

**OAC and Calisphere Assessment 2011-2012:
Executive Summary and Report**

Conducted and Prepared by Sherri Berger
California Digital Library
June 4, 2012



Executive Summary

Background

From October 2011 to May 2012, CDL undertook a series of assessment projects for the Online Archive of California (OAC, <http://www.oac.cdlib.org>) and Calisphere (<http://www.calisphere.universityofcalifornia.edu>). While these services are highly successful and well utilized, we felt some measure of assessment was necessary in order to plan for their continued vitality in a landscape of shifting resources and new technologies. Our key guiding questions were:

- What is the relationship between the OAC and Calisphere platforms?
- Why haven't UC institutions more consistently contributed digital objects?
- What is the best role for Calisphere moving forward, especially within the context of the UC Libraries Digital Collection?

Methodology

The assessment comprised three discrete but related projects:

- "Pop-up" user survey: one multiple-choice question placed on OAC and Calisphere for one week, asking users "who are you?"
- Usage statistics analysis: collection and analysis of four months of web analytics data, cross-analyzed with user survey data where relevant.
- Contributor interviews: in-person and phone interviews with contributors from 10 UC campuses and two non-UC institutions, focused on access needs for digital objects.

Findings

What follows is a brief list of high-level findings from the assessment; a report with more detailed findings and illustrative data is available on request.

- Calisphere beyond K-12 teachers

Although the traditional target audience for Calisphere is K-12 teachers, in reality this user group comprised only 10% of visitors during the week of the survey. The largest identified user group was college and graduate students (35%), who additionally demonstrated significant usage of both the unique digital objects *and* the K-12 curated collections.
- Need for digitization resources

Funding for ongoing digitization was one of the most cited barriers to contributing digital objects to OAC and Calisphere. Grant funding is no longer available for project-based digitization, and few if any contributors have a stable, dedicated funding stream. Most digitization is scan-on-demand. Meanwhile, institutions also lack the tools and workflows to handle increasing quantities of born-digital resources.

- Need for object management

The digital object ingest mechanism for OAC and Calisphere (METS repository, no management system) is out of step with contributor workflows. There is a need for easier metadata changes/object management, as well as accommodation of a broader range of formats.

- Aggregation without hosting

Some contributors have a robust management and display system at the local/campus level, yet feel strongly that they would benefit from the added visibility and aggregation of having their objects appear in the OAC/Calisphere access system(s). In these cases it may be preferable to surface locally-hosted metadata and/or content in the interface, rather than requiring them to also be hosted in OAC/Calisphere.

- Connect content to more content

While our navigational and contextual pages on both the OAC and Calisphere are popular and well utilized, many users arrive first at the unique content in the repository (finding aids and digital objects). We could do more to ensure that “content leads to more content” hosted on OAC/Calisphere and potentially other platforms, for example by implementing subject access, linking to locally hosted content, pointing to curated collections, etc.

- “Pride of place” desires

Contributors would like to “show off” the digital objects they contribute in a more robust way than is currently supported by the platforms. On the OAC, digital objects are presented within the context of finding aids; on Calisphere, there is no real mechanism for directing users to all of the content contributed by the institution. Additionally, contributors do not always have an incentive to contribute materials to Calisphere that are not about California and/or appropriate for K-12 audiences, as these are the kinds of materials that are surfaced in the curated collections.

- User tagging, crowdsourcing

Contributors are very interested in user tagging of digital objects, especially for the purposes of improving metadata and applying subject access.

- Next-gen curated collections

Evidence suggests that there is something of value to the curated collections on Calisphere beyond the K-12 teacher audience. Contributors are also interested in the potential of these for the purpose of undergraduate teaching and/or online companions to in-person exhibits. More research is needed to determine how to continue to build relevant contextual layers for new audiences.

Immediate opportunities

There are several areas in which we can immediately start to address the needs uncovered, namely:

- Conduct an assessment of key user groups identified in the survey to determine their expectations for access, context, and functionality around digital objects.

- Create more paths to more content at the item level.
- Revisit the design of the Calisphere homepage to make clear the breadth of content available.
- Investigate technical solutions for surfacing metadata and/or content within the interface without requiring it to be hosted in the OAC/Calisphere repository.
- Provide contributor “showcase” pages on Calisphere that highlight their digital objects.

Long-range opportunities

The following projects could be realistically implemented in the longer term, due to their complexity and/or contingencies on other projects and decisions:

- Refocus Calisphere design and development activities to respond to the needs of a broader range of users than previously identified.
- Implement a content management system for OAC/Calisphere.
- Streamline workflows and policies between Merritt and OAC/Calisphere.

Questions for further discussion

The assessment raised some additional questions that should be considered, as they may have a bearing on the continued directions of OAC and Calisphere, specifically as related to digital objects:

- How can the CDL assist the UC system with ongoing digitization projects? What funding sources may be available to support this work?
- Are there opportunities for the CDL to assist the campuses with processing and managing born-digital materials?
- How do we prioritize conflicting needs and requirements among OAC and Calisphere’s various users and constituents?
- What is the best model moving forward for hosting and providing access to digital objects within OAC and Calisphere?
- How can the UC Libraries Digital Collection initiative inform our thinking about these issues?

Report on Findings and Opportunities

This report presents an overview of the methodologies, findings and outcomes of three assessments conducted between October 2011 and May 2012:

- “Pop-up” user survey to identify website audiences (p. 5)
- Web analytics statistical analysis to understand patterns of usage (p. 7)
- OAC/Calisphere contributor interviews to understand access needs for digital objects (p. 8)
- Service development and enhancement opportunities, based on assessment findings (p. 12)

The assessments were undertaken at this time for a variety of reasons:

- Next steps for Calisphere: As Calisphere nears its sixth anniversary, it is time to consider how this resource can be updated and/or expanded to meet our users’ evolving needs. Most pertinently, can and should Calisphere function as a resource for audiences beyond K-12 teachers?
- Relationship with OAC: The increasingly efficient and integrated organizational and technical infrastructure at the CDL highlights the complex and interdependent relationship between Calisphere and the OAC. Is there a continued need to maintain two separate websites, or are there other, equally user-centric solutions for providing access to digital objects? How can we most successfully meet the need for archival context through finding aids, as well as the need for direct access to digitized materials across collections?
- Contributor needs: It has been some time since the CDL has spoken with contributors about their needs, specifically regarding digital object hosting and presentation. With a newly formulated group focusing on access systems—and with plans for a UC Digital Collection management and access solution in the works—this is the right time to revisit these questions and understand what the OAC and Calisphere can do to better support digital objects.
- Data-informed strategy: In this time of tight budgets and heightened accountability, there is an increasing demand for data to inform and justify decision-making. We believe this analysis provides a useful model for future service assessment efforts.

We also have produced two full-length reports on the survey/statistics analysis and the contributor assessment, respectively, with more extensive description of methodology and findings. These reports are available for consultation on request.

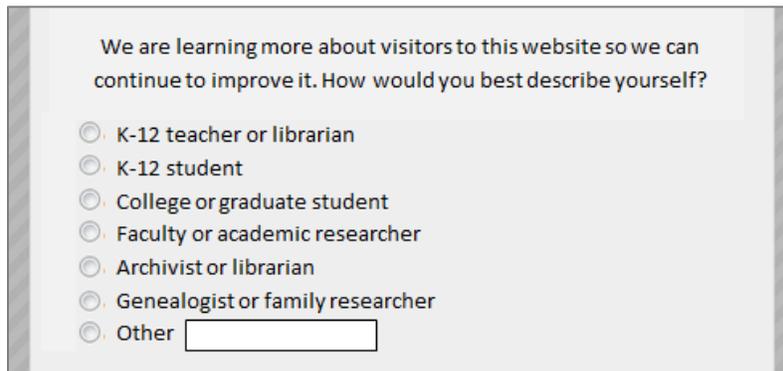
1. “Pop-Up” User Survey

Objectives

- Who uses our websites?
- Do our target audiences match our actual audiences?

Methodology

- We ran a one-question, multiple-choice survey on OAC and Calisphere. The survey “popped up” for every visitor, a few seconds after his/her entrance on any page of either website.
- The survey ran for a period of one week: November 30 – December 7, 2011.
- We added the OAC/Calisphere Google Analytics code to the survey, allowing us to post-facto match user groups with trends and behaviors.
- The survey question read as follows:



We are learning more about visitors to this website so we can continue to improve it. How would you best describe yourself?

- K-12 teacher or librarian
- K-12 student
- College or graduate student
- Faculty or academic researcher
- Archivist or librarian
- Genealogist or family researcher
- Other

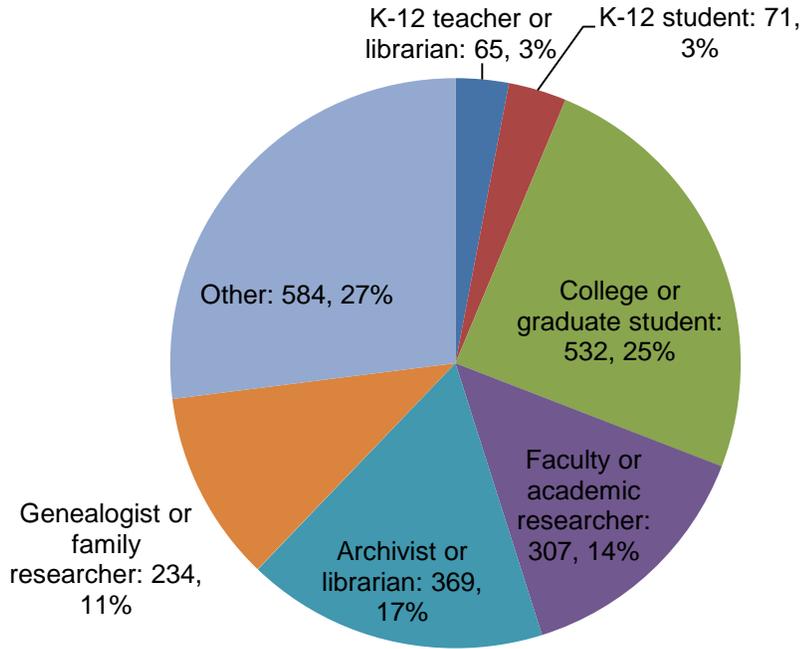
Caveats

- The survey only ran for a period of one week, during a historically robust period of usage. However we cannot necessarily assume that usage was typical during that period.
- We intend to repeat the survey within six months to verify the general applicability of our findings.

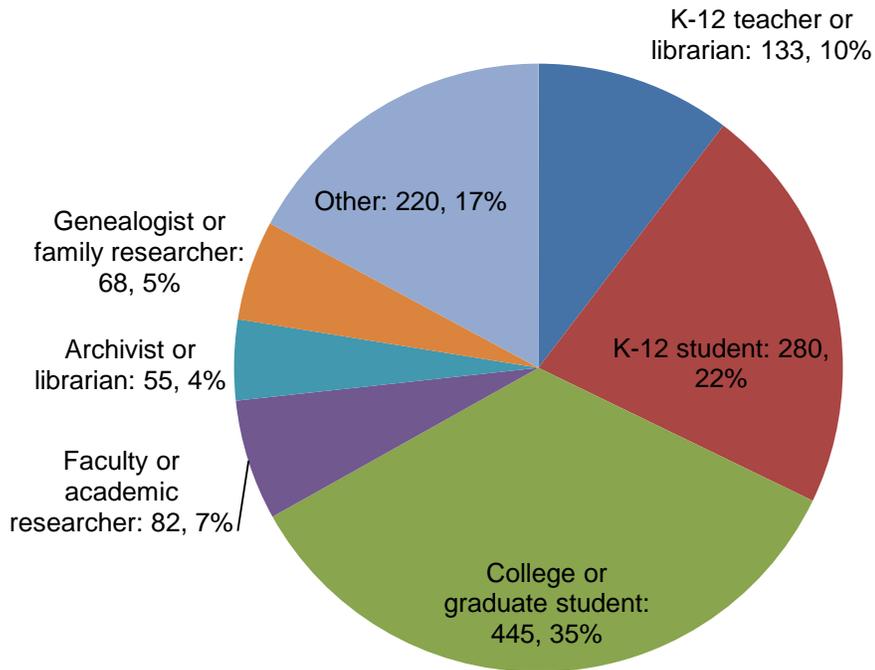
Key findings (visuals on following page)

- OAC users matched our expectations: college/graduate students, archivists/librarians, faculty, genealogists, and mostly professionally oriented “others.”
- Calisphere users were surprisingly diverse, and well beyond the target audience of California K-12 educators:
 - K-12 teachers comprised just 10% of Calisphere users, and only half of those from CA.
 - K-12 students comprised 22%, suggesting the K-12 use case is moving away from mediated experience and towards direct research.
 - College/grad students comprised 35%, and they displayed a broad interest in the site, including JARDA and themed collections
- College/graduate students actually comprised our largest single group of users across both sites.

OAC User Groups Identified by Survey



Calisphere User Groups Identified by Survey



2. Usage Statistics analysis

Objectives

- How engaged are our users?
- What are the most common paths to our websites?
- What content is used and with what frequency?

Methodology

- Extensive analysis of usage statistics for both OAC and Calisphere, as reported by Google Analytics.
- Four months of data: July 1 – October 31, 2011.

Key Findings

- How engaged are our users?
 - We have many first-time visitors, likely due to the large amount of content to which we provide free access. This is a good reminder that we are part of a larger research ecosystem.
 - There is a significant group of power users for both sites, and especially OAC (primarily but not exclusively archivists, as analyzed using pop-up survey segments).
- What are the common paths to our websites?
 - For both sites, just over half of traffic comes from a search engine (Google and others).
 - Wikipedia is the most frequent referrer to both sites.
 - OAC has a particularly high referral rate, due in part to the fact that contributors link to their landing pages and finding aids from their local websites.
 - Both sites, but particularly Calisphere, see a significant amount of direct traffic. Direct traffic is mostly likely the result of a visitor using a bookmark or typing the URL into a browser.
- What content is used and with what frequency?
 - **Content viewed follows a long tail.** The most popular pages are home pages, landing pages, and (for Calisphere) themed collections. However, the majority of usage of both sites is spread across the multitude of finding aids and digital objects in the repository. For Calisphere, this suggests that digital objects are being found and accessed regardless of whether or not they are included in themed collections and special sets.
 - 23% of OAC visits and **only 9% of Calisphere visits used the site search.** This makes sense for Calisphere given its robust browsing interface. However, only some of the content is optimized for browsing (that is, appears in the themed collections and canned search sets), meaning that users may be missing content of interest to them.
 - 75% of OAC visits and 78% of Calisphere visits **enter at repository content** (finding aid, digital object, search result set), rather than a home page, landing or navigational page.

Taken together, these findings make the case for us to **use content to connect to more content**. We cannot assume that users will navigate by way of the homepage, or utilize the site search.

3. Contributor Assessment: Access Needs for Digital Objects

Objectives

- What are the needs and motivations of contributors regarding access to their digital objects?
- What more can the OAC and Calisphere do to meet their access needs, and the needs of users?
- What are perceptions of OAC and Calisphere, jointly and separately?

Scope and Methodology

- We conducted interviews with twelve OAC/Calisphere contributors: one on each of the 10 UC campuses, and two representative non-UC contributors (a historical society and a university).
- We approached the head of special collections or other high-level staff contact for each contributor, but invited this person to bring other staff into the conversation.
- Interviews focused on *digital objects*. Finding aids were discussed in the context of their effectiveness at organizing, providing context for, and connecting users with digital objects. Ingest conversations focused on digital objects rather than EAD and other descriptive formats.

Key Findings

(The interspersed quotations were either taken verbatim from the interviews or were edited slightly to present a composite viewpoint shared by multiple contributors.)

- **Benefits to contributing:**
 - Contributors cited many reasons for contributing digital objects to OAC and Calisphere, or wanting to contribute more in the future
 - Online access, period: For contributors without a way to surface content locally, OAC/Calisphere is *the* access platform for their digital objects, or what they strive for.
 - Visibility: The OAC and Calisphere provide recognizable destinations for researchers, where local websites sometimes do not. Contributors with robust local systems saw OAC/Calisphere as extending the visibility of their digital objects.
 - “One-stop shopping”: Many contributors remarked on the ability for users to search across collections in the OAC and Calisphere; aggregation improves the level of service.
 - Like content: A few contributors mentioned the benefit of Calisphere, in particular, to provide access to “materials with similar content in a common context.”
 - Branding: One UC contributor cited the branding of the UC and the CDL as a benefit.
 - Financial payoff: One UC contributor mentioned that Calisphere generates significant revenue for the institution in the form of reproduction requests.

- **Perceptions of OAC and Calisphere**

- Contributors demonstrated high familiarity with the OAC and Calisphere, how they are organized, and their intended audiences.
- Most, if not all, interviewees said they believe Calisphere provides users with more direct access to digital objects than the OAC. Some contributors direct their undergraduate students there.
- Most interviewees understood that all digital objects on the OAC appear on Calisphere. In fact, some contributors did not know that the OAC contains the digital objects as well, because finding aids are so prominent there. To them, Calisphere is the presumed access point for digital objects.
- One contributor did not realize the extent of materials on OAC and Calisphere, thinking the collection was much smaller. Another contributor made the related comment that the homepages of both sites look rather static, and therefore do not show that the collection is steadily growing.

- **Barriers to contributing:**

- a) Digitization: lack of funding and/or streamlined, cost efficient processes.

- Several contributors—both large and small, at UC and beyond—noted that grant funding for digitization has essentially dried up.
- Some UC contributors were eager to see funding for digitization deemed a priority.

“The OAC is not the issue. Digitization itself is the challenge, the barrier to participation.”

“I would love to see the CDL push a digitization effort for the campuses.”

- b) Ingest: new workflows or better system needed.

- The turnaround for OAC/Calisphere is too slow for scan-on-demand materials.

“We’re using Omeka because we need a quick, lightweight implementation for content.”

- There is a widespread need to change metadata more easily, rather than re-send the METS.

“Getting everything packaged and put in the OAC is onerous.”

“We need ways to move quickly; to modify and take down content.”

- Contributors desire better coordination between Merritt and OAC/Calisphere, as well as between local systems when possible. Duplicate hosting is not generally seen as desirable because then there are multiple files to update and manage; several contributors would prefer a crawl or harvest approach to making content discoverable in one place.

“If we have to worry about multiple versions of content, we have less of an incentive to contribute it.”

- c) Formats: a need to develop additional workflows for more digital object formats.

- There is a perception that OAC and Calisphere simply “do not accept” formats beyond METS, TEI, and PDF texts. While we are glad to work with contributors to take whatever they have, it is true that these are the only formats with established, “routine” workflows (“routine” in quotes because even these require a large degree of customization and support).
 - PDF, A/V, and multi-format materials were specifically cited as high priorities.
 - d) Calisphere branding: lack of incentive to contribute content beyond K-12, California.
 - Contributors said they perceive a disconnect between our collection policy, which is inclusive and accepts objects representing any subject, and the Calisphere homepage and themed collections, which highlight particular topics for users.

“The Calisphere homepage misrepresents the wealth and diversity of the materials in the repository.”
 - They assume that if they were to contribute materials that are not related to California and/or are not appropriate for K-12 audiences, then these would not be surfaced on Calisphere. Therefore there is not always a clear payoff for contributing digital objects to the resource.

“It’s a question of fit when thinking about whether our materials belong in Calisphere.”
 - Contributors showed significant interest in broadening Calisphere’s audience when we discussed the “pop-up” survey results with them.
 - e) Rights: rights appear to be a lesser, but present, barrier to contributing objects. CDL could play a role in having a centralized rights service; this would help with digitization and access decisions.
 - f) Special local needs: sometimes certain users or formats simply demand special interfaces or functions that are not met by OAC/Calisphere (e.g., a highly interactive rare book).
- **Access points:**
 - Contributors want to be able to refer users to one—or at least fewer—access points for their digital content. However, they also want to make their materials visible in the places users are.

“Everything doesn’t have to actually live together, but there has to be a better cohesive access point so I can send everyone to one place for all our stuff.”
 - There is concern about maintenance and workload when comes to “redistribution” (allowing content to be harvested by or hosted on third-party applications), along the same lines as explained above: the more places content is hosted, the more complicated it becomes.

“Everyone likes the idea of multiple access points, but not redundant work.”
- **Desired end-user enhancements:**
 - Better context, paths:
 - Contributors want to provide more context for their digital objects, so users understand their relationships with the collection and the institution. This goal relates to the access points

discussion above, but also has to do with how users understand objects, not just find them. Contributors had several ideas for usability and interface adjustments on this point.

- They also want to provide more paths to other materials and resources that are within the same collection or institution, or on the same subject at another institution. Indeed, some contributors even voiced interest in linking to—or pulling in—non-special collections resources such as Wikipedia content, eScholarship articles, digitized secondary sources, etc.

“We need multiple ways of getting at things to accommodate different levels of expertise in accessing content.”

- Curated sets/exhibits:

- There was interest in building out curated sets or exhibits, especially for the purpose of undergraduate research or for highlighting content (e.g. companion to an in-person exhibit).

“People want you to tell them about the content. It’s no longer enough just to offer up the images.”

- User tagging and sharing:

- Contributors were very excited about the potential for user-generated metadata and subjects. This was raised several times in response to an open-ended question about access improvements.

“We are expecting the community to be able to help us with identification.”

“Why not have users do the work of applying subjects through tags?”

- One contributor expressed a strong interest in better user sharing capabilities, for example via Pinterest and other contemporary platforms.

- Innovative access capabilities:

- One contributor thought more could be done in the way of experimenting with new access technologies, for example making a mobile version of Calisphere, doing more with maps, and pursuing other user-side innovations within the platforms.

- Other: Some contributors mentioned other end-user enhancements such as visual browse, bookbag, print on demand, and monetizing content.

- **Born-digital materials**

- Several contributors mentioned the need for support for processing and managing born-digital materials. While this was somewhat off-topic from the access focus of the discussions, it is worth stating here so the CDL knows this is a challenge our contributor base is facing.

4. Service Enhancement Opportunities

Below is a list of service enhancement opportunities for OAC and Calisphere, gleaned from the assessment work described above. Reflecting the needs and behavior of our various contributor and user communities, this list provides a helpful guide for future resource allocation and technical development for the Calisphere/OAC platforms—with a focus on digital object hosting and access.

Creation / Ingest / Management of Digital Objects

- Consider ways that CDL could help facilitate digitization, especially at UC.
 - The NGTS work currently in progress may propose a UC systemwide approach or program of digitization, which may or may not call upon CDL to participate.
 - CDL could also investigate leveraging mass digitization workflows for special collections.
 - Although there are no longer as many grant opportunities for digitization, we may be more competitive for those that do exist if we apply for them collaboratively (such as how we applied for the successful CLIR Hidden Collections processing grant).
 - Ultimately, robust financial support and solutions for digitization will have to come from UC Libraries leadership.
- Move away from an ingest model and towards a content management system for OAC/Calisphere. Smaller institutions need a DAMS in order to contribute; every institution needs to be able to manage and more easily edit content.
- Work towards supporting new formats. Priority formats are A/V materials and PDF digital objects.
- Support a metadata harvest, crawl, or API that makes locally hosted content discoverable in OAC/Calisphere.
 - Contributors see the value in providing aggregated access, across collections and institutions, to digital objects and other related content.
 - However, institutions with locally hosted content will not contribute it if it is too much effort or demands high-touch management of redundant copies.
- Better integrate OAC/Calisphere with Merritt. Contributors do not want to have to submit twice, nor manage duplicate objects.
 - Streamline submission workflows for these two services.
 - Develop shared digital asset agreements and policies.

Access Interface and Functionality

- Return to the question of how the OAC and Calisphere should be related, both technically and as appears to the user.
 - More research could be done into how users perceive the two sites and movement between them. Is it disorienting or no big deal?

- Different models should be explored and examined, for example combining the sites and keeping the strengths of each, or making Calisphere the authoritative repository for all users and “borrowing” objects in the OAC for the specific research use case.
- Adjust the user interface of both OAC and Calisphere to clarify the relationship between digital objects and collections and improve navigation between them. This modification could include:
 - Simplifying digital object views to raise the visibility of the “collection” link.
 - Investigating options for paging through images via a slideshow-like functionality and/or embedding thumbnails in collection guides on the OAC. This may dovetail with research into alternate ways to model the display of collection guides.
- Create a place where contributors can direct their users to find all of their digital content, both to provide for contributor “pride of place” and enable users to easily find like materials at the institution.
 - This could involve expanding the list of institutions on Calisphere, or refactoring contributor landing pages on OAC to subvert the finding aid structure (or both).
 - This could be as simple as a list of digital objects, or as complex as offering subject headings, thumbnails, featured content, and links to related content on other websites.
 - Eventually it could become a customizable “profile page,” with space for contributors to curate exhibits and highlight new content.
- “Optimize” digital objects so that they can be used as access points to additional like content, either on our systems or beyond. Immediate projects might include:
 - Indicating when an object is part of a themed collection on Calisphere.
 - Giving more prominence to the “more like this” feature on Calisphere.
 - Eventually: linking out to similar objects hosted on local campuses, and/or related content on external websites.
- Work towards true subject access for digital objects on Calisphere and/or OAC (including site-wide facets and flexibility in narrowing and ordering results). Possible starting projects:
 - Allow for subject browsing of objects and/or collections guides within repositories (for contributors that have applied subject headings)
 - Experiment with user-supplied subject tags, either public or behind-the-scenes through controlled crowdsourcing.
 - Implement “browse by decade,” since this does not require significantly normalized data.
- Conduct an assessment to better understand the role of curated collections and exhibits in digital object access platforms, for expanded audiences.
 - Usage statistics and anecdotal evidence suggest that there is something of value in the current themed collection framework, even for audiences beyond the original use case.

- However, it is unclear what of the following functions of this framework are useful: the subject association, the limited selection of objects, the explanatory essay, the institutional authority garnered by the UC brand, and/or unknown characteristics.
- Understanding what is of value will help us more strategically apply resources and prioritize access enhancements, whether towards or away from curated collections.
- Continue to monitor the level of need among the user base for new innovations such as a bookbag, mobile application, content sharing, interactive displays, etc.

Calisphere audience and branding

- Make a policy decision to broaden Calisphere’s target audience beyond the California K-12 teacher community. This will entail:
 - Focusing on surfacing all content equally, through the mechanisms identified above; if we continue to create curated spaces, they should be on any subject, relevant to any user.
 - Shifting the mentality away from “key audiences” and towards “use cases” for the purposes of design, branding, and/or marketing.
- Determine a plan for “downshifting” the K-12 teacher aspects of the site.
 - The level of ongoing support for this audience, even if minimal, must be determined.
 - A short-term solution could be to relegate teaching materials to a “just for teachers” area.
 - A communications strategy should be developed to articulate the changes to the teacher community, especially the individuals CDL partnered with to create and market the site.
- Redesign the Calisphere homepage so that it better represents both the breadth of content on the site and the variety of use cases for it. This should include:
 - Showing the diversity in topics covered by content, especially beyond California history, and making clear that the collection is constantly growing
 - Showing the diversity of institutions that contribute to it. For example, we occasionally see Calisphere referred to on Twitter and blogs as “UC Berkeley’s digital collection.”
 - Using the homepage as a “splash page” to make clear what is available, and shifting the navigational emphasis away from the strict themed collection framework currently in place, towards a future vision of multiple paths to a variety of content and associations.

Potential new directions

- Consider CDL’s role with respect to born-digital materials.
 - Support is needed across the entire digital object lifecycle, not just the hosting and access solution we currently provide.
 - This may include support of hardware and software, as well as development of best practices and protocols.