

Digital libraries: Challenges for evaluation



Tefko Saracevic

Lisa Covi

Rutgers University

tefko, covi@scils.rutgers.edu

Evaluation: what is?

- Questions about **performance**
 - testing, validating, comparing, appraising
- Many approaches & types - making a choice
- In systems approach:
 - **Effectiveness**: how well does a system, or part, perform that for which it was designed?
 - **Efficiency**: at what cost? \$\$\$, time, effort
- Gains insight into behavior & organization
- Always there, willing or not

State of evaluation of digital libraries

- Many projects, some talk & discussion
 - but no evaluation to speak of
- Not high on anybody's agenda
- Related work on metrics proceeding
 - D-Lib Working Group on Digital Library Metrics (an informal, non-funded group)
 - Progress to date: a number of internal discussion papers; overall definitions proposed
 - some criteria & scenarios suggested

In research

■ Dlib Initiative 1 (1995-1998)

- | six projects
- | evaluation talked about around 1995-6, but only some evaluation performed in projects
- | project results as a whole not evaluated
 - what did they actually accomplish? ???

■ Dlib Initiative 2 (1999-)

- | 21 projects + 3 in undergrad education
- | 6 (of 21) mention some evaluation, but no details at all. Evaluation not a even a minor component
- | undergrad projects: one evaluation

Research ... lingering questions

- What, if anything, is meant by evaluation in DLI projects? In dlib research in general?
- Is evaluation considered necessary at all?
 - | Why is no attention paid to evaluation?
 - | Is just something that computes enough for evaluation? Or anecdotes about reactions?
 - | Is this a new kind of science? Or development?
- What of public, overall evaluation?
 - | What of refereed publications? Where are they?

In practice

- Many dlibs built and operating
 - not one evaluated, but improvements made
- Publishers built dlibs
 - e.g Elsevier had use and economic evaluation
- Professional societies have dlibs
 - no evaluation, but improvements made
- Evaluation approaches:
 - internal discussion, observation, experience, copying
 - improvements, redesigns follow

Needed and lacking



■ Overall conceptual framework

■ **Construct** - objects, elements - to be evaluated

- | What is actually meant by a digital library? What is encompassed? What elements to take? What is critical?

■ Evaluation approach

■ **Context** - level - of evaluation

- | What is “evaluation” in dlib context? What approach to use? On what to concentrate?

Needed ... more



■ Criteria for evaluation

- | What to evaluate in that context? What to reflect?
What parameters, metrics to select for evaluation?

■ Measures

- | What measures to apply to various criteria? What metrics can be translated into measures?

■ Methods

- | How to evaluate? What procedures to use?

Required



- These are essential requirements for any evaluation
 - | construct, context, criteria, measures, method
- No specification on each - no evaluation
- Here we talk about first three

Construct:

What is meant by a dlib?

- Two conceptualizations stressing:
 1. distributed objects in various forms, distributed access, representation, operability (computer science)
 2. institution, collection, services, availability (libraries)
- First is research perspective
 - | focus on a range of research problems, with little or no operations; “dlib” very broadly interpreted
- Second is library operational perspective
 - | focus on practical problems of transforming library institutions and services, with little or no research; “dlib” very specifically interpreted

Research perspective

"Digital libraries are **organized collections of digital information**. They combine the **structuring and gathering of information**, which libraries and archives have always done, with the **digital representation** that computers have made possible."

Lesk, 1997

(evaluation constructs or elements are in **bold**)

Library conception

“Digital libraries are organizations that provide the **resources**, including the specialized **staff**, to **select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of**, and ensure the **persistence over time** of collections of **digital works** so that they are readily and economically **available for use** by a defined community or set of communities.”

Digital Libraries Federation (DLF)

Constructs/elements for evaluation

- Digital collection(s), resources
 - Selection, gathering
 - Distribution, connections
 - Organization, structure (physical & intellectual)
 - Representation, interpretation
- Access
 - Intellectual, physical
 - Distribution
 - Interfaces

constructs ... more



- Services
 - Availability
 - Dissemination, delivery
- Preservation, persistence
- Security, privacy, policy, legality
- Users, use, communities
- Management, economics
- Integration

Context - general

- Any evaluation is a tuple t
 - | between a selected element to be evaluated and a selected type of its performance
- Leads to selection of a level of evaluation
 - | What to concentrate on? What level of performance?
- Use-centered & system-centered levels
- Dlib performance can be viewed from a number of standpoints or levels
 - | What are they?

Context - use-centered levels



■ Social:

- How well does a dlib support inf. demands, needs & roles of society, community?
 - | hardest to evaluate

■ Institutional:

- How well does a dlib support institutional, organizational mission & objectives? How well does it integrate with other resources?
 - | tied to objectives of institution, organization
 - | also hard to evaluate

use levels ... more



■ Individual:

- How well does a dlib support inf. needs & activities of people?
 - | most evaluations of many systems in this context
 - | use of various aspects, contents, features by users
 - | task performance

Context - system-centered levels



■ Interface

- How well does a given interface provide access?

■ Engineering

- How well does hardware, networks, configurations perform?

system levels ... more



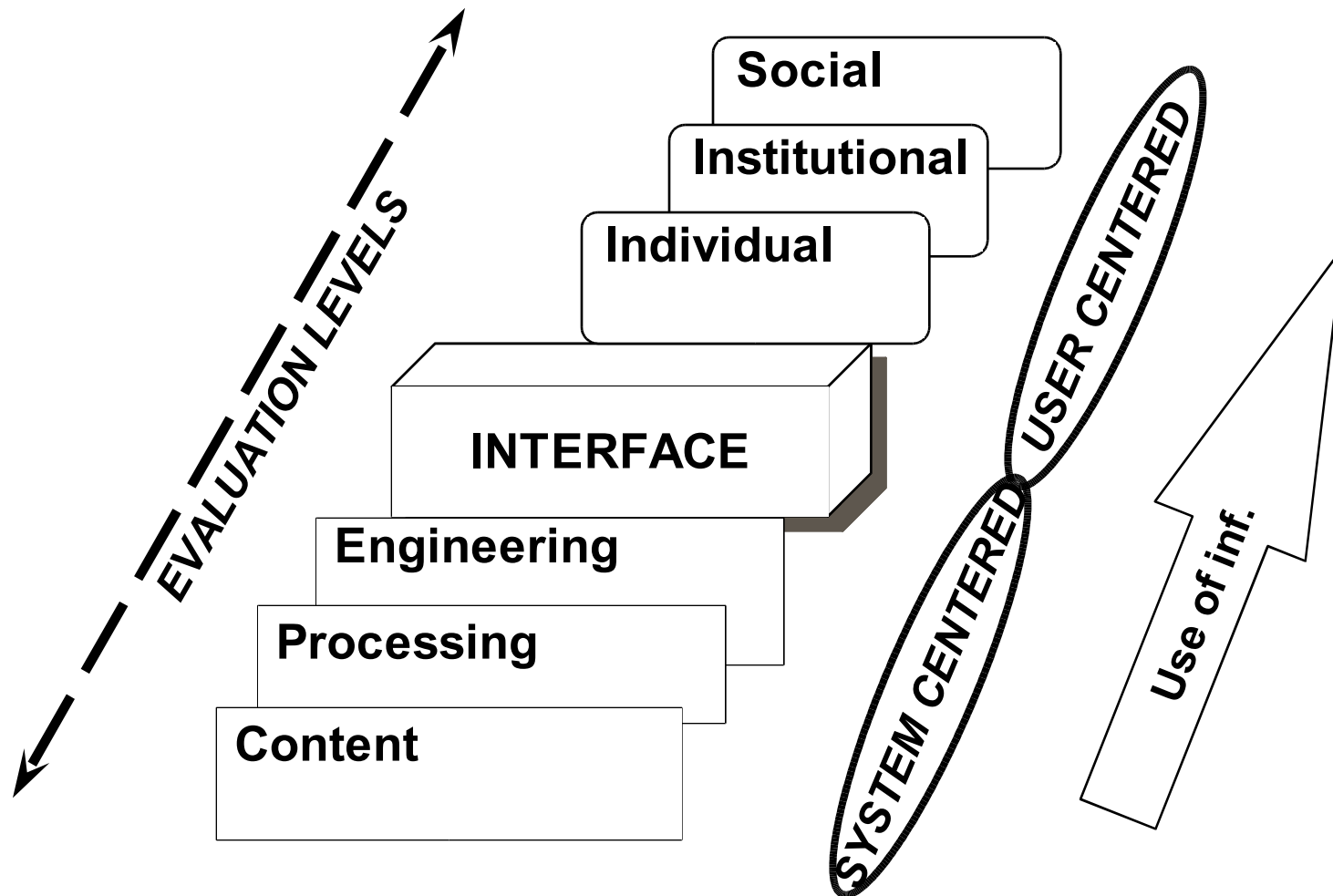
■ Processing:

- How well do procedures, techniques, operations, algorithms ... work?

■ Content

- How well is the collection selected, organized, structured, represented?

Levels of evaluation



Criteria



- For each level criteria have to be determined
- Traditional library criteria:
 - collection
 - | purpose, scope, authority, coverage, currency, audience, cost, format, treatment, preservation ...
 - information
 - | accuracy, appropriateness, links, representation, uniqueness, comparability, presentation ...
 - use
 - | accessibility, availability, searchability, usability ...

criteria ... more



- Traditional human-computer interaction criteria:
 - usability, functionality, effort level
 - screen, terminology & system feedback, learning factors, system capabilities
 - task appropriateness; failure analysis
- Traditional retrieval criteria:
 - relevance: precision, recall measures
 - satisfaction, success, overall value

criteria ... more



■ Value study criteria - value-in-use

■ values users assign to dlib use

- | assessment by users on qualities of interaction with a dlib service & worth or benefits of results of interaction with the dlib as related to reasons for using it

- | multidimensional - composite of

1. Reasons for use
2. Interaction with a dlib service
3. Results or impacts of use

Adaptation



- Traditional criteria have to be adopted to dlibs & expanded
 - | to include unique characteristics of dlibs
- Criteria for research results evaluation have to include some of these, plus:
 - traditional measures of research & design evaluation from systems approach & computer science,
 - and science in general - peer evaluation

Conclusions



- Investment in dlibs very high & growing
- So far investment in evaluation very small
 - How do we know what is accomplished?
 - What works, what does not?
 - What mistakes, practices not to repeat?
- Evaluation of dlibs very complex
 - Needs own methodological investigation
 - Metrics work very important. Funding?

conclusions ... more



- **Critical questions, not yet raised:**
 - **How can dlib efforts proceed without evaluation?**
 - **What are the consequences?**

Thank you
Hvala

Danke

Merci

Gracias

Grazie

