Introduction
An increasingly complex environment

- In many European countries, improvements in quantity and quality of government microdata production have come with better conditions of access for researchers.

- Yet difficulties and inequalities in access remain. In particular:
  - diverse perceptions of confidentiality issues coexist;
  - discrepancies subsist between official procedures and actual practices;
  - documentation is often insufficient or uneven.

- This leads to low transparency and high complexity of the system.

- Complexity at the national level is reinforced at the European level, due to heterogeneities among countries.

The case for a European research infrastructure

- In this context, a European Research Infrastructure can offer pathways for access by acting as a central data broker.

⇒ CESSDA Data Archives have played this role for several years at national level: what can be their place at European level?
What follows aims to provide an overview of recent developments and trends, to position the contribution of CESSDA Data Archives so far, and to outline directions for further improvement.
1. The structure of national statistical systems, and its impact on access
Structure of statistical systems

• Centralized vs. Decentralized System:
  – role of regional/local offices;
  – data produced by ministries, government departments, etc.
  ⇒ one central data provider vs. several providers.

• Register-based data collection vs. questionnaires:
  – today, increasing use of administrative data in all countries
  ⇒ possible changes in what is recognised as official statistics.

• The role of Central Banks in some statistical systems.

• The role of Eurostat.

2. Existing modes of access to national data
Evolving modes of access

- Public Use Files and Scientific Use Files:
  - Researchers vs. general public and/or students (CAMPUS files);
  - Differing degrees of detail (the notion of *de facto* anonymisation);
  - License vs. No license.

- Tabulations:
  - Public tabulations vs. special (bespoke) tabulations;
  - Development of web-based tools to customize tables.

- New solutions for confidential data:
  - Secure on-site data centres;
  - Secure remote connections through the Internet:
    - remote access vs. remote execution.

16/02/2009

Modes of Access

<table>
<thead>
<tr>
<th></th>
<th>Public use Files</th>
<th>Scientific use Files</th>
<th>Extracts (subsets)</th>
<th>Public tabulations</th>
<th>Special (bespoke) tabulations</th>
<th>Secure remote access / execution</th>
<th>On-site safe centres</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Census, Register, or Microcensus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household Surveys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data from Administrative Registries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other economic and financial data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Current modes of access, regardless of distributor, in European countries.

- > 9 countries
- 6-8 countries
- 3-5 countries
- < 3 countries
Current modes of access

• Table 1 provides evidence that:
  
  – There has been significant progress in granting access to anonymised household survey data and (to a lesser extent) to census data;

  – Secure modes of access to more detailed datasets (safe centres and remote access facilities) are undergoing a rapid evolution.

3. Level of access, and who gives access to national data
Three country studies

Different options for sharing dissemination responsibilities between National Statistical Institutes and CESSDA Data Archives.

France

<table>
<thead>
<tr>
<th>National Statistical Institute (INSEE)</th>
<th>CESSDA Data Archive (Réseau Quetelet)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• PUFs</td>
<td>• PUFs</td>
</tr>
<tr>
<td>• Bespoke tabulations</td>
<td>• SUFs</td>
</tr>
<tr>
<td>• Remote access (pilot project)</td>
<td>• Contributes to preparing bespoke tabulations</td>
</tr>
<tr>
<td>• On-site access</td>
<td>• Collaborates to pilot project</td>
</tr>
</tbody>
</table>

⇒ Case 1: A Data Archive disseminates anonymized data and offers mediation services for confidential data.
### Ireland

<table>
<thead>
<tr>
<th>National Statistical Institute (CSO)</th>
<th>CESSDA Data Archive (ISSDA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• SUFs</td>
<td>• PUFs</td>
</tr>
<tr>
<td>• Public tabulations</td>
<td></td>
</tr>
<tr>
<td>• Bespoke tabulations</td>
<td></td>
</tr>
<tr>
<td>• Secure remote access</td>
<td></td>
</tr>
<tr>
<td>• On-site access</td>
<td>• Support to users.</td>
</tr>
</tbody>
</table>

⇒ **Case 2:** Data Archive contributes to disseminating anonymized data.

### Sweden

⇒ **Case 3:** Data Archive does not disseminate NSI data but offers information and support to users.

<table>
<thead>
<tr>
<th>National Statistical Institute (Statistics Sweden)</th>
<th>CESSDA Data Archive (SND)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Extracts (subsets)</td>
<td></td>
</tr>
<tr>
<td>• Public tabulations</td>
<td></td>
</tr>
<tr>
<td>• Bespoke tabulations</td>
<td></td>
</tr>
<tr>
<td>• Secure remote access</td>
<td>• Support to data users.</td>
</tr>
</tbody>
</table>
The global picture

• Different types of mediation by Data Archives:
  1. Provision of information and support services to data users.
  2. Dissemination of anonymized data.
  3. Dissemination of anonymized data, and support for confidential data.
  4. Dissemination of anonymized and confidential data (NO, UK).

• How are these different functions distributed throughout Europe, by mode of access?

<table>
<thead>
<tr>
<th></th>
<th>Access granted by the National Statistical Institute through the Internet</th>
<th>Access granted by the National Statistical Institute in other form (mail, on-site)</th>
<th>Access from a CESSDA data archive</th>
<th>Access from producer, with support from a CESSDA data archive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Use Files</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scientific Use Files</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extracts (subsets)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Tabulations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special (bespoke) Tabulations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secure remote access/execution</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On-site safe centres</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Who distributes data, by mode of access, in countries in which there exists a CESSDA Data Archive.

- > 9 countries
- 6-8 countries
- 3-5 countries
- < 3 countries
The global picture

• Table 2 provides evidence that:

  – Though many forms of access are dealt with directly by NSIs, CESSDA Data Archives play a role in several countries.

  – CESSDA Data Archives not only disseminate anonymized data, but also offer support for users of more detailed data (also including users of safe centres and remote access facilities).

  – Some CESSDA Data Archives are involved in the dissemination of detailed data (the case of Norway will be presented shortly).

4. Access to Eurostat and international data
Access to Eurostat data

- Modes of access: SUFs and on-site safe centre.
- Eurostat negotiates agreements with institutions, not individual researchers ⇒ Role of research and higher education institutions as intermediaries.
- No role for Data Archives so far.
- Similar in other international organisations (OECD).

5. Costs of access
Costs of access

- Many countries have cut costs of access. Where costs subsist, they are offset by a Data Archive or Research Council on behalf of the national scientific community.

- In some countries, researchers still pay fees individually.

- Disparities have repercussions on access to Eurostat data: there is currently no arrangement for Data Archives to cover costs for the European research community.

6. Additional functions of CESSDA Data Archives
Functions of CESSDA Archives

• Direct dissemination to researchers:
  – Anonymised data and (sometimes) confidential data, as shown above.

• In addition, indirect mediation services:
  – Documentation, metadata preparation (GR, FR…)
  – Anonymisation (DE, NO, …)
  – Archiving, deposit of government data (UK, …)
  – Covering costs for access (FR, …)
  – Information and support to data users (SE, SI…).

Conclusions
Conclusions

• Enhanced production of rich and useful data parallels rising complexity of access systems, both at national and (even more) at European levels.

• To satisfy an increasing demand of information services, documentation, and anonymisation, substantial resources are needed.

• A European Research Infrastructure may provide resources and assist NSIs in upgrading data services.

• CESSDA Data Archives have proven their ability to do so at the national level: can a similar collaboration take place at the European level?

16/02/2009  Conclusions

Thank you!

Roxane Silberman, CESSDA-Réseau Quetelet
Paola Tubaro, CESSDA-Réseau Quetelet

16/02/2009