Digital Libraries at Stanford: infrastructure, services, & projects for accessible content

Mark A. Matienzo / @anarchivist
Stanford University Libraries
My background

https://matienzo.org/
Stanford Digital Libraries as infrastructure and services
Stanford Libraries by the numbers

- 12 million volumes
- 75,000 current serials
- 500,000 cartographic resources (maps, images)
- 59 million manuscript pages across Special Collections
- 120 countries from which we acquire
- 400 languages represented
- 110,000 new volumes acquired each year
Stanford University Libraries (SUL)

- 16 Libraries
- ~450 staff
- 68 subject / information specialists
- 8th largest academic unit at Stanford (by budget)
- Plus...
  - Stanford University Press
  - LOCKSS
  - Art & Architecture (Bowes)
  - Archive of Recorded Sound
  - David Rumsey Map Center
  - Earth Sciences (Branner)
  - East Asia
  - Education (Cubberley)
  - Engineering (Terman)
  - Green
  - Lathrop
  - Marine Biology (Miller)
  - Music
  - Stanford Auxiliary Library 1&2 (SAL)
  - SAL 3
  - Science (Li and Ma)
  - Special Collections
  - Stanford Digital Repository
What Does a Library Do?

- Select resources to add to its collections
- Acquire those resources
- Describe the resources to enable discovery
- Store & manage items until such time as they are needed
- Provide access to the items to patrons on request
- Preserve the items so they are available to future scholars
- Support teaching, scholarship, and research needs
The Digital Library: Content, Services & Infrastructure

- Ebooks
- Google Books
- Digitized mss, texts, images, media
- Born-digital materials (data sets, theses, articles, new media, etc.)

Services

- Discovery
- Delivery
- Use
  - Analysis
  - Annotation
  - Citation
- Collaboration
- Publication

Content

Infrastructure

- Digital preservation infrastructure
- Identifiers, Resolvers
- Server, storage and data center facilities
Stanford Digital Repository: Manage, Preserve, Access

SDR at a Glance

- **2006**: 1st year of operation
- **1.6 million** Digital Objects
- **227 million** Files
- **1,800+** Collections
- **467 TB** Unique Data
- **930+** Depositors from **40+** departments
The Stanford Digital Repository (SDR) provides services to make scholarly resources available over the long term by helping ensure their integrity, authenticity, and reusability.

To fulfill its mission, the SDR must be secure, sustainable and trustworthy.
Funder Mandates & Other Use Cases

1. Supplemental data files for published articles
2. Reproducible Research: sharing data & code
3. Complying with data management plans
4. Archiving pre-prints (with embargo support)
5. Publication archives
6. Open access articles
7. Dark archiving patent information
8. Enriched data access (e.g., large file access, video streaming, archived websites)

Distribution of Stanford Community SDR Deposits

- Article (25%)
- Thesis (10%)
- Working paper (12%)
- Dataset (13%)
- Technical report (16%)
- Class project (10%)
- Software (13%)
- Archival mixed material (12%)
- Report (10%)
- Newsletter (13%)
- Other (10%)
- Book (10%)
Digitization

Stanford has digitized millions of items in the last 20 years:

- Books
- Images
- Maps
- Manuscripts
- Audio
- Video
- Film
- Legacy digital files
- 3D objects

To support

- Collection development
- Digital access
- Research projects
- New forms of research (AI, ML, etc.)
- Preservation
Digitization

Stanford has digitized millions of items in the last 20 years:

- Books
- Images
- Maps
- Manuscripts
- Audio
- Video
- Film
- Legacy digital files
- 3D objects

To support

- Collection development
- Digital access
- Research projects
- New forms of research (AI, ML, etc.)
- Preservation

Stanford University by Arthur Lites, 1943 -
https://purl.stanford.edu/jr426dw0774
Preserving legacy research

Hopkins Marine Station CalCOFI hydrobiological survey of Monterey Bay, CA: 1951 - 1974

Project Goals

1. Digitize everything;
2. Rehouse materials in archival boxes;
3. Make content available via Stanford Digital Repository (SDR);
4. Convert select data from PDF to actionable format (tabular text).
Capturing new research

Chris Gerdes
Director, Revs Program
Simplifying publication management
Honoring Student Achievement
Replaying lectures, performances, oral histories
Feeding social science research
Harvesting the web
Supporting Access, Discovery, and Delivery

- SearchWorks
- Spotlight
- EarthWorks
- The International Image Interoperability Framework (IIIF)
- Requests and Course Reserves
SearchWorks

- Discovery environment ("library catalog") for Stanford University
- Allows for searching and browsing of materials from SDR and Stanford's libraries (inc. some outside of SUL), and articles
- Integrated access to digital resources, request management for delivery of materials, and citation generation

https://searchworks.stanford.edu/
EarthWorks

- Discovery of geospatial assets
  - Purchased
  - Licensed
  - Stanford-produced
  - Multi-institutional

- Direct access to geospatial web-services
  - Preview
  - Download
  - Sideload to online services

https://earthworks.stanford.edu/
Spotlight

- Full-featured website for showcasing digital collections:
  - Visually attractive, responsive, and accessible website
  - Self-service exhibit creation
  - Powerful searching and browsing, multiple search result views, include gallery, slideshow, and map views
  - Integrated with SDR and IIIF

Uses: rich digital libraries, digital exhibits, student portfolios and teaching tool

https://library.stanford.edu/research/spotlight

https://exhibits.stanford.edu/
Why Spotlight?
International Image Interoperability Framework

- Open international community building specifications for rich access to image and other resources
- Used to support deep zoom, presentation, annotation, and comparison of images
- Expanding to provide support for audiovisual materials and 3D resources

https://iiif.io/
How digital library projects happen: content and projects
Digital Object Life Cycle

<table>
<thead>
<tr>
<th>Original</th>
<th>Capture</th>
<th>Accession: PURL/SDR</th>
<th>Index: SearchWorks</th>
<th>Showcase: Spotlight</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Original Image" /></td>
<td><img src="image2.png" alt="Capture Image" /></td>
<td><img src="image3.png" alt="Accession Image" /></td>
<td><img src="image4.png" alt="Index Image" /></td>
<td><img src="image5.png" alt="Showcase Image" /></td>
</tr>
</tbody>
</table>

| Accessible: | Limited | ✔ | ✔ | ✔ |
| Preserved: | ✔ | ✔ | ✔ | ✔ |
| Discoverable: | ✔ | ✔ | ✔ | |
| Curated: | ✔ | | | |
Stakeholder Roles and Responsibilities

- Content owner
  - Project champion
  - Selection & decision making
  - Preparation
  - Description
  - Funding

- DLSS
  - Project management
  - Technical resources
  - Content management
  - User experience

- Conservation
  - Material review
  - Treatment
  - Stabilization
  - Housing
  - Handling

- Metadata
  - Planning
  - Metadata mapping
  - Best practices for access & discovery
Digitization/Acquisition

Content
- Digitize or Acquire
- Accession in SDR
- Publish in SW
- Showcase in SL

Metadata
- Plan & Map
- Create & Enhance
- Load to SDR
- Review & Remediate
Digitization/Acquisition is a Team Effort
Accessioning

**Content**
- Digitize or Acquire
- Accession in SDR
- Publish in SW
- Showcase in SL

**Metadata**
- Plan & Map
- Create & Enhance
- Load to SDR
- Review & Remediate
Creating Digital Objects is a Team Effort

- Content Owner
  - Descriptive metadata
  - Delivery/display specifications
  - Remediation

- DLSS
  - Assembly in Digital Object Registry (DOR)
  - Well-formed objects
  - PURL
  - Long-term preservation

- Metadata Department
  - Metadata review
  - Enhancement
  - Metadata loading assistance
Accessioning process
Assembly
Publishing in SearchWorks

Content
- Digitize or Acquire
- Accession in SDR
- Publish in SW
- Showcase in SL

Metadata
- Plan & Map
- Create & Enhance
- Load to SDR
- Review & Remediate
Publishing is a team effort

- **Content Owner**
  - Quality control & remediation on:
    - Content
    - Metadata
    - Display
    - Rights

- **DLSS**
  - Develop, operate & maintain discovery environments
  - Remediation of technical issues

- **Metadata Department**
  - Remediation guidance & support
Minimum Metadata for SearchWorks

- Title
- Creator
- Language
- Type of Resource
- Date

& preferably more!

Metadata supporting discovery
Building Exhibitions in Spotlight

Content
- Digitize or Acquire
- Accession in SDR
- Publish in SW
- Showcase in SL

Metadata
- Plan & Map
- Create & Enhance
- Load to SDR
- Review & Remediate
Building an Exhibit is a Team Effort

**Content Owner**
- Create exhibit concept
- Liaise with service team
- Coordinate text creation & layout

**DLSS**
- Develop, operate & maintain Spotlight
- Service team: document, train & consult

**Metadata Department**
- Metadata display guidance & support
Ongoing Maintenance
Thank You!

Questions?

Mark A. Matienzo / @anarchivist
Stanford Libraries
matienzo@stanford.edu