

Data Curation Program Strategic Plan

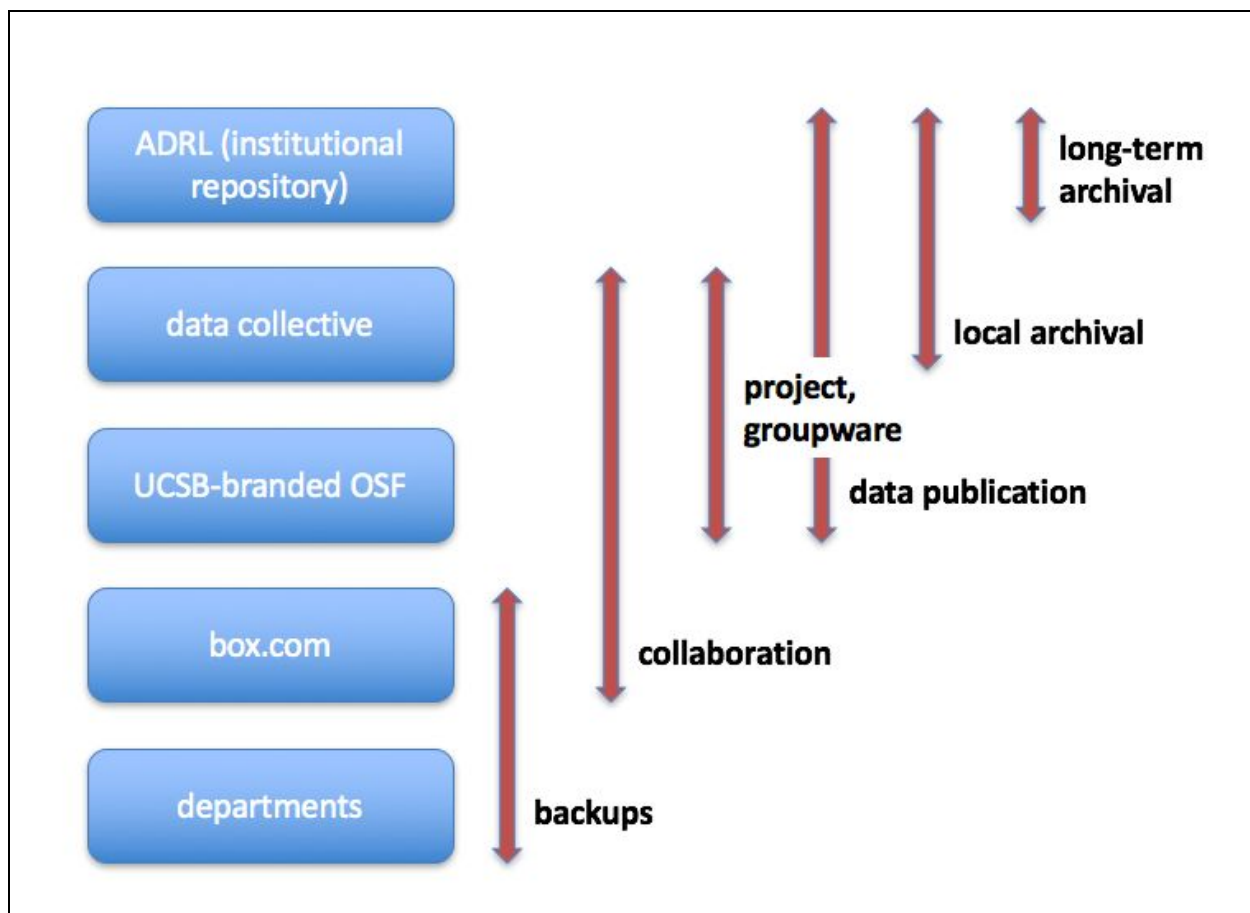
The mission of the Data Curation Program is to support the management and preservability of research data generated on campus. The Program aims to connect researchers with curation services, whether the latter be internally- or externally-provided. The Program engages researchers throughout the research data lifecycle, from providing consultation during the initial phases of projects, to offering hands-on assistance in establishing effective day-to-day data management strategies, to identifying solutions for long-term archiving of final products.

Data curation services

The Program's interactions with faculty over its first year have revealed a critical need for a campus-based repository for faculty research data. For some researchers, there are simply no appropriate discipline-specific repositories available for their work, or external repositories exist but are too costly or not a good fit for their data or goals. For others, discipline-specific repositories exist and deposit may even be required by granting agencies, but the form the repositories require the data to be deposited in, and the limited services the repositories provide, make external deposit a burden, not a help. And for many researchers, external repositories do not fulfill the desire to maintain data locally in an accessible, working form to sustain threads of research across projects and grants, and to support collaboration among project teams both on campus and between UCSB and partner institutions.

A repository by itself is not sufficient, however. As shown in the figure on the next page, a repository exists in a spectrum of overlapping solutions that provide data curation services ranging from simple backup, to data publication, to long-term archiving. Researchers would benefit from the availability of all these services.

- S1. Develop (and support thereafter) the “Data Collective,” a Library-mediated, self-deposit repository for faculty data.
- S2. Contribute to the development of an institutional repository.
- S3. Develop a UCSB-branded instance of the Open Science Framework (OSF), and participate in the OSF Ambassador program.



Data management/curation services and the functions they provide.

Education

While the end goal of data curation is long-term preservation, the actual work of curation involves improving the day-to-day management of data such that the data is maintained in a state that supports future curation activities. This approach requires researcher involvement throughout the research data lifecycle, which in turn requires education. Rapidly evolving tool and service landscapes exacerbate the need for education.

- S4. Continue the Bren pilot project (a collaboration with the Bren School to address preservation of and access to student group project data and other artifacts). Develop a unit on data management standard practices and requirements, to be integrated into the Bren student group project curriculum.
- S5. Invite external speakers to deliver curation-related training. Begin with inviting an OSF Ambassador to deliver training on the use of the Open Science Framework.

- S6. Develop educational offerings in the areas of data management, publication, and citation; investigate delivery methods and venues and potential target audiences. Compare and coordinate with Collaboratory, Center for Scientific Computing, and Office of Research offerings and programs.

Outreach

The Data Curation Program, being only a year old, is new. More fundamentally, the entire field of data curation is relatively new; few comparable programs are more than 5 years old. Thus, to the extent that researchers may be familiar with concepts of data management and the issues of data preservation, they are apt to be less familiar with the solutions available to them and the services the Library offers. Additionally, it cannot be assumed that researchers will flock to new Library data curation services; they will need encouragement and assistance.

- S7. Develop (and deliver, following the launch of the Data Collective) departmental outreach presentations advertising the Library's data curation services, and the Program's services generally.
- S8. Develop outreach materials (e.g., pamphlets) describing the services offered by the Library and by campus, and the benefits of their use.

Consultation

Consultation will remain a mainstay of the program, though with the development of on-campus data curation services, the nature of the consultations will likely change to focus more on assisting faculty in starting to use those campus services. This in turn will likely lead to deeper (and more resource intensive) interactions between curators and researchers.

- S9. Build cooperative partnerships with subject librarians, and engage them in assisting with faculty consultations.

Staffing

The Program requires manpower to operate data curation services, perform consultations, deliver presentations and instruction, and perform outreach. Additionally, the Program requires curation expertise to be able to present researchers with the most appropriate and up-to-date solutions. In addition to the geospatial expertise already in house, the Program requires expertise in: the social sciences (because statistical and social science data are cross-cutting, and engender unique curation challenges such as privacy and de-identification); the life sciences (which has been at the vanguard of informatics and ontology research for many years now); and the humanities (where researchers are newly embracing computational methods).

- S10. Continue efforts to recruit curation expertise, whether in the form of dedicated curator positions or split curator/librarian positions.
- S11. Advocate for the hiring of subject librarians with significant education and outreach experience, to join the Data Curation Program and serve as curation liaisons to their subject department(s); and identify existing librarians who may be well-positioned to develop into this area.

Coordination

The Program occupies a unique place in the campus organizational structure. Departments and ORUs work closely with their researchers, and hence understand their researchers' needs well, but they typically do not address curation of research data. ETS provides some research-related, campus-wide computational support, but as with ORUs, not in the area of data curation. The Library has always supported curation activities, but historically its efforts (e.g., ADRL) have been restricted to supporting its own collections. The Data Curation Program can be viewed as operating at the nexus of these activities: leveraging campus-wide, ORU, and Library computational support to bring data curation services to researchers.

- S12. Take a leadership role in coordinating a unified, campus-wide strategy for the management and curation of research data.