The Library
Research, Teaching, and Learning

Berkeley, CA 94720-6000

24 March 2022

Dear Jordan Jacobs:

We write in response to Associate Vice Chancellor for Research Rugg's 17 February 2022 e-mail that distributed and sought comment upon a revised draft UC Research Data Policy (Revised Draft Policy). Thank you for your help in stewarding this feedback process from UC Berkeley stakeholders.

The Research Policy Analysis and Coordination (RPAC) group has made meaningful and positive changes to the Revised Draft Policy. We appreciate the careful consideration and inclusion of most of our previous comments and specific edits. In light of a few minor issues introduced by the most recent revisions proposed by the committee, below we offer final suggestions that will strengthen and clarify the final policy for UC faculty, researchers, and students. We also include an attachment with edits and suggestions to the Frequently Asked Questions (FAQs) document that accompanied this latest version of the Draft Policy.

Definitions

1. ***In brief: The definition of "Principal Investigator(s) (PI)" should exclude University-affiliated students conducting independent research, so that the UC does not claim ownership of students' independently-created research data.***

Under the revised policy, there is a seemingly narrow (yet remarkably common) circumstance in which the University could arguably claim ownership in Research Data created solely by students. We do not think the Research Data Policy intends this result.

The Revised Draft Policy sets forth that the Regents own Research Data "created by or at the direction of **[1]** University Researchers or **[2]** Principal Investigators during the course of University Research…" (enumeration added). It is important to consider each of these two encompassed categories separately to assess their impact on student-generated research:

    a. "University Researchers"

Likely crafted to avoid encumbering student-generated research data unnecessarily, the Revised Draft Policy provides that students can be considered "University Researchers" only to the extent the students "participate in the design, conduct or reporting of a Principal Investigator's research project." As such, under this definition of "University Researchers," the UC Regents will own student-created research data only when a student creates that data in the context of supporting a Preliminary Investigator's research. This result aligns with the

University's general positions regarding student-created materials (*see, e.g., Copyright Ownership Policy*).[1]

    b. "Principal Investigators"

However, what happens if the student is not a "University Researcher" working for a "Principal Investigator," but rather is the "Principal Investigator," itself? Students otherwise can satisfy the prerequisites of the definition of "Principal Investigator," in that they are "University-affiliated" and can be "primarily responsible for a research project…regardless of the source of funding." In turn, if such a student leading an independent research project shall be considered a "Principal Investigator," then the University would be claiming ownership of research data created by students solely in the context of their independent research. We do not think this was or should be the policy's intent.

This is not merely a hypothetical issue, because students regularly serve as their own principal investigators. Our Library stewards an open access publishing fund covering journal publication fees for research articles by University faculty, staff, and students. Many of the papers whose publishing we fund are authored by graduate students serving as both the Principal Investigator and corresponding author. Thus far in FY2021-22 alone, we have covered publishing fees for 17 open access journal articles in which a UC Berkeley graduate student was the sole or lead author. And the overall number of student-led independent research projects generating research data that have not sought our funding for open access publishing is likely much higher.

If the new Research Data Policy treats such students as "Principal Investigators," then the University would in turn claim ownership of all of the research data generated by these independent investigations in a manner inconsistent with all other University policies that expressly waive University ownership claims to student creations. To avoid establishing University ownership over independently-generated student research data, we suggest the addition of the following text under the definition of PI:
- "For purposes of this Policy, University-affiliated students conducting independent research shall not be considered Principal Investigators."

2. ***In brief: In the definition of "University Research," the phrase "University resources" should be replaced with "University Research Facilities" to clarify scope and align with FAQs.***

---

[1] *University of California Copyright Ownership Policy,* Section III(3): "As between the University and its students, copyright ownership of works prepared by University students (including graduate students) resides with such students, unless the work: (a) was created primarily in the course or scope of the student's University employment; (b) involved the use of Significant University Resources; (c) is a Sponsored Work, Contracted Facilities Work, or Commissioned Work; or (d) was created under a separate agreement that specifies a different copyright owner. For the purposes of this section, a student's financial aid is not considered Significant University Resources. Absent extraordinary circumstances, copyright ownership of theses or dissertations authored by University students resides with such students." Available at
https://copyright.universityofcalifornia.edu/resources/copyright-ownership.html.

The Revised Draft Policy reintroduces a University ownership claim over any research data generated by Principal Investigators and University Researchers "using University resources." In an earlier round of comments we had suggested that this broad "University resources" phraseology exceeds rights granted to the University in parallel policies, such as the *Copyright Ownership Policy* (which requires that an author must take advantage of "*Significant* University Resources" for the UC to claim copyright in the work) and the *Policy on Ownership of Course Materials* (which requires that an instructor have received "*Exceptional* University Resources" for the UC to claim ownership). Further, each of these comparable University policies delineates the difference between such "ordinary" and "significant" or "exceptional" University resources, to further resolve ambiguity and narrow the scope of University-claimed ownership. Unfortunately, this Revised Draft Policy does not.

From a practical perspective, the more expansive "University resources" term used in the Revised Draft Policy would sweep under University ownership any Research Data created pursuant to things like course assignments, library search queries, classroom experimentation, etc.—all of which could involve the requisite "scientific inquiry" to be considered "Research Data"[2] and all of which could be generated in ordinary campus settings like classrooms, meeting rooms, or libraries. Based on prior rounds of revision to this draft data policy, we believe such data was never intended to be subsumed within the four corners of the policy.

The FAQs now attached to the Revised Draft Policy further support our interpretation. The first three FAQs address research conducted with or by third parties, for which the University makes efforts (as we had proposed in earlier comments) to disclaim ownership of third-party generated research data, as well as to disclaim ownership of research data created in "normal" University facilities like classrooms, meeting rooms, and libraries. Yet, while the FAQs recognize that research data should not include data created using "only general campus facilities," the Revised Draft Policy's definition of "University Research" is not so-limited. Under the policy language, the University would claim ownership of data involving "scientific inquiry" created *anywhere* on campus using *any* campus resources at all.

To establish consistency between the FAQs and the Revised Draft Policy, and to bring the Revised Draft Policy in line with the limitation of "exceptional" or "significant" University resources used in parallel ownership policies, we suggest the following two modifications:
- Import the definition of "University Research Facilities" from the FAQs into the definitions section of the Revised Draft Policy; and
- In the definition of "University Research," replace the term "University resources" with the now-defined "University Research Facilities."

3. ***In brief: The definition of "Research Data" should be modified to state that research data includes recorded information that "documents, supports, and validates research findings."***

---

[2] We address the use of the phrase "scientific inquiry" in point three, below.

The current Revised Draft Policy's definition of "Research Data" can be improved by providing more guidance as to what kinds of recorded information are subject to the Policy.

A previous version of the Draft Policy offered the following definition of "Research Data": "Recorded information reflecting original observations and methods **resulting from a research study or necessary to document, support, and validate research findings**…" (emphasis added)[3]. In the new draft, "Research Data" is instead defined as "Recorded information reflecting original observations and methods **resulting from scientific inquiry, regardless of the form or the media on which they may be recorded**" (emphasis added). One can reasonably debate whether shifting from observations emerging from "a research study" to observations emerging from "scientific inquiry" renders the definition of "Research Data" broader or narrower. In our view, the use of either "research study" or "scientific inquiry" is of some worry, as it could potentially include data generated pursuant to course assignments, library search queries, classroom experimentation, and the like. However, we would be less concerned about either of those terms if the new policy simply reinstated the second part of that earlier definition—i.e. data that is "necessary to document, support, and validate research findings." The reinclusion of that explanation would make clearer the scope of what constitutes "Research Data," and would eliminate the potential for inclusion of unintended data related to typical student and scholarly functions. Further, the phrasing would better scope researchers' (and campus support units') data preservation and management obligations under the "Compliance/Responsibilities" portion of the new policy.

We also note that reinstating the "necessary to document, support, and validate research findings" terminology would add context for why laboratory notebooks, in particular, are subsequently called out as an example of the type of data subject to the policy. Perhaps an even clearer approach, though, would be to remove the specific example of laboratory notebooks and instead point to a more comprehensive set of examples that could be offered in the Frequently Asked Questions document.

Accordingly, we suggest the following modifications to the current "Research Data" definition:
- Reinstate previously-included language so that the definition reads: "Recorded information reflecting original observations and methods resulting from scientific inquiry, *including that which documents, supports, and validates research findings*, regardless of the form or the media on which they may be recorded"; and
- Remove the specific reference to laboratory notebooks in the current definition and build out an FAQ with a more comprehensive list of examples of what qualifies as "Research Data."
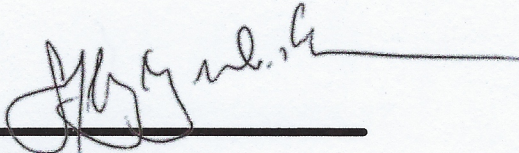
4. *Updates to Frequently Asked Questions.*

Finally, our suggested edits to the FAQs are redlined in the accompanying document.

---

[3] Draft Policy version from 07/16/2021

As always, thank you for considering and incorporating as possible our comments throughout this process. We reiterate what we have said in the past about this Draft Policy: Having a UC Research Data Policy will provide needed clarity and guidance across the UC system. We are happy to speak with the RPAC directly if there are questions about our remarks.

Sincerely yours,

Jeff MacKie-Mason
University Librarian
Chief Digital Scholarship Officer
Professor, School of Information and Professor of Economics

Shawna Dark
Chief Academic Technology Officer & Assistant Vice Provost, Undergraduate Education
Executive Director of Research, Teaching and Learning