Topics in RDM #1 - ICPSR and Depositing Data

1. Introduction to ICPSR from Harrison: see “10 Things You Need to Know about ICPSR” in the Discussion 1 folder in Box
   1. Funding for ICPSR runs on a consortia model - members pay dues based on their Carnegie rating
   2. At Berkeley, Jon Stiles is the official representative (OR) and Harrison Dekker is the designated representative (DR).
   3. CAUTION: ICPSR serves as an archive for some data sets (census, etc.) that are much easier to access from other sources.

2. Discussion:

1. Question for people with experience using ICPSR: when researchers come for consultations with ICPSR, how much information do you provide?
   * 1. General overview - if students learn about ICPSR on their own, it’s common for them to reach a dead end. If they have a general topic, ICPSR may be one place they look at.
     2. For any reference librarian, take them as far as you’re comfortable with. When you don’t know what you’re looking at, send them to the Data Lab. Basic census statistics are available in ICPSR, but those are much more easily accessible in other places. This is a good example of when NOT to use ICPSR.
     3. Repos like ICPSR or Dataverse are great for discovery. Well-cited data on a faculty web page is good, too, however. Not everyone has begun to deposit in a repository yet, though. We’re in a transition period. We ought to be helping our faculty make these deposits, so we will be able to find it; so it doesn’t go away when the faculty member retires and their web site goes away.
   1. Question for Jon Stiles and Harrison Dekker: how would you recommend students/researchers use ICPSR?
      1. We’re not actively encouraging students to upload to ICPSR, should we be? Or to OpenICPSR?
         1. ICPSR is a curated collection, whereas OpenICPSR is not curated. Your dataset may not fit ICPSR’s curation vision. If your data is in the curated collection, then it needs to fit their collection guidelines.
         2. The benefit of ICPSR is long term curation, but it’s not fully open!
         3. No fee to members to upload to OpenICPSR. For more information on pricing for OpenICPSR, please visit [Plans & Pricing](https://www.openicpsr.org/home/pricing)

3. What data repository should we use? Should UC Berkeley recommend to faculty and students that they upload data to ICPSR? What about OpenICPSR? If we establish our DataVERSE platform, is that an alternative to ICPSR?

Answers from Harrison Dekker and Jon Stiles:

For now, Dataverse has only been approved at UC Berkeley for hosting datasets that the library acquires. It’s currently out of our scope to suggest that faculty deposit their own data there (though not a bad idea). It is really easy to underestimate the time and effort required to help a researcher get data into an archivable format. To what extent do we have time and staff to provide help?

DASH could be an alternative, as would be other open platforms. Dataverse handles DDI (Data Documentation Initiative, <http://ddialliance.org/>) metadata, so we could choose to curate our datasets locally.

4. Data Seal of Approval (basic introduction - 5 minutes)

Present [Guidelines](http://datasealofapproval.org/en/information/guidelines/) related to data producers, repositories, and consumers as a handout to guide discussion attendees through this portion of the session (handout is in the Box folder). If a repository has the Data Seal of Approval or fits another certification or standard, will that make users less critical of the data they retrieve from that repository?

* How is this better than the unwashed internet for students? (I.e., does ICPSR vet the content and findings of the study, or only the metadata and the process by which the data has been managed?)
* This dovetails into a much larger question of data literacy. There is a difference between searching for data and searching for statistics. Sometimes these lines can become blurred, especially since we have such a short period of time with many of our students.

5. . Conclusion/next steps: next workshop; feedback (2 minutes)

Idea for Topic #2: Data literacy / statistical literacy for the social sciences

Jess has a short bibliography Celia mentions presentation by Adam Beauchamp at this year’s ACLR, asks, “Did anyone see it?”

Jesse: “ICPSR training taught us to understand the levels of service that we can provide, [and that it was totally ok to work at the level comfortable to you, given your level of knowledge.]"

Afterwards, Susan Edwards remarked that she liked Jamie’s mental “checklist” of which repository fits which cases.