OSR Fair 2016: Data Use Agreements

**What is a DUA?**
A Data Use Agreement (DUA) is an agreement that must be entered into before there is any transfer or use of information by or on behalf of Stanford, or its agents, of any data that are subject to a DUA. A DUA is required for any research, teaching, or other work involving the use of information that is not publicly available or is not released by the provider of the information.

**What is NOT a DUA?**
A research Data Use Agreement (DUA) is not necessary when data are to be used for internal business, clinical trials, or administrative business in support of University and/or Hospital operations. DUA are not required when the data are to be used for internal training, education, or internal business. DUA are not required for collaborative discussions with outside entities regarding a proposed study in a DUA or NIH.

**Memo from Dean Arvin**
**Dated 2/5/15**
Access to large data sets has become a key component of research at Stanford. Often, the data provider, or recipient, requests the researcher or Stanford to sign a written or verbal agreement. This memo outlines when Stanford researchers may sign such agreements themselves and when Stanford’s University office to review and sign the agreement.

Stanford’s research contracting and procurement offices can answer questions and help you recognize these agreements:
- Agreements with government or non-profit entities are handled by the University’s Office of Sponsored Research.
- Agreements with industry are handled by the Industrial Contracts Office.
- Agreements with commercial or other data are handled by the Membership Office.

**Agreements for incoming Data**
You may sign a data agreement in your individual capacity under the following conditions:
- The nature of the data and the agreement is a matter of common agreement, for example, common prior practice, including online “click” agreements.
- The data do not include “personal identifying information” (PII), Protected Health Information (PHI), protected research data, or any data that would enable an individual to be identified.
- The data do not include any information that would enable an individual to be identified.

**Common Issues with DUAs**
- Data are not considered to be public records and are not subject to Freedom of Information Act (FOIA) requests. However, the identification of an individual must be removed from the data before the data can be distributed to an individual.
- All data must be de-identified before the data can be distributed to an individual.
- All data must be distributed to an individual in a form that does not enable the identification of an individual.
- All data must be distributed to an individual in a form that does not enable the identification of an individual.
- All data must be distributed to an individual in a form that does not enable the identification of an individual.
- All data must be distributed to an individual in a form that does not enable the identification of an individual.

**Common Data Providers**
- The most common data providers are usually other U.S. educational institutions.
- The University of California and the University of Texas at Austin are major data providers.
- The University of California and the University of Texas at Austin are major data providers.
- The University of California and the University of Texas at Austin are major data providers.
- The University of California and the University of Texas at Austin are major data providers.

**Common Questions from OSR**
- What is the purpose of the data?
- Who will use the data?
- What are the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
- What is the data to be used for?
Data Definitions and Guidance

**Definitions**

**Data**: Factual information used as a basis for reasoning, discussion, or calculation.

**Research Activities**: The conduct of scholarly inquiry as it relates to the performance of a study.

**Non-Research Activities**: Traditional business, clinical, financial, or administrative functions in support of University and/or Hospital operations.

**Stanford Memo from Dean Arvin regarding Data Use Agreements**

All Stanford Faculty and Staff are subject to the limitations and requirements regarding the right to enter into a Data Use Agreement. The OSR decision tree is intended to help determine in tandem with DoR’s policy not only which Office at Stanford will handle the Data Use Agreement, but also when a Data Use Agreement is either required or recommended.


**Resources and Links**

- Stanford webpage for PHI policies and compliance: https://acp.stanford.edu/hipaa/hipaa
- Stanford Industrial Contracts Office: https://sites.stanford.edu/ico/
- Stanford Office of Sponsored Research: https://doresearch.stanford.edu/research-offices/osr-office-sponsored-research

Data Classifications

**Data Risk Classifications**

**Low Risk Data**: The data is intended for public disclosure and any loss of the data would have no adverse impact on Stanford. Examples: Research data (at PI’s discretion), information in the public domain.

**Medium Risk Data**: The data is not generally available to the public, or the loss of the data could have a mildly adverse impact on Stanford. Examples: Unpublished Research data (at PI’s discretion), Stanford internal memos or reports, student records.

**High Risk Data**: Protection of the data is required by law, or the loss of the data or system could have a significant adverse impact on Stanford. Examples: SSN’s, financial account numbers, export controlled information, donor contact information.

**Health-Related Data Definitions and General Classifications**

**De-identified Data Set** (Low Risk). Health information that does not identify an individual.

**Limited Data Set** (Medium Risk). Health information that includes up to three of the following individual identifying elements: dates (admission, DOB, DOD, service, etc.), location (state, city, zip code), and age.

**Protected Health Information** ("PHI") (High Risk). Individually identifiable health information consisting of 18 specific identifying data elements that identifies the individual.