

# exploring perspectives on the evaluation of digital libraries

Giannis Tsakonas & Christos Papatheodorou

{gtsak, papatheodor}@ionio.gr

Database and Information Systems Group,  
Laboratory on Digital Libraries and Electronic  
Publishing, Department of Archives and Library  
Sciences, Ionian University,  
Corfu, Greece.

Digital Curation Unit,  
Institute for the Management of Information Systems,  
'Athena' Research Centre,  
Athens, Greece



introduction

a few things about this tutorial



# what are the benefits?

- Viewing things in a coherent way.
- Familiarizing with the idea of the interdisciplinarity of the domain and the need of collaboration with other agents to perform better evaluation activities.
- Forming the basis to judge potential avenues in your evaluation planning.

# how it is structured?

- Part one: fundamentals of DL evaluation (1,5 hour)
  - reasons to evaluate
  - what to evaluate
  - agents of the evaluation
  - methods to evaluate
  - outcomes to expect
- Break (0,5 hour)
- Part two: formal description and hands on session
  - a formal description of DL evaluation (0,5 hour)
  - hands on session (0,5 hour)
- Discussion and closing (0,5 hour)

part one

# fundamentals of digital library evaluation

# evaluation tutorials in DL conferences

Evaluating, using, and publishing eBooks	2001	ECDL JCDL
Evaluating Digital Libraries for Usability	2002	ICADL
Usability Evaluation of Digital Libraries Usability Evaluation of Digital Libraries	2003	
Evaluating Digital Libraries Qualitative User-Centered Design in Digital Library Evaluation	2004	
Evaluating Digital Libraries Evaluation of Digital Libraries	2005	
	2006	
	2007	
	2008	
	2009	
Lightweight User Studies for Digital Libraries	2010	

# what is evaluation?

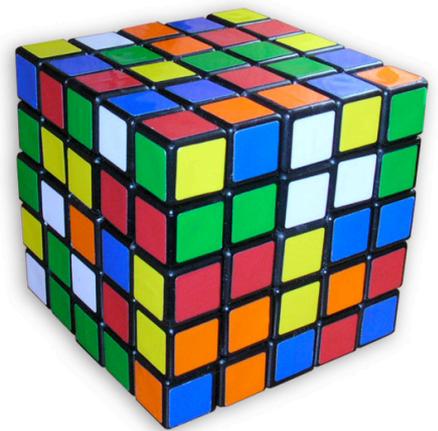
- Evaluation is the process of assessing the value of a definite product or an operation for the benefit of an organization at a given timeframe.

# how to start evaluating?

- From objects?
  - digital libraries are complex systems
  - depending on the application field their synthesis increases
- ...or from processes?
  - evaluation processes vary because of the many agents around them
  - many disciplines, each one having many backgrounds
- This dualism will escort us all the way.

# metaphor: a usual Rubik's cube

- Roles: content managers (librarians, archivists, etc.)  
computer scientists, managers, etc.
- Research fields: information science, social sciences, human-computer interaction, information retrieval, visualization, data-mining, user modeling, knowledge management, cognitive psychology, information architecture and interaction design, etc.





# evaluation framework

- Why: the first, but also the most difficult, question.
- What: the second question, with a rather obvious answer.
- Who: the third question, with a more obvious answer.
- How: the fourth question; a quite puzzling one.

questions

why

# reasons to evaluate [a]

- We evaluate in order to improve our systems; a generic beneficiary aim.
- We evaluate in order to increase our knowledge capital on the value of our system:
  - to describe our current state (recording our actions)
  - to justify our actions (auditing our actions)
  - to redesign our system (revising our actions)
- We evaluate to make other understand what is the value of our system:
  - to describe how others use our system
  - to justify why they are using it this way (or why they don't use it)
  - to redesign the system and its services to be better and more used

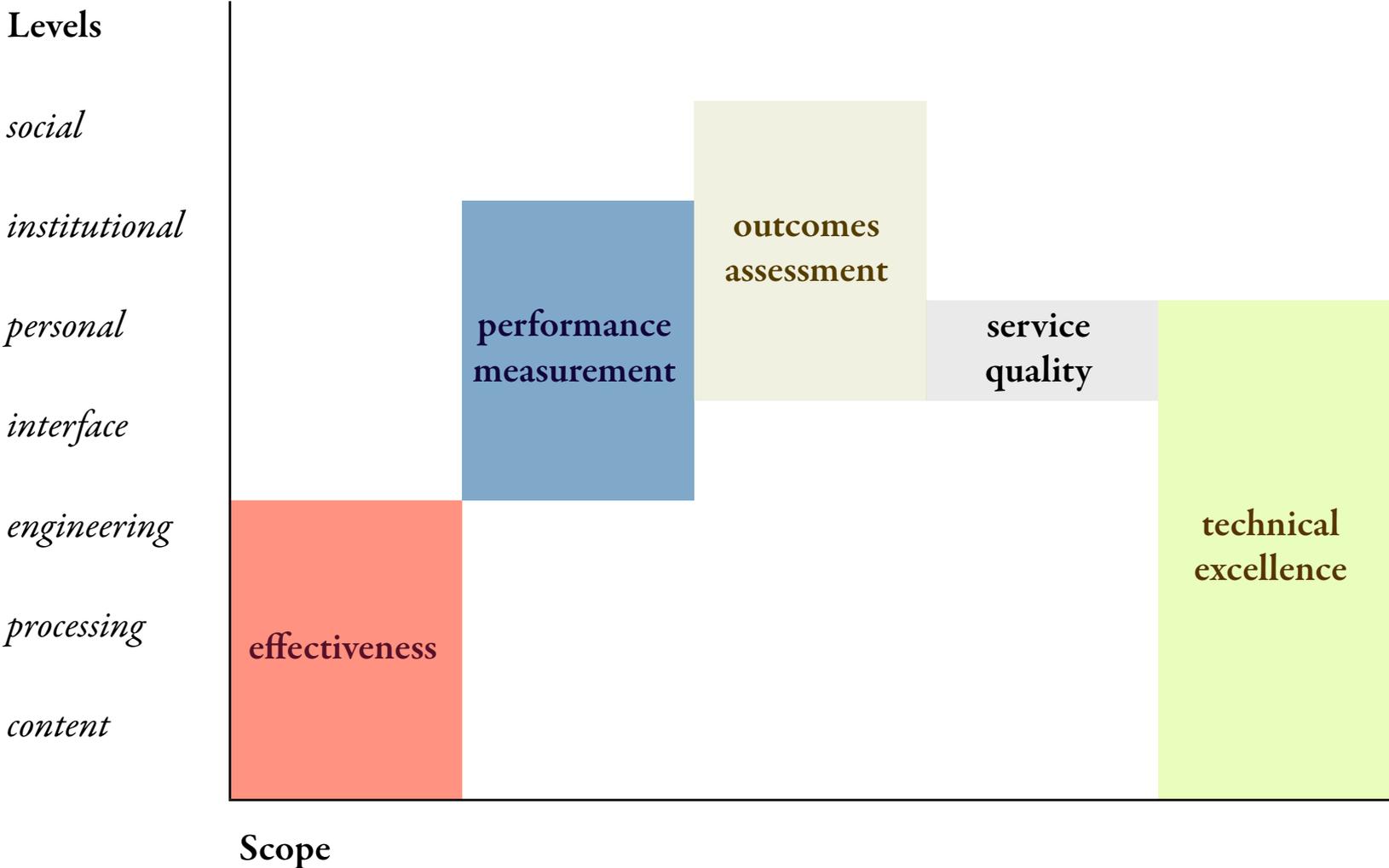
## reasons to evaluate [b]

- But, is it clear to us why we are evaluating?
  - the answer shows commitment
  - is it for internal reasons (posed by the organization, e.g. monitoring), or external reasons (posed by the environment of the organization, e.g. accountability)?
- Strongly related to the context of working.

# scope of evaluation

- Input-output evaluation
  - about effective production
- Performance measurement
  - about efficient operation
- Service quality
  - about meeting the goals of the served audience
- Outcomes assessment
  - about meeting the goals of the hosting environment
- Technical excellence
  - about building better systems

# defining the scope



# find a reason

- Understanding the reasons and the range of our evaluation will help us formulate better research statements.
- Therefore, before initiating, we need:
  - to identify the scope of our research,
  - to clearly express our research statements,
  - to imagine what kind of results we will have,
  - to link our research statements with anticipated findings (either positive, or negative).

questions

what

# what to evaluate?

- Objects
  - parts or the whole of a DL.
  - system or/and data.
- Operations
  - the purposive use of specific parts of DLs by human or machine agents.
  - usage of system or/and usage of data.

# objects

- Interfaces
  - retrieval interfaces, aesthetics, information architecture
- Functionalities
  - search, annotations, storage, sharing, recommendations, organization
- Technologies
  - algorithms, items provision, protocols compliance, preservation modules, hardware
- Collections
  - size, type of objects, growth rate

# operations

- Retrieving information
  - precision/recall, user performance
- Integrating information
- Usage of information objects
  - patterns, preferences, types of interaction
- Collaborating
  - sharing, recommending, annotating information objects
- Harvesting
- Crawling/indexing
- Preservation procedures

questions

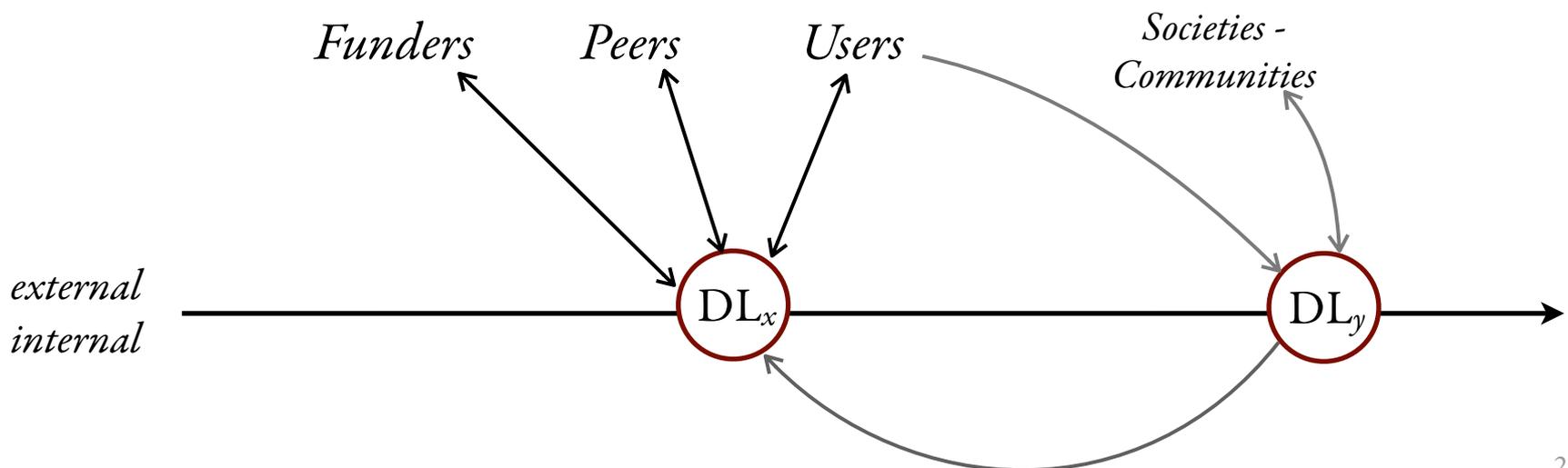
who

# agents of evaluation

- Who evaluates our digital library?
- Our digital library is evaluated by our funders, our users, our peers, but most important *by us*.
  - we plan, we collect, we analyze, we report and (‘unfortunately’) we redesign.

# we against the universe

- We evaluate and are evaluated against someone (~thing) else.
  - this can be a standard, a best practice, a protocol, a verification service, a benchmark
- Often, we are compared against ourselves
  - we in the past (our previous achievements)
  - we in the future (our future expectations)



# we and the universe

- We collaborate with third parties.
  - for instance, vendors providing usage data, IT specialists supporting our hardware, HCI researchers enhancing our interface design, associations conducting comparative surveys, librarians specifying metadata schemas, etc.
- We need to think in inter-disciplinary way
  - to be able to contribute to the planning, to check the reliability of data and to control the experiments, to ensure the collection of comparable data, etc.



questions

how

# methods to evaluate

- Many methods to select and to combine.
- No single method can yield the best results.
- Methods are classified in two main classes:
  - qualitative methods
  - quantitative methods
- But more important is to select ‘methodologies’.

# methods, methods, methods...

- interviews
- focus groups
- surveys
- traffic/usage analysis
- logs/keystrokes analysis
- laboratory studies
- expert studies
- comparison studies
- observations
- ethnography/field studies

# metaphor: the pendulum

- Often is hard to decide if our research will be qualitative or quantitative.
- Sometimes it is easy; predetermined by the context and the scope of evaluation.
- Quantitative try to verify a phenomenon, usually a recorded behavior, while qualitative approaches try to explain it, identifying the motives and the perceptions behind it.

# metaphor continues: the pendulum

- However, it is not only about the data. It is about having an approach during the collection, the analysis and the interpretation of our data.
  - for instance, microscopic log analysis in deep log analysis methodology can provide qualitative insights.
- There are inherent limitations, such as resources.
  - for instance, interviews are hard to quantify due to time required to record, transcribe and analyze.

# selecting methods

- Various ways to rank these methods:
  - resources (to be able to ‘realize’ each method)
  - expertise (to support the various stages)
  - infrastructure (to perform the various stages)
  - data collection (to represent reality and be relieved from bias), see census data or sample data.
- Triangulating methods is essential, but not easy.
  - “...using MMR allows researchers to address issues more widely and more completely than one method could, which in turn amplifies the richness and complexity of the research findings” Fidel [2008].

# criteria

- Criteria are ‘topics’ or perspectives of measurement or judgement.
- Criteria often come grouped.
  - for instance, the categories of usability, collection quality, service quality, system performance efficiency, user opinion in the study of Xie [2008]
- Criteria often have varied semantics between domains.

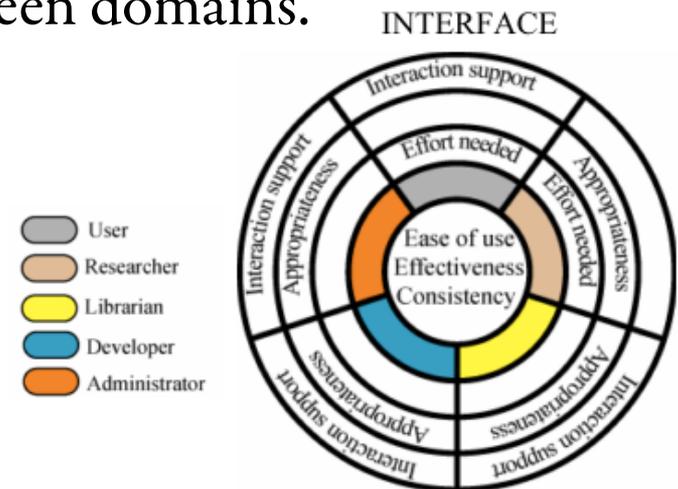
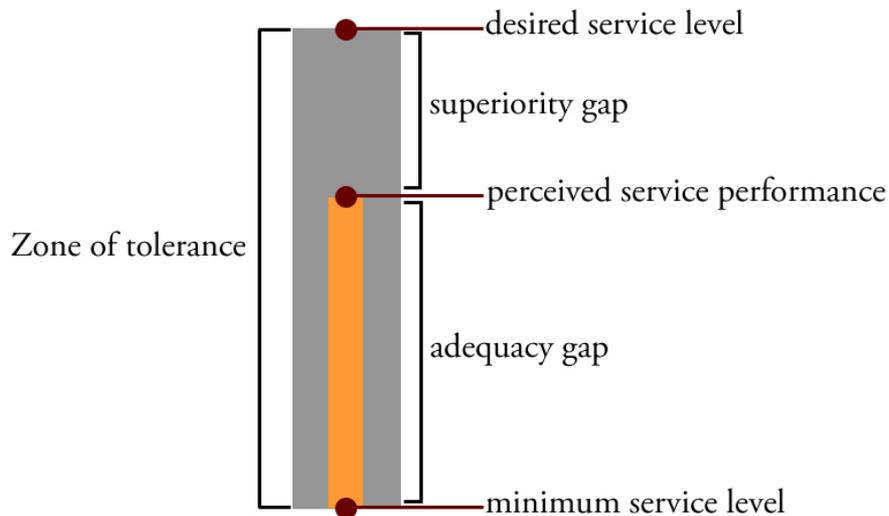


Figure taken by Zhang [2007]



# metrics

- Metrics are the measurement units that we need to establish a distance between -at least - two states.
  - one ideal (target metric)
  - one actual (reality metric)
- An example, LibQUAL's scale



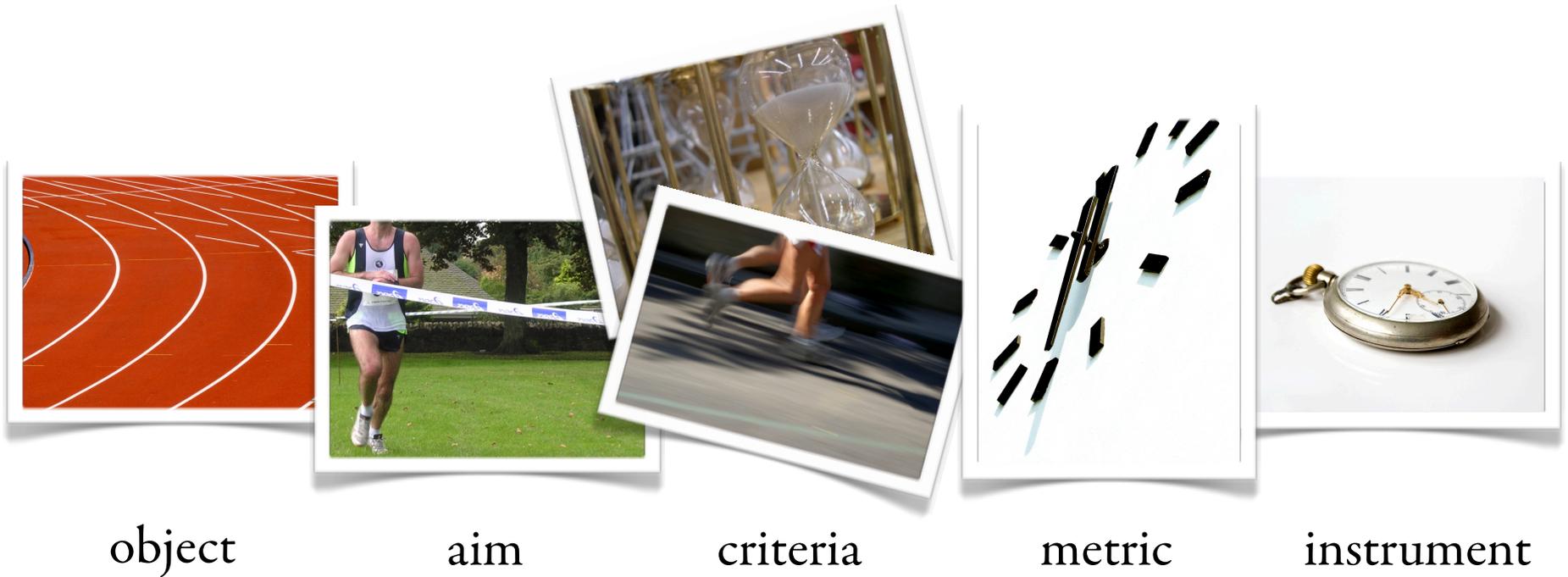
# examples of metrics

Qualitative (insights)

	Unstructured measurement of opinions and perceptions	Observed aspects of performance, e.g. selections, patterns of interactions, etc.	
<b>Goals &amp; Attitudes</b> (what people say)	Scaled measurement of opinions and perceptions	Recorded aspects of activity, e.g. time or errors, via logs or other methods	<b>Behaviors</b> (what people do)

Quantitative (validation)

# the loneliness of the long distance runner



Example taken by Saracevic [2004]

answers

plan

# evaluation planning from above

- Questions
  - like these we have asked
- Buy-In
  - what is invested by each agent
- Budget
  - how much is available
- Methods
  - like those already mentioned



Figure taken by Giersch and Khoo [2009]

# practical questions

- Having stated our research statements and selected our methods, we need:
  - to make an ‘inventory’ of our resources
  - to define what personnel, how skilled and competent is
  - to define what instruments and tools we have
  - to define how much time is available
- All depended on the funding, but some depended on the time of evaluation.

# planning an evaluation

- Upper level planning tools
  - Logic models: useful to have an overview of the whole process and how is linked with the DL development project.
  - Zachman's framework: useful to answer practical questions for setting our evaluation process.

# logic models

- A graphical representation of the processes inside a project that reflects the links between investments and achievements.
  - inputs: project funding and resources
  - activities: the productive phases of the project
  - outputs: short term products/achievements
  - outcomes: long term products/achievements

# logic models: an instance

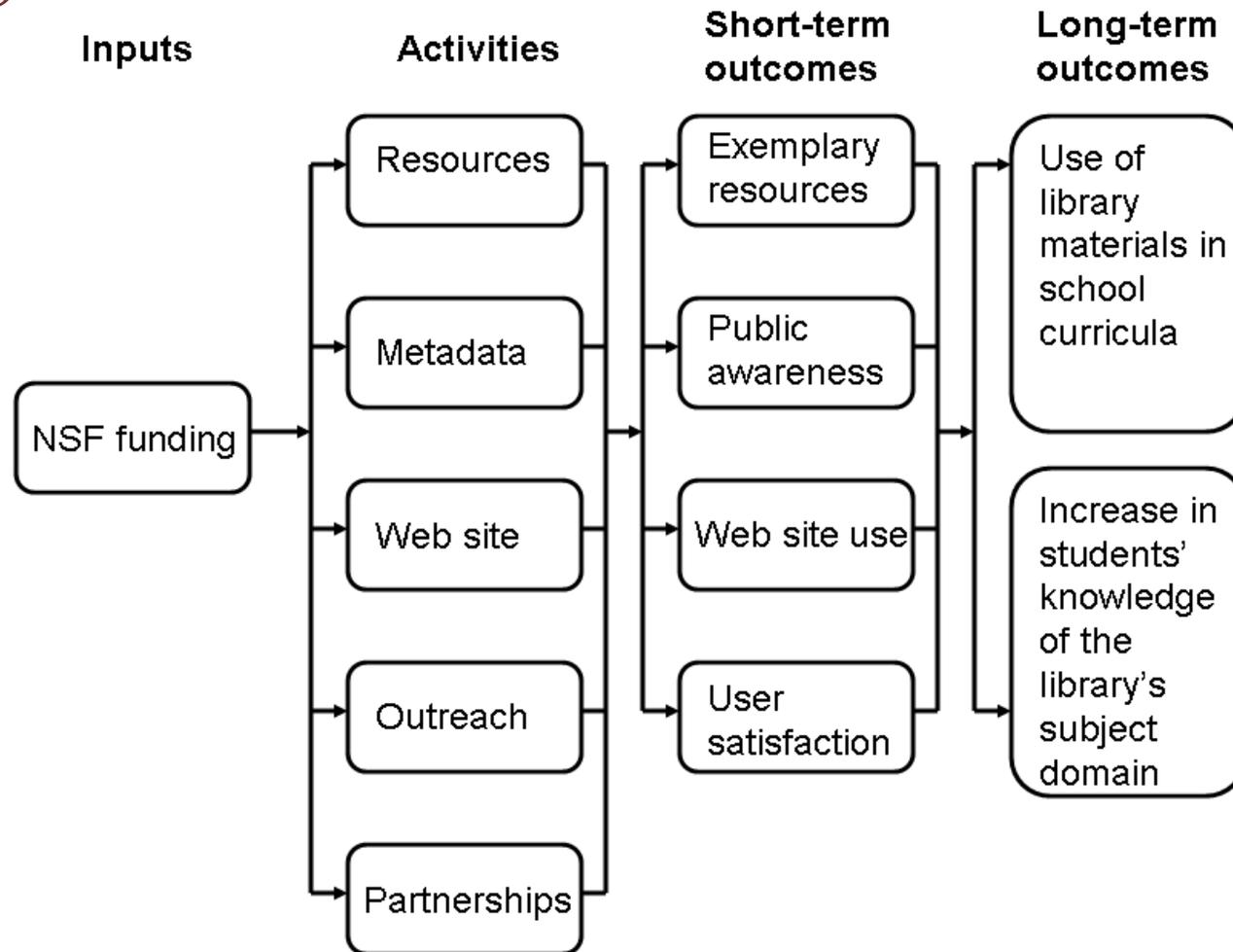


Figure taken by Giersch & Khoo [2009]

# Zachman Framework

- Zachman Framework is a framework for enterprise architecture.
- It depicts a high-level, formal and structured view of an organization; a taxonomy for the organization of the structural elements of an organization under the lens of different views.
- Classifies and organizes in a two-dimensional space all the concepts that are essential to be homogeneous and are needed to express the different planning views.
  - according to participants (alternative perspectives)
  - according to abstractions (questions)

# Zachman's matrix

	<i>What Data</i>	<i>How Process</i>	<i>Where Location</i>	<i>Who Work</i>	<i>When Timing</i>	<i>Why Motivation</i>
<b>Scope [Planner]</b>	Core Business Concepts	Major Business Transformations	Business Locations	Principal Actors	Business Events	Mission & Goals
<b>Business Model [Owner]</b>	Fact Model	Tasks	Business Connectivity Map	Workflow Models	Business Milestones	Policy Charter
<b>System Model [Evaluator]</b>	Data Model	Behavior Allocation	Platform & Communications Map	BRScripts	State Transition Diagrams	Rule Book
<b>Technology Model [Evaluator]</b>	Relational Database Design	Program Specifications	Technical Platform & Communications Design	Procedure & Interface Specifications	Work Queue & Scheduling Designs	Rule Specifications
<b>Detail representation [Evaluator]</b>	Database Schema	Source Code	Network	Procedures & Interfaces	Work Queues & Schedules	Rule Base
<b>Functioning Bus [Evaluator]</b>	Operational Database	Operational Object Cod	Operational Network	Operational Procedures & Interfaces	Operational Work Queues & Schedules	Operational Rules

# guiding our steps

- Assuming that we decided what is best for us, can we find out how far we are from it?
- Is there a roadmap?

# a roadmap

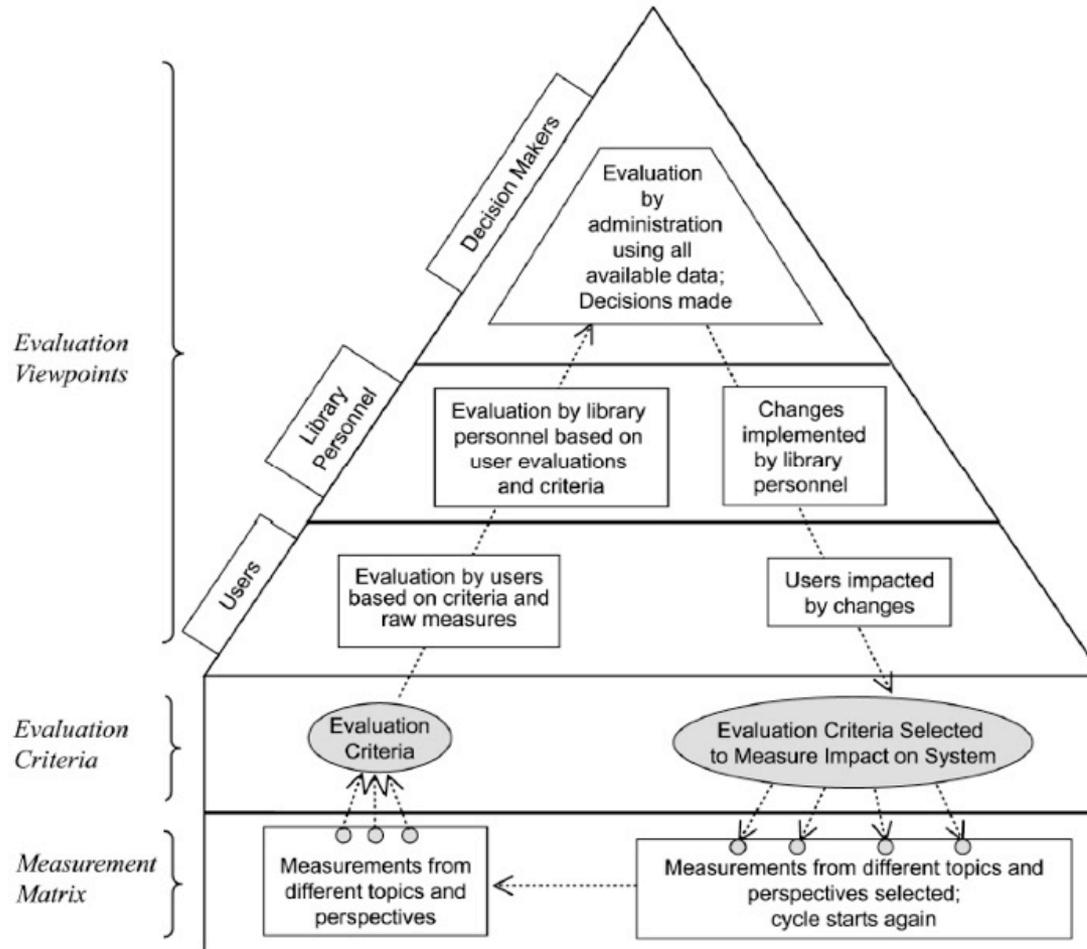
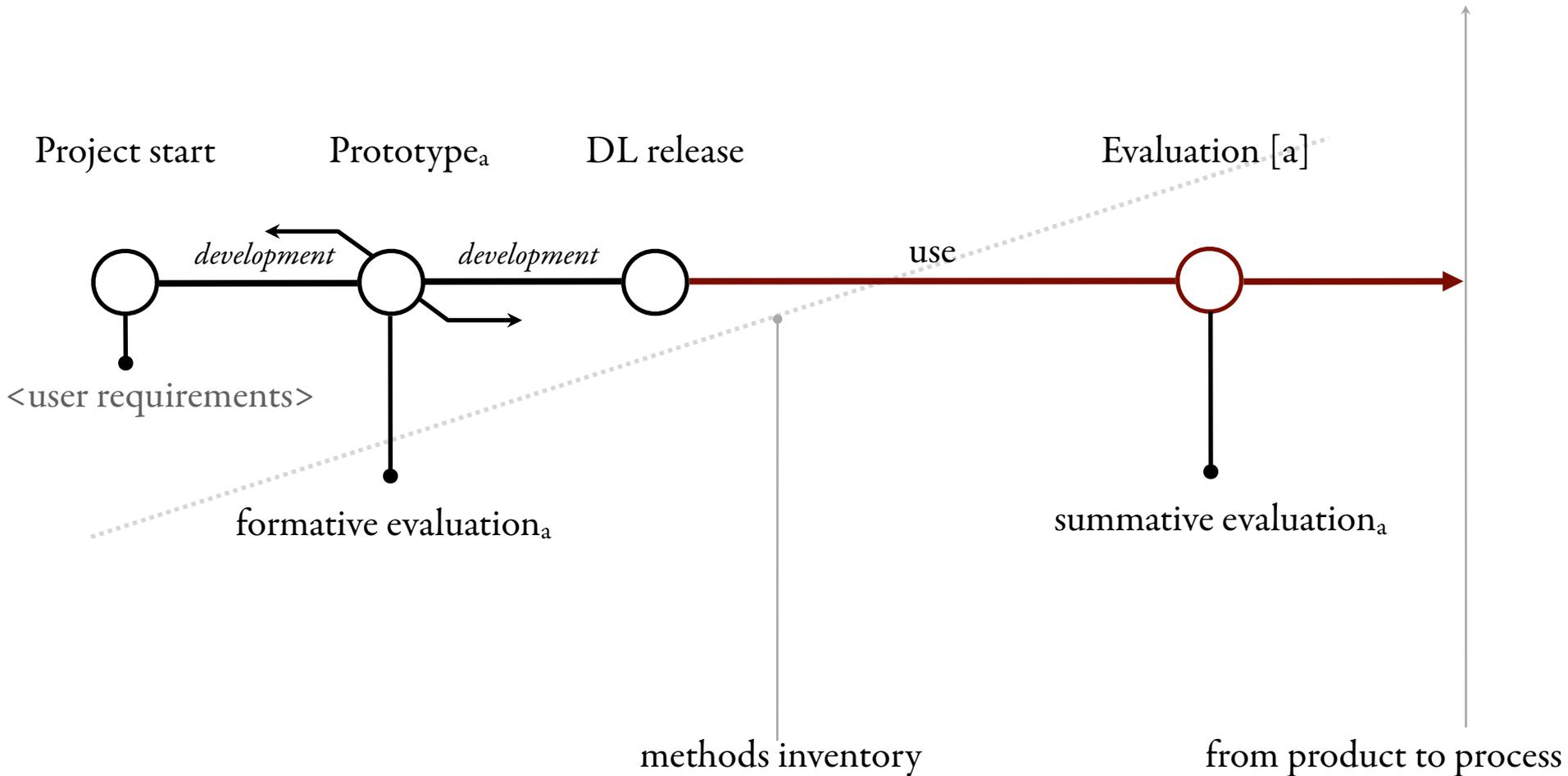


Figure taken by Nicholson [2004]

# when do you evaluate?



# how much is... many?

- Funding is crucial and must be prescribed in the proposal.
- Anecdotal evidence speaks of 5-10% of overall budget.
- Budget allocation usually stays inside project gulfs (also anecdotal evidence).
- Depending on the methods, e.g. logs analysis is considered a low cost method, as well as heuristic evaluation techniques are labeled as 'discount'.

answers

outputs

# outcomes to expect

- Usually evaluation in research-based digital libraries is summative and has the role of ‘deliverable’.
- What should we expect in return?
  - some positive and some negative results
- Positive
  - a set of meaningful findings to transform into recommendations (actions to be taken).
  - dependent on the scope of evaluation.
- Negative
  - a set of non-meaningful findings that can not be exploited
  - highly inconsistent and scarce
  - non applicable and biased

# careful distinctions

- Demographics and user behavior related data are not evaluation per se.
- Assist our analysis and interpretation of data, describing thus their status, but they are not evaluating.

part two

# a formal description of the digital library evaluation domain

## by now

- You have started viewing things in a coherent way,
- Familiarizing with the idea of collaborating with other agents to perform better evaluation activities and
- Forming the basis to judge potential avenues in your evaluation planning.
- But can you do all these things better?
- And how?

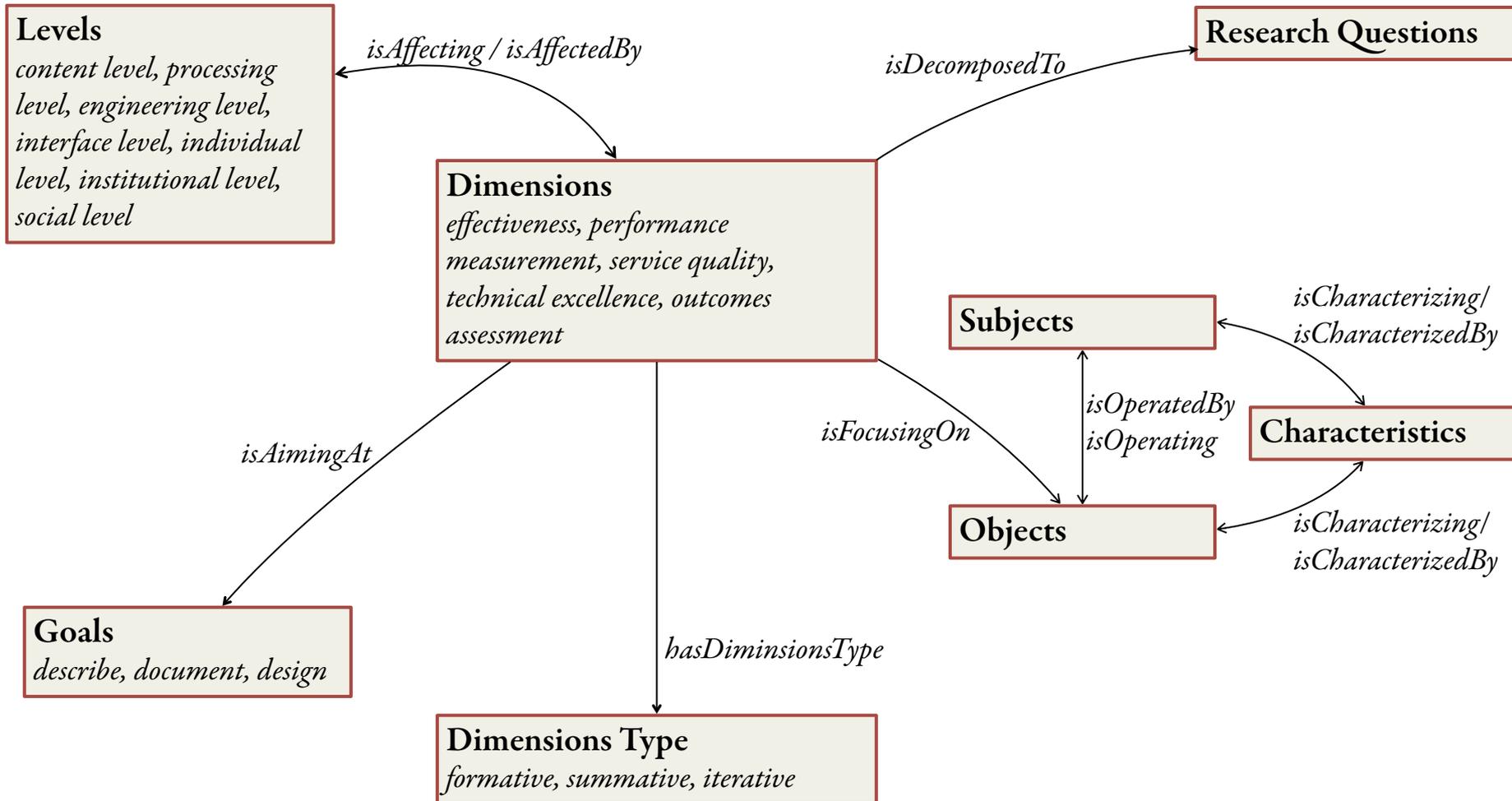
# ontologies as a means to an end

- Formal models that help us
  - to understand a knowledge domain and thus the DL evaluation field
  - to build knowledge bases to compare evaluation instances
  - to assist evaluation initiatives planning
- Ontologies use primitives such as:
  - classes (representing concepts, entities, etc.)
  - relationships (linking the concepts together)
  - functions (constraining the relationships in particular ways)
  - axioms (stating true facts)
  - instances (reflecting examples of reality)

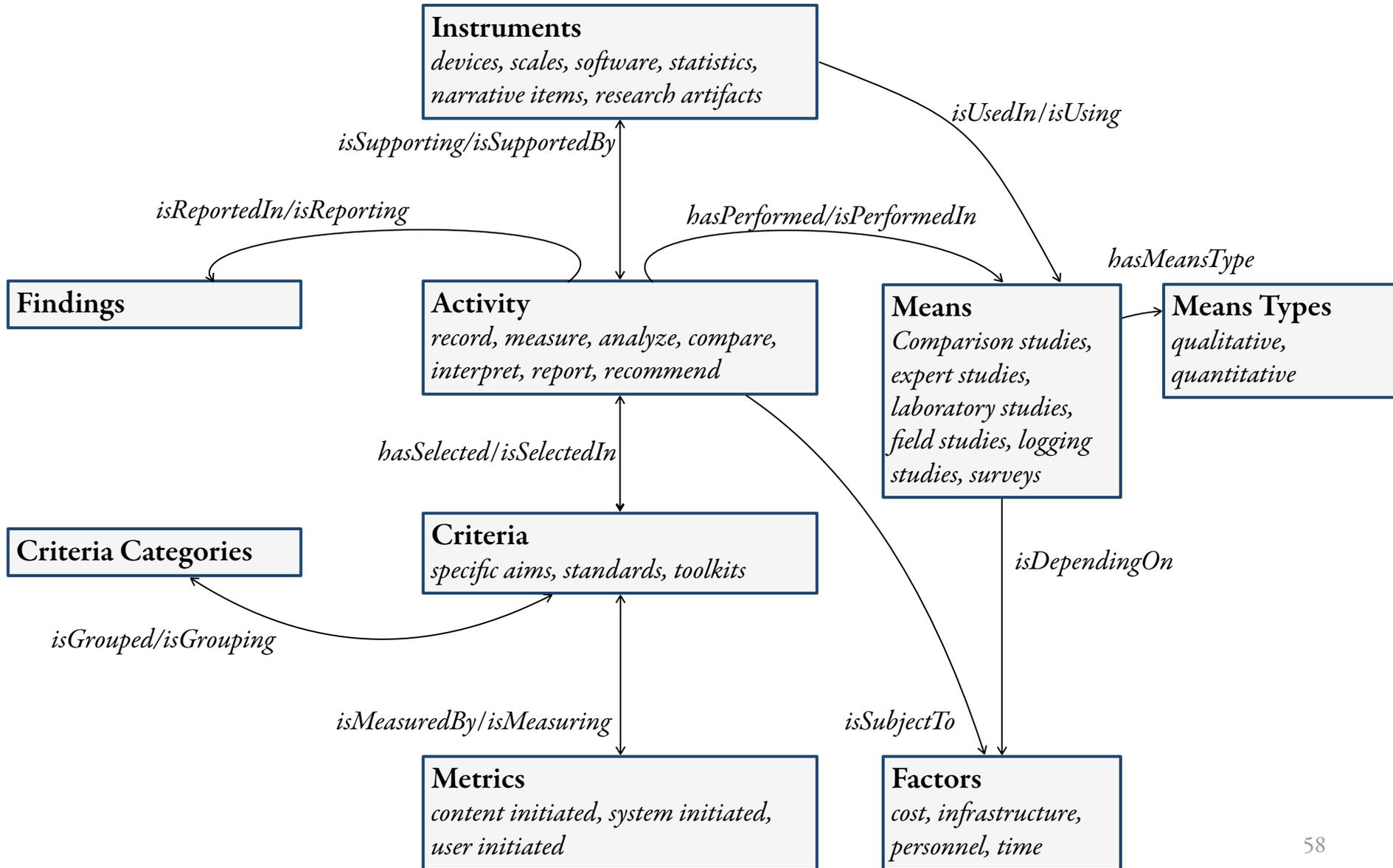
# a formal description of DL evaluation

- An ontology to
  - perform useful comparisons
  - assist effective evaluation planning
- Implemented in OWL with Protégé Ontology Editor

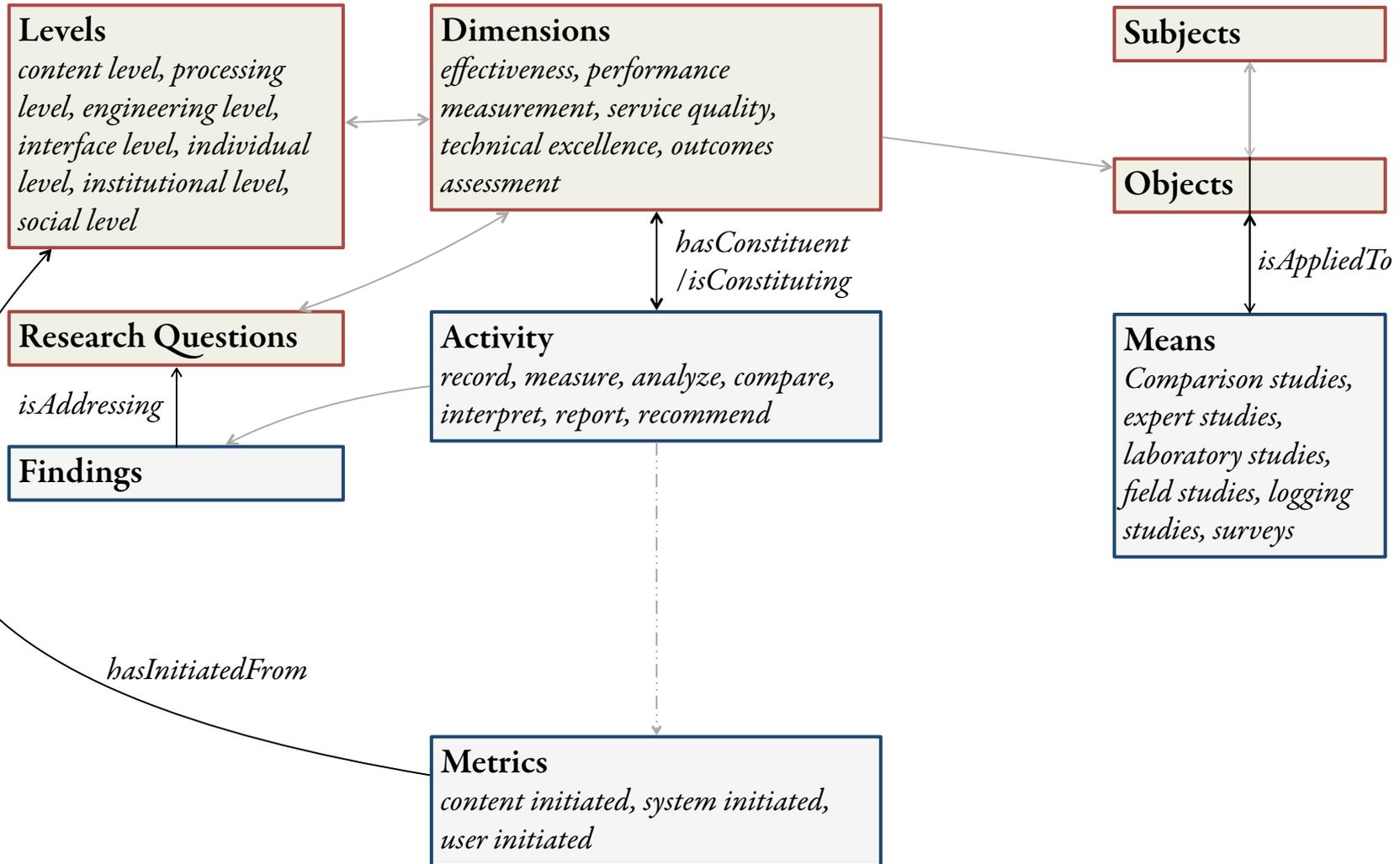
# the upper levels



# the low levels



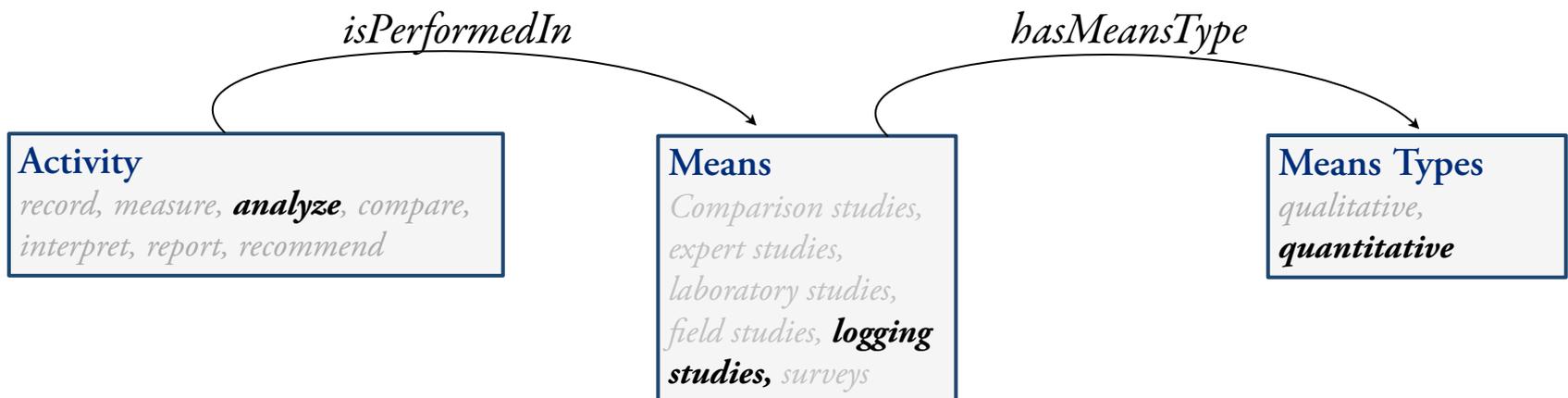
# connections between levels



# use of the ontology [a]

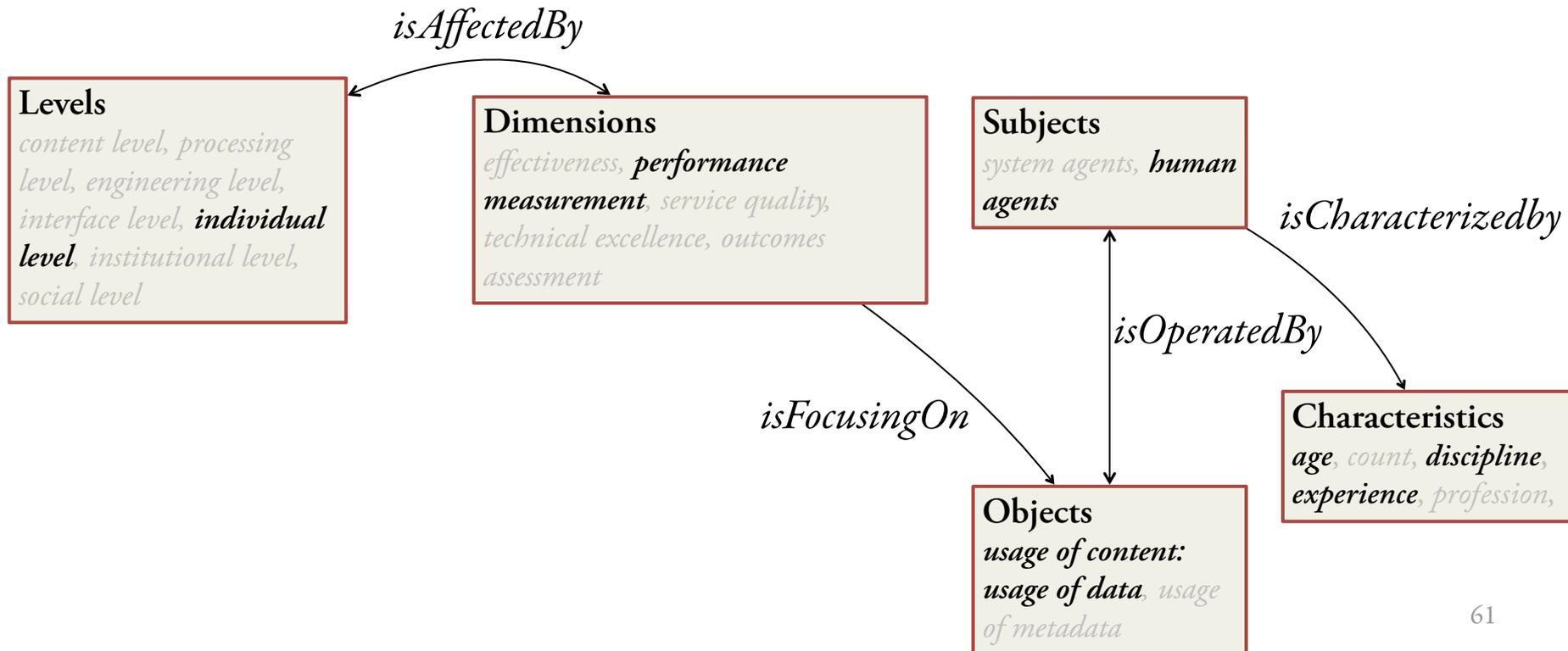
- we use ontology paths to express explicitly a process or a requirement.

Activities/analyze - *isPerformedIn* - Means/logging studies- *hasMeansType* - Means Type



# use of the ontology [b]

Level/content level - *isAffectedBy* - Dimensions/effectiveness - *isFocusingOn* - Objects/  
usage of content/usage of data - *isOperatedBy* - Subjects/human agents *isCharacterizedby* -  
Characteristics/human agents-age, human agents-count, human agents-discipline, human  
agents-experience



hands on session

things to do

## exercise

- Based on your experience and/or your evaluation planning needs, use the ontology schema to map your own project and describe it.
  - If you do not have such experience, please use the case study outlined in Hand Out 5.
  - Use HandOuts 3a and 3b as an example and Hand Out 2 to fill the fields that you think are important to describe your evaluation.
  - Furthermore, report, if applicable, using Hand Out 4:
    - what is missing from your description
    - what is not expressed by the ontology

ending part

conclusions & discussion

# summary

- Evaluation must have a target
  - “...evaluation of a digital library need not involve the whole – it can concentrate on given components or functions and their specific objectives”. Saracevic, 2009
- Evaluation must have a plan and a roadmap
  - “An evaluation plan is essentially a contract between you ... and the other ‘stakeholders’ in the evaluation...”. Reeves et al., 2003
- Evaluation is depended on the context
  - “...is a research process that aims to understand the meaning of some phenomenon situated in a context and the changes that take place as the phenomenon and the context interact”. Marchionini, 2000

# but why is it so difficult to evaluate?

- Saracevic mentions:
  - DLs are complex
  - Evaluation is still premature
  - There is no strong interest
  - There is lack of funding
  - Cultural differences
  - Quite cynical: who wants to be judged?

# can it be easier for you?

- We hope this tutorial made easier for you to address these challenges:
  - DLs are complex (you know it, but now you also know that DL evaluation is equally complex)
  - Evaluation is still premature (you got an idea of the field)
  - There is no strong interest (maybe you are strongly motivated to evaluate your DL)
  - There is lack of funding (ok, there is no funding in this room for your initiative)
  - Cultural differences (maybe you are eager to communicate with other agents)
  - Who wants to be judged? (hopefully, you)

addendum

resources

# essential reading [a]

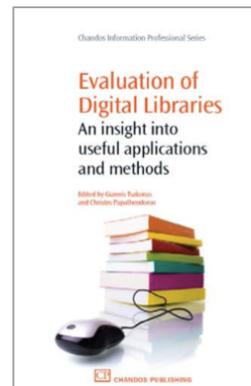
- Bertot, J. C., & McClure, C. R. (2003). Outcomes assessment in the networked environment: research questions, issues, considerations, and moving forward. *Library Trends*, 51(4), 590-613.
- Blandford, A., Adams, A., Attfield, S., Buchanan, G., Gow, J., Makri, S., et al. (2008). The PRET A Rapporteur framework: evaluating digital libraries from the perspective of information work. *Information Processing & Management*, 44(1), 4-21.
- Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., et al. (2007). Evaluation of digital libraries. *International Journal on Digital Libraries*, 8(1), 21-38.
- Gonçalves, M. A., Moreira, B. L., Fox, E. A., & Watson, L. T. (2007). “What is a good digital library?”: a quality model for digital libraries. *Information Processing & Management*, 43(5), 1416-1437.

# essential reading [b]

- Hill, L. L., Carver, L., Darsgaard, M., Dolin, R., Smith, T. R., Frew, J., et al. (2000). Alexandria Digital Library: user evaluation studies and system design. *Journal of the American Society for Information Science and Technology*, 51(3), 246-259.
- Marchionini, G. (2000). Evaluating digital libraries: a longitudinal and multifaceted view. *Library Trends*, 49(2), 4-33.
- Nicholas, D., Huntington, P., & Watkinson, A. (2005). Scholarly journal usage: the results of deep log analysis. *Journal of Documentation*, 61(2), 248-289.
- Powell, R. R. (2006). Evaluation research: an overview. *Library Trends*, 55(1), 102-120.
- Reeves, T. C., Apedoe, X., & Woo, Y. H. (2003). *Evaluating digital libraries: a user-friendly guide*. University Corporation for Atmospheric Research.
- Saracevic, T. (2000). Digital library evaluation: towards an evolution of concepts. *Library Trends*, 49(3), 350-369.

# essential reading [c]

- Wilson, M. L., Schraefel, M., & White, R. W. (2009). Evaluating advanced search interfaces using established information-seeking models. *Journal of the American Society for Information Science and Technology*, 60(7), 1407-1422.
- Xie, H. I. (2008). Users' evaluation of digital libraries (DLs): their uses, their criteria, and their assessment. *Information Processing & Management*, 44(3), 27.
- Zhang, Y. (2010). Developing a holistic model for digital library evaluation. *Journal of the American Society for Information Science*, 61(1), 88-110.
- Giannis Tsakonas and Christos Papatheodorou (eds). 2009. *Evaluation of digital libraries: an insight to useful applications and methods*. Oxford: Chandos Publishing.



# tutorial material

- Slides, literature and rest of the training material are available at:
  - <http://dlib.ionio.gr/~gtsak/ecdl2010tutorial>,
  - <http://bit.ly/9iUScR>,  
link to the papers' public collection in Mendeley

