9-c: Digital Library Evaluation, User Studies

Draft, Barbara Wildemuth, 10/07/2009

1. **Module name: Digital Library Evaluation, User Studies**

2. **Scope**

   While a number of kinds of evaluation/research studies may be conducted during the design and development of a digital library (e.g., usability testing), this module is concerned with methods for evaluating the outcomes, impacts, or benefits of a digital library, including cost/benefit analyses. It also includes methods that are useful for general user studies (i.e., studies that intend to more fully understand people's interactions with digital libraries). While some methods covered here are useful for usability testing, usability inspections and usability testing are explicitly covered in module 6-d, Interaction Design, Information Summarization and Visualization, and Usability Assessment.

3. **Learning objectives**

   Students will be able to:
   - Understand the importance of DL evaluation;
   - List and describe the strengths and weaknesses of multiple approaches to evaluation; and
   - Apply an appropriate evaluation method to a particular DL.

4. **5S characteristics of the module**

   Streams: N/A
   Structures: N/A
   Spaces: N/A
   Scenarios: Scenarios may form the basis of an evaluation plan, by describing particular situations of use that must be supported effectively by the DL.
   Societies: The concept of societies may be useful in planning an evaluation because it will support the evaluator in more systematically consider the potential stakeholders of the DL.

5. **Level of effort required (in-class and out-of-class time required for students)**
   - In-class time: 2-2½ hours
   - Out-of-class time: 1½ hours for assigned reading
   - Learning activities (optional): See notes on timing with each activity or assignment.

6. Relationships with other modules (flow between modules)

It is expected that this module will follow other modules on digital libraries, and will be in the final portions of the module sequence.

7. Prerequisite knowledge required (completion optional)

Students will *not* be expected to have had prior training in social science research methods.

8. Introductory remedial instruction: None

9. Body of knowledge

Evaluation and user studies:

Definition of evaluation: "An appraisal of the performance or functioning of a system, or part thereof, in relation to some objective(s)" Saracevic, 2000, p.359

Evaluation incorporates the making of value judgments about whether performance is adequate
Its purpose is to inform decision making (Reeves et al., 2003)

Evaluation is <u>critical</u> to any project
NSF recommended that at least 10% of the project budget be devoted to evaluation in their early DL initatives

User studies may be more general, in terms of the types of questions asked
They do not necessarily incorporate the making of value judgments about performance quality
User studies may be more specific, in that they involve users
Evaluations may be conducted on the DL collection or other aspects of the DL without involving users

This module will focus particularly on evaluations that involve users

The object of the evaluation or user study: digital libraries and their processes/functions

A particular aspect of a digital library
Individual digital libraries
Multiple digital libraries

DL processes that may be evaluated (based on Saracevic, 2005), and criteria for
evaluation
Information representations, metadata and surrogates used in the DL
Task appropriateness
Usability
User satisfaction
Particular tools available in the DL
Index, search, and output features
Navigation, browsing
Failures in functionality or usability
User satisfaction
Particular services offered by the DL
Collection quality
Retrieval performance (recall, precision)
Reliability
Human-intermediated services (e.g., reference services)
User satisfaction with individual services or with collection of
services
User behaviors when interacting with the DL (may or may not be
evaluative)
Information seeking/searching behaviors
Use of information retrieved
Work patterns

"All efforts to design, implement, and evaluate digital libraries must be rooted in the
information needs, characteristics, and contexts of the people who will or may use
those libraries." Marchionini, Plaisant, & Komlodi, 2003, p.1

Questions that may be asked during an evaluation/user study

Frame the study questions based on the decisions that must be made about the
DL's functions/processes (Reeves et al., 2003)
Focus on those questions that are most important for making the decisions
that are most important
Focus on impacts of DL functions/services (Marchionini, Plaisant, &
Komlodi, 2003)
What types of impacts are there? On whom?
Who and what influence those impacts?

Formative versus summative evaluation
Formative evaluation focused on decisions about how to modify/change
the DL's functions/services
Summative evaluation focused decisions about the worth or value of the
DL's functions/services

Stages/steps in the evaluation/research process

Develop an evaluation/research plan (Chowdhury & Chowdhury, 2003; Reeves et
al., 2003)
Clarify the decisions to be addressed and the questions they generate
Identify appropriate evaluation methods, including sampling procedures,
data collection procedures and data analysis procedures
Carry out the evaluation/research plan
Report the results to appropriate stakeholders
Primarily the decision makers
Also other constituencies of the DL

Review of an example evaluation, in terms of these steps
Possible example evaluations:
Bishop, A.P. (1998, December). Measuring access, use, and success in
digital libraries. The Journal of Electronic Publishing, 4(2). Retrieved
February 8, 2006, from www.press.umich.edu/jep/04-02/bishop.html.
Marchionini, G. (2000). Evaluating digital libraries: A longitudinal and
multifaceted view. *Library Trends, 49*(2), 304-333.

Evaluation design strategies

Naturalistic studies
For some evaluation studies, it is critical to conduct them in a natural or
naturalistic setting/context
The constraints of the setting usually imply that fewer experimental
controls can be applied to the study design
Usually, the evaluator will need to take into account aspects of the setting
as part of the data collected for the evaluation study

Experiments
Usually conducted in a lab setting, or a setting in which control over the
conditions of the evaluation study can be exerted
The researcher attempts to control all the potential effects on the results of
the study, other than those effects being intentionally manipulated as
the focus of the evaluation

Some important concepts in designing an experiment
Randomization is a key tool for control: random sampling and/or
random assignment to treatment and control groups
Variables: the researcher will manipulate the independent variables
(e.g., whether the DL has a particular feature or not) and will
evaluate the outcomes based on the dependent variable

The design may be a within-subjects design (where each participant interacts with all the variations of the independent variables, and so comparisons on the dependent variable are made "within" each subject's performance) or a between-subjects design (where each participant interacts with only one version of the system and comparisons are made "between" groups of subjects)

Avoiding the effects of researcher bias

It's easy for a researcher's biases to influence the design of a study and, thus, its outcomes
Identify your biases
Ensure that your study design and procedures will allow you to avoid any influence on the study outcomes

Data collection and measurement methods

Collecting data from people requires ethical treatment of those people as study participants
Each institution will require review of the research proposal by an Institutional Review Board that verifies that study participants are being treated ethically

Observation of user behaviors, including transaction logs

To see what the user is doing as he or she interacts with the system

Observation of work (e.g., via contextual inquiry)
Special kind of interview

Observe person while performing task to be supported
Interrupt with questions about how and why, as needed

Think-aloud protocols
During DL use, the participant is asked to verbalize their thought processes
Allows you to observe "unobservable" cognitive behaviors
Usually videotaped or audiotaped

Indirect observation of work
Logging and metering techniques embedded in the software of the current system or intermediate versions

Diaries
> For detailed descriptions of tasks
>> How much time they take
>> Sequential dependencies between tasks

> Allows observation over longer periods of time than contextual inquiry interviews

Interviews and focus groups

> Augmenting other data collection methods, or on their own

> Uses during DL evaluation
>> For identifying problems in the DL design
>> For additional features needed in the DL
>> For other improvements in the DL which the user can suggest

> Individual interviews or group interview (focus groups)
>> Focus groups require a skilled facilitator

Questionnaires

> Surveys: typically one item/question per construct

> Measures: intended to measure constructs that are not directly observable and not easily measured with a single item
>> The more subjective the construct, the more likely that you will need a multiple-item measure for it
>>> Find a measure in the literature, rather than developing your own

> Print vs. online administration
>> Print allows people to annotate (can be good or bad)
>> Online eliminates the need for a separate data entry step

Study sample: Who should be participants in your evaluation study?

Define the population of interest
> Current users or a subset of them
> Potential audiences (who are not current users)

Consider sample size
> Usually a tradeoff between small sample (cheaper) and generalizability

> Intensive versus extensive studies

Intensive studies: to thoroughly understand a phenomenon within
its context
Extensive studies: to understand the extent of a phenomenon
within a population

Develop a sampling plan

Random sampling
Supports statistical inferences to the population

Identify a population to which you want to generalize your
findings
Enumerate the population
Draw a random sample

Problem: enumerating the entire population
Not necessarily problematic, but often is

Other methods of sampling
Quota sampling, purposive sampling, accidental/convenience
sampling

Strive for representativeness
In range, as well as central tendency

Develop a plan for recruiting the sample you want
May need to offer incentives

Analysis and interpretation of data

Reporting the results and interpreting the results are two distinct steps

Interpretation should address the questions, "What do the results mean? How
should they be understood?"

All results must be interpreted in the context of:

Prior empirical work and relevant theoretical frameworks
Situation
What is happening in the particular situation in which the study
was done?

        Weaknesses in the research
                Measurements: level of reliability; validity
                Design: attrition, external events; internal and external threats to
                   validity
                Analysis method: assumptions violated

Recommend particular actions, based on the interpretation of the results

10. Resources

Assigned readings for students:

Nicholson, Scott. (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation, 60*(2), 164-182.

Reeves, Thomas, Apedoe, Xornam, & Hee Woo, Young. (2003). Evaluating digital libraries: A user-friendly guide. University Corporation for Atmospheric Research; National Science Digital Library. Retrieved 3/1/2007 from http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0004.207.
    Chapter 1, Why evaluate? (p.1-6)
    Chapter 2, Evaluating planning (p.7-21)

Recommended background reading for instructor:

Chowdhury, G.G., & Chowdhury, S. (2003). Digital library evaluation. In *Introduction to Digital Libraries.* London: Facet Publishing, 267-283.

Frechtling, J. (2002). *The 2002 User Friendly Handbook for Project Evaluation.* National Science Foundation, Directorate for Education & Human Resources, Division of Research, Evaluation, and Communication.

Marchionini, G., Plaisant, C., & Komlodi, A. (2003). The people in digital libraries: Multifaceted approaches to assessing needs and impact. In Bishop, A.P., Van House, N.A., & Buttenfield, B.P. (eds.), *Digital Library Use: Social Practice in Design and Evaluation.* Cambridge, MA: MIT Press, 119-160.

Reeves, Thomas, Apedoe, Xornam, & Hee Woo, Young. (2003). Evaluating digital libraries: A user-friendly guide. University Corporation for Atmospheric Research; National Science Digital Library. Retrieved 3/1/2007 from http://www.dpc.ucar.edu/projects/evalbook/EvaluatingDigitalLibraries.pdf.

Saracevic, T. (2000). Digital library evaluation: Toward evolution of concepts. *Library Trends, 49*(2), 350-369.

Additional potential readings include:

Bishop, A.P. (1998, December). Measuring access, use, and success in digital libraries. The Journal of Electronic Publishing, 4(2). Retrieved 2/8/2006 from www.press.umich.edu/jep/04-02/bishop.html.

Bishop, A.P., Mehra, B., Bazzell, I., & Smith, C. (2003). Participatory action research and digital libraries: Reframing evaluation. In Bishop, A.P., Van House, N.A., & Buttenfield, B.P. (eds.), *Digital Library Use: Social Practice in Design and Evaluation.* Cambridge, MA: MIT Press, 161-189.

Bollen, J. and R. Luce. (2002). Evaluation of digital library impact and user communities by analysis of usage patterns. *D-Lib Magazine, 8*(6) June 2002. Retrieved 3/1/2007 from http://www.dlib.org/dlib/june02/bollen/06bollen.html

Bryan-Kinns, Nick & Blandford, Ann. (2000). A survey of user studies for digital libraries. RIDL Working Paper. Retrieved 3/1/2007 from http://www.cs.mdx.ac.uk/ridl/DLuser.pdf.

Choudhury, G.S.; Hobbs, B.; M Lorie, Flores, N.E. (2002, July/August). A framework for evaluating digital library service. *D-Lib Magazine, 8*(7/8). Retrieved 3/1/2007 from http://www.dlib.org/dlib/july02/choudhury/07choudhury.html.

Marchionini, G. (2000). Evaluating digital libraries: A longitudinal and multifaceted view. *Library Trends, 49*(2), 304-333.

Rieger, R., & Gay, G. (1999, June 15). Tools and Techniques in Evaluating Digital Imaging Projects. *RLG DigiNews*.

Saracevic, Tefko (2005). How were digital libraries evaluated? Presented at Libraries in the Digital Age (LIDA), Dubrovnik and Mljet, Crotia, May 30-June 3. Retrieved 3/1/2007 from http://www.scils.rutgers.edu/~tefko/DL_evaluation_LIDA.pdf.

Thong, J. (2002). Understanding user acceptance of digital libraries: What are the roles of interface characteristics, organizational context, and individual differences? *International Journal of Human-Computer Studies, 57*(3), 215-242.

11. Concept Map (created by students)

12. Exercises / Learning activities

a. *Analyze a DL evaluation report*

Exercise 13.a, "Analyze a DL evaluation report," could be adapted to an in-class small-group discussion exercise. If so, the results of the each group's analysis could be reported orally or could be posted to a class wiki or discussion forum.

If used as an in-class exercise, assign the groups to read a particular evaluation report before class, to prepare for their in-class discussion and report.

Time requirements: 2 hours of preparation outside of class; 25-30 minutes for discussion in class; 20-30 minutes for report presentation in class, depending on the number of groups.

b. *Develop an evaluation plan*

Based on a DL that is familiar to all the students in the class (e.g., flickr, MySpace, a music collection, the university's OPAC or a special collection that is well-known), have students work in small teams (3-4 people each) to develop a draft evaluation plan. They can play the role of an evaluation consulting firm, designing an evaluation study for their client, the DL managers.

Each plan should include the following:

The evaluation questions to be addressed
     Stated briefly, in one sentence (preferably ending with ?)
The sample to be included
     How they will be selected
     How they will be recruited
The methods for data collection
     The types of data to be collected, and how each pertains to the evaluation question
     The procedures for collecting the needed data

Have each team of students present their plan to the class, as if it were an initial presentation to the client (the DL managers).

Time requirements: It is expected that the students will prepare their report after this module has been presented in class. Students should expect to spend 4-5 hours outside of class, preparing their reports. Each report should be presented in 7-10 minutes, during the next class session.

c. *Interview a digital librarian about evaluation*

Note: This exercise is only possible if there are a number of robust local digital library projects, and the students will have access to their directors/administrators.

For this exercise, students should work in pairs; each pair will be assigned to investigate a particular digital library. Prior to the interview, each pair should read the available documentation on the digital library on which they're focused. Using the following interview guide, they should interview the director/administrator of the digital library.

Interview guide:
When was the digital library first established?
What are the primary goals of the DL?

In what ways do you evaluate whether you're achieving those goals?
Do you evaluate any other aspects of the DL's operations? If so, how?
How and to whom are the evaluation results reported?

Each pair should write up a brief (1-3 page) summary of their interview findings. In addition, they should be prepared to orally report on the most interesting aspects of those findings at the next class session.

Time requirements, outside of class: 1-2 hours for preparatory reading; 1 hour for conducting the interview; 2-3 hours for writing up the interview report.
Time requirements, in class: 30-40 minutes for the class to discuss the findings from the interviews.

13. Evaluation of learning objective achievement

a. *Analyze a digital library evaluation report*

Using Saracevic's (2005) meta-analysis of digital library (DL) evaluations as a framework, evaluate an additional DL evaluation report. The report can be selected from the following[1]:

> Byrd, S., et al. (2001). Cost/benefit analysis for digital library projects: The Virginia Historical Inventory project (VHI). *The Bottom Line: Managing Library Finances, 14*(2), 65-75.
> Gambles, A. (2001). The HeadLine personal information environment: Evaluation Phase One. *D-Lib Magazine, 7*(3).
> www.dlib.org/dlib/march01/gambles/03gambles.html.
> Palmer, D., & Robinson, B. (2001). Agora: The hybrid library from a user's perspective. *Ariadne, 26.* www.ariadne.ac.uk/issue26/case-studies/intro.htm.
> Zhang, Y., Lee, K., & You, B.-J. (2001). Usage patterns of an electronic theses and dissertations system. *Online Information Review, 25*(6), 370-377.
> List of possibilities; still need to be viewed.

Analyze the evaluation report in terms of the following aspects:
- "**Construct** for evaluation.
  What was evaluated? What was actually meant by a "digital library"? What elements (components, parts, processes…) were involved in evaluation?
- **Context** of evaluation - selection of a goal, framework, viewpoint or level(s) of evaluation.
  What was the basic approach or perspective? What was the level of evaluation? What was the objective(s)?
- **Criteria** reflecting performance as related to selected objectives.
  What parameters of performance were concentrate[d] on? What dimension or characteristic [was] evaluated?

---

[1] Additional reports can be added to this list or reports can be deleted, at the instructor's discretion.

- **Methodology** for doing evaluation.
  What measures and measuring instruments were used? What samples? What procedures were used for data collection? For data analysis?
- **Findings** from evaluation studies
  Only a single generalization is provided." (Saracevic, 2005, p.2-3)

Prepare a report (2-5 pages, single-spaced) summarizing the findings of your analysis.

The report should be evaluated in terms of its demonstration that the authors understood the DL evaluation conducted, its coverage of the five aspects of evaluations posed by Saracevic, its identification of strengths and weaknesses in the DL evaluation, and its clarity (organization, grammar, etc.).

Time requirements: approximately 6-8 hours outside of class.

b. *Develop an evaluation plan*

Class exercise 10.b, "Develop an evaluation plan," could be adapted as a graded assignment. Each team would be expected to develop their plan over the week after the class's discussion of evaluation. If class time is available, the final plan can be presented orally; or, if preferred, the evaluation plans could be turned in as an evaluation proposal (2-4 pages).

The evaluation plans would be evaluated in terms of their completeness (were all the major components of an evaluation study addressed?), their feasibility (could the evaluation study actually be conducted, given reasonable resources?), and their clarity.

Time requirements: 6-8 hours outside of class, preparing and writing the evaluation plan.

14. Glossary

Between-subjects design: A research design in which "each research participant receives only one level of the independent variable" (Schmidt, 2000).

Dependent variable: "A variable that may, it is believed, be predicted by or caused by one or more other variables called independent variables." (U.S. Dept. of Justice, n.d.)

Evaluation: "An appraisal of the performance or functioning of a system, or part thereof, in relation to some objective(s)" Saracevic, 2000, p.359

Formative evaluation: An evaluation that is intended to "strengthen or improve the object being evaluated. Formative evaluations are used to improve [information systems] while they are still under development." (Trochim, 2001, p.347)

Independent variable: "A variable that may, it is believed, predict or cause fluctuation in an dependent variable." (U.S. Dept. of Justice, n.d.)

Research design: "A plan of what data to gather, from whom, how and when to collect the data, and how to analyze the data obtained." (U.S. Dept. of Justice, n.d.)

Sample: "The actual units you select to participate in your study." (Trochim, 2001, p.351)

Stakeholders: "People who have a vested interest in the success of the project or are involved in the implementation of the project." (California State University, Monterey Bay, n.d.)

Summative evaluation: An evaluation that "examine[s] the effects or outcomes of [an information system]." (Trochim, 2001, p.352)

Within-subjects design: A research design in which "each research participant provides data for all the levels of the independent variable" (Schmidt, 2000).

References for glossary:

California State University, Monterey Bay, Information Technology. (n.d.) Data warehouse glossary. Retrieved 5/23/2007 from http://it.csumb.edu/departments/data/glossary.html.

Saracevic, T. (2000). Digital library evaluation: Toward evolution of concepts. *Library Trends, 49*(2), 350-369.

Schmidt, S.R. (2000). Research methods for the digitally inclined. Retrieved 11/16/2007, from http://www.mtsu.edu/~sschmidt/methods/design/design.html.

Trochim, W.M.K. (2001). *The Research Methods Knowledge Base.* Second ed. Cincinnati, OH: Atomic Dog Publishing.

U.S. Dept. of Justice, Office of Justice Programs, Bureau of Justice Assistance, Center for Program Evaluation. (n.d.). Glossary. Retrieved 5/23/2007 from http://www.ojp.usdoj.gov/BJA/evaluation/glossary/glossary_r.htm.

15. Additional Useful links

None

16. Contributors

Barbara Wildemuth (UNC, developer), Jeff Pomerantz (UNC, evaluator), Sanghee Oh (UNC, evaluator), Mike Christel (CMU, evaluator), Flora McMartin (Broad-based Knowledge, evaluator), Liz Liddy (Syracuse, evaluator)