

The Role of Licensing and Enforcement Mechanisms in Promoting Access and Protecting Confidentiality

Marilyn Seastrom,
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education;
Candice Wright,
Carnegie Mellon University;
John Melnicki,
Harbor Lane Associates



Introduction



∅ Legislative mandates require the collection and dissemination of specific data elements, including individually identifiable data.



∅ Confidentiality laws require the protection of individually identifiable data.



- ∅ U.S. Privacy Act of 1974, as amended

- ∅ Article 285 of the European Union Treaty





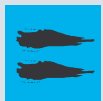
Public Dissemination



∅ Limited to aggregate data, models, and sometimes anonymised public use data files.



∅ Anonymised—by removing all direct identifiers and by coarsening the data.



∅ Leads to a tension between the dual needs of data protection and data access.





Protection vs. Access



∅ In the European Union, 2002 and 2003 laws grant access to qualified external researchers to some Eurostat confidential data



∅ In the U.S., OMB Information Quality Guidelines require that federal agencies ensure and maximize the:



- ∅ Quality, objectivity, utility, integrity, transparency and reproducibility of influential information.





Increased Access



∅ US—2002—Title V, Subtitle A, Confidential Information Protection (CIP 2002) of the E-Government Act includes provisions for sharing confidential statistical data with qualified external researchers.



∅ Mechanisms:

- ∅ Data centers,
- ∅ Remote access, and
- ∅ Data licensing/sharing agreements.





Data Agreements



∅ This report will explore the use of data sharing agreements in the United States and abroad.



∅ These agreements allow external researchers to use protected original data files in a secure environment at their home institutions, subject to the terms and responsibilities specified in the agreement.



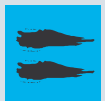


Data Agreements



∅ Application

- ∅ Describes requested data,
- ∅ Designates site for analysis,
- ∅ Specifies length of requested use,
- ∅ Describes proposed research,
- ∅ Explains why restricted-use data are needed, and
- ∅ Identifies all members of the research team.





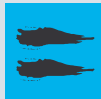
Data Agreements



∅ Data Security Plan



- ∅ Specify the exact location where data/printouts will be used and stored,
- ∅ Describe the physical security arrangements,
- ∅ Describe computer security, both hardware and software,
- ∅ Agree that the data will not be backed up on a routine basis, and
- ∅ Agree that no remote access will be allowed.





Data Agreements

∅ Additional Features



- ∅ All authorized users submit confidentiality pledge,



- ∅ Institutional concurrence,



- ∅ Statistical agency review of researcher's products,



- ∅ Cell size limits on published/released statistical aggregates,

- ∅ Notification of changes in personnel,

- ∅ On-site security inspections, and



- ∅ Terms for termination—return or destruction of data.



US Data Agreements



∅ Identified 16 different data agreements

- ∅ Provide access to confidential or restricted/limited access data files that have been de-identified, but contain data that could be used for indirect identification of individuals.



∅ Most are for government data, although some are administered by third party research centers.

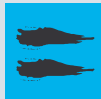




US Data Agreements



- Ø AHRQ: Agency for Healthcare Research and Quality
- Ø BJS/NIJ: Bureau of Justice Statistics/Nat. Institute of Justice
- Ø BLS: Bureau of Labor Statistics
- Ø CMS: Centers for Medicare and Medicaid Services
- Ø CDS: Duke University Center for Demographic Studies
- Ø ICPSR: Community Tracking Study
- Ø ICPSR: National Election Studies/PSID
- Ø NCES: National Center for Education Statistics
- Ø NCI: National Cancer Institute
- Ø NHLBI: National Heart, Lung, and Blood Institute
- Ø NIA: National Institute on Aging
- Ø NICHD: Nat. Inst. of Child Health and Human Development
- Ø NSF: National Science Foundation
- Ø OJJDP: Office of Juvenile Justice/ Delinquency Prevention
- Ø RAND: RAND
- Ø CPC: University of North Carolina Population Center





EU Data Agreements

Ø Eurostat

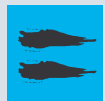


Ø Denmark



Ø Finland

Ø Ireland



Ø Netherlands

Ø Portugal



Ø Sweden

Ø United Kingdom





Other Data Agreements

∅ Australia

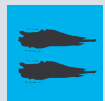


∅ Canada



∅ Hong Kong

∅ India



∅ Israel

∅ New Zealand

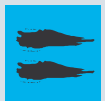




Comparisons



∅ EU external access is limited to anonymised data files, US and other countries provide external access to de-identified files.



∅ Individual differences exist, but the basic terms of data agreements are similar across organizations within the US and across countries internationally.



∅ Success is tied to the strength of the agreement.

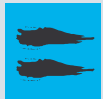


Data Agreements

∅ Applications all require:



- ∅ Description of requested data,
- ∅ Indication that the data will be used for research/statistical purposes,
- ∅ Agreement is not transferable to other persons or institutions; and
- ∅ Explanation of why restricted-use data are needed (US only),
- ∅ Description of proposed research (30—all but 2 other countries), and
- ∅ Length of requested use (30—all but 2 other countries).





Elements of Enforcement

∅ Conditions influencing compliance:

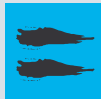


∅ Researcher's signature (all 32),



∅ Access limited to research team (all 32),

∅ Attempts to identify individuals prohibited (all 32),



∅ Institutional concurrence (23),

∅ Security pledges from all users (23),



∅ Minimum cell size requirements for publishing (10), and



∅ Review/notification of publications (20).



Elements of Enforcement

∅ Data Security:



- ∅ Data security plan (27),



- ∅ On-site data security inspections (17),

 - ∅ The agency reserves the right to inspect.



 - ∅ Limited information as to whether inspections are used.



 - ∅ In some cases the inspections are not done routinely.



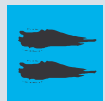
 - ∅ Inspection clause carries more weight with users if they know that inspections regularly occur.



Termination/Close Out



Ø All 32 reserve the right for the loaning agency to revoke agreement and demand return of data if a violation occurs.



Ø 25 include requirements for the return or destruction of the data at end of the agreement, in accordance with specified procedures.





Penalties



∅ US, 15 include revocation clauses,

- ∅ Privacy Act—misdemeanor, \$5,000
- ∅ CIP 2000—Class E felony, \$250,000/5 years
- ∅ Individual agencies penalties range from \$5,000 to \$250,000 with prison terms from 5 to 10 years, or revocation of grants



∅ EU, country supervisory authorities engage in legal proceedings, e.g.,

- ∅ Revocation of agreement and denial of future access,
- ∅ Legal proceedings, and
- ∅ Detention or imprisonment.





Penalties

∅ Other countries:



∅ Australia—\$5,000/2 years



∅ Canada—legal action



∅ Israel—violation of copyright laws



∅ New Zealand—\$1,000





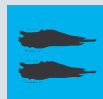
Measuring Effectiveness



∅ External Audits: 1999 GAO report on compliance of one agency with the Privacy Act.



∅ Failure to routinely monitor external users was cited as a weakness.



∅ Failure to track or monitor the return or destruction of data was also cited.





Measuring Effectiveness

∅ Self monitoring: limited information.



- ∅ Swedish paper submitted to the UN (2003) on data use agreements states that “a violation of confidentiality regarding microdata use has in fact hardly ever occurred in the National Statistical Institute data based research projects.”



- ∅ In the US, one agency (NCES) published the results of a 1998 analysis of security inspection reports, noting no known disclosures, and violations that posed varying degrees of risk of exposure.





Security Inspections

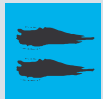
∅ Four types of violations:



- ∅ 1) Minor—easily corrected during security inspection:



- ∅ Failure to use proper signage of facility and/or on computer terminal,



- ∅ Addition of junior staff without adequate training on the protection of restricted access data.





Security Inspections

∅ Four types of violations:

- ∅ 2) Minor—but not corrected by one-time intervention, due to lack of oversight

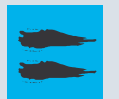
- ∅ Computers unattended,

- ∅ No log for check in/out of data,

- ∅ Data not properly stored when not in use,

- ∅ Designated space lacks basic security, and

- ∅ Designated space not located.





Security Inspections

∅ Four types of violations:

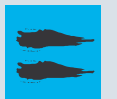
- ∅ 3) More immediate risk of disclosure due to access by unauthorized users through computer configurations:

- ∅ Any access to unsecured, public networks,

- ∅ LAN/WAN connections to unsecured server,

- ∅ Use of restricted data on unauthorized terminals, and

- ∅ Improper unsecured data distribution procedures.

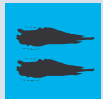




Security Inspections

∅ Four types of violations:

- ∅ 3) More immediate risk of disclosure due to access by unauthorized users through improper use of data:
 - ∅ File sharing with unauthorized users (e.g. new RAs),
 - ∅ Use of data at an unauthorized location (e.g. private home),
 - ∅ Researcher moves data to different institution without notification, and
 - ∅ Failure to return or document destruction of data at end of agreement.





Security Inspections

∅ Four types of violations:



∅ 4) Serious violation:



∅ An individual is identified and disclosed by someone using restricted access data.





Next Steps



∅ More routine use of security inspections.



∅ Implement termination procedures to help ensure that the data are not used for unauthorized purposes.



∅ Maintain complete and accurate records for all data agreements.



- ∅ Electronic record system can be used to maintain contact with users (e.g., reminders of current users and expiration dates, data updates).





**For further information,
contact:**



∅ Marilyn.Seastrom @ed.gov



This paper is intended to promote the exchange of ideas among researchers and policy makers. The views expressed in it are part of ongoing research and analysis and do not necessarily reflect the position of the U.S. Department of Education.