



Current developments in access to administrative data for research

Roxane Silberman
CASD-GENES & CNRS, Saclay, FRANCE
2018 IASSIST & CARTO Conference
Montreal, May 29-June 1, 2018

INTRODUCTION

- Objective of this presentation
 - Some thoughts on the developments and challenges encountered in research access to administrative data in non registered-based countries as France
- Intensive use of administrative data once mainly the case in the registers-based countries (e.g Nordic Countries in Europe)
 - With National Statistical Institutes in central position
 - Also in position to provide access for research (eg Statistics Denmark, CBS)
- Now at the top of the agenda in all countries
 - IT development, decreasing response rate in surveys, budget issues, open data policies, rich resources for evidence-based policies
 - For official statistics, public policies and research
 - EU: Code des bonnes pratiques European Statistical System 2011
 - US: Report on *The Promise of Evidence-Based Policy making* 2017
 - France: Conseil national de l'information statistique (CNIS) Bozio & Geoffard, *L'accès des chercheurs aux données administratives*, 2017
- A great opportunity for research, requiring changes in the legal frameworks however also raising other challenges

OUTLINE

- The French and EU contexts
- Changing the legal frameworks in France re. research access to administrative data
- Further challenges
 - Metadata issues
 - Harmonisation of procedures
 - Centralisation vs decentralized secure access
 - Linkage issues
 - Long term preservation issues
 - Transnational access issues



I. THE FRENCH AND EU CONTEXTS

THE FRENCH CONTEXT FOR ADMINISTRATIVE DATA

The French case: an example where:

- Administrative data are held by various government bodies: ministries statistical departments, government agencies with various statutes in charge of retirement, families policies, unemployment ...
- Official statistics and administrative data are under different legal frameworks
- The national statistical Institute (INSEE)
 - ✓ Coordinates the statistical system (census and surveys)
 - ✓ Not the administrative data
 - ✓ Even if can mobilize them to enrich official surveys

EU LEGAL CONTEXT FOR DATA PROTECTION

- Best use of administrative data requires highly detailed data and linkage possibilities
 - For a large part “confidential” data
 - If data on persons, the EU legal framework applies
- 2018: General Data Protection Regulation (GDPR)
 - From the 1995 Directive unevenly translated in the EU countries national laws
 - To a Regulation applying to all EU members states
 - Entry in force in May 2018
- Compatible with research use of highly detailed administrative data
 - Provisions regarding research use of confidential and sensitive data
 - Some issues re. long term preservation
 - A unified EU space that should facilitate transfer within EU & conditions for transfer out of the EU



2. CHANGING THE LEGAL FRAMEWORKS FOR RESEARCH ACCESS TO ADMINISTRATIVE DATA IN FRANCE

PREVIOUS CHANGES IN FRANCE RE. ACCESS TO CONFIDENTIAL MICRODATA A FAVOURABLE CONTEXT TO ENLARGE ACCESS TO ADMINISTRATIVE DATA

- ***Informatique & Libertés Law (Privacy Protection law)*** : Successive changes to implement **provisions for research use of confidential and sensitive data**: in the 80ies for epidemiologist, then enlarged in 2004 in reference to the 1995 EU Directive, updated in 2018 to be in line with the EU GDPR
- ***Statistical Law***: Successive changes to implement **provisions for research use of confidential microdata** : in the 80ies for business data, in 2008 for household and persons data (in conjunction with the National Archives law)
- **An independent authority, *Comité du Secret Statistique* and accreditation procedures for confidential official microdata**, first set up for business data, then enlarged to household and persons data
- **A Secure Remote Access Centre**: set up in conjunction with the 2008 changes in the legal frameworks
 - ✓ Initially for providing access to Insee data
 - ✓ Progressively enlarged to Ministries statistical departments and other confidential data

HOWEVER NOT COVERING ADMINISTRATIVE DATA

- A rather global framework ensuring trust and security
- However, administrative data not covered by the Statistical Law
 - When INSEE surveys started using tax data instead of survey questions on income: no longer possible for INSEE to provide these information to researchers
 - Held by various government bodies under specific regulations and kept under strong barriers though some exceptions were possible
 - Most difficulties had to deal with the notion of “professional confidentiality”
 - Different names
 - Specific regulations/legal frameworks depending on domains
 - Various interpretations by the administrations
 - In principle, long term preservation and access under the National Archives law, however uneven

STEP BY STEP ...

- 2014 Change in the *Livre des procédures fiscales (Tax regulation)*
 - Tax data available via the CSS advice and CASD secure access
- 2015 *Loi sur la santé (Health Law)*
 - Medico-administrative data (French exhaustive database)
 - Criteria: Public interest
 - A specific procedure and authority
 - Some access via CASD
- 2016 Decision of the Central Bank (that is under the ECB legal frameworks) to open an On site Secure Access

... AND AN ATTEMPT FOR A GENERAL FRAMEWORK

- 2016 *Loi pour une République Numérique (Digital Law)*
 - An attempt to avoid changing each specific regulation step by step
 - Has benefited from the strong involvement of researchers (use cases) and from the context of the Open Data Policies
 - An article implementing a global framework to deal with the specific obstacle re. research access raised by the “professional confidentiality” in the various regulations for administrative data
 - Another article in the same law facilitates data linkage based on the national identifier (NIR) both for the official statistics needs and the research needs.
- A large perimeter of highly detailed administrative data now in principle accessible though some legal issues remain (e.g justice)



3. FURTHER CHALLENGES

FROM THEORY TO PRACTICE

While these new regulations and decisions in principle have open access to a wide perimeter of administrative data for research

- The decision re. accreditation process and access system remains in the hands of the various data holders
- With a number of challenges
 - ✓ Where to find information and what about metadata ?
 - ✓ Centralized vs decentralized systems for accreditation and access ?
 - ✓ How to organise linkage ?
 - ✓ What about the long term preservation issues for such multiple data sources as well as for those resulting from linkage ?
 - ✓ What about transnational access ?

WHERE TO FIND INFORMATION AND WHAT ABOUT METADATA ?

- What is available, how, where ?
 - Currently not much information on administrations websites and no central point in place
 - Geoffard and Bozio report (CNIS) recommendation to set up a Vade Mecum for researchers
- What about metadata ?
 - A huge investment required
 - Need for priorities
 - How to finance the investment ?
 - Involving the researchers: the experience for the tax data

CENTRALIZED VS DECENTRALIZED SYSTEM FOR ACCREDITATION AND ACCESS ?

- Researcher accreditation
 - Will the administration decide to go
 - via the existing authority (CSS) in charge of providing advice for researcher accreditation for official microdata
 - or set up their own procedure
 - Multiple accreditation for research is various sources and harmonisation issues
- Secure mode of access
 - We may see a growing number of secure access systems
 - Currently CASD, Central Bank, health data ...
 - Will the administrations choose setting up their own secure access ?
 - Security requirements need to be harmonized
 - Will also require to have rules for exchanges and harmonisation for research projects using data from various sources (e.g Central Bank and Insee data currently not possible)
 - Also financial issues
- A public service for providing secure access for research ?

LINKAGE ISSUES

- The legal framework facilitates linkage for research using the NIR
 - Example: Health status impact on employment
 - With possibilities for longitudinal data requiring re-use of the key
- Linking data while protecting confidentiality
 - Combining 2 datasets
 - With nobody having
 - Identity information
 - and
 - Data information from the 2 datasets
- Use of the HASH processing: no possibility to retrieve the NIR from the HASH code
- Also requires to have 2 Third Trust Party
 - Key Management Authority
 - Secure access provider
- Which institution ? One system for all projects ?

WHAT ABOUT LONG TERM PRESERVATION ?

- What about long term preservation for such multiple data sources as well as for those resulting from linkage ?
- Current situation unclear for administrations
- Should National Archives Law apply or Privacy Protection that set up limits ?
 - It was agreed that National Archives in charge of selection
 - A CASD collaboration with the National Archives now set up

WHAT ABOUT TRANSNATIONAL ACCESS ?

- Increasing researchers needs to use administrative data from various countries
 - A number of H2020 projects requiring recurrent linkage with administrative data from various countries in their specific domain (poverty, immigration etc...)
- The GDPR should facilitate transnational access within the EU
- However requiring setting up a network of secure access providers
 - IDAN (International Data Access Network) project coordinated by CASD (France, UK, Netherlands and Germany)



Sommaire

CONCLUSION

CHALLENGES AND DATA IMPACT

- Changes in the legal frameworks in France now facilitate use of the huge amount of administrative data
- However a number of challenges remain
- Many issues seems similar to those raised in other countries though each context is different
 - UK and the ADRN Administrative Data Research Network
 - US: *The Promise of Evidence-Based Policy making 2017*
- Data impact for research and evidence-based policies huge
 - Inequalities analysis and access to the highly detailed microdata as the tax data
 - Health: the Mediator scandal in France and access to the SNIIRAM, the exhaustive Social Security data basis
-



Sommaire

THANKS !

roxane.silberman@casd.eu

<https://casd.eu>