ArcLight and beyond
Illuminating the problem of "seamless" archival discovery and delivery

Mark A. Matienzo / @anarchivist / 7 August 2017

Collaboration & Interoperability Architect, Stanford University Libraries

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

Overview & history
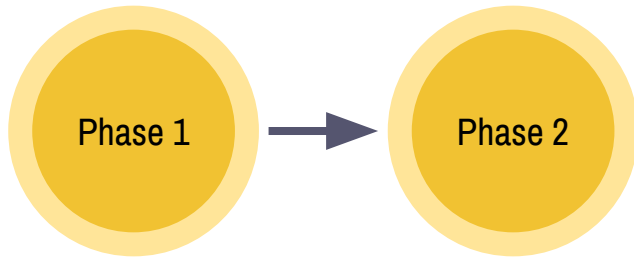
Description and objectives

- Project initiated by Stanford University in 2014 to address a long-standing interest in discovery/delivery of information in archives

- **Support discovery of physical and digital objects**

- Compatibility with and intended for integration with other systems, e.g. ArchivesSpace and Hydra-based repositories

- Development, enhancement, and maintenance by the Blacklight/Hydra communities

- Maintain a community focus throughout the project

Design/development process

- Design process led by 2 UX designers in Stanford Libraries' Digital Library Systems & Services group (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** was an intentional choice and existed from the beginning
  - Informed by Stanford's participation in open source projects
  - Opportunity for other institutions to identify needs and participate in work
  - Build interest and identify potential commitments for software development

Stanford | LIBRARIES

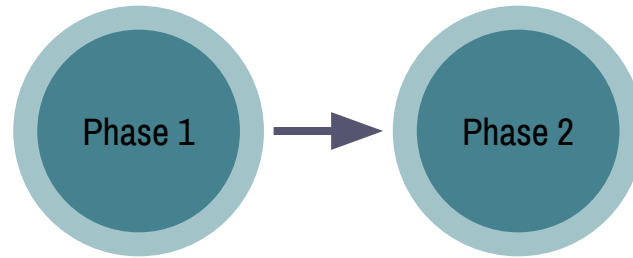# Process structure/timeline

Discovery

Information Architecture

MVP Development and Visual Design

Phase 1 | **Phase 2** | **Phase 1** | **Phase 2** | **Phase 1**

Environmental scan
Stakeholder goals
Interview planning

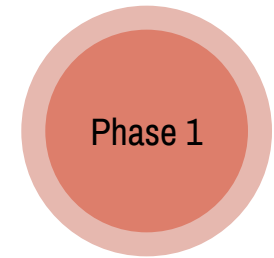Archivist interviews
End-user interviews
Interview analysis

User needs
User personas
Requirements prioritization

Conceptual models
Conceptual sitemaps
Wireframes

Minimum viable product
Visual design mockups
Visual design style guide

Nov 2014 — Feb 2015 ···· May 2015 — Summer 2016 — Oct 2016 ···· Nov 2016 — Jan 2017 — Feb 2017 ···· Feb 2017 — Mar 2017 — Apr 2017 ···· Apr 2017 — Jun 2017

Digital materials and ArcLight

Early origins of requirements

- Improving discovery and delivery of digital objects was identified as an early goal for ArcLight
- 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; desire for integration with Aeon or other existing registration/request systems

Stanford | LIBRARIES

User interview quotes (1)

Sample quotes from on delivery in context and embedded viewers

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)**

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)**

ArcLight probably can't have an embeddable viewer, accomplish delivery of every file format. The great thing about finding aids, is we have this link. For the most part, for crawls and [digital repository], [you] probably get a better view. **(Archivist)**

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

Stanford | LIBRARIES

User interview quotes (2)

Sample quotes from on siloing

I think the distinctions [where content is stored] are important for us … for knowing where things are, especially if there is a reprocessing project or we need to verify something. That's what we use our current collections management database for… But probably I feel researchers don't care where things live as long as they can have access to them. **(Archivist)**

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)**

I don't think that in terms of presenting digital material, 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)**

You need to do the search three different times, the fields are not auto-populated, and by the time you get the book, they're like, 'You need to return it next week, no renewals.' That is a disincentive. [...] That sort of siloing exists in so many levels in the library's digital space. **(Researcher)**

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

User interview quotes (3)

Sample quotes from on access controls/registration integration

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)**

There have been so many times I have been frustrated with the [digital library] with their restrictions on digital content. It's like, there's a magazine, that's not printed anymore, that was last published in 1920, and I want the 1919 edition, and [digital library] is like, '... I cannot give you the digital content. You need to drive to [library] and request it.' **(Researcher)**

[M]ost of the borndigital [...] they have to come to the Reading Room to actually look at the full text. [...] Because, I think nobody feels quite comfortable releasing [...] everything [...] because there's no way to know if there's private information, no matter how much due diligence you've done, and because we don't have ways for people to register and become somewhat accountable if they look at data and do something wrong with it. **(Archivist)**

The integration with Aeon is first and foremost of definite importance to us. **(Archivist)**

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

Sample MVP requirements

Must Have

- Display/link digital material at various levels: item, folder, series, and/or 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 (i.e., a Hydra repository, Archive-It web email in ePADD, etc.).
- Gain access to digital content in various formats directly from finding aids/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 rendering 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
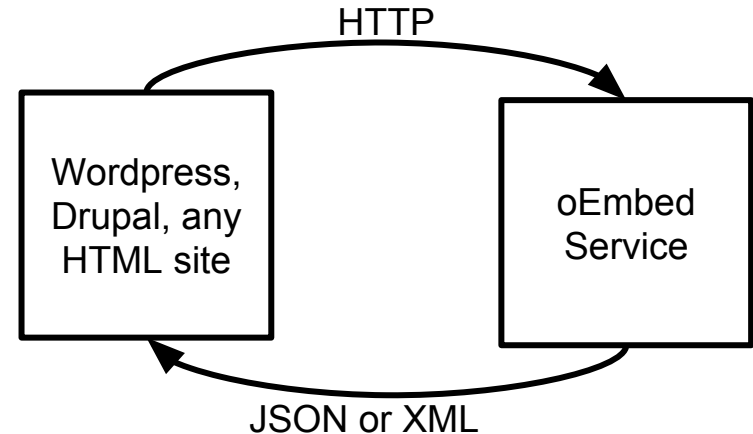
Demonstration

Implementation

Technical affordances

- We should leverage existing technologies from both within and beyond the cultural heritage IT sector to make this easier
- 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

oEmbed

- Simple format/API for sharing embedded Web content on
- 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\">
                ...
}
```

HTTP

Wordpress, Drupal, any HTML site

oEmbed Service

JSON or XML

IIIF

… or the International Image Interoperability Framework

- A **Community**
- That develops **Shared APIs for** web-based image delivery
- Implements them in **Software**
- And exposes interoperable **Content**

http://iiif.io/

What does IIIF provide?

Image API

- Transfer image pixels
- Image manipulation

Presentation API

- Presentation of an object (pixels + navigation and metadata)
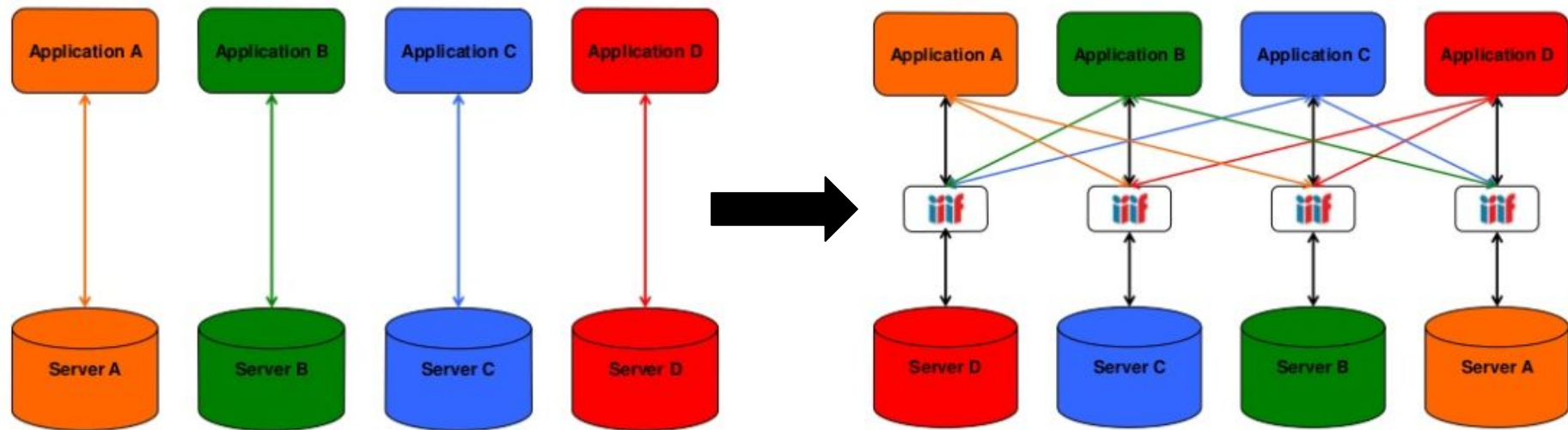- Easily share and re-use, mix and match content
- Annotate content

Search API

- Search annotations

Authentication API

- Provide interoperability for access-restricted content

IIIF APIs

Challenges

- Adoption of some of these standards still requires work
- Integration of request management systems is still challenging because of implementation differences
- Permission models of digital object delivery systems
- Coordinating community effort

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

Thank you!

Mark A. Matienzo / @anarchivist / 7 August 2017

Collaboration & Interoperability Architect, Stanford University Libraries

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

STANFORD | LIBRARIES

Discussion questions

- How do you want to provide access to digital materials from your special collections? What are the barriers to doing that?
- What would improved delivery of those materials across platforms mean to you and your users?
- How would you want a registration and request management system like Aeon to help with delivery of digital materials?
- How can developers of archival discovery platforms or finding aid interfaces help you achieve any of these goals?

Notes document: **http://bit.ly/xpsu-arclight-notes**