Data Archives and the ethics of Research Data

Presentation at the 40th IASSIST Conference
Toronto, June 3-6, 2014

Katharina Kinder-Kurlanda
GESIS - Leibniz Institute for the Social Sciences
katharina.kinder-kurlanda@gesis.org

Laurence Horton
The London School of Economics and Political Science
@laurencedata, L.Horton@lse.ac.uk
Intellectual Property Rights

• Picture is confused and confusing
  – Within EU nations
  – Across EU nations
  – Internationally
# Comparing IRP

<table>
<thead>
<tr>
<th>Germany</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No IPR for facts. Uncommon for datasets</strong></td>
<td><strong>No IPR for facts. Possible for data.</strong></td>
</tr>
<tr>
<td><strong>Requirement</strong></td>
<td><strong>IPR for “original” works: “skill, judgement, labour”.</strong></td>
</tr>
<tr>
<td>• Originality: Gathering multiple works, data, independent elements</td>
<td></td>
</tr>
<tr>
<td>• Individuality: personal creation – structured and subjective choices</td>
<td></td>
</tr>
<tr>
<td><strong>Use</strong></td>
<td><strong>Copying facts, “fair dealing” (private and non-commercial research)</strong></td>
</tr>
<tr>
<td><strong>Copying facts, quotes, personal use</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Third party</strong></td>
<td><strong>Copyright is transferable. Employer rights.</strong></td>
</tr>
<tr>
<td>Copyright is non-transferable. Right holder is always the author.</td>
<td></td>
</tr>
</tbody>
</table>
EU Database directive 1996

- Applies to independent works, data or other materials
  - Systematically or methodically arranged;
  - Accessible by electronic or other means;
  - Result of a substantial investment (in the creation of the database rather than in the creation of facts).
- The person or entity which has made the substantial investment qualifies as the database producer and enjoys the exclusive database rights.
- Using a non-substantial portion of the database is allowed without consent.
- Consent required for:
  - Copying it, or its essential parts (reutilization);
  - Making it, or its essential parts, available to the public (extraction);
  - Retrieving (i.e. copying or downloading) substantial portions of the database;
  - Repeatedly and systematically retrieving non-substantial portions of the database;
  - Reusing (i.e. publishing) substantial portions of the database.
Data protection legal context in the UK

• Data Protection Act (DPA)
• Information Commissioner Office (ICO) guidance and commissioner precedents
Data protection legal context in Germany

• Federal Data Protection Act = Bundesdatenschutzgesetz (BDSG)
  – “The purpose of this Act is to protect individuals against infringement of their right to privacy as the result of the handling of their personal data.”
  – “Personal data” = any information concerning the personal or material circumstances of an identified or identifiable natural person
Anonymisation in the UK

"Anonymisation is the process of turning data into a form which does not identify individuals and where identification is not likely to take place."

See: http://www.ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation
Anonymisation in Germany

BDSG, §3,6 on 'factual anonymisation':

“’Rendering anonymous’ shall mean the alteration of personal data so that information concerning personal or material circumstances cannot be attributed to an identified or identifiable natural person or that such attribution would require a disproportionate amount of time, expense and effort. “
Legal context of the EU

1995 Data Protection Directive

• personal data is **any** information relating to an individual
• apply when a person can be identified, directly or indirectly
• implemented (unevenly) through member states
Current issues for researchers

Lack of harmonization in data access regimes due to legal uncertainty

• What's fine for a Belgian is not fine for the Dutch

• Gaps, inconsistencies and contradictions
  – Under what circumstances are small cell values problematic...?
  – What degree of security measures is 'enough'?
  – Does the door of a safe room need to be closed...?
Current issues (2)

Technological developments are ahead of legal ones

- Therefore 'best practice' has become very important as everybody is cautious (Avoiding unnecessarily conservative approaches)
  - e.g.: the proposed EC current reforms
Current Issues (3)

Balancing researchers' needs against a public mood of (understandable) suspicion

• Example: high profile cases of personal data leakage (not research data, but still...)

• Example: Albrecht report: Legitimate concerns regarding marketing interests pushing for cutting down on privacy lead to researchers being 'punished'
  – But: Data becomes more transparent through research!
Archives to the rescue?

• Data archives are expanding their role within data infrastructures. Not just end point data providers, but support services for those creating data
  – We have an interest in ensuring good quality data is produced

• Archives build up a store of expertise through experience
  – We encounter a lot more data and a diversity of projects that most researchers never encounter
Where archives help

Combating ignorance in the research community about data protection laws, by...

• Active help with specific queries
• Proactive help: Creating training and support materials
Providing the technological capacity to share data that could not ordinarily be shared
• Legally binding user licenses
• Secure data services for rich but sensitive data
Example: GESIS SDC - Secure Data Center

- Contractual safeguards
- Technical safeguards
- Organisational safeguards
Thank you!