Community Efforts to Develop Best Practices in Digital Library Assessment: One Year of Progress

DLF Assessment Interest Group

Southeastern Library Assessment Conference
Atlanta, Ga.
November 17, 2015
Presentation Outline

1. Background
2. Users Studies
3. Analytics
4. Citations
5. Cost Assessment
6. Questions

Created by ARudmann from Noun Project
DLF AIG: Important Links

You do not need to be a member of DLF to join/participate!

DLF Assessment Interest Group Google Group and listserv:

DLF Assessment Interest Group wiki, including links to all the white papers and best practice documents and tools discussed today:
http://wiki.diglib.org/Assessment

Founded in 2014, the AIG is currently co-chaired by Joyce Chapman (Duke) and Santi Thompson (University of Houston). There were 4 Working groups in 2014/15: User studies, Analytics, Citations, Costs.
User Studies Working Group

Evaluating users of digital libraries and their needs

1. Goals of project
2. Methodology
3. Results and recommendations
4. Next steps
Goals

Guidelines/best practices for assessing facets of digital library value by analyzing user interaction with digital libraries.
Scope

User and Usability Studies

Return on Investment

Content Reuse

Created by Lemon Liu from Noun Project
Examining the literature

1. What research strengths exist in the areas of usability, ROI, and reuse assessment in digital libraries?
2. What gaps exist in these areas of focus?
3. What are possible next steps for the community to address?
# User and Usability Studies

<table>
<thead>
<tr>
<th>Strengths</th>
<th>User-centered design and assessment</th>
<th>Design strategies through user search behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research Gaps</strong></td>
<td>Behavioral observations and examination of user's task context</td>
<td>Over-reliance on standard testing tasks and user feedback</td>
</tr>
<tr>
<td><strong>Recommendations</strong></td>
<td>Users’ research needs</td>
<td>User’s role in system development</td>
</tr>
</tbody>
</table>
## Return on Investment

<table>
<thead>
<tr>
<th><strong>Strengths</strong></th>
<th>Measurement of time and cost for processing</th>
<th>Theoretical application of ROI to library project management</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research Gaps</strong></td>
<td>Benefits of cost/benefit analysis</td>
<td>Limited corpus of cost data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No standard methodology for implementation</td>
</tr>
<tr>
<td><strong>Recommendations</strong></td>
<td>More studies</td>
<td>More data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>More tools</td>
</tr>
</tbody>
</table>
## Content Reuse

| Strengths                  | Reuse among humanities-focused digital repositories |  |  |
|----------------------------|-----------------------------------------------------|  |  |
| Research Gaps              | Patterns of reuse                                   | Web log analysis | Non-digital reuse |
| Recommendations            | User groups -- science and social science           | Digital repository interface and reuse | Assessment framework |
Next steps

Documenting best practices?

Generating assessment toolkit(s)?

Or…?

Wanna get involved?

Contact Santi Thompson: sathomp3@central.uh.edu

http://bit.ly/1KzCU6S
Analytics

1. Goal and scope
2. White paper methods and recommendations
3. Next steps
Goal

*Develop guidelines and best practices for using analytics in digital libraries*

Scope

Focus on Google Analytics

Guidelines for all experience levels
Literature review

Literature review reveals gap in best practices

“How-to” analytics literature covers: use analytics to improve usability and discoverability, set-up google analytics (case studies), and complete transactional log analysis
White Paper: Metrics Methodology

1. Choose metrics to recommend

2. Define each metric

3. Importance of metric

4. Bolster definition with library-centric examples

Caveat: metrics require interpretation by local organization to be relevant and actionable.
Baseline Google Analytics Metrics Recommendations

A. Content Use and Access Counts
   1. Content Use and Access Counts Defined
   2. Site Content Reports
   3. Bounce Rate
   4. Download Counts
   5. Time
   6. Pageviews
   7. Sessions

B. Audience Metrics
   1. Location
   2. Mode of Access
   3. Network Domain
   4. Users

C. Navigational Metrics
   1. Path Through the Site
   2. Referral Traffic
   3. Search Terms
White Paper: More than Metrics

Approaching analytics: know thyself

Alternative tools and methods: consider trade-offs between tools and methods

Going beyond the baseline: customization and platform specific considerations
Next steps

Options options options

More platform specific metrics?

More examples?

Share metrics?

Discontinue the analytics group?

Citations Working Group

1. Background discussion
2. Methodology
3. Recommendations
4. Next steps
Citations Working Group

DLF AIG Charge

1. What should a citation consist of?

2. How can we best support appropriate citations?

3. To what extent do common citation formats support this?

4. What are the limitations of current digital library software systems for displaying citation information?

5. What are best practices for displaying citation information for reference manager software capture?
Draft citation standards, based on what can and can't be incorporated into APA, Chicago, and MLA, that incorporate the necessary elements for digitized special collections and institutional repository content
Citations Working Group

Methodology

Data set citations
Archival/manuscript citations
Web/electronic document citations
Institutional citation recommendations

Citation Styles
APA, Chicago (Notes and Bibliography), MLA
Citations Working Group

Recommendations

Include:

- Item name or title
- Collection name
- Repository information (physical and/or digital)
- Unique identifier (Digital Object Identifiers (DOI), Handle, Persistent Uniform Resource Locators (PURL), citable URL, etc.)
Citations Working Group

Next Steps for these guidelines:

Publication
Adoption
Advocacy
Next steps for citations working group:

1. What are the limitations of current digital library software systems for displaying citation information?

2. What are best practices for displaying citation information for reference manager software capture?

3. More? Do we as a group start creating scripts/plugins so that the most common DL & IR platforms can automatically generate statements for preferred citations? Do we need to do a bibliometric study of DL items in scholarly literature? Etc.
White Paper: “Guidelines for citing library-hosted, unique digital assets”

http://bit.ly/1MNJ3Ci
Cost Assessment Working Group

Goals

Process & products

Call for data
Cost Assessment Working Group

Goals

Process & products

Call for data

To aggregate and make freely available a large set of time/cost data on the performance of various tasks involved in the digitization process, in order to assist organizations in digitization project planning and benchmarking.
Cost Assessment Working Group

Library Digitization Cost Calculator

This calculator aggregates available data on the cost and time it takes to perform various activities associated with library digitization. The calculator provides estimates not accuracy, as each institution is different. The tool provides average time and cost information from other institutions who have donated their data to this project. Data is currently limited. This calculator was built by Joyce Chapman, please contact joyce.chapman@duke.edu for more information, or use the feedback form.

- **Types of scanner**
  - Assume: 1 f = 1,200 scans
- **Extent (linear feet)**
- **Hourly pay of student workers**
- **Student benefits as % of pay**
- **Annual salary of staff member**
- **Staff benefits as % of pay**
- **Metadata creation**
  - Assume: To calculate metadata creation time (by item, not scan)
  - Assume: Calculations will be for removing fasteners from 100% of items
  - Assume: Flagging/bulling materials to be sent to conservation department
  - Assume: Calculations will be for review of 100% of items
  - Assume: E.g., locate pp. 1-5 of a discrete document and place them together
  - Assume: This will be calculated as if it were done for 100% of items
  - Assume: This will be calculated as if 100% of items are fragile
  - Assume: This will be calculated as if 100% of items need disbinding

**Calculate**
Cost Assessment Working Group

Goals

Process & products

Call for data

Created by Takao Umehara from Noun Project
Cost Assessment Working Group

Performed a review of existing literature (published and unpublished) in the areas of

• Collection of time and cost data for digitization

• Existing best practices in quality control and metadata creation

• Found fewer than 20 resources!
Cost Assessment Working Group

Guidelines and definitions (1 of 2)

- Preparation of original materials
  - Condition review
  - Disbinding
  - Fastener removal
  - Flattening
  - Rights review
  - Sorting materials into items
  - Supporting
  - Unique identifier assignment

- Image capture
  - Film or transparency scanner
  - Flatbed scanner
  - Manual DSLR camera
  - Medium format camera
  - Overhead scanner
  - Sheet feeding scanner
Quality control
  Level 1, 2 or 3
Descriptive metadata creation
  Level 1, 2 or 3
Post-processing
  Alignment/rotation
  Background removal
  Clean up / dust removal

Color correction and tonal adjustments
Cropping images
Stitching
Post-preparation
  De-sorting
  Re-binding
  Re-fastening
**Digitization Cost Calculator**

<table>
<thead>
<tr>
<th>People performing the work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salaried 1</td>
</tr>
<tr>
<td>Annual salary</td>
</tr>
<tr>
<td>% benefits</td>
</tr>
<tr>
<td>Hourly 1</td>
</tr>
<tr>
<td>Hourly pay</td>
</tr>
<tr>
<td>% benefits</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Preparation of original materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition review</td>
</tr>
<tr>
<td>% materials</td>
</tr>
<tr>
<td>Performed by</td>
</tr>
<tr>
<td>Disbinding</td>
</tr>
<tr>
<td>% materials</td>
</tr>
<tr>
<td>Performed by</td>
</tr>
<tr>
<td>Fastener removal</td>
</tr>
<tr>
<td>% materials</td>
</tr>
<tr>
<td>Performed by</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Image capture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capture device</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quality control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
</tr>
<tr>
<td>Level 2</td>
</tr>
<tr>
<td>Level 3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Post-processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alignment/rotation</td>
</tr>
<tr>
<td>% materials</td>
</tr>
<tr>
<td>Performed by</td>
</tr>
<tr>
<td>Background removal</td>
</tr>
<tr>
<td>% materials</td>
</tr>
<tr>
<td>Performed by</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Local Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Field 1</td>
</tr>
<tr>
<td>% materials</td>
</tr>
<tr>
<td>Performed by</td>
</tr>
<tr>
<td>Local Field 2</td>
</tr>
<tr>
<td>% materials</td>
</tr>
<tr>
<td>Performed by</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Activity</th>
<th>Time (hours)</th>
<th>Cost (Salary 1)</th>
<th>Cost (Hourly 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faceover removal</td>
<td>10.3</td>
<td>N/A</td>
<td>$10</td>
</tr>
<tr>
<td>Image capture</td>
<td>3.47</td>
<td>$10</td>
<td>$10</td>
</tr>
<tr>
<td>Mending/repair</td>
<td>0.2</td>
<td>$10</td>
<td>$10</td>
</tr>
<tr>
<td>Total</td>
<td>10.5</td>
<td>$30</td>
<td>$30</td>
</tr>
</tbody>
</table>
Cost Assessment Working Group

Call for data submissions

Call for data submissions (a document that explains what we’re doing and tells you how to contribute your data)

Data submission form (the actual online form for submitting data), or just copy this link [http://bit.ly/1LV9oxI](http://bit.ly/1LV9oxI)

All data submitted to this project will be publicly available, both via aggregate calculations made by the calculator, and by institution on a separate reference page of the calculator’s website.
Cost Assessment Working Group

Call for data submissions

What amount of data do I have to submit?

What format do I submit the data in?

How does the calculator work / how will my submitted data be used?
Questions?

Contact: joyce.chapman@duke.edu


Download documents and follow progress on our wiki: http://wiki.diglib.org/Assessment