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](http://bit.ly/arclightproject)) 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
- Environmental scan
- Stakeholder goals
- Interview planning

Phase 2
- Archivist interviews
- End-user interviews
- Interview analysis

Information Architecture

Phase 1
- User needs
- User personas
- Requirements prioritization

Phase 2
- Conceptual models
- Conceptual sitemaps
- Wireframes

MVP Development and Visual Design

Phase 1
- Minimum viable product
- Visual design mockups
- Visual design style guide

Timeline:
- Nov 2014
- Feb 2015
- May 2015
- Summer 2016
- Oct 2016
- Nov 2016
- Jan 2017
- Feb 2017
- Mar 2017
- Apr 2017
- Jun 2017
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 straightforward 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)*

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
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

FileNotFoundException

Line 171
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

Browser

ArcLight webapp
- ArcLight
- ArcLight Indexer
- BL Range Facet plugin
- Blacklight 7 alpha

MySQL
- oEmbed Service (digital object display)

Solr
- Google Forms (request form impl. For demo)

EAD files

Can replace or override

AWS Elastic Beanstalk (demo only)

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 … and Stanford Digital Repository
- Allows us to model a pattern of reuse of objects

```json
{
    "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\(^4\), Tom Cramer\(^2\),
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\(^{1,2}\), 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/
Thank you!

Mark Matienzo, Stanford University / @anarchivist
Coalition for Networked Information, 11 December 2017
For more information on ArcLight, visit http://bit.ly/arclightproject