# DataTags: Legal Research & Development


## Alexandra Wood (Fellow) & Olga Slobodyanyuk (Law Student Intern)

Berkman Center for Internet & Society


**Privacy Tools for Sharing Research Data**

*A National Science Foundation Secure and Trustworthy Cyberspace Project*

## Data Use Agreements

**Objective**

- To analyze how different organizations interpret and apply statutory and regulatory requirements and common practices related to the sharing of research data in their data use agreements and policies.

**Approach**

- Review of a sample of data sharing agreements and policies from over 100 data repositories, data enclaves, individual studies or data series, and organizations, such as academic institutions, government agencies, commercial entities, and nonprofit organizations.
- Clustering of common terms into general categories and typologies of approaches within each category and formulation of corresponding DataTags questions and tags for each approach.

**Common terms and approaches at the data ingestion and storage stage**

- Eligible data and depositors
- Rights to an ownership of the data
- Data contents (including identifiability and sensitivity of data contents)
- Compliance with consent and human subjects protection rules
- Repository rights and responsibilities
- Acceptable end user access, use, and transfers of data (including: public access, restricted access, enclave access, embargoed access, and no access)
- Data retention and withdrawal
- Liability

**Common terms and approaches at the data access and use stages**

- Data ownership
- Access, use, sharing, and reuse restrictions (including tiered access models)
- Data confidentiality and security procedures
- De-identification requirements
- Data retention requirements
- Enforcement practices and procedures
- Liability provisions
- Attribution requirements

### Sample DataTags Question

**Under the agreement, with whom are users permitted to share the data? [Check all that apply.]**

- ❑ Any person named in the agreement (list of individuals provided as an attachment)
- ❑ Collaborators working on the same research project
- ❑ Any person named in the user's application for the data
- ❑ Any person upon obtaining written authorization from {the owner, the depositor, Dataverse}
- ❑ Any person under the same terms and conditions as the data were provided to {Dataverse, the user}
- ❑ Any person who has entered into an agreement with {the owner, the depositor, Dataverse}
- ❑ Any person

- ❑ The agreement prohibits users from sharing the data with anyone.
- ❑ The agreement does not specify with whom users may share the data

## University Data Classification & IRB Policies

**Objective**

- To better understand how universities interpret and apply the range of human subjects protection regulations, data privacy laws and regulations, and data security best practices at the federal and state levels.

**Approach**

- Review of IRB and data classification policies from 32 of the top research universities in the United States
- Compilation of excerpts of the privacy-related language from the various policies in a 180+ page appendix
- Analysis of the common features of the IRB and data aclassification policies, their relationship to each other, and their relationship to information privacy statutes and regulations

**Observations**

Common features of IRB policies

- Definitions of terms such as privacy, confidentiality, anonymous, de-identified, sensitive information, personally identifiable information, and protected health information
- Definitions typically reference regulations such as the Common Rule, HIPAA, FERPA, and state data protection standards
- Explicit terms such as data minimization, linkage, security, and retention requirements

Variations among IRB policies

- Identification of different typologies of approaches to privacy in IRB policies, ranging from policies that include no or minimal explicit requirements for privacy protection, to policies that cover comprehensive guidelines that extend beyond the minimum privacy protections required under the law

| | |
|---|---|
| Level 1 | **Non-confidential research information** *Public information* |
| Level 2 | **Benign information to be held confidentially** *Information the disclosure of which would not cause material harm, but which the University has chosen to keep confidential* |
| Level 3 | **Sensitive or confidential information** *Information that could cause risk of material harm to individuals or the University if disclosed* |
| Level 4 | **Very sensitive information** *Information that would likely cause serious harm to individuals or the University if disclosed* |
| Level 5 | **Extremely sensitive information** *Information that would cause severe harm to individuals or the University if disclosed* |

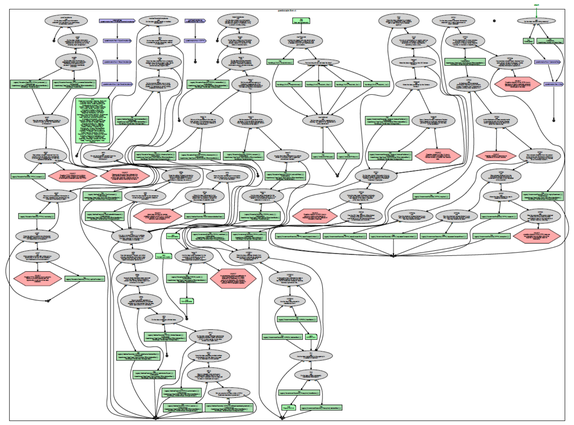
Common features of data classification policies

- Data classification policies typically include 3-5 tiers of categories and provide a short list of examples of the types of information that fall within each category
- Some university data classification policies are integrated with the university's IRB policies, and, at other universities, these policies are distinct.
- University data policies are broader in scope than IRB policies. They often address privacy and security issues beyond human subjects research protection, including GLBA, PCI-DSS, and other laws and standards related to the protection of information related to financial transactions, intellectual property and proprietary considerations, and laws related to export controlled data

## Privacy Statutes & Regulations

**Objective**

- To conduct legal research and draft memoranda analyzing the laws and regulations governing the sharing of research data.
- To translate the legal requirements into a set of inference rules for determining, based on the properties of a dataset and its holder, which legal provisions apply and how the dataset should be handled as a result.

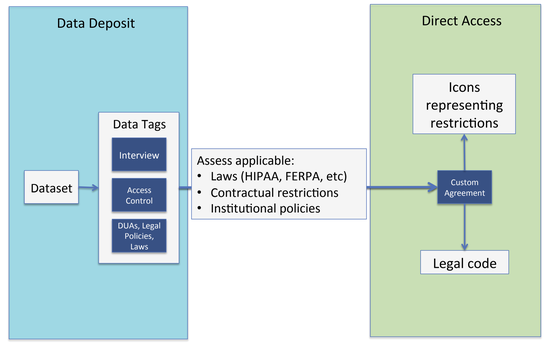


**Approach**

- Overview of each law: legislative history, definitions of key terms (including judicial and administrative interpretations), key provisions related to the collection, use, and sharing of data, and any relevant exceptions to the general provisions
- Legal analysis of each law and how it affects researchers and data repositories
- Identification of the different categories of information that are covered under each law, the properties of the data and its holder that are relevant to determining whether it falls into one of the categories identified, and the data handling requirements for each category of information
- Drafting of annotated questionnaire and formulation of corresponding tags and license terms

**Statutes and Regulations Covered**

- The Common Rule
- HIPAA Privacy Rule & Security Rule
- Substance abuse confidentiality regulations
- Family Educational Rights and Privacy Act (FERPA)
- Protection of Pupil Rights Amendment (PPRA)
- Education Sciences Reform Act
- Privacy Act of 1974
- Confidential Information Protection and Statistical Efficiency Act (CIPSEA)
- Title 13, US Code
- Driver's Privacy Protection Act




## Privacy Definitions

**Objective**

- To compare definitions of privacy and confidentiality (and related terms such as personally identifiable information and deidentification) across different areas of law and across different disciplines
- To cluster and conduct a type analysis of the different categories of approaches

**Approach**

- Survey of federal and state information privacy laws, regulations, and standards in the United States, including a spreadsheet of the definitions (and judicial and administrative interpretations) of the terms
- Mapping of the definitions of personally identifiable information to the terms presented in Paul M. Schwartz & Daniel J. Solove, "The PII Problem: Privacy and a New Concept of Personally Identifiable Information," 86 *N.Y.U. L. Rev.* 1814 (2011).

**Tautological Approaches**

- Cable Communications Policy Act of 1984: "The term 'personally identifiable information' does not include any record of aggregate data which does not identify particular persons," 47 U.S.C. § 551(a)(2)(A).
- Confidential Information Protection and Statistical Efficiency Act
- Education Sciences Reform Act of 2002

**Non-public Approaches**

- Gramm-Leach-Bliley Act: ". . . personally identifiable financial information; and any list, description or other grouping of consumers, and publicly available information pertaining to them, that is derived using any personally identifiable financial information that is not publicly available information." 17 C.F.R. 160.3(u)(1), (2).
- The Common Rule
- Federal Agency Data Mining Reporting Act of 2007
- Video Voyeurism Prevention Act of 2004

**Specific Types Approaches**

- HIPAA Privacy Rule: 18 identifiers enumerated under the definition of de-identified information, 45 C.F.R. § 164.514(b)(2)(i).
- Family Educational Rights and Privacy Act of 1974
- Privacy Act of 1974
- Driver's Privacy Protection Act of 1994