Towards a European distributed remote access
Main issues and future challenges

Paola Tubaro    Roxane Silberman
Centre National de la Recherche Scientifique, Paris

ISI 2011 Conference, STS 65, 26 August 2011
Data access as a key issue

- Access to confidential (or highly detailed) microdata for scientific purposes is now a major issue on the table.

- NSIs and other official data producers (Eurostat, ECB, Central Banks etc.) increasingly recognize it as central.

- Use of data for research ensures maximum use of data to:
  - raise returns from public investment;
  - support evidence-based policy decision-making.
Progress, open questions and uncertainties

- Recent significant improvements both at national and European levels.

- Growing number of initiatives, often in dialogue with data archives and research community.

- Demand for comparative studies makes trans-border access crucial.

- Remaining open questions and uncertainty about final outcomes.
Different perspectives of data producers and users

- Data producers need to ensure data protection and to maintain respondents’ trust.
  Backed by a complex legal framework:
  - Statistical law;
  - Privacy / Personal data protection law;
  - Archival / records management law.

- Data users (researchers) need to access detailed microdata for higher quality of analyses and more informed policy advice.

⇒ Needs to protect personal information and to provide data access to researchers may collide.
Emerging needs

- Cutting-edge statistical tools and increased computational capacity call for detailed microdata.
- Additional data needs also include:
  - administrative data;
  - combined survey and administrative data;
  - merging of several microdata files.

⇒ Growing researchers’ demand to access ever more confidential / highly detailed official microdata.
Towards a European distributed remote access

---

**Background**

**New challenges ahead**

**Increasing constraints**

- A tendency to an increasingly higher degree of *anonymisation* of datasets, from the 1990s onwards.

- It is meant to protect respondents and respond to social anxiety about privacy.

- Yet heavy anonymisation may jeopardize usefulness of information contained in the data.

⇒ Researchers’ uneasiness towards what they feel is a move backwards.
Changes in legislation and regulations

- Statistical laws and/or data protection laws increasingly recognise research data needs and offer derogations.

- Major step forward when 1995 European Directive on protection of personal data recognised research purpose.

- Several countries have followed suit (e.g. France 2004).
Changes in legislation and regulations (cont.)

- Similar developments with statistical laws: UK 2007, France 2008, Portugal 2008 etc.

- 2002 European Regulation to organise research access to European datasets; a new version is now under discussion.

⇒ Despite remaining restrictions, legal frameworks are evolving in a direction more favourable to research.
New IT solutions for secure access

- **Data labs** on the premises of NSIs, where access can be controlled:
  - no possibility to copy, print, download data;
  - output subject to checks.

- **Remote execution**:
  - as above in terms of security settings;
  - but researchers can connect remotely from their institution;
  - no actual access to data is allowed (submission of batch jobs only).

- **Remote access**: as above, but researchers can actually see the data.
New IT solutions for secure access (cont.)

- Many European countries have adopted one or more of these solutions.
- In some cases, system is managed by a trusted academic institution: UK (UK Data Archive), France (GENES).
- Typically in conjunction with statistical disclosure controls and rigorous accreditation processes:

  ⇒ Safe settings, safe data, safe people.
Recognition of trans-border access issues

- Initiatives to improve European and cross-country access within Europe:
  - Ongoing revision of research access regulation at ESS level.
  - Need for enhanced trans-border access to both European and national datasets.
  - Discussions around need for enhanced access to Eurostat data and harmonisation of rules and practices across Member States.
Improving dialogue between stakeholders

- Reinforced cooperation at national and European levels between NSIs, data archives, and the research community.

- Past and present projects and initiatives, in particular:
  - FP7-funded CESSDA PPP (2008-9): partly dedicated to exploring conditions for multi-stakeholder dialogue, at European level;
  - FP7-funded Data without Boundaries (DwB) (2011-15): trust-building and piloting partnerships between NSIs, data archives, and researchers.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Census</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Social Survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Person Register</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Business Survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. Business Register</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1.** Modes of access in European countries (2008).

- **no country**
- **1-3 countries**
- **4-6 countries**
- **7-9 countries**
- **10-12 countries**
- **≥ 13 countries**
The uneven landscape of data access throughout Europe

- Widespread but uneven progress in access to anonymised social survey and even census data.

- Tabulations (both standard and bespoke) increasingly popular thank to web-based tools.

- Rapid spread of secure IT solutions for access to confidential data — still evolving.
The uneven landscape of data access throughout Europe (cont.)

- Business survey data typically subject to restrictions, with some exceptions.
- Administrative data (both person and business) generally not widely available, with exceptions.

⇒ Despite progress in access to confidential data, difficulties remain and access to even highly anonymised data is limited in some countries.
More subtle differences

- Degrees of **anonymisation** differ across countries, data-providing institutions, and datasets.
  ⇒ Impact on conditions of access to confidential data too.

- **Accreditation** rules and procedures differ widely across countries; no standard exists for cross-border accreditation.
More subtle differences

Lack of a uniform terminology:

- e.g. PUFs and SUFs do not always refer to the same degree of anonymisation, or the same access conditions;
- this may hide further differences in access opportunities across countries;
- resulting inequalities for researchers based in different countries;
- further difficulties for comparative cross-country research, and for research on Europe.
Remaining general issues and difficulties

- **Accreditation** process often very long, sometimes untransparent.

- **Standardisation** (in terms of data formats, documentation etc.) still to be achieved.

- **Cross-border access** remains challenging and (at best) time-consuming.
Architecture of access system: which choices?

- future **New European regulation and ongoing developments**:
  - moving towards a **unified, harmonised system** for access, or to juxtaposition of parallel systems (ESS / CESSDA)?
  - to what extent will **data archives** and representatives of **researchers** be involved as partners with **NSIs**?
  - to what extent will national systems be integrated into a European one?
Towards a European distributed remote access

Future uncertainties

Technical solutions: which options?

- Safe data labs;
- Remote execution;
- Remote access.

Various issues: Costs, security standards, legal conditions for trans-border access...
Output controls: which approach(es)?

- **manual, exhaustive**: expensive and time-consuming;
- **manual, on a sample**: do not fully eliminate disclosure risks;
- **automated**: if threshold of acceptance is set too high, they may re-introduce anonymisation by other means.
Which future for anonymised data (SUFs)?

SUFs:

- will they be made *obsolete* by secure access systems?
- or will they survive as an *intermediate* solution, still useful for a number of data needs?
- disappearance of SUFs might raise demand for access to confidential datasets: how to cope with it?
Data without Boundaries (DwB)

The DwB project (2011-15) as a way forward:

- participation of 1/3 NSIs, 1/3 CESSDA data archives, 1/3 research institutions and universities;
- start pilot projects with secure access systems, both onsite and remote, allowing trans-border access to national data;
- to test technical, legal, practical solutions for a system that may subsequently be extended at national and (possibly) at European levels;
- organise a stable, systematic structure for dialogue and collaboration between all stakeholders including NSIs, data archives, and the research community.
- enhance collaborations at European level.
Thank you!

Paola Tubaro, Roxane Silberman

Contact: paola.tubaro@ens.fr