

Criteria and methods in evaluation of digital libraries: use & usability

Tefko Saracevic, Ph.D.



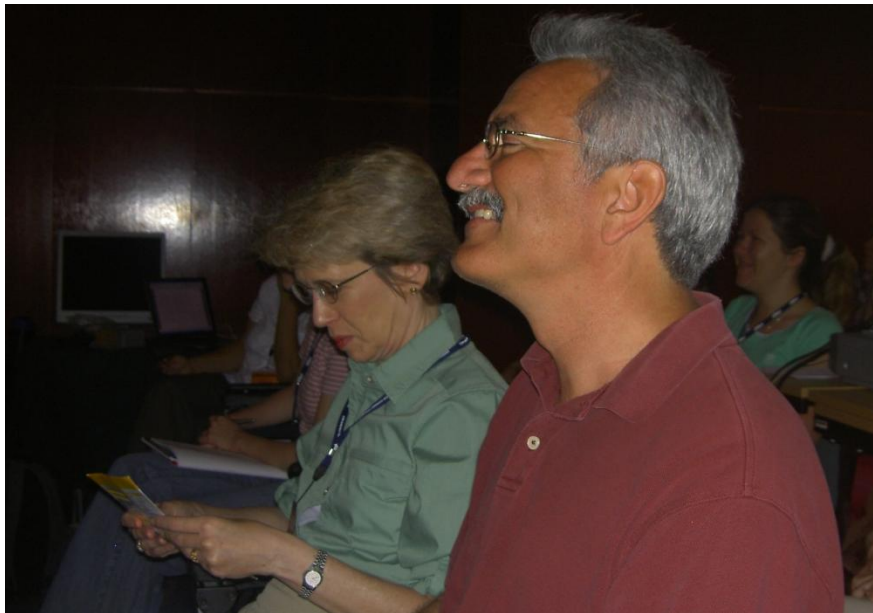
This work is licensed under a

[Creative Commons Attribution-NonCommercial-ShareAlike 3.0 United States License](https://creativecommons.org/licenses/by-nc-sa/3.0/)



“Evaluating digital libraries is a bit like judging how successful is a marriage”

(Gary Marchionini, 2000)



([Gary](#) with [Chris Borgman](#) – taken at the conference [Libraries in the Digital Age](#) (LIDA) held every two years in Zadar, Croatia)

ToC

- introductory musings
- on the scene: several perspectives
- what is needed for evaluation?
- criteria for DL evaluation
- methodologies for DL evaluation
- the versus hypothesis
- toward conclusions

evaluation: definition

Dictionary:

assessment of value

the act of considering or examining something in order to judge its value, quality, importance, extent, or condition

in systems:

assessment of **performance** in terms of effectiveness and/or efficiency

- **effectiveness**: how well did a system (or part thereof) do that for which it was designed – related to objectives
- **efficiency**: at what cost - \$\$\$\$, effort, time

evaluation ...

in digital libraries:

assessment of performance (effectiveness, efficiency)
on basis of given criteria

- performance could be related to usability
- criteria may be specified by users or derived from professional practice, other sources or standards
- at issue:
 - what criteria to use?
 - what methods to employ?

digital libraries

- since emergence in early/mid 1990's
 - many institutions & fields got involved
 - great many practical developments
 - many research efforts & programs globally
 - large & growing expenditures in practice
 - applications & use growing exponentially
- everything about digital libraries is explosive
- **except evaluation**
 - relatively small, even neglected area

literature reports on DL evaluation

- two distinct types:
 - meta or “about” literature
 - suggest approaches, models, concepts, questions;
 - useful for establishing a framework, guide in work
 - (e.g. [Evaluating digital libraries: A user friendly guide](#))
 - object or “on” literature
 - actual evaluations, contains data
 - (e.g. [How Usable are Operational Digital Libraries – A Usability Evaluation of System Interactions](#) (done at Rutgers))
- but we are concentrating here on object literature only

boundaries of DL evaluation

- difficult to establish, apply
 - particularly as to process – e.g.
 - crossing into information retrieval (IR): where does IR evaluation stop & DL evaluation start?
 - or any technology evaluation?
 - or evaluation of web resources and portals?
 - is every usability study evaluation as well?
- brings up the perennial issues:
 - what is a digital library? what are all the processes that fall under DL umbrella?

on the scene

- as we discussed already

- several different communities involved in digital libraries, each with quite different
 - perspective, concepts, meanings in dealing with DL
 - concentration, emphasis, approach, models
 - thus, different perspective in evaluation
- many disciplines, institutions involved
 - bringing different perspectives to evaluation

computer science perspectives: emphasis in evaluation

- concentrates on research & development (R&D)
- **technology** centered
 - distributed & organized **knowledge resources** in digital formats
 - how to collect, store, organize, diverse types of information - texts, images, sounds, multimedia ...
 - new kind of **distributed database services to manage unstructured multimedia resources**
- **and they want to evaluate those aspects**

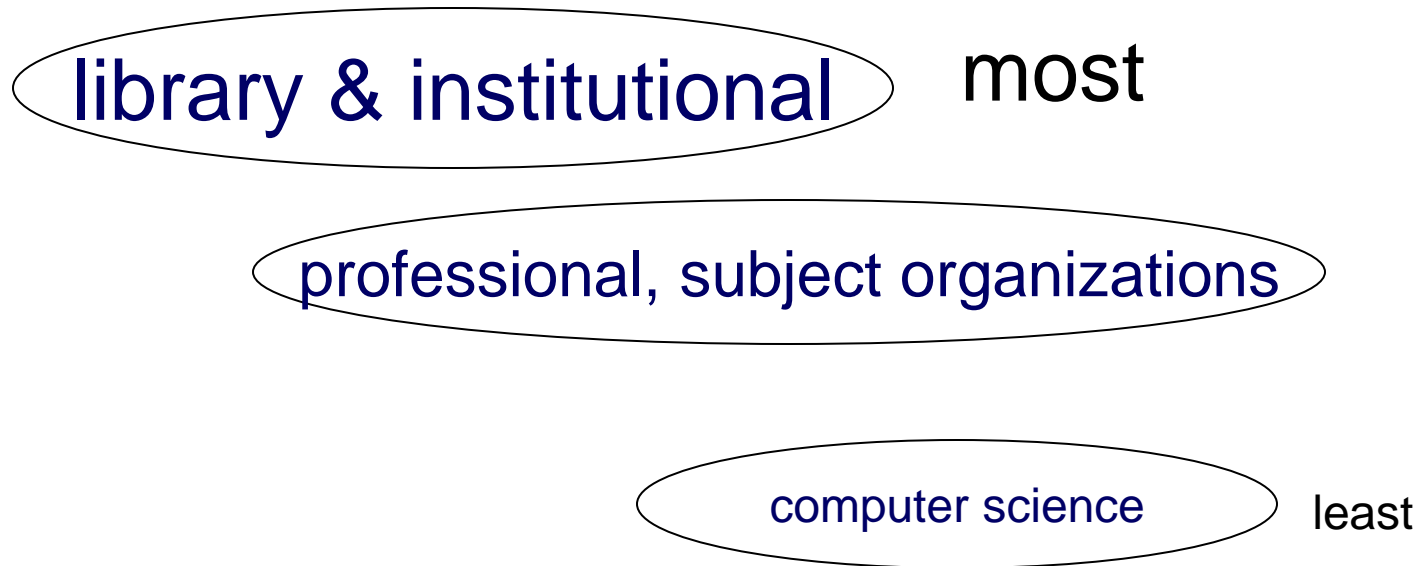
library & institutional perspective: emphasis in evaluation

- concentrates on institutions, service, **practice**
 - logical extension of libraries
- **content, collection, service** centered
 - **creation** of digital collections
 - **access** to & **use** of collections
 - **services** provided
- guided by service **mission**
 - various environments, user communities
 - various degrees of integration or separation
- and they want to evaluate that

organizational, subject perspective: emphasis in evaluation

- variety of organizations involved
 - scientific & technical societies
 - various fields, academic units
 - projects - institutions, consortia
 - museums, historical societies
 - government agencies
- concentrate on **collections & their uses in specific areas, subjects**
 - new forms of publishing in their area
- services to communities or perceived needs
- and they want to evaluate that

amount of evaluation in different communities



what is needed to evaluate performance?

1. **construct** - system, process, part to be evaluated
2. **objectives** - reasons, desires for evaluation
3. **criteria** - standards, base for reflecting objectives
4. **measure** - units for recording & comparing criteria
5. **measuring instruments** - devices, tools that record a measure
6. **methodology** - way of doing evaluation
 - assembling, applying, analyzing

examples

Element	Athletic event	Digital library
Construct	10 km race	a given DL or part e.g. searching as in inf. retrieval
Objective	winner?	effectiveness – how well did it perform?
Criteria	speed - time	relevance
Measure	minutes, seconds	precision, recall
Instrument	stopwatch	people, judges
Method	timing from start to finish	borrowed from Text REtrieval Conference (TREC) laboratory

criteria in DL evaluation

- reflect performance of a digital library (or part) as related to selected objectives
 - in studies: *what parameters of performance were concentrated on?*
- in digital library evaluation: no basic or standardized criteria, no overall agreement
 - many have been used
 - even for the same objectives

usability

- International Standards Organization - [ISO 9241-11 \(1998\)](#)
“Extent to which a user can achieve goals with effectiveness, efficiency and satisfaction in context of use”
- Jacob Nielsen (usability guru) [definition](#):
“Usability is a **quality attribute** that assesses how easy user interfaces are to use. The word "usability" also refers to methods for improving ease-of-use during the design process.”

usability in digital libraries

- widely used, but no uniform definition for DL
- general, meta criterion, covers a lot of ground
- umbrella for many specific criteria used in DL evaluations

six classes of criteria for DL evaluation derived from literature

- **content**
 - how well are digital collections selected, developed; objects created, organized, represented, presented?
- **technology**
 - how well do hardware & software support library functions?
- **interface**
 - what is available for users to interact & how much is interaction supported or hindered?

classes of criteria (cont.)

- **process/service**
 - what processes & assistance are provided; what range of services is available; how well are they functioning; (carrying out tasks as: *search, browse, navigate, find, evaluate or obtain a resource*)?
- **user**
 - what are the outcomes of DL use – changes in human information behavior, cognitive state, decision-making, problem-solving; impact on accomplishing tasks; broader impact/benefit in research, professional work?
- **context**
 - how well does a DL fit into, respond to, follow larger context – institutional, economic, legal, social, cultural; effects on context?

sample of criteria from literature

Content	Technology	Interface
completeness, size coverage, overlap quality, accuracy validity, authority adequacy, diversity informativeness freshness accessibility, availability complexity-organizational structure transparency, clarity effort to understand ...	response time processing time speed capacity load accessibility effectiveness efficiency compatibility quality reliability robustness...	attractiveness consistency representation of concepts - labels communicativeness of messages display, attractiveness appropriateness consistency ease of use effort error detection, personalization ...

sample ... (cont.)

Process/ Service	User	Context
<p>learnability, effort/time, support, convenience ease of use lostness (confusion) completion (achievement of task) interpretation difficulty sureness in results error rate responsiveness reliability,...</p>	<p>satisfaction, success relevance, usefulness of results impact, value quality of experience barriers, irritability preferences learning effect productivity use/reuse,...</p>	<p>institutional fit, usefulness productivity of & impact on community members sustainability interoperability rights management, copyright abidance organizational usability, ...</p>

methodologies

- digital libraries are complex entities
 - many methods appropriate
 - each has strengths, weaknesses
- range of methods used is wide
 - there is no “best” method
 - but, no agreement or standardization on any methods
- makes generalizations difficult, even impossible

methodologies used

- surveys (most prevalent)
- interviews
- observations
- think aloud
- focus groups
- task performance
- log analysis
- usage analysis
- record analysis
- experiments
- economic analysis
- case study
- ethnographic analysis

general results from all evaluation studies

- not synthesized here
- hard to synthesize anyhow
- generalizations are hard to come by
- except one!

users and digital libraries

- a number of studies reported various versions of the same result:

users have many difficulties with DLs

- usually do not fully understand them
- they hold different conception of a DL from operators or designers
- they lack familiarity with the range of capabilities, content and interactions
- they often engage in blind alley interactions

a nice quote from an evaluation study

“It’s like being given a Rolls Royce and only knowing how to sound the horn”

quote from a surgeon in study of digital libraries in a clinical setting over a decade ago but it still holds for many users (Adams & Blanford, 2001)

analogy

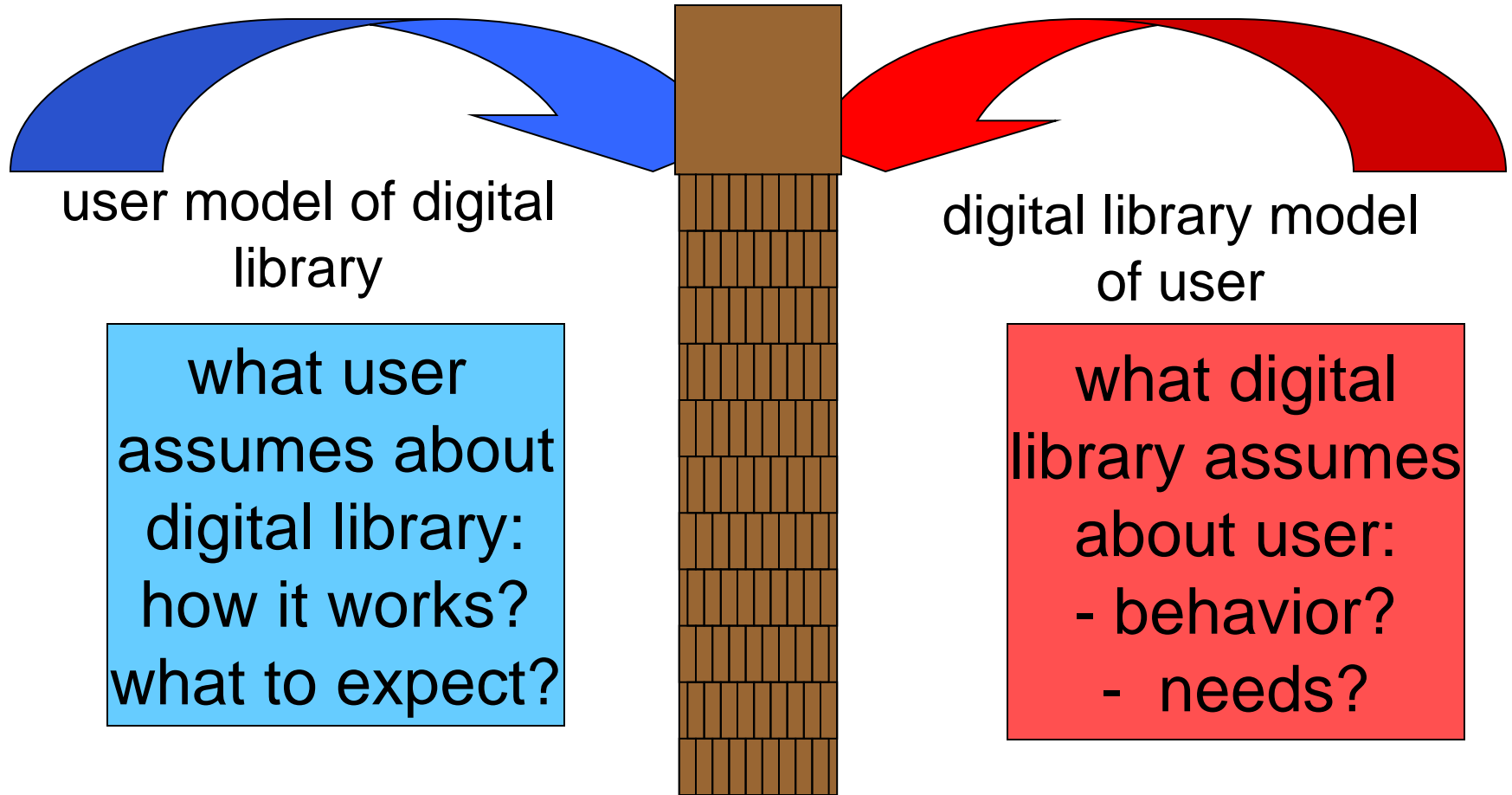
- perceptions of users and perceptions of designers and operators of a DL are generally not very close
- users are from Venus and digital libraries are from Mars (or is it vice versa?)
- leads to the **versus** hypothesis

is it:

user AND digital library
or
user VERSUS digital library

- why VERSUS?
 - users and digital libraries see each other differently

how close are they?
user VERSUS digital library model



the versus hypothesis

in use, more often than not, digital library users and digital libraries are in an adversarial position

- hypothesis does not apportion blame
 - does not say that digital libraries are poorly designed
 - or that users are poorly prepared
- adversarial relation may be a natural order of things

toward conclusions: evaluation of digital libraries

- impossible? not really
- hard? very
- could not generalize yet
- no theories
- no general models embraced yet, although quite a few proposed
- in comparison to total works on DL, only a fraction devoted to evaluation

why? – some speculations

- **complexity:** DLs are highly complex
 - more than technological systems alone
 - evaluation of complex systems is very hard
 - just learning how to do this job
 - experimenting with doing it in many different ways
- **premature:** it may be too early in the evolution of DL for evaluation on a more organized scale

why? (cont.)

- **interest:** there is no interest in evaluation
 - R&D interested in doing, building, implementing, breaking new paths, operating ...
 - evaluation of little or no interest, plus there is no time to do it, no payoff
- **funding:** inadequate or no funds for evaluation
 - evaluation time consuming, expensive requires commitment
 - grants have minimal or no funds for evaluation
 - granting agencies not allocating programs for evaluation
 - no funds = no evaluation.

why? (cont.)

- **culture:** evaluation not a part of research and operations of DL
 - below the cultural radar; a stepchild
 - communities with very different cultures involved
 - language, frames of reference, priorities, understandings differ
 - communication is hard, at times impossible
 - evaluation means very different things to different constituencies

why – the end

- **cynical:** who wants to know or demonstrate actual performance?
 - emperor clothes around?
 - evaluation may be subconsciously or consciously suppressed
 - dangerous?

ultimate evaluation

- the ultimate evaluation of digital libraries:
 - assessing transformation in their context, environment – how did digital libraries affect them?
 - determining possible enhancing changes in institutions, learning, scholarly publishing, disciplines, small worlds ...
 - and ultimately determining effects in society due to digital libraries

final conclusion finally

- after all those years evaluation of digital libraries still in formative years
- not funded much, if at all
- but necessary for understanding how to
 - build better digital libraries & services &
 - enhance their role

evaluation perspective – Rockwell



still another one ...

Wine Evaluation

1.) color



ring - clarity, brilliance
high acid? region?

2.) legs - very little → low alcohol
significant → 12.5 - 13%

3.) nose - A.) Aroma (grape smell)
B.) Bouquet (2nd ageing)

- pepper
- blackberry
- oak
- truffles
- honey (sauvignes)
- apple (malic acid)
(young wine)

4.) taste A.) acidity
B.) Body



