

---

# Cascading compliance requirements for thesis-related data and software: A survey of institutional policy at American (AAU) research universities

Gail Clement\*<sup>1</sup>

<sup>1</sup>Caltech Library – 1200 E California Blvd, Pasadena, CA, 91125, United States

## Abstract

According to the CODATA/Research Data Alliance draft International Principles on the Legal Interoperability of Research Data (<https://rd-alliance.org/group/rdacodata-legal-interoperability-ig/wiki/legal-principles-data.html>), the distribution and dissemination of research data (including software) is, to a significant extent, governed by law and institutional policy that specify ownership of these research outputs and the rights such ownership conveys. In the United States, copyright does not affix to purely factual information, but may govern sharing and reuse of protectable elements in a dataset (selection, arrangement, database design, and documentation) and software (original code, comments, and documentation). Moreover, institutional policy may establish that ownership of research data and software belongs to the University, not the creator. Researchers may be obligated to cede control of their outputs to the University's technology transfer or legal counsel's office. Additionally, research funding agencies may require, as a condition of awarding a grant, that outputs created with their support be disseminated in a particular manner.

This confluence of circumstances leaves graduate students in a complex and confusing compliance framework requiring clear understanding and careful navigation of cascading requirements. As enrolled students, does their work fall under the institute's copyright policy statements? Are the textual components of their theses handled differently than the associated research data and software? Is the research data and software they produce to fulfill graduation requirements owned by anyone? If so, who? And what rights and choices do students have to control, share, and invite reuse of their research?

The author, as a member of the CODATA/Research Data Alliance Legal Interoperability of Research Data interest group and a coauthor of the draft principles, has investigated the legal and policy provisions surrounding research data and software created at research universities in the United States. Analysis of the policies of the institutions belonging to the Association of American Universities reveals a highly variable and confusing landscape for American graduate students at this time. The seemingly incomplete guidance and at times contradictory assertions being made around ownership of graduate student data and software yields uncertainty and risk for non-compliance with legal and policy requirements. This presentation will summarize the chief areas of American law and policy that impact the sharing of graduate research outputs and point out areas of uncertainty and confusion. The presenter will share summary data from the full text analysis of AAU institutions' posted IP policies, theses manuals, and repository deposit licenses that represent assertions about

---

\*Speaker

student rights in their research data and software. The discussion component of the presentation recommends model language that could be adopted to ensure greater clarity and compliance to help graduate students and their advisors make sound dissemination decisions.

On the author:

Gail Clement is an academic/research librarian with extensive experience in scientific and grey literature publishing, scientific repositories, copyright and authorship education, and E-research support. She serves as Head of Research Services at the Caltech Library, leading a team of 12 subject librarians, repository, metadata and licensing specialists to develop quality information resources and authorship services for the campus community. She initiated the innovative Author Carpentry program at Caltech to provide training and tools in responsible, reproducible, and reuseable research information handling, building capacity for scientists to craft the fully executable Science Paper of the Future.

In her previous position as Professor and Scholarly Communication librarian at Texas A&M, Gail led the University's ORCID integration initiative, managed the Online Access to Knowledge (OAK) fund covering authors' open access publishing fees, and oversaw the retrospective digitization and online re-publication of many thousands of theses, dissertations, and engineering reports. Her curriculum on academic authorship and research publishing established the Library's close partnership with the Graduate School and the 'Grad Aggies' professional development series. It has also been offered through academic publishing groups such the US Electronic Thesis and Dissertations Association (USETDA).

Gail's professional service includes membership on the CODATA/Research Data Alliance's Legal Interoperability of Research Data Interest Group; the Overleaf Steering Group; and past service on ORCID Outreach Steering Committee. Gail is a member of the Editorial Board for the Journal of Librarianship and Scholarly Communication (JLSC) where she co-edited the 2015 special issue on data sharing, data publication, and data citation. She has assisted with peer review for several scholarly journals including College & Research Libraries. Gail writes and speaks regularly on scholarly communication, scientific knowledge management, and research publishing topics.

**Keywords:** copyright, research data, institutional policy, compliance