ArcLight

illuminating discovery to delivery for archives & special collections

Mark Matienzo, Stanford University / @anarchivist

Coalition for Networked Information **#cni17f**

11 December 2017

Description and objectives

- Project initiated by Stanford Libraries in 2014 to address needs related to discovery/delivery of information in archives
- Support discovery of physical and digital collections
- Compatibility with and intended for integration with other systems, e.g. ArchivesSpace and Hydra/Samvera-based repositories
- Development, enhancement, and maintenance by the Blacklight and Hydra/Samvera communities
- Maintain a community focus throughout the project

Design and development process

- Design process led by 2 UX designers in at Stanford Libraries (Gary Geisler and Jennifer Vine)
- Followed a model for **user-centered design** developed and refined over time (see DLF 2014 presentation) and leveraged existing practices for community-based open source software development
- **Community-oriented, collaborative design/development process**
  - Intentional choice, informed by DLSS open source participation
  - Opportunity for other institutions to identify needs and participate
  - Build interest and identify potential commitments for software development

Overview and timeline of work to date

Discovery

Phase 1 → Phase 2

Environmental scan
Stakeholder goals
Interview planning

Archivist interviews
End-user interviews
Interview analysis

Information Architecture

Phase 1 → Phase 2

User needs
User personas
Requirements prioritization

Conceptual models
Conceptual sitemaps
Wireframes

MVP Development and Visual Design

Phase 1

Minimum viable product
Visual design mockups
Visual design style guide

Designing integrated delivery

Early origins of requirements

- Improving discovery for archives and delivery of digital objects
  - Broad interest from all stakeholders early on
  - Demand to remediate marginalization of archival discovery
- Stakeholder goals emphasized specific needs
  - Delivery of digital materials in context of description
  - Address “siloing” of different kinds of digital content
  - Access controls for both discovery and delivery, including for digital materials
  - Shared need for integration with Aeon or other existing registration/request management systems

User interview quotes

- **Embedding:** ArcLight, to the extent that it's feasible, we want to give the user access to the digital objects within ArcLight… Pretty straight forward for something like images, even for video…Even an embeddable viewer…at some point. **(Archivist)**

- **Conveying context:** The other issue, I do think when things are digitized it's easy for them to… get the sense that they're not in a continuum next to other things or in folders or together in a way. Sometimes, it's very crucial how things are or left or either reorganized or whatever it is. The things that are nearby. **(Researcher)**

- **Content siloing:** People click on a link in the finding aid and go to DSpace. It's not particularly user friendly and requires people to download materials and access them on their local machines. We aren't serving our researchers well. **(Archivist)**

- **Access controls and registration:** We don't allow direct download, because we have no control. There's no registration… […] Without that, people when they come into the Reading Room, they sign a form saying that they understand copyright, and we're indemnified. If they could do that online, then we'd be indemnified to some degree. **(Archivist)**

http://bit.ly/arclight-design-documents

Personas

Composite user sketches intended to demonstrate common needs

Clockwise from top left:

Archivist who serves as ArcLight administrator

Processing archivist

Advanced researcher / faculty user

ArcLight
Administrator Persona

Marcia Garza
Systems Archivist
Archival Collections department at an independent research institution

After working as a junior web developer during grad school, Marcia is enjoying her first year working as an archivist. The work is not without its frustrations, however, and she's excited enough about the potential of ArcLight to improve access to the archive's collections that she volunteered to be their first ArcLight administrator. She's more tech-savvy than most of the other archivists here, but she's not a programmer and she's a bit nervous about whether she'll be in over her head.

"I'm more than willing to spend whatever time it takes to configure ArcLight if it will tame the confusion and chaos that currently limits access to our fantastic collection of archival materials."

Frequency of ArcLight use: Daily

Technical proficiency: High. Marcia suffers from repetitive stress disorder and prefers to use only a keyboard.

	Motivation	Scenario	Specific Goals
❶	Easily customize ArcLight to be institution-specific	When she logs into the ArcLight administration page, Marcia is pleased to find it even more straightforward than the Drupal-based websites she's worked on. She quickly uses a few forms to change the basic appearance of the site to reflect the Institute's name, logo, and colors. It doesn't take much more effort to update the site's footer, to specify the recipients of the feedback form, and to customize the template of a virtual reading room consent form to reflect the Institute's policies. Despite the efforts of the Institute's public service archivists, Marcia has noticed that many researchers don't understand how to effectively use the Institute's archives or how to request items from the collections. To help address these issues, Marcia creates support pages that outline her Institute's reading room policies, explain how to make digitization requests, and describe the Institute's local conventions and descriptive practices.	• Customize labels and appearance options • Add support pages with institution policies, etc. • Customize an integrated feedback form • Customize a template for a virtual reading room consent form
❷	Selectively activate and configure desired features	Marcia had concerns about how well ArcLight would accommodate some of the idiosyncratic ways the Institute describes and makes archival materials discoverable. As she works through the Configuration section of the administrative interface, she finds that many features can be turned on or off, and some features can be customized. She can select to expose only some of the available social media features, for example, associate specific content viewers with different types of content, and make some detailed choices about how search results are displayed. She can also enable and customize some options related to full-text search of digital objects. She sees that some defaults can be overridden at the collection level so that an archivist can determine whether the settings are appropriate for a given collection.	• Selectively enable features of interest, such as specific social media features • Configure details of complex features, such as content viewers, full-text search, and search results display
❸	Configure viewers by content type	When she first set up ArcLight, Marcia accepted the default viewers for all digital content types. The support developer for her department has recently created and installed a new email viewer that is closely integrated with the Institute's authentication service. Marcia updates the configuration to use the new viewer for all email content. She gives the heads-up to an archivist that a web-based text-mining plugin created by an ArcLight community member has been installed. The archivist will want to update one of her recent collections to make the writer's Word documents available to this plugin. As a rule, though, archivists don't need to revisit collection descriptions to update viewers: once Marcia sets a default for a content type, it applies to all objects of that type automatically.	• Configure viewers for rendering digital material within (or one-click-away-from) ArcLight • Archivists don't need to revisit collections to make sure the viewers are up to date

ArcLight
Advanced Researcher Persona

Anna Chandler, Ph.D.
Associate Professor
Department of American Culture at a large public university

Dr. Chandler's research and teaching interests are centered in the Arab and Muslim American Studies Program. She uses physical and digital collections at the archives to introduce her students to primary source materials, and for her own publications and conference presentations. She is experienced with various public access catalogs, finding aid interfaces, and online collections. She frequently takes digital photographs of archival materials with her cell phone, and uses *Zotero* to manage citations.

"As I'm engaged in both teaching and research, I have to make the most of my time when I'm at the archives. I browse finding aids and access online content when I'm at home or my office while preparing for class or conducting research."

Frequency of ArcLight use: Weekly

Technical proficiency: Moderate

	Motivation	Scenario	Specific Goals
❶	Find primary source materials relevant to pedagogical and research interests	Dr. Chandler is beginning work on an article about the contemporary Muslim-American experience in Detroit. She begins by doing a keyword search across the entirety of the archives' holdings, then facets the results by geographic location and date range. The search results include collections, specific collection components, and digital objects. She selects a new collection, the Midwest Ecumenical Forum record, opens its finding aid, and starts to browse its intellectual hierarchy before conducting several keyword searches within the collection. She bookmarks a few items of possible interest. She returns to the initial results, and selects a relevant folder title from the Nabeel Hassan papers. When the component opens, she sees that it is part of the collection's "Islamic Center" subseries (which in turn belongs to a "Topical Files" series) and that there are no restrictions on access or use of the materials.	• Employ multiple search/discovery strategies across and within collections • Search full text where available • Receive search results that include collections, components, and digital content • Facet and refine search results • Browse/navigate the hierarchy of an archives
❷	Request items or digitization in advance of visits to the archives	Dr. Chandler has identified a potentially interesting folder in the Nabeel Hassan papers. She clicks the "Request This" button next to the folder title so that the material will be ready for her next visit to the reading room, and then continues to review the finding aid for relevant content. As she peruses the scope and content note for a series of "Articles, Manuscripts, and Poems," she realizes that a volume of Hassan's diary will be essential to her research; she therefore goes to the component in ArcLight and clicks the "Digitization Request" button. By the time she returns to her initial search results, she is alerted to two notifications in her account: the box she requested will be ready for her in two days (and will be held at the reading room for a week) and her digitization request has been queued for processing.	• See clear and unambiguous information regarding applicable use or access restrictions • Submit item requests and digitization orders • Receive notifications about requests • Save collections, components, digital objects to a bookmark list
❸	Gain direct access to digital content (as permitted by access/use restrictions)	Dr. Chandler facets her search results to show only digital content and discovers a broad range of relevant information in multiple content types. She notices that items associated with a particular email account in the Khalil Alawan papers are restricted to reading room-only access due to copyright issues and makes a note to view them on her next trip to the archives. Moving on, she finds a number of oral histories from Muslim community leaders in the Jawad Maznui papers that she can stream remotely in ArcLight via an embedded player. She then turns her attention to some born-digital office files in the Charles Hassan papers that are only accessible to members of the Midwest University community. After clicking on the item and authenticating using her registered username and password, she gains access to a virtual reading room. She navigates through the directory structure and is able to render common file formats in-browser via an embedded player.	• Access digital content: directly from finding aid/discovery interface; from various sources (Hydra, Archive-It, ePadd, etc.); in-browser as much as possible • Easily find information regarding access/use restrictions (including copyright) • Access the virtual reading room for materials with access/use restrictions

ArcLight
Archivist Persona

Coretta Rey
Arrangement and Description Archivist
Special Collections department in a large public university

With 12 years of experience, Coretta is an expert at producing archival metadata. She's passionate about making archival collections discoverable and accessible, but frustrated with workflows that require her to produce description in multiple systems, export to several others, and wait days or months in between. As her repository's backlog of unprocessed hybrid and born-digital collections grows, Coretta is excited about exploring new ways to provide access to digital materials in a user-friendly discovery environment.

"I want to get archival information to the user easily, quickly, and completely—giving researchers streamlined access to all sorts of digital material within ArcLight, as much as possible."

Frequency of ArcLight use: Daily

Technical proficiency: Moderate to high

	Motivation	Scenario	Specific Goals
❶	Easily publish and update archival descriptions and digital collections	Coretta just finished processing a large hybrid collection—the papers of a well-known writer. She created description in ArchivesSpace, and used Forensic Toolkit Imager, BitCurator, Archive-It, and Archivematica to process and package the born-digital materials and ingest them into her Fedora/Hydra repository. In ArcLight, she selects her data sources, and selects the indexing options she wants. She knows researchers will want to search the content of the writer's drafts and correspondence, not just her description, so she selects full-text indexing. When the indexing is complete, she previews what the collection will look like to the end-user. She catches an error in the description, and corrects it in ArchivesSpace; pushing the small update to ArcLight is quick. When she's happy with the preview, she clicks "Publish" to make the collection discoverable in ArcLight.	• Create and update archival description in any tool; index and publish it in ArcLight • Workflow optimized for integration with ArchivesSpace • Integrate digital material from different sources • Make indexing choices based on content • Preview the collection in context
❷	Link digital content to the description at any level	The writer's collection includes over 2,000 digital photos. Coretta creates description for the bulk of these images at the series level, rather than the item level, knowing that in ArcLight the images will be presented as a grid of thumbnails for the user to browse or sort by embedded metadata. For a few important images the writer has annotated (ones used for covers of the writer's books), Coretta creates (a separate series with) item-level description to make them more easily discoverable. Most of the writer's manuscripts are in Word files, which display as static documents by default. Coretta's Support Developer recently installed a web-based text mining plugin created by an ArcLight community member; Coretta overrides the default viewer so the Word files open in the plugin.	• Integrate digital material into archival description at item, folder, series, and/or collection level • Override default viewer configuration for a collection • Expose embedded content metadata to end users
❸	Configure discovery and access restrictions	The writer has stipulated that a series of drafts can only be accessed in the reading room, but when Coretta demonstrates ArcLight's virtual reading room functionality, he consents to virtual access. Coretta configures access so that researchers must request permission to view the drafts online, and once approved, must agree to terms of use before each access. She specifies that all emails in the collection are embargoed for 5 years; when this period expires, the emails will be automatically released in ArcLight.	• Support access restrictions at the item, folder, series, and/or collection level • Enforce embargo until the specified date has passed • Allow access to specific users/groups
❶	Enhance the archival description via linked data	The archives received this writer's collection largely due to its relationship to other collections in the archives. At the donor's request, Coretta has taken the extra step to create a rich agent record, including bio notes from multiple sources, and links to other writers and collections. She would love to see network graphs and other visualizations of these rich connections in the discovery environment.	• Support EAC or other linked-data connections to other collections and contextual information • Display graph or other visualizations of these connections

Sample MVP requirements

Must Have

- Display/link digital material at various levels: item, folder, series, collection
- Display of AV in context of description
- Display of images in context of description
- Support for user access to digital content
- Communicate level of description for digital objects

Should Have

- Bring together elements of the archival collection that might be in different silos
- Gain access to digital content in various formats directly from discovery interface.
- Integrate digital material from different locations or systems outside of ArcLight

Could Have

- Provide layered/tiered access
- Users can sign consent form/waiver
- Configure viewers for digital material within (or one-click-away-from) ArcLight
- Define user group(s) that can access items, folders, series, and/or collections

Won't Have

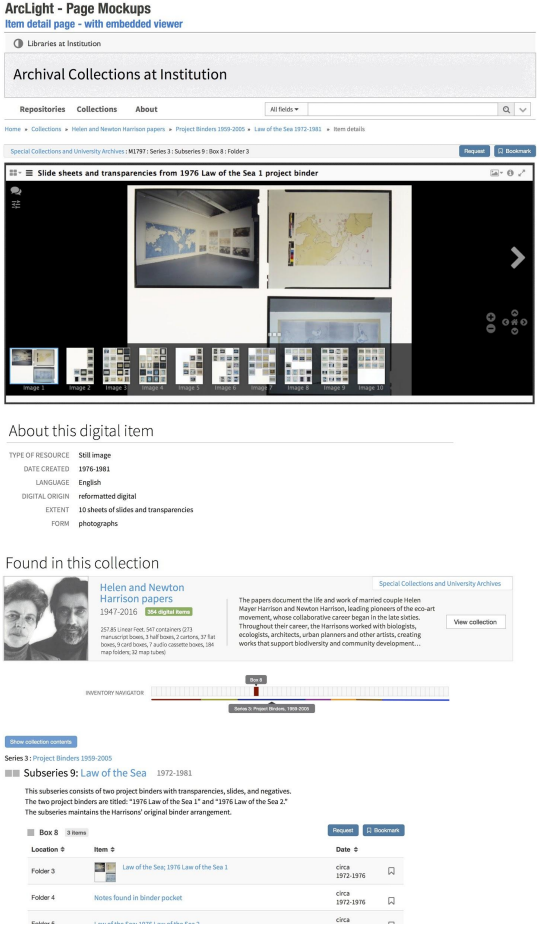
- Stream AV side-by-side with a transcript (e.g. oral histories)
- Provide "virtual reading room" access to authenticated users for materials with restrictions
- Restrict access to digital objects by IP
- Staff can upload on-demand scans

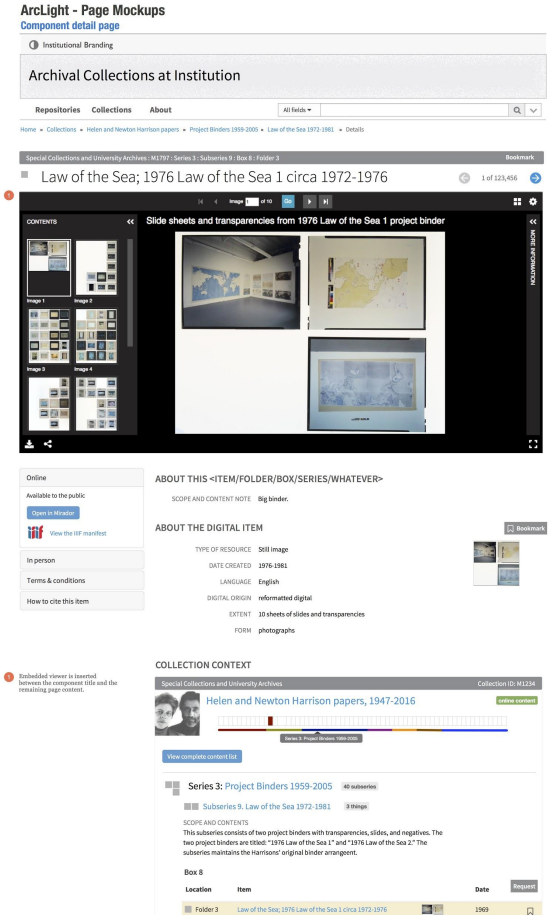
Wireframes

Tracking the evolution of design for integrated digital object delivery

Left: early iteration of wireframe

Right: revised design based upon feedback





The ArcLight MVP

https://arclight-demo.projectblacklight.org/

Screenshots: discovery

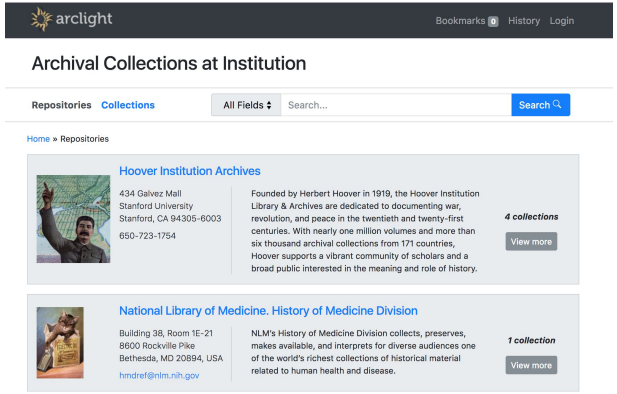
Clockwise from top left:

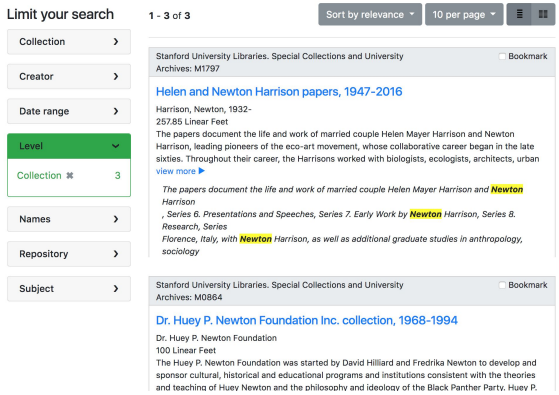
Repository/department listing page

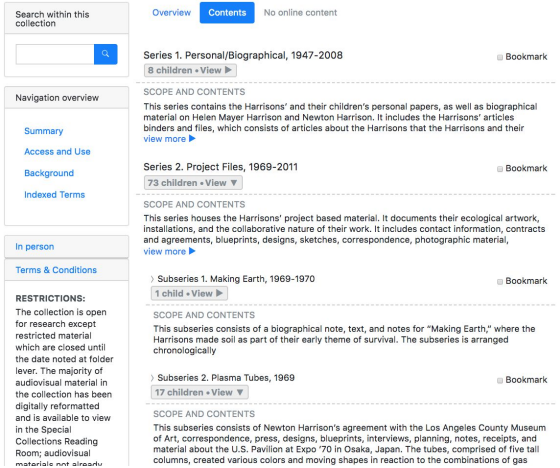
Keyword search and hit highlighting

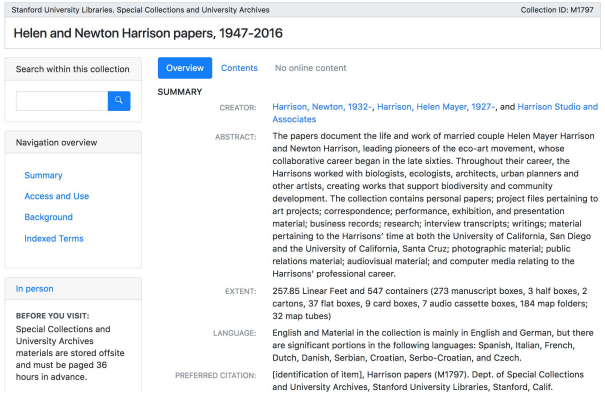
Finding aid frontmatter

Hierarchical inventory display









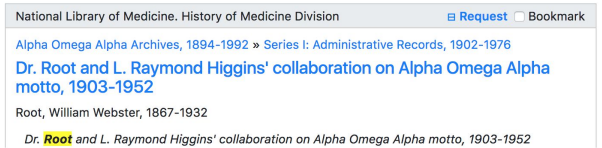
Screenshots: delivery

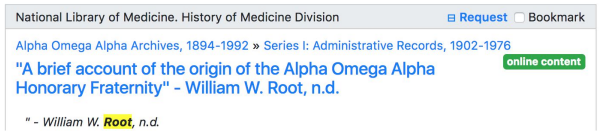
Clockwise from top left:

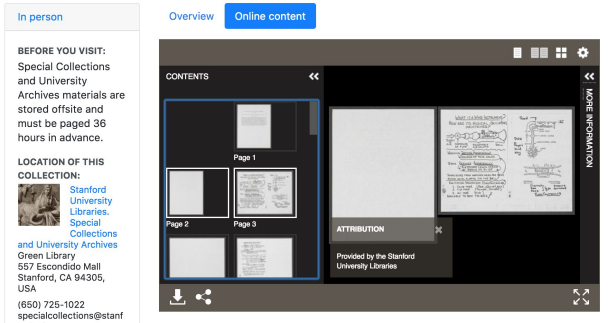
Integrated, configurable request links

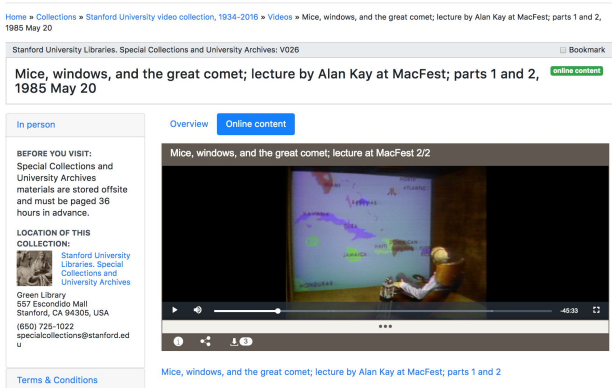
Image/paged content display

Integrated AV playback





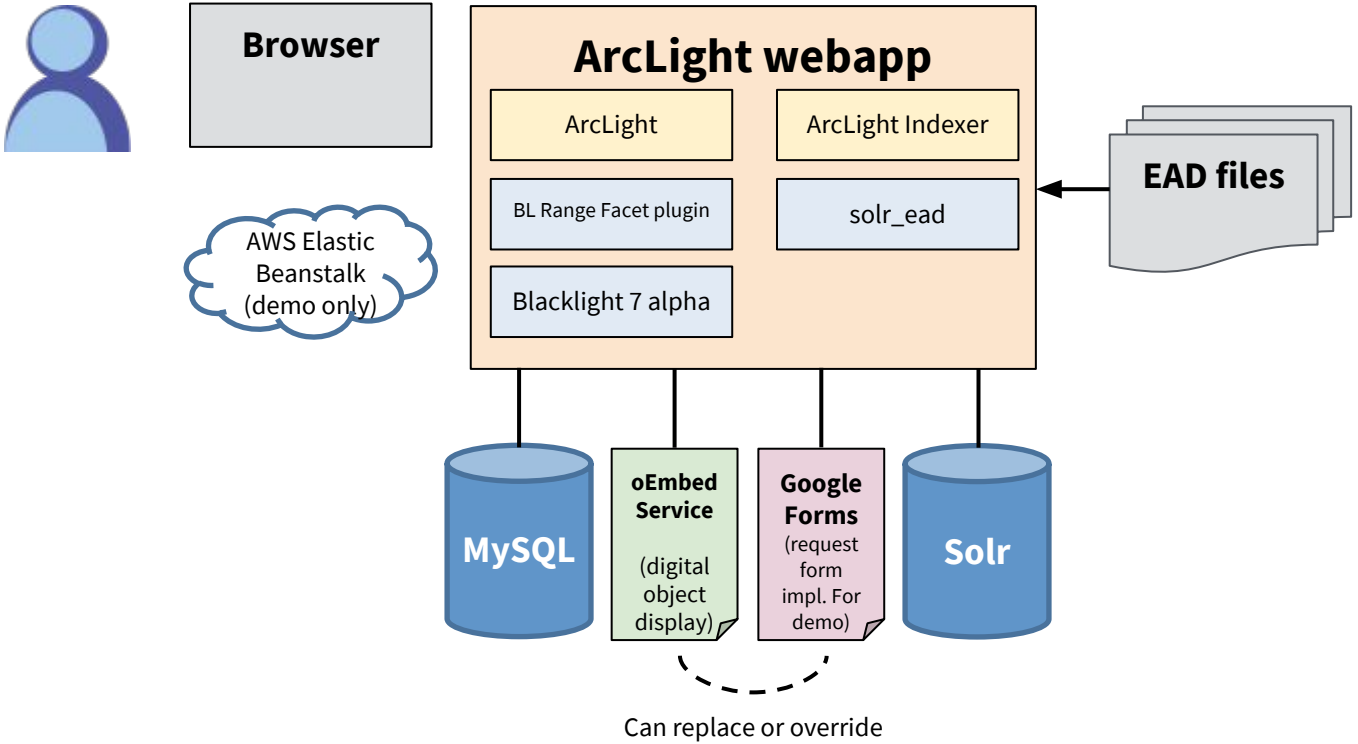




Implemented features

- Presentation of archival description, hierarchically and by individual components
- Repository information
- Integrated delivery of digital objects (using oEmbed)
- Indexing of EAD 2002
- Keyword searching and faceting by collection, creator, date range, level of description, names, creators, repository, etc.
- Sorting by date, creator, title, relevance
- WCAG 2.0 Level AA conformance
- Search within collections
- Hit highlighting
- Component-level views with contextual information
- Proof of concept request management integration
- Bookmarks
- Configurable repositories and departments
- Basic inheritance of descriptive metadata (e.g. for access and use restrictions)

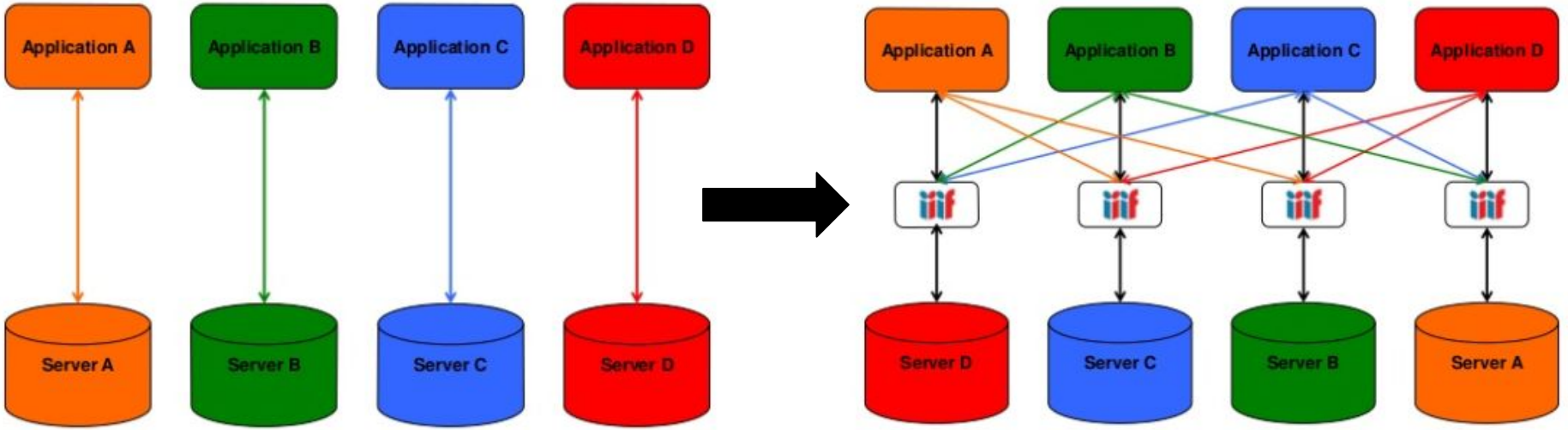
Architecture



Technical affordances

- We should leverage existing technologies from both within the cultural heritage IT sector and beyond to make this easier
  - International Image Interoperability Framework (IIIF)
  - oEmbed
- We should not limit ourselves to custom delivery mechanisms only for ArcLight - any system should be able to reuse viewers
- We need to allow other users or developers to build custom viewers when necessary or to cover specific content types

IIIF reduces some delivery barriers

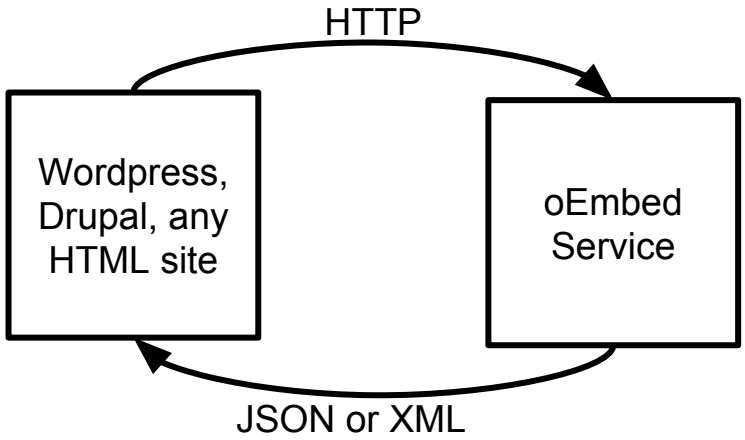


http://iiif.io/

oEmbed

- Simple format/API for sharing embedded Web content
- Existing specification used by many platforms: YouTube, Flickr, Hulu, Slideshare, Twitter … & **Stanford Digital Repository**
- Allows us to model a pattern of reuse of objects


```
{
        "version": "1.0",
        "type": "video",
        "provider_name": "YouTube",
        "provider_url": "http://youtube.com/",
        "width": 425,
        "height": 344,
        "title": "Amazing Nintendo Facts",
        "author_name": "ZackScott",
        "author_url": "http://www.youtube.com/user/ZackScott",
        "html": "<object width=\"425\" height=\"344\">
                ...
}
```

Topics for further thought

- oEmbed adoption is a huge opportunity for cultural heritage and education sectors but needs platform-level support
- Integration of request management systems is not easily reproducible given differences in implementations
- Broader need to consider how permissions/restrictions for repositories and content delivery integrate with request management systems
- Opportunity to improve front-end systems integration not just within archives, but across our institutions

Current status and next steps

- Development currently on hold, other than minor maintenance and support inquiries
- Other institutions experimenting with implementations and considering additional work
- Looking for more user input, especially from researchers or scholars
- Anticipating potential work in 2018 calendar year
- Investigating funding opportunities

Acknowledgments

Individuals

Hillel Arnold[3], Tom Burton-West[3], Tom Cramer[3], Max Eckard[3], Roger Espinosa[3], Erin Fahy[2], Frank Ferko[3], Charles Fosselman[3], Patrick Galligan[3], Gary Geisler[1,2], Bonnie Gordon[3], Darren Hardy[2], Wendy Hagenmaier[1,3], Nabeela Jaffer[1,3], Jenny Johnson[3], Jessie Keck[2], Gordon Leacock[2], Mark Matienzo[1,2], Sarah Newhouse[1], Kayla Ondracek[1], Michelle Paquette[3], Sarah Patton[3], Dallas Pillen[3], Chris Powell[1,3], Jack Reed[2], John Rees[1,3], Mike Shallcross[1,3], Stu Snydman[3], Camille Villa[2], Jennifer Vine[1,2], Laura Wilsey[1,3]

(1) Design phase contributor
(2) MVP development team member
(3) MVP development stakeholder

Institutions

Chemical Heritage Foundation
Columbia University
Georgia Tech
Getty Research Institute
Indiana University
National Library of Medicine
New York University
Pennsylvania State University
Rockefeller Archives Center
Stanford University
United States Holocaust Memorial Museum
University of Michigan
Yale University

More information

- Demo site: https://arclight-demo.projectblacklight.org/
- Design documentation: http://bit.ly/arclightproject
- Demo videos: http://bit.ly/arclight-demo-videos
- Google Group: http://bit.ly/arclight-community

Thank you!

Mark Matienzo, Stanford University / @anarchivist

Coalition for Networked Information, 11 December 2017

For more information on ArcLight, visit http://bit.ly/arclightproject