



CROSS-NATIONAL
DATA CENTER
in Luxembourg

LIS: Access to and Work with Cross-national Micro Databases

Thierry Kruten, Director of Operations
First European Data Access Forum - March 2012



Our Mission

LIS – **Cross-National Data Center in Luxembourg** – was funded in 1983

- to **enable, facilitate, promote, and support cross-national comparative research on socio-economic outcomes**
- by **collecting** and **harmonizing** cross-national micro-datasets and **making these data accessible** to a global community of researchers, educators, and policy makers at their remote locations under privacy restrictions

LIS includes micro-data from **45 countries**, mostly in Europe and North America, but also in Middle East, Asia-Pacific and Latin America

- The **LIS Database** includes harmonised income, household- and person-level characteristics, labor market outcomes collected over a period of four decades (1967 – 2010)
- The **LWS Database** is the first cross-national database of harmonised wealth microdata in existence containing variables on assets and debt, behavioural indicators and also on income and labor market variables



Accessing micro-data: What are the challenges?

DATA PROVIDERS

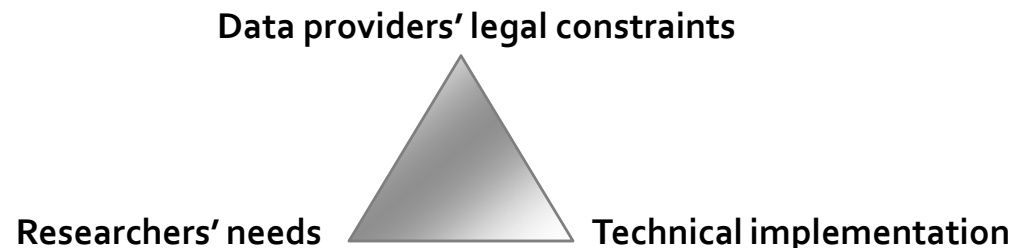
LIS negotiates with each country separately to acquire microdata under the constraint of being compliant with the existing national privacy restrictions

RESEARCHERS

Needs of individual information in order to test the effectiveness of policy or to set new policies implies to obtain access to micro (and possibly sensitive) data

LIS Cross-national Data Center

- ❑ A remote execution system called **LISSY** offers researchers the possibility to analyze microdata at their own work of place by submitting statistical queries
- ❑ The trade-off between an **operational and user-friendly** system and a **safe** system is evidently present for remote systems since disclosure control limits the details of the output.



The challenge consists in defining the appropriate administrative, legal, operational and technical processes to provide researchers with the best trade-off



LIS data access eligibility

- ❑ Use of the data is restricted to **Social Science research purposes only**. Access is limited to researchers affiliated with an academic, governmental or supra-national organization research department

- ❑ Users must register with LIS and acknowledge to obey the rules governing the use of the microdata
 - *Applicant information (coordinates, supervisor information, project etc...)*
➔ **to check the affiliation**

 - *Pledge in which are defined the terms of the agreement and the rules on data security and confidentiality. Must be (electronically) signed*
➔ **to enforce a contract**

 - *LIS grants access to databases for a limited time period (1 year) renewable annually*
➔ **to get up-to-date information on users**

 - *Registered users receive an user account (id and password) that is strictly personal*
➔ **to uniquely identify user**

**Over 4,000 users from 55 countries had ever been registered
[600 applications per year over the last 5 years]**



Security and confidentiality concerns

There are no trade-off in relation to the IT security

- ❑ The LISSY system consists in a secure architecture comprising a series of (physically separated) software working together to run users' statistical requests. At no moment a user is in direct contact with the microdata

➡ **IT security audit**

Flexibility to configure the desired level of confidentiality

- ❑ Security settings defined for an automatic scan of each incoming users' request
 - *Adaptable by project, user profile, datasets to match **data providers** legal constraints*
 - *Suspicious jobs are sent to a review queue for a **manual review***
 - *Blank outputs are returned for any job that breached security settings to the sender along with an error message explaining the violation*

➡ **55,000 requests per year to monitor**

- ❑ All incoming jobs and outputs are stored allowing to **trace back** the researchers' job history

➡ **signing a pledge is not enough**



And COMMUNICATION

Extensive documentation is available for public on-line access on LIS website in order to support researchers working with LIS data

- ❑ For each LIS and LWS micro-dataset, detailed information on original survey, LIS variables' content and availability, etc... are supplied to allow users understanding **the context** in which LIS outcomes should be analyzed
- ❑ Information on how to access to and work with microdata
 - *Data accreditation (register and renew users access, confidentiality rules...)*
 - *Data access system (guidelines on how to use the LISSY system, FAQ...)*
 - *Learning materials (digital libraries with job templates, self-teaching packages...)*



Constant update of the documentation is time-consuming but essential to users

Support facilities as a mean to improve researchers' ability to work with LISSY and to reduce risks of breaching confidentiality rules

- ❑ ***User support*** (500 emails per year) and training sessions through ***workshops***



Conclusion

- ❑ Still challenges to face such as revising the LIS databases documentation system by supplying a new meta-data system that will allow LIS users to create tailored documentation extracts fitted to their individual needs
- ❑ The key objective to work on : **constantly** adjusting the microdata access **services** to fulfill researchers' needs while maintaining the same level of security and communication

Thank you for your attention
Any question is welcome !