



CROSS-NATIONAL
DATA CENTER
in Luxembourg

The challenges of harmonizing data from middle income countries

UNECE Expert meeting on measuring poverty and inequality
29-30 November 2018, Vienna

Teresa Munzi

LIS – Cross-National Data Center in Luxembourg

Funded in 1983

MISSION: To enable, facilitate, promote, and conduct cross-national comparative research on socio-economic outcomes and on the institutional factors that shape those outcomes.

Since 2001

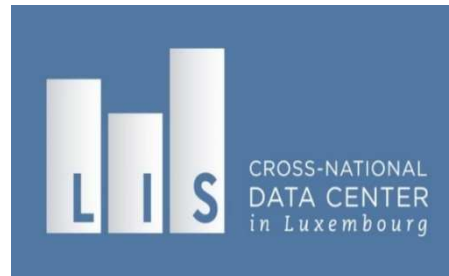
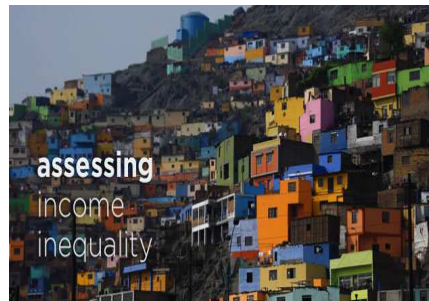
Independent non-profit institution, financed by country contributions and international organisations

Today

Two offices:

- Luxembourg (headed by Daniele Checchi, staff of 10)
- New York (headed by Janet Gornick, staff of 2)

President François Bourguignon



Overview of the LIS data

Deliverable

Two cross-national harmonised databases that allow international comparative research using micro-data:

- LIS (focus on income): 339 datasets – **this presentation will focus on LIS**
- LWS (focus on wealth): 39 datasets

Scope

Initial focus on high-income countries, successively extended to middle-income countries

Time span

From the late 1960s to 2016

Geographical coverage

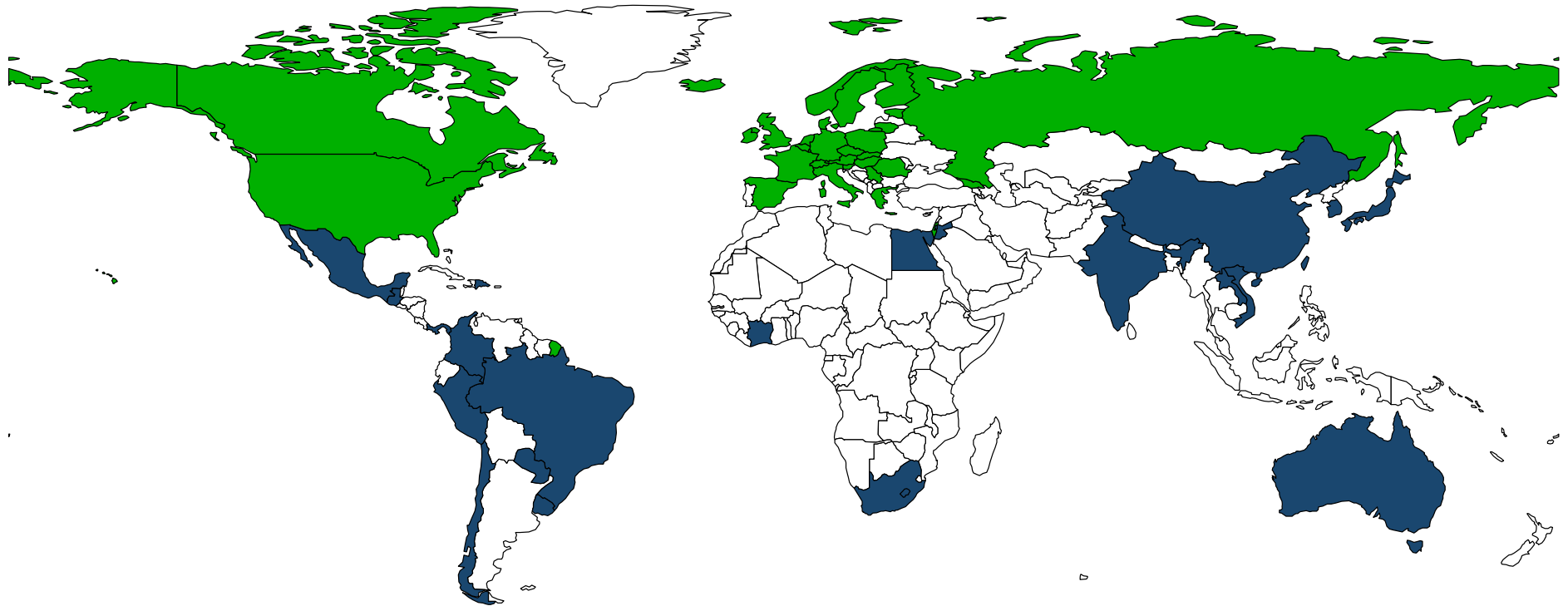
World-wide, but some regions are less covered (Africa, EECCA...)

Main contents

- household composition and characteristics
- socio-demographic characteristics of household members
- extensive set of labour market data
- detailed breakdown of household and individual income data
- household consumption data
- a detailed set of wealth and behavioural variables (LWS only)



LIS Coverage of UNECE Member countries



-  Not covered
-  LIS UNECE country
-  LIS other country



What we do at LIS

Core activity: **Data work**

Step 1. Data acquisition

We identify appropriate datasets (*neutral, reliable, and high-quality data*)

We negotiate with each data provider

Step 2. Data harmonisation

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*

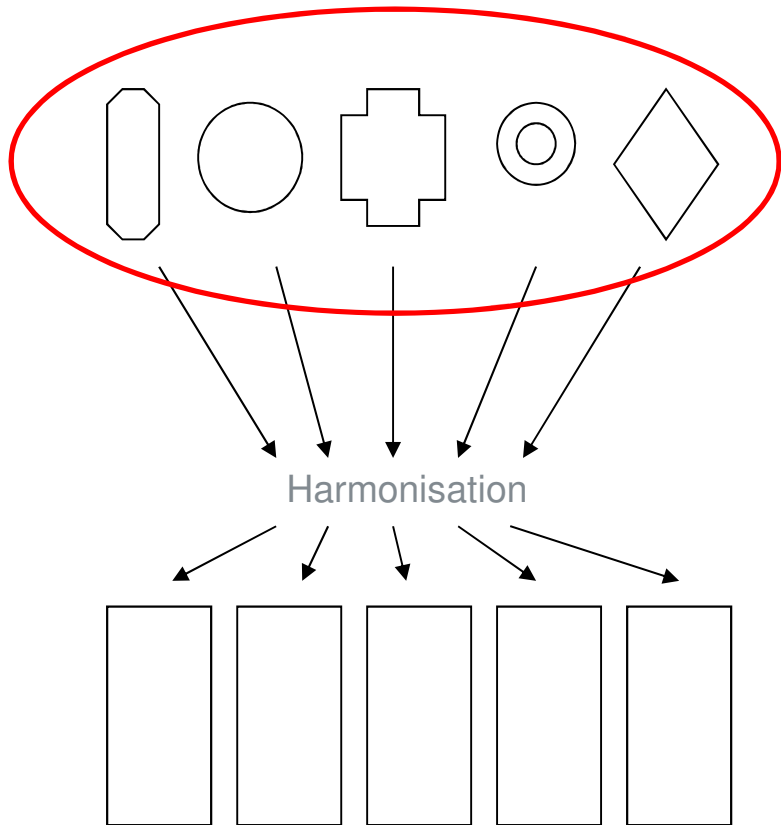
Other activities

Research-promotion activities (conferences, work on methodological issues, collaborations with networks/users/journalists, newsletter, individual research)

Support (user support, research visits)



Ex-post harmonisation at LIS

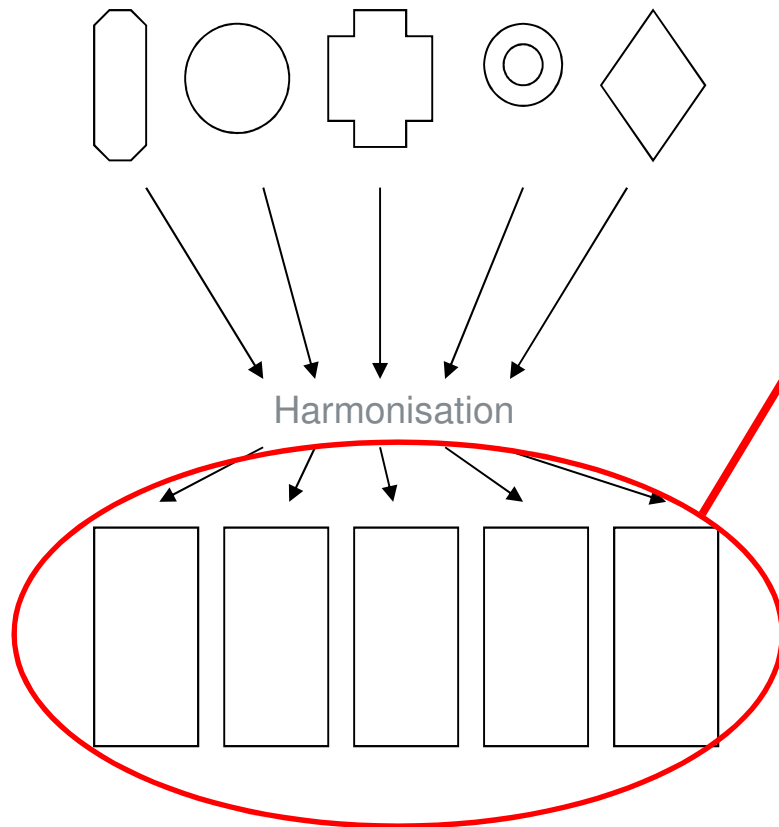


The origins of the LIS data

- LIS does not organise 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*

Ex-post harmonisation at LIS



Final output:
the LIS/LWS datasets (CCYY)

- Technical harmonisation: 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 harmonisation
 - Based on the same definitions
 - Comparable concepts

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

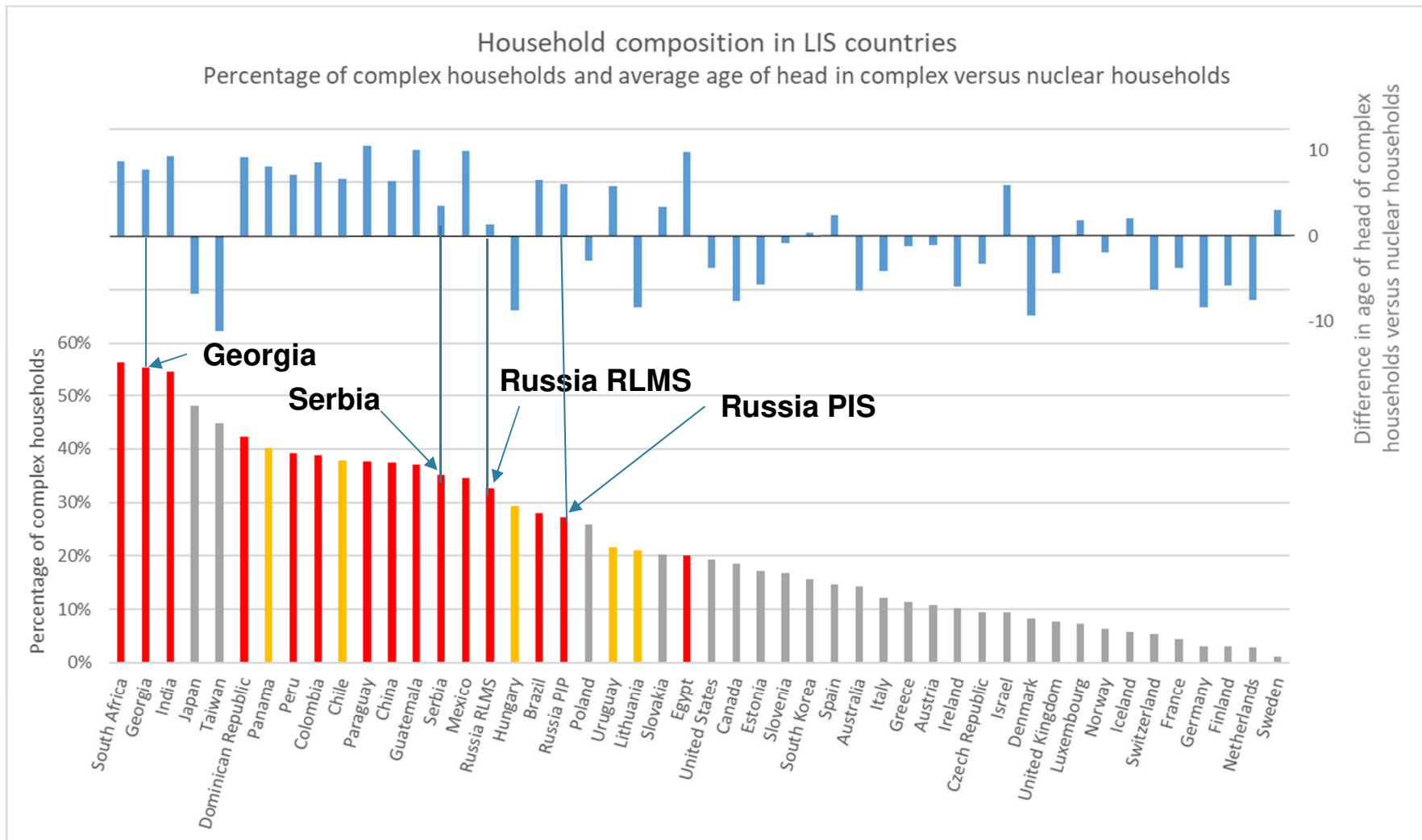
➔ BUT: very challenging due to very different inputs!

Inclusion of middle-income countries in LIS

- **Historically:** focus on high-income countries
- **2007:** pilot project on the feasibility of including middle-income countries
- **2011: new template** adjusted to better accommodate middle-income countries
 - *inclusion of non-monetary income into DHI*
 - *adjustment of the concept of household member and addition of living arrangement variables*
 - *new topics: rural/urban indicator, farming activity indicator, type of dwelling, involvement in marginal/informal work, characteristics of second job*



Household composition: prevalence of complex household structures

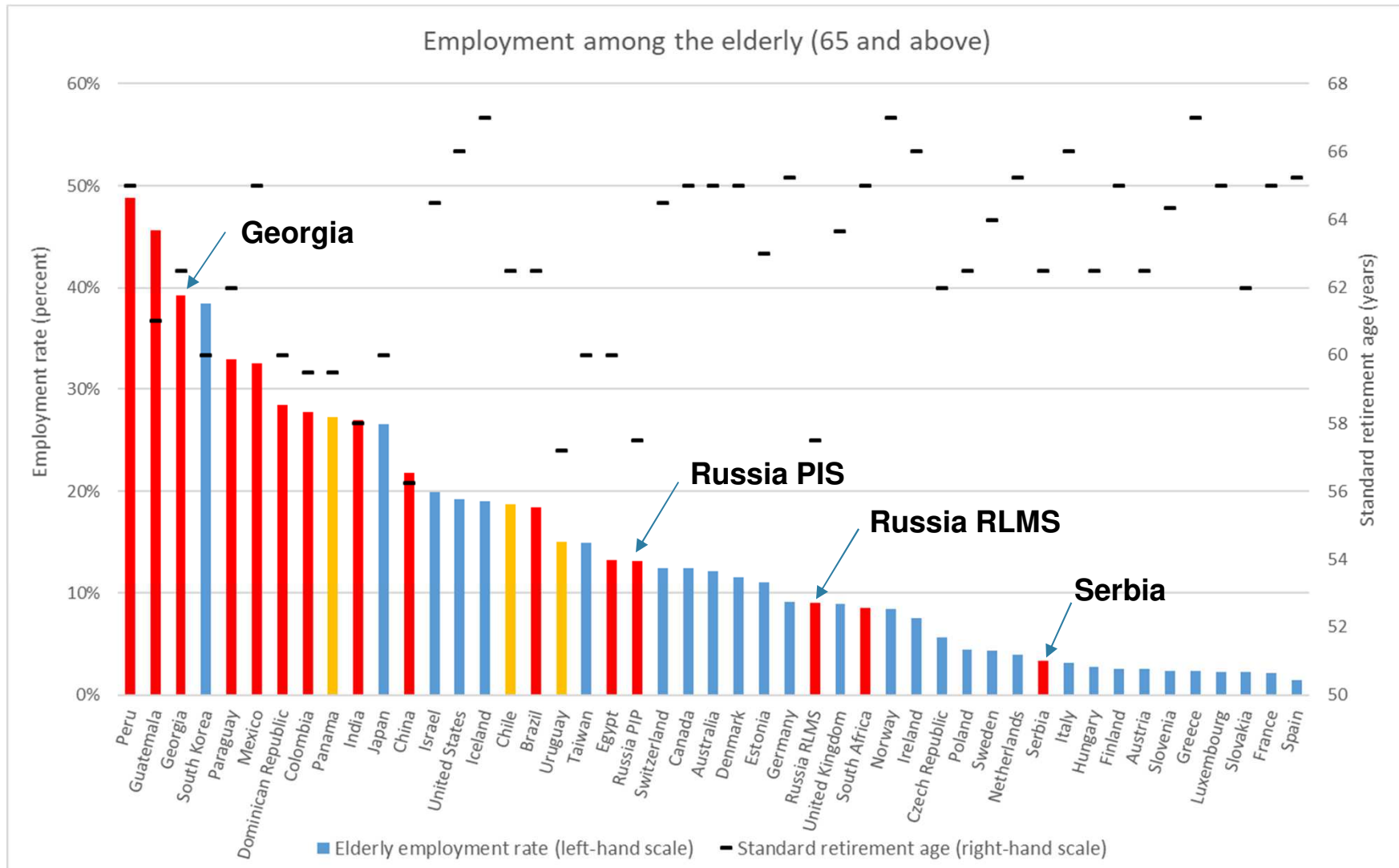


Issue: analyses centered around the head/spouse may become biased

Recommendation: importance of providing detailed living arrangements for all household members (possibly pointers to partners and parents)



Employment: prevalence of informal employment (elderly, child, unpaid)



Issue: measurement of employment (artificially high), inconsistency with earnings

Recommendation: importance of providing detailed information about marginality / informality of employment



Treatment of non-cash incomes

- ❑ Relative importance of non-cash incomes (consumption from own production, in-kind individual public goods) w.r.t. high income countries
- ❑ LIS now includes those in total disposable income

BUT

- ❑ Not always available (**incomparability**)
- ❑ **Risk of double-counting** when in-kind incomes are collected in the consumption module and added to the incomes

RECOMMENDATION

- ❑ *If income and consumption survey*: be very careful at the collection stage about where in-kind incomes/consumptions are collected
- ❑ *If income only survey*: collect also in-kind incomes



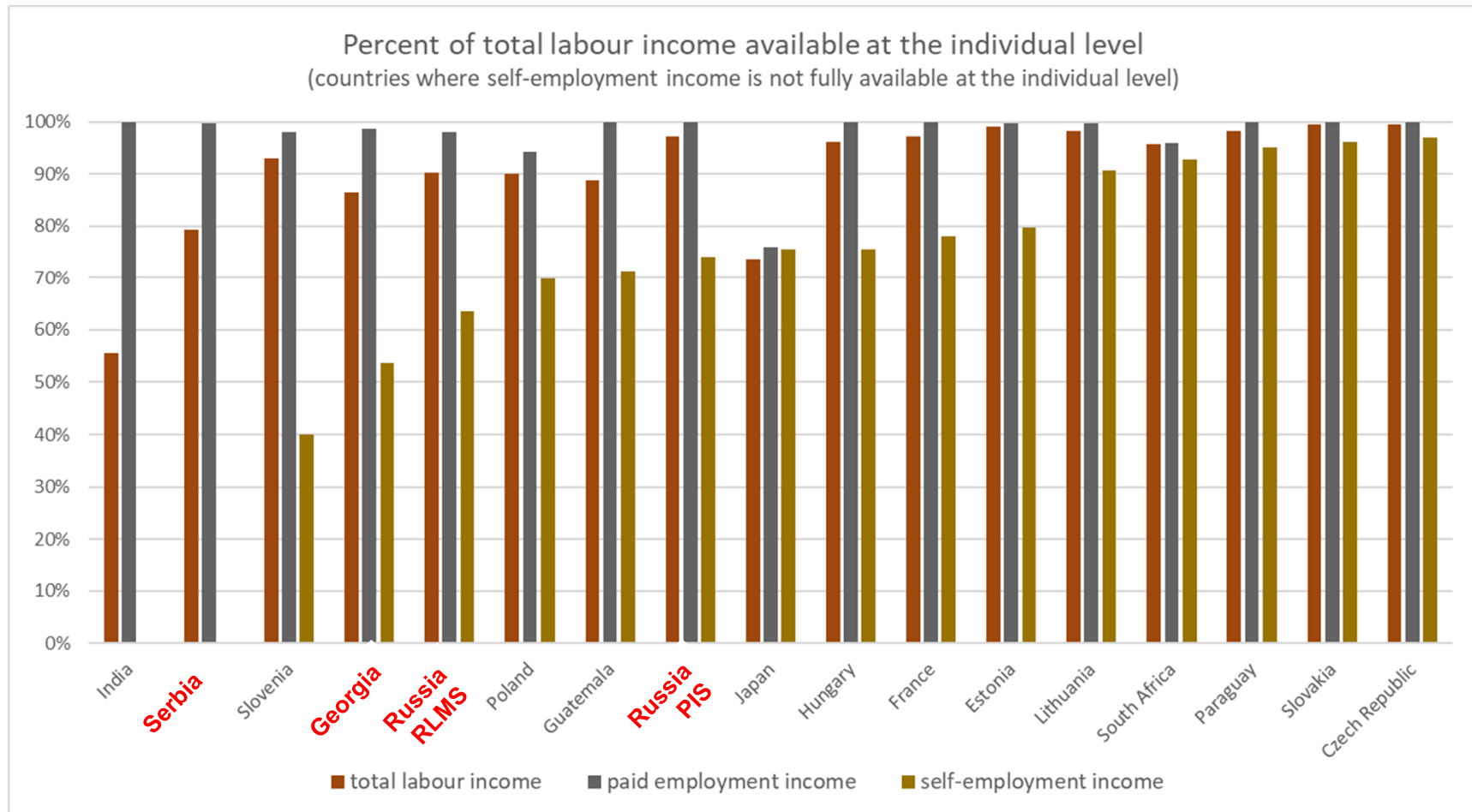
Measurement of self-employment incomes

Issue is of particular relevance in middle income countries as self-employment is much more common

- ❑ **Underestimation**: self-employment incomes are much harder to capture in general, even more so in presence of informal/marginal employment
- ❑ **Risk of double counting**: when collected both at the individual level and in the household business sections of the questionnaire
- ❑ **Individual level self-employment income not measurable**: small portion collected at the individual level (possibly only the incomes of the liberal professions), while other self-employment incomes are collected at the household level

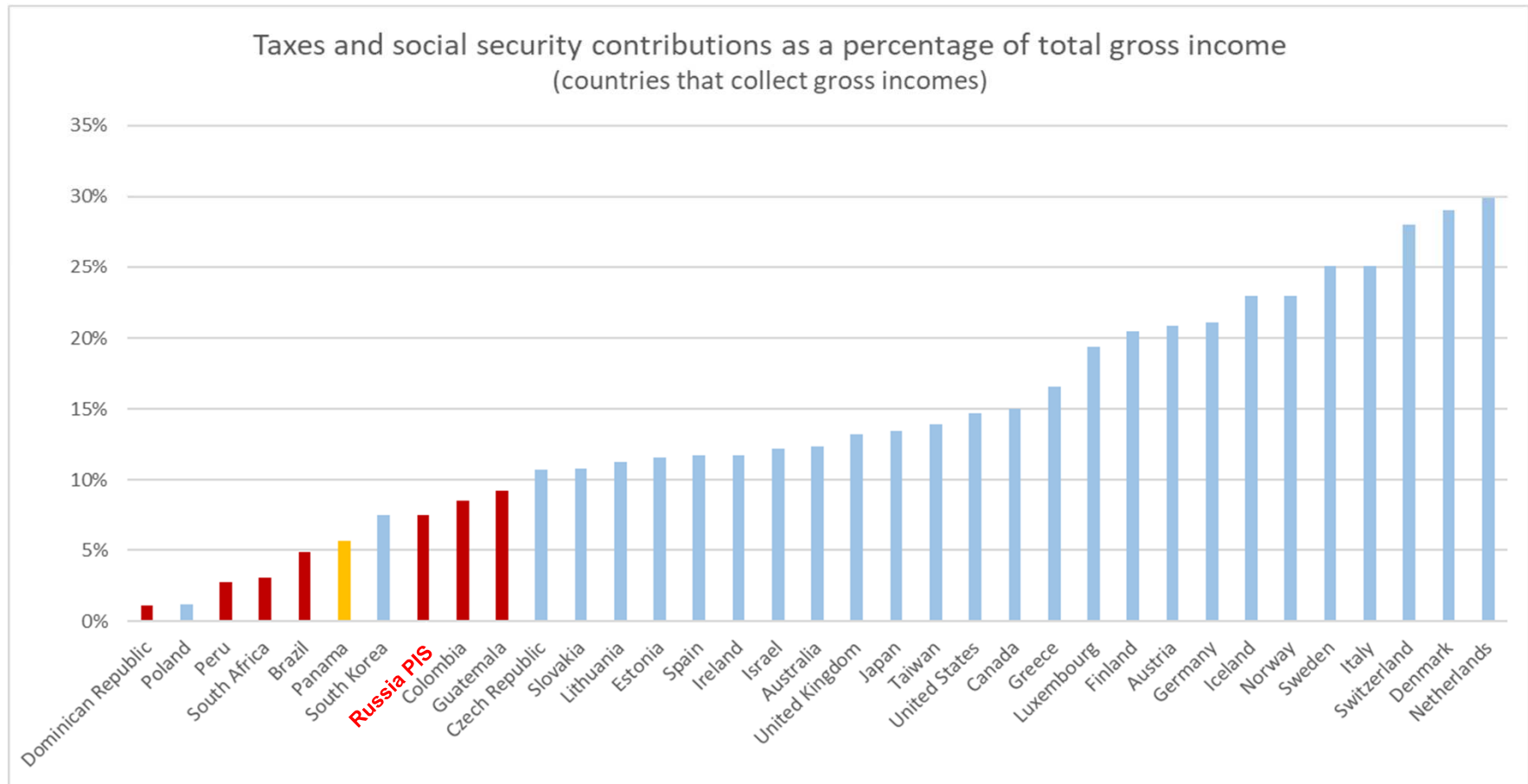


Measurement of individual level self-employment income



With the exception of France and Japan, these are all countries that are or have recently been middle-income countries!

Measurement of direct taxes and social security contributions

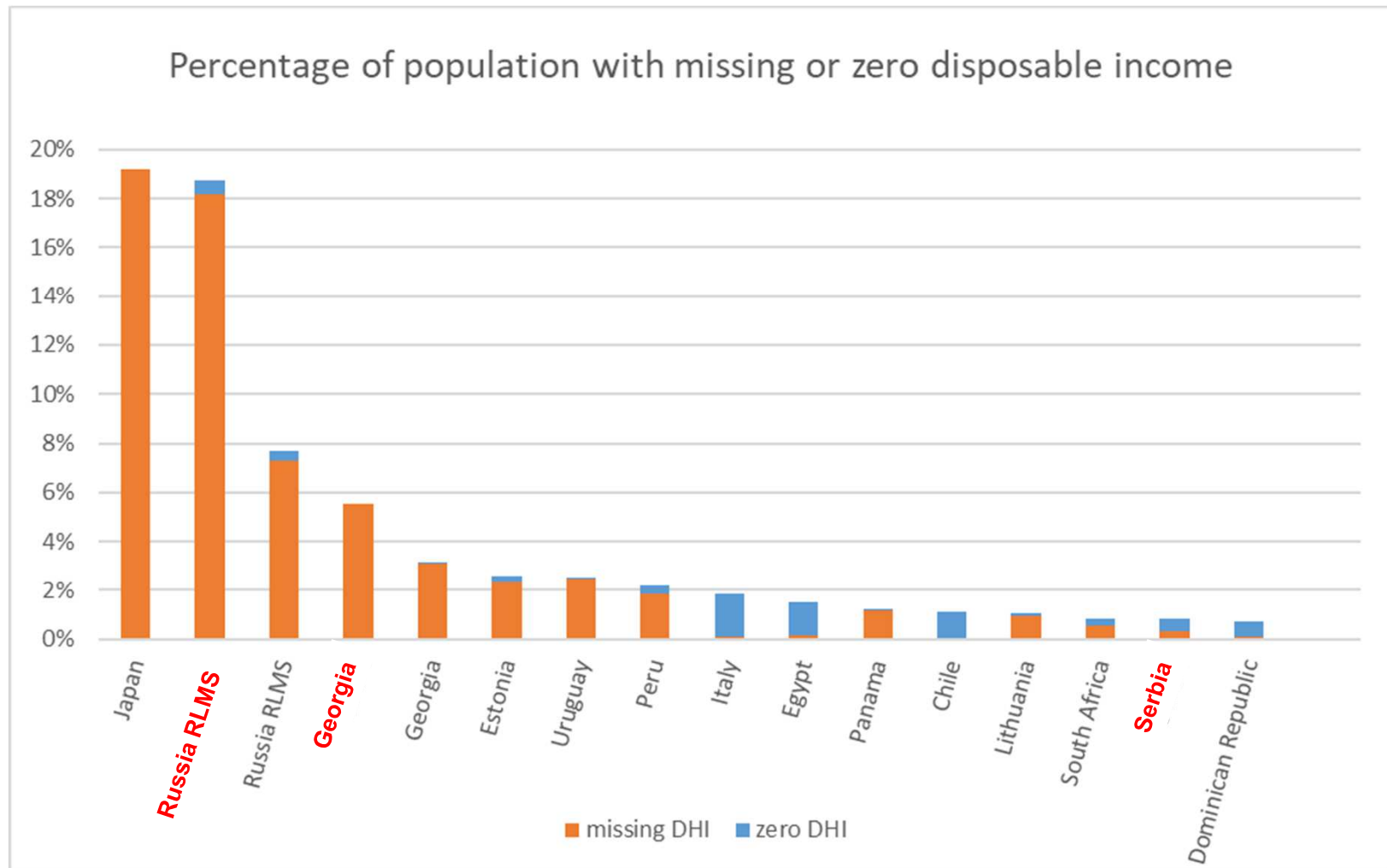


Issues:

- difficult to measure (eg PL)
- higher reliance on indirect taxes (which are typically not as redistributing as the direct ones) biases downwards inequality when compared to countries that rely more on direct taxation



Incomplete income data is much more prevalent



Issue: bias due to non-random distribution of households with no income information

Recommendation: calculation of the weights and imputation

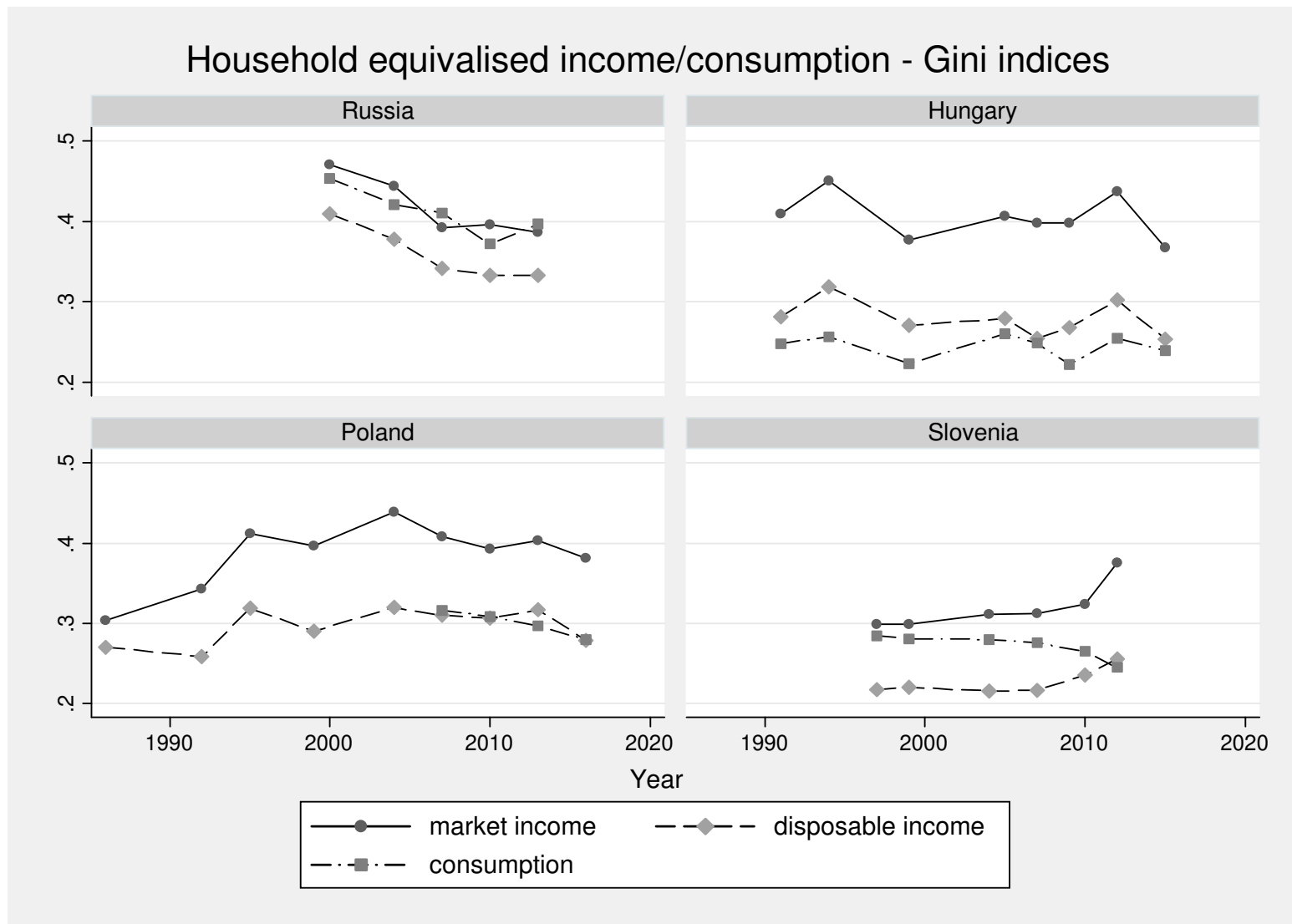


Additional recommendations

- ❑ In case of differential sampling (eg urban versus rural), importance of providing accurate weights
- ❑ Data providers know best their data: data cleaning, editing and possibly imputation should happen at that stage, and not during the harmonization
- ❑ Importance of providing good documentation



Inequality trends in Central and Eastern Europe



Evolution of income shares in Central and Eastern Europe

