Program 2018

Note: The program is moving to Sched, https://ucdlfx2018.sched.com all future updates will happen here

• Tuesday Morning Pre Conference Meetings (Feb 27)
  • Data Curation Unconference
  • Archiving AV: Analog to Born Digital
  • Git and Github for Metadata
• Tuesday Morning Pre Conference Tours (Feb 27)
  • UCR Library Creat'R Lab
  • UCR Library Special Collections & University Archives 4 to Explore
• Thursday Post Conference Meetings (March 1)
  • Beyond copyright: the where, what, why, and when of the public domain
  • Intro to Library Carpentry
• Birds of a Feather Lunches (Wed Feb 28)
  • Digital Collection Assessment Birds of a Feather
  • UC Linked Data Project Birds of a Feather
  • Web Archiving Birds of a Feather
• Full Sessions (Tuesday and Wednesday, Feb 27 and 28)
  1. Participatory digital collections: engaging opportunities for enhancement, experimentation and transformation
  2. DAMS Panel
  3. Demystifying Data Curation
  4. Digital Scholarship
  5. Developing Technologies Together: The Secret Sauce for Collaboration
  6. Identifiers
• Combined Programs (Tuesday and Wednesday, Feb 27 and 28)
  7. 3D / Makerspace
    • Publishing Molars and Mandibles: Challenges and Opportunities Associated with Emerging Technologies
    • The Creat'R Lab: Starting and Sustaining a Makerspace in the Library
  8. Born Digital
    • Appraisal and Description of Social Media Collections
    • Cobweb: Collaborative Digital Collection Development for Web Archives
  9. Collaboration
    • Clear Collaboration
    • "Out of Many, One": Creating a UC-Wide Descriptive Standard for Born-Digital Archival Material
  10. Copyright
    • Copyright Complexities: Promoting copyright education on the digitized campus
    • Copyright in the digital library: where we stand in 2018
  11. Digitization approach / theory
    • Digital Conversion in the Modern Research Ecosystem
    • Building the Discography of American Historical Recordings into a 78rpm Mass Digitization Project
    • Fed Doc Big Data Big Items
  12. Program Introductions
    • Library Carpentry: data and software skills training for librarians
    • Launching the digital lifecycle program @ UCB
  13. Project updates
    • ResourceSync at UCLA: PRL Use Case
    • IIIF Tiling on Demand with AWS Lambda
    • Building & Managing a Large Collaborative Open GIS Project
  14. Storage and Preservation Networks
    • UC Data Network: A Systemwide Solution for Free Research Data Management
    • Beyond the Repository: Exploring Integration Between Local and Distributed Digital Preservation Systems
• Other
  15. Lightning Talks

Note: The program is moving to Sched, https://ucdlfx2018.sched.com all future updates will happen here

Tuesday Morning Pre Conference Meetings (Feb 27)

Data Curation Unconference

This half-day unconference will address data management problems facing libraries across California, providing a forum for participants to bring up difficulties and unsolved issues in their work and receive suggestions based on experiences from different campuses. Challenges discussed can range from infrastructure to shifting cultural practices.

To make the most of our three hours, we will solicit problem statements before the meeting. In the spirit of a true unconference, topics can also be added at the very start of the meeting. The top up-voted problems will be discussed during the first session. After group discussions, the outcomes will be reported back to the whole group. Since the outcome of these discussions may influence the preferences for other topics, the voting process will be repeated prior to the second session. Based on previous experience at the California Data Librarians unconference in 2017, we consider this problem-solving session of great value to librarians and analysts involved with data management.
Archiving AV: Analog to Born Digital

Audiovisual media is ever increasing and contributing to backlogs in libraries and archives where analog materials experience degradation, deposited hard drives sit on shelves, and digital files languish on servers. Many librarians and archivists consider audiovisual collections to be a challenge to preservation due to rapid format obsolescence, barriers to digitization, and popular software-as-a-service archiving models that limit practitioners’ knowledge and participation in the digital archiving process.

In this half-day workshop fundamental questions will be addressed including:

- What constitutes an archive of audiovisual media and how does one address digitization?
- How can an archive curate and steward a born digital AV collection for preservation and access?
- How can archiving AV be practical, efficient, and cost-effective?

This workshop curriculum will include characteristics of analog and born digital AV, methods of digitization and digital curation, efficient and practical digital tools, and preservation strategies for retention and access.

Git and Github for Metadata

This is a two-part half-day workshop. The first part will cover a modified version of the “Version Control with Git” Software Carpentry lesson, tailored for a non-developer audience, with more focus on metadata (typically taught as a half-day, 3 hour workshop). The second part will focus on the use of Git and GitHub in the context of metadata workflow.

We will present examples and strategies, taken from recent work by UC Santa Barbara and UC San Diego, of version control, pull requests, and automated hooks and integrations as they relate to moving metadata through a workflow and into our repositories.

Tuesday Morning Pre Conference Tours (Feb 27)

UCR Library Creat’R Lab

UCR Library Special Collections & University Archives 4 to Explore

Thursday Post Conference Meetings (March 1)

Beyond copyright: the where, what, why, and when of the public domain

(half day)

The public domain is made of works that never received copyright protection, works for which copyright protection has lapsed, and items that fall outside copyright’s scope altogether. Given how copyright protection can limit digital library activity, understanding where copyright does not apply, or where it no longer applies, can be profoundly enabling.

This half-day workshop covers:

- The basics of copyright scope. Where does non-copyrightable subject matter find itself in library collections? How does copyright apply to data and metadata?
- The basics of public domain determinations. When and how does copyright lapse? How can we make copyright determinations and how confident can we be about any given public domain determinations we make?
- Public domain determinations at scale. What can be done? What is the Copyright Review Management System? What are its advantages and weaknesses?
- Operationalizing public domain determinations. How can we take what we know about the public domain status of works in digital collections and put it use?

Intro to Library Carpentry

(full day)

Intro to Library Carpentry

Note: registration is capped at 30 people. If we get less than 10 registrants, we will not hold a workshop and reschedule a Carpentry in Southern California at another date.

This full-day workshop will cover “Introduction to Data” and “Unix Shell” lessons of Library Carpentry. You don’t need previous knowledge of the tools presented. The lessons bust technical jargon, demonstrate the power of the command line interface and give attendees the power to work with directories and files, find and manipulate data, and do complex search/replace. Library Carpentry is made by librarians, for librarians to help you: automate boring repetitive tasks; create, maintain, and analyze sustainable and reusable data; work effectively with IT and systems colleagues; and better understand the use of software in research. The classes introduce fundamentals of computing, provide a platform for further self-directed learning, and are relevant to all areas of librarianship. More information on Library Carpentry classes is available at https://librarycarpentry.github.io/

Presenters: Laura Smart (UCI), Danielle Kane (UCI), Noah Geraci (UCR), Kat Koziar (UCR), Scott Peterson (UCB), Reid Otsuji (UCSD)

Literate Programming Workshop
A fundamental challenge for open science is how best to create and share documents containing computational results. Traditional methods involve maintaining the code, generated tables and figures, and text as separate files and manually assembling them into a finished document. As projects grow in complexity, this approach can lead to procedures which are error prone and hard to replicate.

Fortunately, new tools are emerging to address this problem. The half-day workshop will introduce a solution that’s gaining popularity in the R community utilizing the freely available RStudio development environment and other open source components. In the workshop we’ll demonstrate how to create a “compilable” document containing all the text elements (including bibliography), as well as the code required to create embedded graphs and tables. We’ll demonstrate how the process facilitates making revisions when, for example, a reviewer has suggested a revision or when there has been a change in the underlying data. We’ll also demonstrate the convenience of integrating version control into the workflow using RStudio’s built-in support for git.

The following tools will be covered:

- RStudio
- Markdown
- Zotero
- BibTex
- GitHub

Presenters: Harrison Dekker (University of Rhode Island), Tim Dennis (UCLA), Juliane Schneider (Harvard Catalyst, Clinical and Translational Science Center)

Birds of a Feather Lunches (Wed Feb 28)

Lunch will be provided on the second day. Exact logistics tbd, maybe box lunches

Digital Collection Assessment Birds of a Feather

UC Linked Data Project Birds of a Feather

Web Archiving Birds of a Feather

Kathryn Stine, Manager, Digital Content Development and Strategy, California Digital Library

Members of the Cobweb team (including staff from both CDL and UCLA) will facilitate discussion and information sharing about web archiving at and across UC campuses. Topics may include workflows for crawl/capture and metadata, existing and potential collaborative digital collection development activities, and opportunities for developing and supporting tools for researchers and curators, collection development, and discovery and access of web archives.

Full Sessions (Tuesday and Wednesday, Feb 27 and 28)

1. Participatory digital collections: engaging opportunities for enhancement, experimentation and transformation

Amy Azzarito, Assistant Director of Online Strategy, UC Davis Library
Dawn Childress, Librarian, Digital Collections and Scholarship, UCLA Library
Carl Stahmer, Director of Digital Scholarship, UC Davis Library
Peter Broadwell, Academic Projects Developer, UCLA Digital Library
Paul Fogel, Mass Digitization Manager and Technical Lead, Co-Technical Lead for HathiTrust project, California Digital Library

Library digital collections have the potential to move beyond access and serve as engaging platforms for enhancement, experimentation, and transformation. This panel will highlight a few such collections/projects that illustrate ways in which we are engaging our communities and creating data from materials at our respective University of California libraries. The panelists will discuss the merits of projects that scale vs. high-touch examples (and why boutique isn’t necessarily a dirty word) and will explore how smaller-scale projects enable us to model digital collection interactions and how exploratory work with collections data provide opportunities to demonstrate use cases for enhancing, re-mixing, and re-using our digital collections. The panel will close with a discussion on issues related to digitization selection, and suggest (for broader discussion) selection criteria that can facilitate more participatory, data-rich (or potentially data-rich) digital collections.

2. DAMS Panel

Description tbd

3. Demystifying Data Curation

Vessela Ensberg, Associate Director Data Management, UC Davis
Ho Jung Yoo, UCSD
Emily Lin, UC Merced
Amy Naseer, UC Berkeley
Curation, particularly of data, can feel like an amorphous process. The lack of clear definition may prevent library staff from being willing to engage with it. Since data management is a new frontier in digital libraries, and data curation is an important component of it, it is important to unpack the process to enable more library staff to engage with it.

In this presentation we define data curation in a set of specific steps and actions. We give concrete examples of how we have applied them and discuss the difficulty level of each step as well as the skillset required to undertake it. We discuss how these steps contribute to our goals of making data discoverable, citable, comprehensible, and reusable.

4. Digital Scholarship

Zoe Borovsky, Ph.D.
Librarian for Digital Research and Scholarship
Lead for Research Partnerships
UCLA Library

Stacy Reardon
Literatures and Digital Humanities Librarian
The University Library
University of California, Berkeley

Quinn Dombrowski
Service Manager for Research Computing Consulting
Research IT
University of California, Berkeley

Mary W. Elings
Assistant Director and Head of Technical Services
The Bancroft Library
University of California, Berkeley

Carl Stahmer, Director of Data and Digital Scholarship, UCD
Zoe Borovsky, Librarian for Digital Research and Scholarship, UCLA
Erik Mitchell, Associate CIO, UCB
Laura Smart, Head of Digital Scholarship Services, UCI

5. Developing Technologies Together: The Secret Sauce for Collaboration

Kathryn Stine and Michael Thwaites; Manager, Digital Content Development and Strategy/Senior Developer; CDL
Emily Lin and Alisak Sanavongsay; Head of Digital Curation and Scholarship and Digital Assets Programmer; UC Merced
Brian Riley and Stephanie Simms; DMP Tool Tech Lead and Product Manager; CDL
Barbara Hu; Software Developer/Technical Committee Chair; CDL
Marisa Strong; Tech Lead/App Dev Manager; CDL
Mark Matney; Digital Library Software Developer; UCLA DL
Declan Fleming; Chief Technology Strategist; UC San Diego
Rachael Hu; UX Manager; CDL

Are you considering a collaborative project? In today’s digital library climate with increased interest in community-based development work, technology collaboration is an enticing, seemingly efficient and cooperative way to build digital systems that meet the needs of many. A sound, productive community and development structure is core to any successful technology collaboration. In this session, we will share a number of different co-development project scenarios including different communication strategies, project management approaches, and management tools. We hope to examine patterns for success and note lessons learned that can help you shape your future co-development experiences. Newcomers and seasoned collaborators are all welcome. Join us for lively and informative retrospectives of co-development stories.

6. Identifiers

Greg Janée, CDL/UCSB; John Kunze, CDL; Marisa Strong, CDL; Judy Dobry, CDL
Daniella Lowenberg, Research Data Specialist & Product Manager (CDL)
Perry Willett, Digital Preservation Service Manager (CDL)
Joan Starr, EZID Service Manager (CDL)
Lisa Schiff, Technical Lead (CDL)

Combined Programs (Tuesday and Wednesday, Feb 27 and 28)

7. 3d / Makerspace
Publishing Molars and Mandibles: Challenges and Opportunities Associated with Emerging Technologies

(25 mins)

Doug Daniels, Lux Lab Coordinator, UCLA Library
Deidre Whitmore, Digital Archaeology Lab and Data Publication Manager, Cotsen Institute of Archaeology at UCLA

Emerging technologies such as three-dimensional printing and scanning are becoming increasingly available in higher education, including in many libraries. By utilizing these technologies, librarians and researchers can make fragile and often impossible to access artifacts readily available to anyone with an internet connection. We will be using a small collection of Neolithic bones from Masis Blur, Armenia (ca. 6200 – 5400 cal.BC), as a case study to examine the ins and outs of 3D scanning as a means of preserving and publishing work. To start, a brief overview of the scanning technology will be given, with a few pointers about some of the limitations of the technology. Select samples from the collection will be showcased, both the 3D scan and a 3D printed copy of the artifact. After the scanning hardware and workflows have been discussed, we will discuss how this partnership has promoted the preservation and sharing of research data. We will also discuss designing best practices and a workflow for scholars to deposit 3D models and descriptive metadata into UCLA’s institutional repository. We will end with an examination of the different academic applications and some of the challenges associated with publishing digital 3D content online.

The Creat’R Lab: Starting and Sustaining a Makerspace in the Library

(25 mins)

Brianna Marshall, Director of Research Services, UC Riverside
Michele Potter, Open Research Librarian, UCR

This presentation will highlight the case study of the Creat’R Lab, a makerspace formed in partnership between the UCR Library and the Office of Research and Economic Development. Presenters will describe decisions made in the first year of the makerspace, with a focus on three key areas: policies and staffing, campus outreach, and 3D printing and makerspace technologies. Attendees will hear our lessons learned and leave the session with an understanding of how to start a makerspace at their institution. Additionally, interested UC DLF attendees will be able to take a tour of the Creat’R Lab during their time at the conference.

8. Born Digital

Appraisal and Description of Social Media Collections

(25 min)

Bergis Jules, University Archivist, UCR
Eric Milenkiewicz, Digital Initiatives Program Manager, UCR

Social media has become a primary space archivists are beginning to explore as an avenue to building rich collections of primary source materials. It is a complex space where topics around politics, culture, entertainment, and social issues are frequently discussed. In addition to the large volume of data present in social media, several different types of content formats can be found there as well, including images, photographs, videos, and websites. This makes it difficult for archivists to make sense of the content. While the collection of social media data has received a lot of attention around software development and large scale collecting issues, less attention has been paid to the meaningful collection of that content, i.e. appraisal, with the goal of building research collections. This session will focus on two projects aimed at appraisal and description of social media content in special collections. One is around the work of the Documenting the Now project which aims to build tools for archivists to help them appraise social media content and the other is a pilot project in the UCR library to represent a social media collection through a finding aid and within Calisphere.

Cobweb: Collaborative Digital Collection Development for Web Archives

(25 min)

Kathryn Stine, Manager, Digital Content Development and Strategy, California Digital Library
Stephen Abrams, CDL
Andrew Wallace, UCLA

The demands of archiving the web in comprehensive breadth or thematic depth easily exceed the capacity of any single institution. As such, collaborative approaches to web archiving are necessary, and their success relies on curators understanding both what has already been archived, by whom, and how. With funding from the Institute of Museum and Library Services, Cobweb, a joint project of the California Digital Library, UCLA, and Harvard University, supports three key functions of collaborative collection development: nominating, claiming, and holdings. Curators establish thematic collecting projects in Cobweb and encourage nominators to suggest relevant web sites as candidates for archiving. For any given project, archival programs can claim their intention to capture a subset of nominated sites. Cobweb interacts with external data sources to populate a holdings registry, aggregating metadata about existing collections and crawled sites to support curators in planning future collecting activity and researchers in exploring archived web resources useful to their research.

We’ll share recent project activities, including a walkthrough of the most current Cobweb prototype.

9. Collaboration

Clear Collaboration

(25 min)
Collaborations in digitization projects have become commonplace. Institutions work together on thematic projects, often in pursuit of grant funding, in efforts to reveal hidden collections and unite related collections. Projects involving multiple institutions and vendors highlight the need for effective communication and tracking methods for achieving accurate results, and the ability to make changes without losing data. We will discuss several digitization projects - The State Medical Society Journal project, The Eric Berne papers, The Bay Area Response to the AIDS Crisis project, and UC Merced’s work with Yosemite National Park and the Merced County Historical Society. We will discuss how we established shared expectations around key issues, including intellectual property rights management, as well as project management and communications throughout the project lifecycle for these collaborations.

"Out of Many, One": Creating a UC-Wide Descriptive Standard for Born-Digital Archival Material

Kate Tasker, Digital Archivist, The Bancroft Library, UC Berkeley
Annalise Berdini, Digital Archivist, UC San Diego
Charles Macquarie, Digital Archivist, UC San Francisco
Shira Peltzman, Digital Archivist, UCLA
(representing the Born-Digital UC Common Knowledge Group)

Currently there’s no descriptive standard that adequately addresses born-digital archival material, and institutional practices for creating finding aids to these collections vary substantially.

A major challenge is that DACS, the official content standard of the U.S. archival community, does not provide specific guidelines for born-digital materials. This is compounded by the relationship DACS has to other standards, such as MARC and EAD, which do not always align in the context of born-digital. The lack of clear guidance results in finding aids that do not fully describe the quality, quantity, and usability of digital material. This diminishes accessibility and ensures that each organization must 're-invent the wheel' when describing this content. To improve the clarity and usefulness of finding aids, and to promote consistency across campuses, digital archivists from the UC system have developed a descriptive standard for born-digital archival material. Presenters will provide an overview of their work, including the project’s impetus and methodology, discuss how it has contributed to the sustainability of the team’s respective digital archive programs, and share lessons learned. Particular emphasis will be placed on the collaborative nature of this work, and on opportunities to put the standard into use beyond the UC system.

10. Copyright

Copyright Complexities: Promoting copyright education on the digitized campus

Peggy Tahir, Education & Copyright Librarian, UCSF Library

Background:
The need for copyright expertise in academic libraries gained importance with the increasing reliance on digital resources and the use of electronic course management systems and e-learning platforms. The advent of MOOCs (massive open online courses) and other types of online courses necessitated that librarians be able to answer faculty questions and educate them on copyright law as it applies to teaching and learning in the digital environment.

Methods:
This presentation will explore and discuss the promotion of copyright education on campus. It will discuss outreach, the importance of fair use, and how to encourage and promote fair uses of material. Images can be particularly tricky; there will be discussion of open images, images and resources in the public domain, and works with creative commons licenses.

Results:
Those attending the session will learn:
copyright basics, focusing on fair use and digitized resources
how creators can license their work for use by others
useful open resources
ideas for promoting copyright education

Conclusions:
Developing copyright law expertise expands the librarians’ role when working with faculty, instructional designers, and course management/collaborative learning environment teams. Copyright education promotes best practices among faculty using copyright protected materials for their courses.

Copyright in the digital library: where we stand in 2018

Michael Wolfe, J.D., Scholarly Communications Officer, UCD

Digital library projects are wound up inextricably with the state of an ever-changing copyright law that governs the usage of everything from books to software. What can be digitized? In what conditions? How can collections be used, to whom can they be made available, and subject to what restrictions? Court decisions, especially those about fair use, can change the landscape without much notice, and so can the much slower movements of Congress and the Copyright Office. This talk will review the current state of affairs for copyright and the digital library; highlight the initiatives, cases, bills, and administrative actions to watch; and venture one or two humble predictions about where things are headed.
11. Digitization approach / theory

**Digital Conversion in the Modern Research Ecosystem**

*(12 min)*

Stefan Elnabli; Media Curation Librarian \ Supervisor, Digital Reformatting Operations; UC San Diego

Conversion to or within the digital domain is integral to the preservation, access, and stewardship of digital collections, but how does one measure its value in the modern research ecosystem beyond its ability to provide a means for access? Decisions made in any conversion phase have great influence on outcomes of preservation, access, and digital scholarship. As such, conversion in the context of digital preservation models and theories about knowledge production plays an important role in digital curation and the wisdom produced out of humanities research. This talk will explore how to think about librarians in the context of digital humanities and illustrate how digital conversion and scholarship have a tendency to influence each other’s goals, aims, abilities, and outcomes toward sustainable and innovative stewardship of digital collections.

**Building the Discography of American Historical Recordings into a 78rpm Mass Digitization Project**

*(25 min)*

David Seubert, Curator, Performing Arts Collection, UCSB Library

TBD

The UC Santa Barbara Library, with support from the Packard Humanities Institute, is building out its audio digitization program to enable the library to digitize historical 78rpm sound recordings at scale. The project integrates the Discography of American Historical Recordings database with a new mass digitization workflow, to increase efficiency, dramatically lower costs, and reduce errors. The library is currently digitizing 200 titles per week and ramping up to digitize 500+ titles per week. This talk explores our unique philosophy of digital library creation, issues we’ve faced in building the project out to scale, and issues with providing online access.

**Fed Doc Big Data Big Items**

*(12 min)*

Lynne

12. Program Introductions

**Library Carpentry: data and software skills training for librarians**

*(25 min)*

John Chodacki
Tim Dennis, Director, Social Sciences Data Archive, UCLA Library

Library Carpentry is an open education volunteer network and lesson organization dedicated to teaching librarians data and software skills. The goal is to help librarians better engage with constituents and improve how they do their work. This talk will serve as an introduction on how Library Carpentry formed in 2015, evolved as a global community of library professionals and will continue as a future sibling of the Carpentries, an umbrella organization of distinct lesson organizations, such as Data and Software Carpentry. We’ll cover existing collaborative lesson development, curricula coverage, workshop activities and the global instructor community. We’ll then talk about the future coordinating activities led by the UC system to align and prepare for a merging with Data and Software Carpentry.

**Launching the digital lifecycle program @ UCB**

*(25 min)*

Erik Mitchell, Associate University Librarian / ACIO, University of California, Berkeley
Lynne Grigsby, Head of Library IT, University of California, Berkeley

In FY2018 UC Berkeley is launching an expanded program to convert, preserve and publish our physical collections as digital collections. Called the Digital Lifecycle Program, the effort aims to digitize and publish the major collections of UC Berkeley over the next decade. This session will recap our planning process and present the operational framework that is being created.

13. Project updates

**ResourceSync at UCLA: PRL Use Case**

*(20 min)*

Rosalie Lack, Project Manager, UCLA
Kristian Allen, Digital Library Architect, UCLA
Mark Matney, Digital Library Software Developer, UCLA
Pacific Rim Library provides access to aggregated metadata records from the Pacific Rim Research Library Alliance (PRRLA), an organization of academic libraries surrounding the Pacific. With grant funds from PRRLA, the UCLA Library has completed the first phase of the ‘next generation’ PRL with a redesigned interface and an upgrade to ResourceSync. We have developed and contributed to open source software that enables institutions to deploy ResourceSync capabilities that plug into their existing OAI-PMH repository. The presentation will provide an overview of the project and a forum for discussing opportunities for ResourceSync development and implementation at UC.

### IIIF Tiling on Demand with AWS Lambda

(12 min)

**Kevin S. Clarke, Digital Library Software Developer, UCLA**

Some digital library image projects require a full-fledged IIIF image server that can generate images of varying dimensions on demand. Others just need to provide an interactive image zooming functionality. For the latter types of projects, it's sufficient to pre-generate image tiles to be served from a IIIF “Level 0” tile server (which might be a IIIF-specific tile server or just a generic Web server like Apache or Nginx). This approach saves disk space, simplifies the system architecture, and may not require an additional caching layer to make the serving of tiles performant. At UCLA, we’ve written a IIIF tiling server that pre-generates the tiles for IIIF viewers like OpenSeadragon, Mirador, and the UniversalViewer. We started using it by running the tiling process on AWS EC2 instances; that gave us processing power on demand without the need to provision and maintain a local server. That approach worked well, but in time we realized we could simplify the process even further by extracting the tiling functionality onto AWS Lambda. We propose to talk about our experiences doing this and why we think Lambda is a good fit for IIIF tile generation.

### Building & Managing a Large Collaborative Open GIS Project

(20 min)

**Michele Tobias, GIS Data Curator, UC Davis Library**

**Tom Brittnacher, Geospatial Data Curator, UCSB Library**

Libraries are increasingly involved in providing data management and GIS services, consultations, project collaboration, and creating open digital data collections. In this talk, Michele and Tom discuss the American Viticulture Areas (AVA) Project and address how the tools and workflows used to create and manage this dataset can be repurposed for other data collections with multiple contributors with limited infrastructure and funding. The AVA Project began as a collaboration between UC Davis and UCSB to create digital spatial data for all 239 boundaries of the American Viticulture Areas as described in the US ATPF Code of Regulations in support of key research conducted at our institutions. The AVA Project utilizes user-contribution practices in the GIS community that have successfully generated open-licensed global datasets like Open Street Map combined with US government documents, open data formats (geoJSON), open source GIS software (QGIS), and collaboration tools (GitHub) to provide a high-quality open and accessible research resource. The project provides the dataset free of charge through a system that accepts contributions, corrections, and updates. The methods employed in the AVA Project provide a framework for managing other collaborative data creation projects, including contributions from multiple campuses as well as entities outside the UC.

### 14. Storage and Preservation Networks

#### UC Data Network: A Systemwide Solution for Free Research Data Management

(25 min)

**Stephen Abrams; Associate Director, UC Curation Center; CDL**

The UC Data Network is a partnership between CDL and campus libraries, CIOs and VCRs to build a sustainable and reliable network for sharing research data. In the UCDN, campuses contribute preservation storage to a common pool for use by CDL’s Dash data publication service and Merritt preservation repository for managing, preserving, and providing access to a growing collection of UC research data. UC researchers face new obligations for proactive and sustainable research data management (RDM) in the form of funder mandates, publication requirements, institutional policies, and evolving norms of academic best practice. While some researchers have access to appropriate disciplinary repositories, these opportunities are not uniformly available across the UC community. By relying upon in-kind contribution of storage capacity rather than direct campus recharge, UCDN removes the primary barrier to wider campus adoption of Dash and Merritt, which can now be offered for use at no cost to individual researchers. This presentation will provide an update on UCDN activities and solicit suggestions for potential improvements or added-value research data management services.

#### Beyond the Repository: Exploring Integration Between Local and Distributed Digital Preservation Systems

(25 min)

**Sibyl Schaefer / Digital Preservation Analyst and Chronopolis Program Manager / UCSD**

Many institutions have established digital repository systems in order to preserve the valuable scholarship and cultural heritage that is either generated or collected by their constituencies. In addition, many of these same organizations have distributed copies of these digital materials to multiple locations in order to mitigate the risks associated with lack of geographic diversity, lack of technological diversity, and loss of data related to human activities and systems failures. Services like APTrust, DPN, and Chronopolis have developed, in part, to provide this geographical diversity. As these services have matured, the problem of tracking data from a local repository to a distributed preservation service has not been resolved. Northwestern and University of California, San Diego were awarded an IMLS Planning Grant (LG-72-16-0135-16) to explore the integration between local repositories and distributed digital preservation systems. This grant seeks to answer questions like, “How does one curate objects to ingest into a long-term dark preservation system?” as well as questions regarding managing multiple copies and versions of digital objects in multiple systems and the implications of varying storage structures on data restoration. To uncover answers, the grant team distributed a survey and also conducted a series of in-depth interviews with cultural heritage institutions. This talk will describe the findings from both, highlighting results regarding criteria used in curation decisions, versioning practices, common workflows and workarounds, and the use (or not) of preservation policies.
15. Lightning Talks