



# Managing Digitization Priorities with Metrics

Robin C. Pike  
MARAC, S1: Archival Impact, April 15, 2016



## History of Digitization

- Ad hoc
- Office of Digital Collections and Research
- Digital Collections Unit
- Digital Conversion and Media Reformatting

MARAC, S1: Archival Impact, April 15, 2016



Before 2005, digitization occurred mostly ad hoc in special collection areas.

The University of Maryland Libraries made a decision in 2005 to create a digital repository using the open source Fedora digital repository. At the time, other solutions, such as CONTENTdm, were not as fully realized, and the University of Maryland Libraries wanted a scalable and flexible solution that could enable the long-term preservation of digital objects.

Hired a programming and librarian support necessary to develop such a system - at the time, the staffing of the then Office of Digital Collections and Research (DCR) included an Assistant Dean, two digital librarians, one database administrator, one developer, one user interface specialist, and the time of a metadata librarian - DCR embarked on a long-term development project, designed to be completed in stages and to meet a number of needs throughout the Libraries.

Digital Collections Unit

2010: one manager and four students. Developers moved to ITD

DCMR

2012: Split to digitization production and digital preservation activities

Mission:

The Digital Conversion and Media Reformatting Department (DCMR) seeks to support the Libraries' collection development goals, and strategic priorities for preservation and access by working with collection managers and subject specialists to digitize collections of all formats through a centralized, production-based environment.

## Development of Workflows

- Four-year process
- In-house
- Vendor



MARAC, S1: Archival Impact, April 15, 2016



Like many of the digitization programs, now at point where refining and revising current guidelines and practices, especially what can be done in-house  
Standardized workflows across collection areas, projects  
Upscaled digitization capacity by outsourcing  
Current Capacity  
In-house digitization: text, image, audio requests and small projects  
Outsourced: large projects, audio and moving image requests

## Development of Digital Collections

- Seven campus libraries
- Broad collection interests

MARAC, S1: Archival Impact, April 15, 2016



### Role:

Provide the digitization operations for the seven College Park Libraries, and serve as a leader for digitization in the community. Outside of Special Collections department.

## Production/Collection Growth

- In-house
  - Requests
    - Patrons
    - Staff (exhibits, classes, etc.)
  - Small projects
    - Second/third tier workflow
- Vendor
  - Large projects
  - Special projects
  - Requests of specialized formats



## Digitization Vendors

- KEY: grouping "like" formats together for better rates
- Lyris Digitization Collaborative
  - Internet Archive
- Rationale
  - One overarching vendor agreement
  - Discounted rates (from normal vendors' rates)
  - Set SOW, with some modifications
  - Spend more time on projects, not contracts
- Ad hoc vendors
  - Specialized work
  - Deadlines

## Selection Process

- Access need
  - Patron/staff requests
  - Formats are barrier to unmediated access
- Preservation need
  - Assessments
- Institutional priorities

## Digitization Initiatives Committee (DIC)

- In lieu of budget line
- Scalable process
- Cross-divisional expertise/interests

## DIC Criteria

Area	Considerations	Weight	Score
Relation to Teaching and Research Priorities	Intellectual value, collection assessment score (if available), previous use of material, etc.	1	High=3 Moderate=2 Low=1 None=0
Relation to Library Administration Priorities	Political, space (disposition or sending to offsite storage), development, public relations, etc.	1	High=3 Moderate=2 Low=1 None=0
Department/Unit Prioritization	When a department/unit submits more than one project, DIC will contact you to request prioritization information.	2	High=3 Moderate=2 Low=1 None=0
Continuation from Previous FY	Multi-year proposal.	2	Yes=1 No=0
Preservation Need	Preservation assessment score; consult with Preservation & Conservation.	2	High=3 Moderate=2 Low=1 None=0
Project Readiness	Available description for metadata; status of preservation work; material selection; other resources needed; plans for prep work or prep work completed. People should have spoken to MSD or P&C and prep work should be on work plans.	2	High=3 Moderate=2 Low=1 None=0
Available Funding	Gifts, endowments, donors, partners, grants, etc.	2	Full=2 Some=1 None=0
Partnerships	Campus, grants, CIC, etc. Relationships or funding.	3	Yes=1 No=0
Copyright status	On-campus access only will not be a limiting factor to digitization.	4	Public=2 On-campus=1 Dark archive=0

## Funding Projects

- Digitization Initiatives Committee
  - Operations (state line)
  - Gifts
  - Endowments
  - Foundation
  - Fundraising
- Grants

Grants: NDNP and other applications NEH Humanities Collections and Reference Resources Implementation

## How We're Tracking Usage

- Better representation across collection areas
- Google Analytics
- Web scraper used for Internet Archive
- Asking for statistics from Chronicling America

## Usage Trends 2013-2015

- Region
  - 70-80% Non-Metro
  - 80-90% Non-College Park
- 60-80% New Users (though each College Park public computer session is tracked as new)
- Languages of highest use
  - English (US), Japanese, European languages (various)
- 3-4 pages visited, on average
- As social media referrals increase
  - 60-70% Leave after arriving on initial page
  - Average session decreasing from 3-4 minutes to 2-3 minutes
- Pageview trends ranking by semester: Fall, Spring, Summer, Winter
- Increasing mobile users

MARAC, S1: Archival Impact, April 15, 2016



### Region

70-80% Non-Metro

80-90% Non-College Park

60-80% New Users (though each College Park public computer session is tracked as new)

Languages of highest use

English (US), Japanese, European languages (various)

3-4 pages visited, on average

As social media referrals increase

60-70% Leave after arriving on initial page

Average session decreasing from 3-4 minutes to 2-3 minutes

Pageview trends ranking by semester: Fall, Spring, Summer, Winter

Increasing mobile users

## Content Trends 2013-2015

- Most frequent landing page is "digital collections home page"
- Users follow search-result-search pattern
- Popular digital exhibits/ collection landing pages
  - Appear to be consistent interest among semesters, inferring classes, though not necessarily from UMDCP
- Most popular:
  - Baroness Elsa von Freytag-Loringhoven Digital Library (Dada movement, poems and papers)
  - Films at UM (educational licensed film content)
  - University AlbUM (UMD photos from various collections)
  - Prange Collection (censored publications from US Occupation of Japan)
  - Jim Henson Collection (licensed videos)

## Plans for Data Use

- Integrate into DIC prioritization rubric
- Outreach to collection areas
  - Project planning
- Fundraising/grants for collection areas
- Repository development
  - Access restrictions due to copyright
  - Need to be more friendly for mobile users (Zoomify image viewer uses flash)

MARAC, S1: Archival Impact, April 15, 2016



Access models governed by repository, a/v streaming server  
 Internet Archive, Chronicling America: public (public domain or with permission)  
 HathiTrust: public or on-campus only through brittle book clause  
 Fedora:  
 Public to all  
 Public on campus/VPN off campus (IP restrictions), metadata public to all  
 Restricted to admin/private (dark archive), metadata restricted  
 Public link, metadata restricted (only used for one collection)  
 Problematic if high-access collection areas are restricted to campus usage only due to copyright or donor agreements; causes extra work for us to deliver files directly to patron if they just see metadata  
**We want to prioritize digital collection production based on what people will use or cannot access otherwise. We also want to build collections that are accessible to the most amount of users, and now have supporting documentation to undertake this as a priority.**