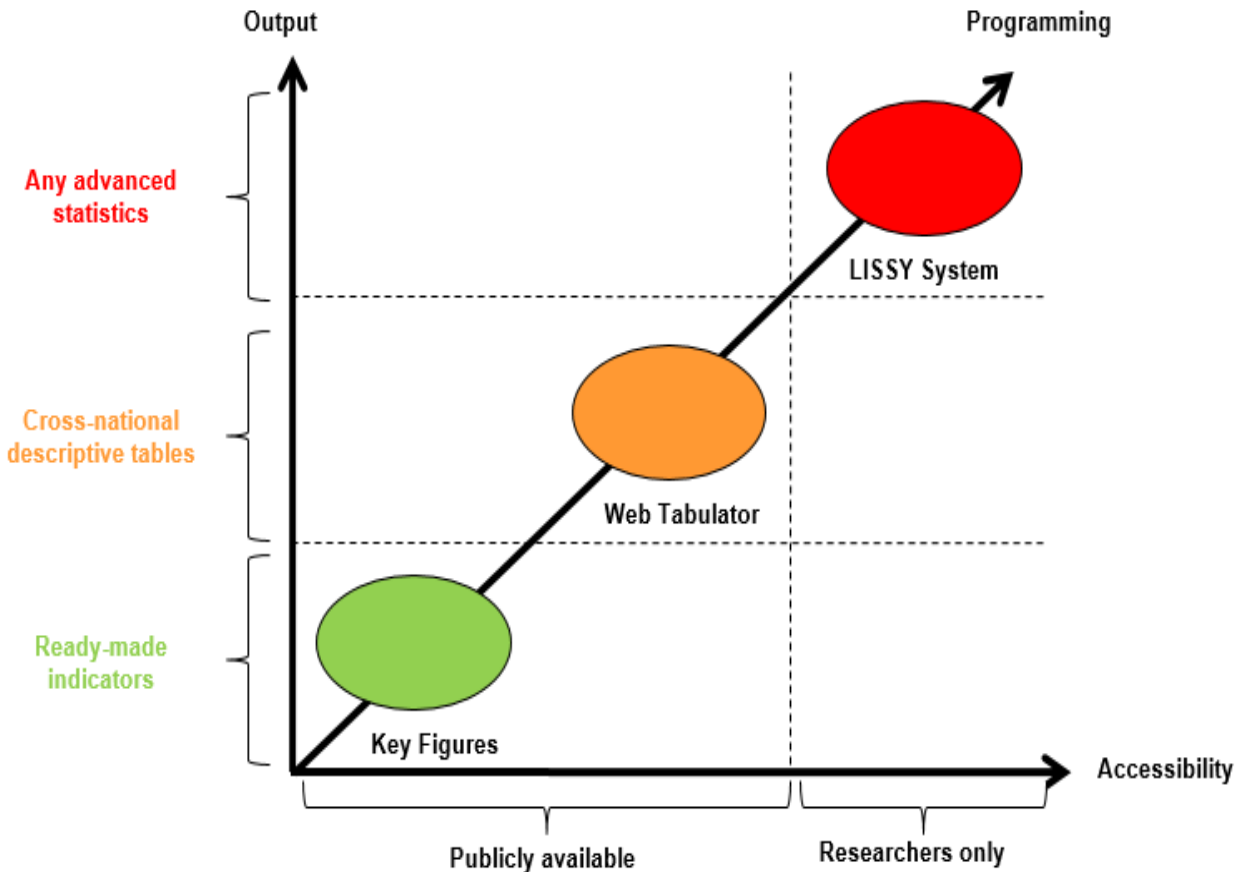
LIS

Luxembourg Wealth Study (LWS) Database

- <u>First</u> available database of harmonized wealth data, available at the household level

- In existence since 2007

- 52 datasets from 16 countries – up or in process

- Revised and updated in 2016 (in coordination with Euro system's Household Finance and Consumption Survey - HFCS)

- Used to study: household assets and debt, wealth inequality, wealth portfolios, financial behavior, policy effects

Research: tools

- Conducting research with LIS/LWS data requires using less to more complex tools
  - *Key Figures*
  - *Web Tabulator*
  - *DART (data visualization tool, forthcoming)*
  - *LISSY*
  - *METIS (**MET**adata **I**nformation **S**ystem)*

Research: working papers
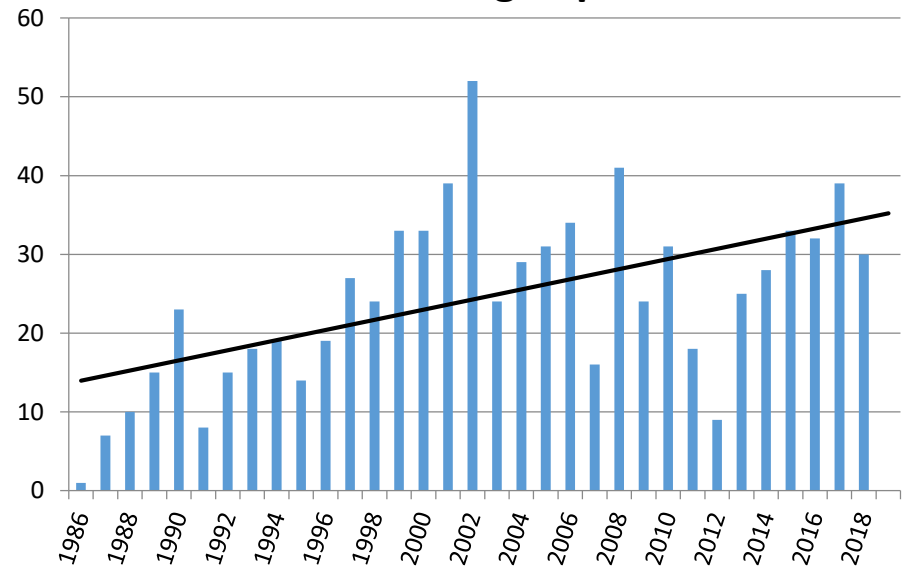
Working Paper Series

- *Over 800 WPs of which over 400 published in peer-reviewed journals and book chapters (e.g. AER, ASR, Econometrica, QJE, Lancet, REStat, Science, etc.)*
- *Included in RePec (Research Papers in Economics)*
- *Aldi Award: best paper by an under 40 author*

- **New WP:** *69 new WP added in 2017 and 2018.*

- **RePEC:** *Since May 2017, LIS/ LWS working papers have been downloaded more than 9,000 times*

- **H-index:** *LIS = 166*
  LWS = 50

LIS Working Papers

Part II :
Regional gaps…

Regional gaps

- We observe very low number of LIS users (not counting students) from CEE countries

- Main reasons
  - *Country coverage in LIS/LWS?*
  - *Training?*
  - *Funding / accessibility?*

CEE country coverage

LIS country coverage
☐ Country not in LIS
■ Country in LIS

9 CEE countries: Czech Republic, Estonia, Hungary, Lithuania, Poland, Romania, Serbia, Slovakia, Slovenia + Russia

Training

- To conduct research with LIS/LWS data requires certain level of technical skills (programming in LISSY, METIS, etc.)
- Hard to evaluate the quality of training in particular countries – let's have a look at the LIS workshop participants

	2013	2014	2016	2017	2018
Total # of participants	28	26	21	28	21
Out of which from CEE	0	0	0	2	3
%	0%	0%	0%	7%	14%

Funding

- The overall functioning of LIS heavily depends on country contributions (~50% of the budget)

- Users coming from LIS contributing countries represent around 90% of the total number of registered users

RANK	COUNTRY
1	United States
2	United Kingdom
3	Canada
4	Germany
5	Italy
6	South Korea
7	Poland
8	Switzerland
9	France
10	Sweden
11	Netherlands
12	Finland
13	Spain
14	Luxembourg
15	Japan
16	Australia
17	China
18	Belgium
19	Israel
20	Denmark

LIS contributing country

Funding (cont'd)

- Historically, no country contributions from CEE region

- Virtual Access to LIS microdata for non-contributing EU researchers has been secured by the InGRID2 program (2017-2021)

- In 2018, out of 40 researchers registered through Virtual Access coming from LIS non-contributing countries (Students are not taking into account), 15 were from CEE

- Overall, 1073 registered users in 2018 (around 50% students)

InGRID2 researchers in 2018	
Austria	2
Belgium	8
Denmark	1
Greece	3
Ireland	1
Poland	6
Portugal	1
Romania	7
Slovakia	2
Spain	2
Turkey	10

What can be done?

- Improve outreach and communication (Twitter, LIS newsletter, annual report)

- Provide training to CEE students / researchers (annual workshop, user self-teaching package)
  - *Application to the 2019 LIS summer workshop (in cooperation with Uni Lux and LISER) is now open!*

- Last but not least, secure funding after 2021
  - *InGRID3?*
  - *Other centralized way of funding?*

Thank you for your attention!

cupak@lisdatacenter.org