Digital Library Evaluation
Tech Talk

Nathan Hall
nathan.hall@unt.edu
March 9, 2011
Introduction

Various approaches to digital library evaluation

(Saracevic, 2000)

But few ways are established

(Borgman et al, 2001)
Problems:

Digital collections play an increasingly vital role in academic libraries

No widely accepted method to evaluate impact/benefit of a digital library to its institution

(Marchionini, 2000); (Zhang, 2007)

Evaluating only the service quality aspects of traditional library operations is no longer sufficient in improving library service

(Kyrillidou, Cook, and Lincoln, 2009, 188)
Problems:

We do not know how to evaluate a digital library as a whole.

(Saracevic, 2009)
Good place to start

Regardless of content or technology, libraries exist to provide a community with access to information.

Buckland (1992)

Here are the books you wanted, Master. I asked the library for your “CD-ROM” and they said this was the closest they had.

(Kyrillidou et al’s quote from 2 slides ago not withstanding)
desiderata for evaluation planning

- Construct
- Framework
- Context
- Criteria
- Methodology

(Saracevic, 2005)
Evaluation Construct

• Specific collection (e.g. Cyber Cemetery)

• Specific digital library (e.g. UNT DL, Portal)

• Specific process (e.g. keyword extraction, server load balancing)
Evaluation framework

• Bottom-up
  Starts with input from stakeholders to develop evaluation/design criteria

• Top-down
  Uses existing criteria based on theory/construct/perspective
Evaluation Contexts – and examples

• Systems-centered – efficiency of some component
• Human-centered – “information behavior”
• Usability-centered – assesses human use of technical features
• Ethnographic – e.g. recent presentation from Prof. Lisa Henry’s class
• Anthropological – ditto
• Sociological – assessment of user community in DL setting
• Economic approach – cost/benefit analysis
Evaluation Criteria

Criteria for digital library evaluation drawn from other research domains (Zhang, 2007, 14)

<table>
<thead>
<tr>
<th>Traditional library</th>
<th>Information Retrieval</th>
<th>User Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Collection: purpose, subject, scope, <em>etc.</em>&lt;br&gt;• Information: accuracy, appropriateness, presentation, timeliness and ownership&lt;br&gt;• Use: accessibility, availability, usability&lt;br&gt;• Standards</td>
<td>• Relevance&lt;br&gt;• Satisfaction&lt;br&gt;• Index, search and output features</td>
<td>• Usability, functionality, effort&lt;br&gt;• Task appropriateness, failures&lt;br&gt;• Connectivity, reliability&lt;br&gt;• Design features&lt;br&gt;• Navigation, browsing&lt;br&gt;• Services, help</td>
</tr>
</tbody>
</table>

Also: ethnographic and anthropological criteria (Saracevic, 2005)
Methodology

- Interviews
- Focus groups
- Observations
- Task accomplishment
- Think aloud
- Case studies
- Log analysis
- Experimentation
- Records analysis
- Usage analysis
- Communications analysis (documents, meetings)
- Economic analysis
Other types of evaluation

Longitudinal evaluation
Systemic evaluation

Similarity is that they both may involve multiple studies
Typical evaluations examine:

- Digital object
- Metadata
- Information
- Collection

Using conventional criteria
(listed on evaluation criteria slide)

(Zhang, 2010)
A few generic evaluation tools that address digital libraries

- ACRL Statistics Worksheet
- ARL DigiQUAL
ACRL (2008-2009) Statistics Worksheet

• Developed for evaluating university libraries
• Asks the following about digital collections and access:
  – number of digital collections,
  – the size in gigabytes,
  – the number of items.
  – number of times unique files were accessed,
  – number of queries conducted.
  – cost of personnel, equipment, software, and services.
DigiQUAL

- Examines DL service quality from user perspective
- Tool measures DL user perceptions and expectations of service quality
- Identifies general 'best practices' that permit generalizations across operations and platforms
- Enhance student learning by managing user perceptions and expectations of digital library services
- Identify opportunities and barriers in establishing a digital library service quality assessment program

(Kyrillidou et al, 2009)
DigiQUAL Development

• Focus groups identified 250 items describing different aspects of service quality in digital libraries.

• Analysis yielded 12 themes of digital library service quality...
ARL themes of digital library service quality

1. Web Attractiveness; Design Features
2. accessibility/navigability;
3. Other Technical Aspects of the Sites
4. interoperability;
6. DL as community for users, developers and reviewers; [might be good, but not requisite for DL in most definitions]
7. collection building;
8. role of federations; [as in federated search across institutions]
9. copyright;
10. resource use;
11. evaluating collections; *
12. DL sustainability.
DigiQUAL

- Authors feel identified themes were sufficiently different from LibQUAL.
- Authors feel DigiQUAL is technically scalable
- Digital libraries must have a critical mass of users willing to fill in survey data before conducting such evaluation as a total market survey.
- DL community collaborative values hampered by competitive funding structure that provides no incentive for sharing evaluation data
Missing criteria

• How well digital library meets users’ information needs
• How well digital metadata reflects ontology of the discipline represented by digital library
• How well DL meets the goals of the institution
Interesting top-down approach

• Yuka Koremura (2008)
  – Systemic evaluation of library performance at 6 interactive levels
  – Use of
    • Thomas Gilbert’s Human Performance Technology
    • Patrick Wilson’s Library and Information Science Policy
<table>
<thead>
<tr>
<th>Vantage Level</th>
<th>Accomplishment Variables</th>
<th>Context</th>
<th>Purpose</th>
<th>Accomplishment Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Philosophical</td>
<td><em>Ideals</em> that relate to the quality of life, transcend specific cultures or politics, and require specific goals if they are to be achieved.</td>
<td>Human identity</td>
<td><em>Raison d’être</em></td>
<td>Physical pleasure</td>
</tr>
<tr>
<td>II. Cultural</td>
<td><em>Goals</em> of the particular culture that give performance its meaning, and require policies if they are to be reached</td>
<td>Culture, the State</td>
<td>Fulfillment of ideals</td>
<td>Good food</td>
</tr>
<tr>
<td>III. Policy</td>
<td><em>Missions</em> that define the basic purpose of institutions and subcultures, and require programs of action</td>
<td>Institutions, organizations (subcultures)</td>
<td>Attainment of specific cultural goals</td>
<td>Tasty snacks manufactured and sold</td>
</tr>
<tr>
<td>IV. Strategic</td>
<td><em>Responsibilities</em> that define the roles of the members of an institution, and require plans for fulfilling them</td>
<td>Roles, jobs, etc.</td>
<td>Completion of missions</td>
<td>Potato chips put in packages</td>
</tr>
<tr>
<td>V. Tactical</td>
<td><em>Duties</em> that must be fulfilled to discharge the responsibilities of any role or job, and require tools for execution</td>
<td>Tasks, skills, etc.</td>
<td>Discharge of responsibilities</td>
<td>Wrapping paper mounted on packaging machines</td>
</tr>
<tr>
<td>VI. Logistic</td>
<td><em>Supplies</em> of resources needed to execute the tasks required by a duty</td>
<td>Implementation schedules</td>
<td>Execution of duties</td>
<td>Quantities of wrapping paper available</td>
</tr>
</tbody>
</table>
Components of PM: Levels of Vantage Point

I. Philosophical Level
   Levels that set fundamental guide for the organization/group/project

II. Cultural Level

III. Policy Level
   Levels that set internal guides for operation of the organization/group/project

IV. Strategic Level

V. Tactical Level

VI. Logistic Level
   A level that set the logistics of the operation that involves time factor

[PM = Performance Matrix] (Koremura, 2008, 22)
(Koremura, 2008, 27)
## Performance Matrix of Ideal Library in general

<table>
<thead>
<tr>
<th>Stage</th>
<th>A. Models</th>
<th>B. Measures</th>
<th>C. Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Philosophical Level</td>
<td>Ideals</td>
<td>Integrity</td>
<td>Commitment</td>
</tr>
<tr>
<td></td>
<td>Utilization of knowledge</td>
<td>People are not utilizing knowledge</td>
<td>Improvement of knowledge utilization</td>
</tr>
<tr>
<td>II. Cultural Level</td>
<td>Goals</td>
<td>Conformity</td>
<td>Policy</td>
</tr>
<tr>
<td></td>
<td>Stock Knowledge</td>
<td>People are not aware that knowledge is available</td>
<td>Create a library or information center</td>
</tr>
<tr>
<td>III. Policy Level</td>
<td>Missions</td>
<td>Worth</td>
<td>Programs</td>
</tr>
<tr>
<td></td>
<td>Dissemination of information</td>
<td>People are not informed</td>
<td>Procedures</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>To make knowledge accessible</td>
</tr>
<tr>
<td>IV. Strategic Level</td>
<td>Responsibilities</td>
<td>Value</td>
<td>Strategies</td>
</tr>
<tr>
<td></td>
<td>Storage of information</td>
<td>People are not finding the information they need</td>
<td>Theories and procedures to find needed information</td>
</tr>
<tr>
<td></td>
<td>Retrieval of information</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Service of information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V. Tactics Level</td>
<td>Duties</td>
<td>Cost</td>
<td>Tools</td>
</tr>
<tr>
<td></td>
<td>Tasks for Cataloging</td>
<td>There are gaps between exemplary task performance and not exemplary task performance</td>
<td>Create programs to fill the &quot;gap&quot; of incompetent task performances</td>
</tr>
<tr>
<td></td>
<td>Tasks for Reference Service</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tasks for Bibliographic Instructions</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tasks for Administration</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tasks for Collection Development</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VI. Logistic Level</td>
<td>Schedules</td>
<td>Material Needs</td>
<td>Supplies</td>
</tr>
<tr>
<td></td>
<td>Things needed to complete the tasks: schedule and supply</td>
<td>Find a gap between the material needed and the timeline to complete the operation</td>
<td>Create programs of material needs and fill the &quot;gap&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Create programs to adjust &quot;gap&quot; time to complete an operation</td>
</tr>
</tbody>
</table>

(Koremura, 2008, 45)
Interesting Bottom-up approach

• Zhang’s Holistic Model for DL Evaluation

• Criteria drawn from
  – interviews ➔
  – card sorting ➔
  – questionnaire rating importance of criteria

(Zhang, 2010)
Conclusion

Very few previous evaluations exist to examine best practices

Different methodologies in use have relative strengths and weaknesses for giving accurate pictures of specific components

No evaluation tool yet widely operationalized or accepted by community as a “holistic” or “systemic” evaluation

There are a few interesting potential options to explore
References


