

evaluation of cultural heritage digital collections: the DiLEO perspective

Christos Papatheodorou & Giannis Tsakonas

Database and Information Systems Research Group, Ionian University, Greece

19-22 September 2011 | Cultural Heritage in CLEF Workshop | Amsterdam, The Netherlands



background

- Evaluation principle: evaluation needs a clear description of the context of the evaluation target (system, service, object, etc.)
- Are CH collections different? Probably yes...
 - CH reflect different cultures' needs and practices
 - CH host mostly non-textual information
 - CH host semantically diverse resources
- Cultural heritage and evaluation
 - do we have evidence that CH collections are evaluated?
 - how are evaluated?
 - do we have results? In which sector?

cultural heritage in Europe

- Several projects for the development of CH collections, such as:
 - NET-HERITAGE, DC-NET, ATHENA, INDICATE, CALIMERA and of course... Europeana and its siblings.
- Most of them governed by business models.
 - coordinating policies and forming best-practices
- CH collections with varied characteristics:
 - system-wise heterogeneous
 - content-wise disconnected, diverse multilingual

in quest of quality

- Calimera Project
 - best practices in a wide range of topics, including multilingualism and evaluation
- Minerva Project
 - guidelines for digitization
 - usability guidelines
- Europeana
 - user studying orientation

Quoting from the Future Agenda on Multilingualism (circa 2005)

- *The Cross-Language Evaluation Forum (CLEF) and CLEF 2004 ... have done a lot of research into multilingual information retrieval. It is to be hoped that such work will form the basis for future developments.*

Also on the future agenda on evaluation:

- *to have maximum impact, **standards need to be understood and applied at national level;***
- *...*
- *the **topics are considered “difficult” and are not widely understood;***
- *good practice is **slow to travel and catch on** (between countries and between sectors or domains);*
- *measuring activity in networked environments is particularly difficult, yet this will be **absolutely key to future use of both statistics and performance measures;***
- *...*

(bold characters signaling our emphasis)

Minerva guidelines on usability



Europeana



- Research constructs:
 - the Europeana Personas Catalogue, a list of archetypical figures built “using input from Europeana partners and research on behavior and search patterns”
 - the Europeana Clickstream Logger, a customized logging schema with emphasis on multilingualism like interface language changes, use of language facets, etc.
- Research studies:
 - surveys investigating awareness, motivation of use, context of use, etc.,
 - user studies, such as focus groups, discussing content and functionality,
 - usability studies, such as eye-tracking studies, focusing on interface architecture and design.

the Tower of Babel

- Which kind of “multilingualism” would impede the construction of the Tower of Babel?
 - the *linguistic* or the *semantic*?
- Evaluation is threatened by different conceptualizations of —often the same— tools, constructs, methods, approaches and so on.



The Tower of Babel by Pieter Breugel,
Kunsthistorisches Museum, Wien, Austria

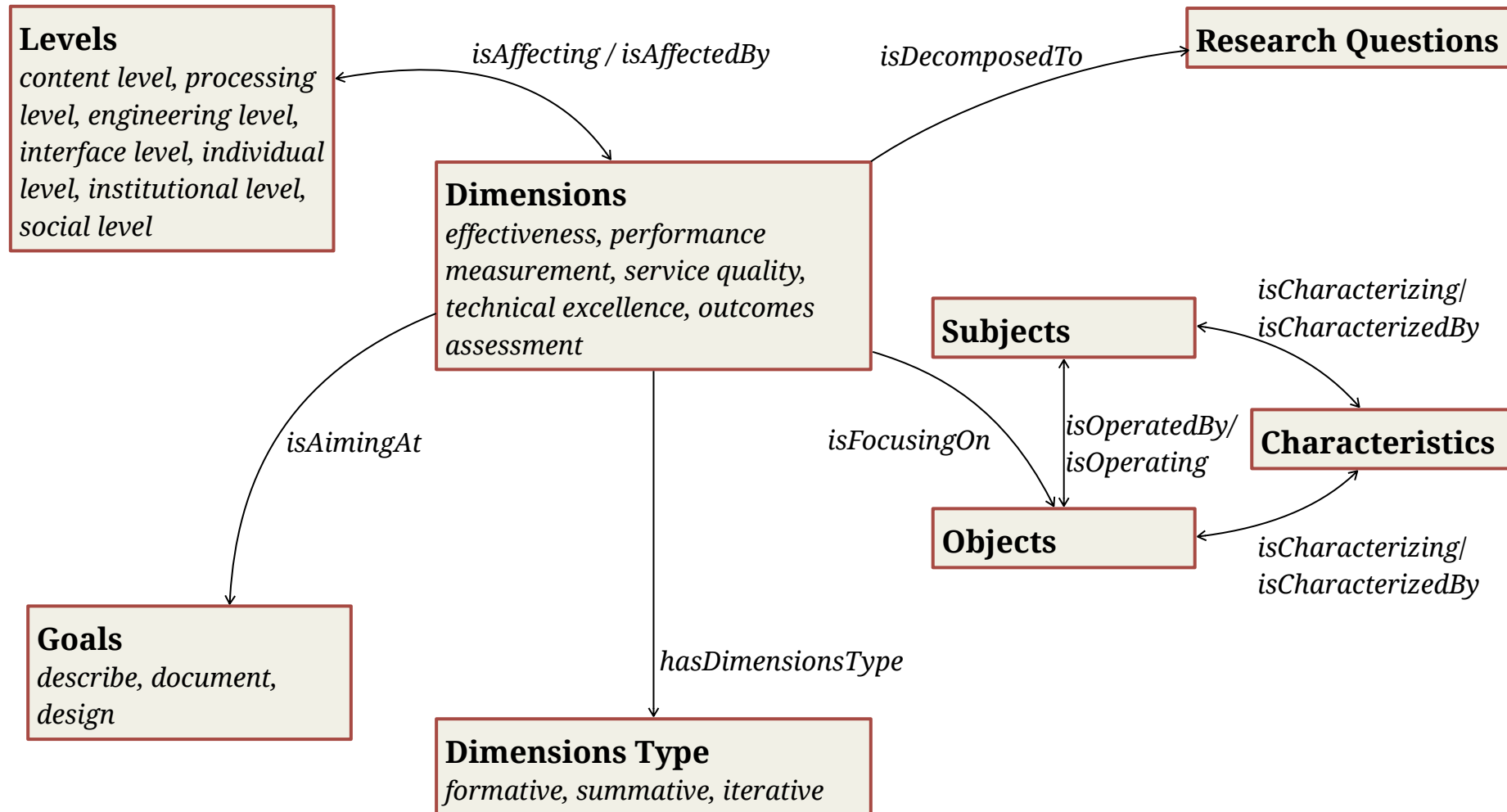
modeling evaluation

- We have developed DiLEO
- A domain ontology, a formal model, that help us
 - to understand the knowledge domain of the DL evaluation field
 - to build knowledge bases to explore evaluation instances
 - to assist the planning of forthcoming DL evaluation initiatives
- Implemented in OWL with Protégé Ontology Editor

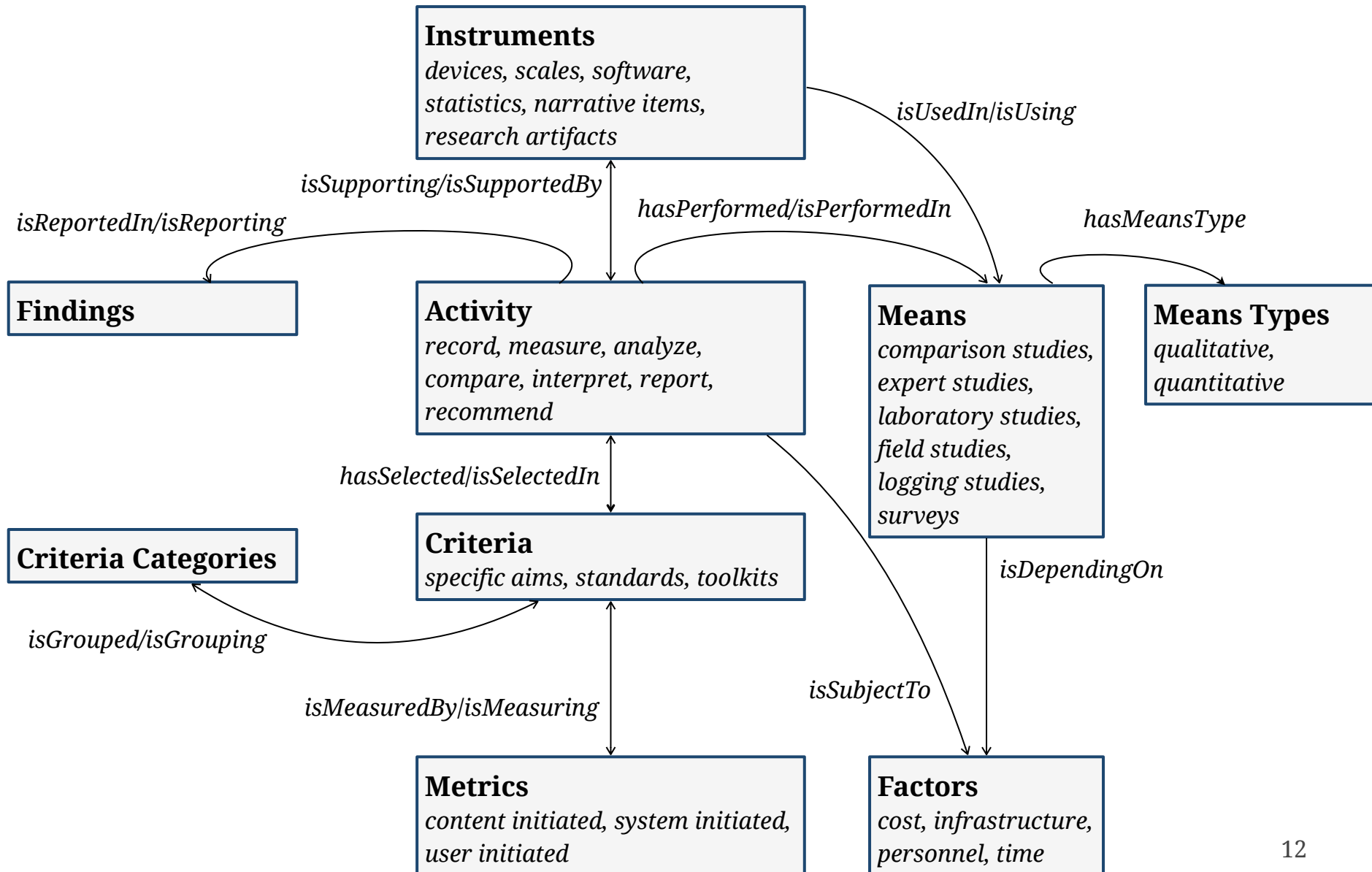
DiLEO in a nutshell

- DiLEO addresses the semantic diversity in the evaluation of digital libraries.
- It provides a vocabulary of concepts and defines the properties that govern their relationships.
- It defines—in a two-level structure—the context of evaluation, its scope and aims, as well as details the practical aspects of an initiative.

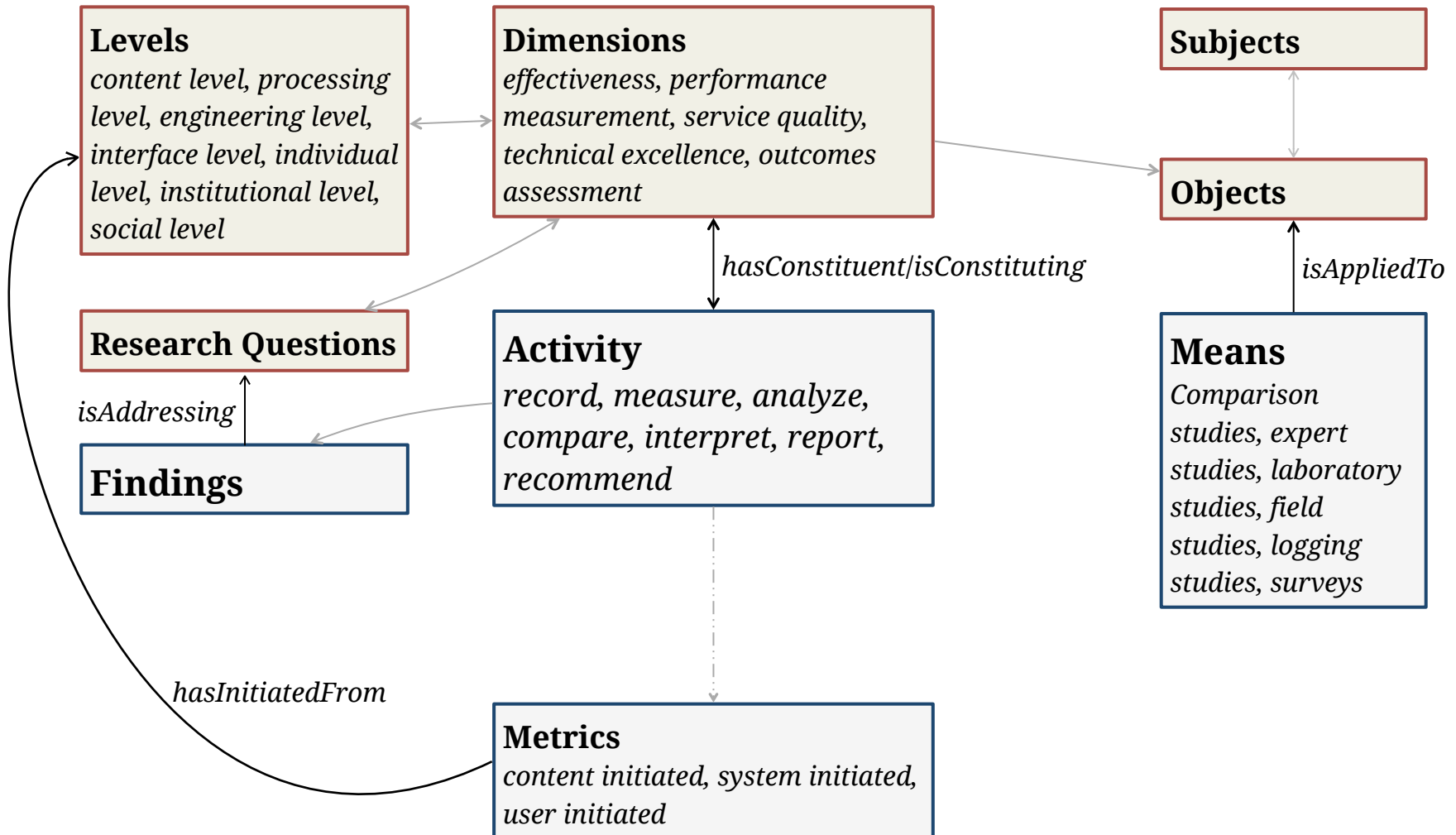
the upper levels



the low levels



connections between levels



use of the ontology

- we use DiLEO
 - to represent knowledge; information based on the analysis of facts
 - to plan evaluation activities; to base future steps on this knowledge
- to do so we use SPARQL queries

use of the ontology - SPARQL queries

Query

```
SELECT DISTINCT ?Research_QuestionsInst ?Means
WHERE {
  ?Research_QuestionsInst a <http://www.owl-ontologies.com/Ontology1256756350:wm2008c_ExplorationOfSessionLenghtAsMetric>.
  ?Dimensions a <http://www.owl-ontologies.com/Ontology1256756350:wm2008c_UsabilityTestingOfSeachResultsPage>.
  ?Activity a <http://www.owl-ontologies.com/Ontology1256756350:nzdl2000_DescribeUserActions>.
  ?Research_QuestionsInst <http://www.owl-ontologies.com/Ontology1256756350:wm2008c_ExplorationOfSessionLenghtAsMetric>.
  ?Dimensions <http://www.owl-ontologies.com/Ontology1256756350:wm2008c_UsabilityTestingOfSeachResultsPage>.
  ?Activity <http://www.owl-ontologies.com/Ontology1256756350:nzdl2000_DescribeUserActions>.
}
```

Results

Research_QuestionsInst	Means
Ontology1256756350:wm2008c_ExplorationOfSessionLenghtAsMetric	Ontology1256756350:wm2008c_Logs
Ontology1256756350:wm2008c_UsabilityTestingOfSeachResultsPage	Ontology1256756350:wm2008c_Logs
Ontology1256756350:nzdl2000_DescribeUserActions	Ontology1256756350:nzdl2000_Logs

research questions (first column) from two studies (*wm2008c* and *nzdl2000*) which used log (second column)

SELECT DISTINCT

```
?Research_QuestionsInst
?Means
WHERE
{
  ?Research_QuestionsInst a<Research_Questions>.
  ?Dimensions a<Technical_Excellence>.
  ?Activity a <Record>.
  ?Means a <Logs>.
  ?Research_QuestionsInst<isBelongingTo> ?Dimensions.
  ?Dimensions<hasConstituent> ?Activity.
  ?Activity<isPerformedIn> ?Means
}
```

use of the ontology - knowledge

- We want to learn about the instruments that were used in survey studies.

Query

```
SELECT DISTINCT
?Instruments WHERE
{
?Means a <Surveys>.
?Instruments <isUsedIn>?Means
}
```

Answers

Instruments/Software

- I/sftw:{eco2002_AccessDatabase}
- I/sftw:{eco2002_SPSS}
- I/sftw:{eco2002_SurveySolutions}
- I/sftw:{nric2009_Software}

Instruments/Narrative Items

- I/nrvi: {eco2002_Questions}
- I/nrvi:{nric2009_Questions}
- I/nrvi:{nric2009_Tasks}

Instruments/Statistics

- I/stat:{eco2002_Statistics}
- I/stat: {nric2009_Statistics}

use of the ontology - planning

- In planning the evaluator need to submit more queries and to synthesize the answers to proceed to decision planning.
- We want to learn about the Research Questions in logging studies (Q1) and the possible limiting Factors (Q2).

Query 1

```
SELECT DISTINCT
?ResearchQuestionsInst
?Means WHERE
{
?ResearchQuestionsInst a <ResearchQuestions>.
?Dimensions a <TechnicalExcellence>.
?Activity a <Record>.
?Means a <Logs>.
?ResearchQuestionsInst <isBelongingTo>?
Dimensions.
?Dimensions <hasConstituent>?Activity.
?Activity <isPerformedIn>?Means
}
```

Answers

```
RQ:{wm2008c_ExplorationOfSessionLengthAsMetric}
RQ:{nzdl2000_DescribeUserActions}
```

Query 2

```
SELECT DISTINCT
?Factors WHERE
{
?Means a <Logs>.
?Means <isDependingOn>?Factors
}
```

Answers

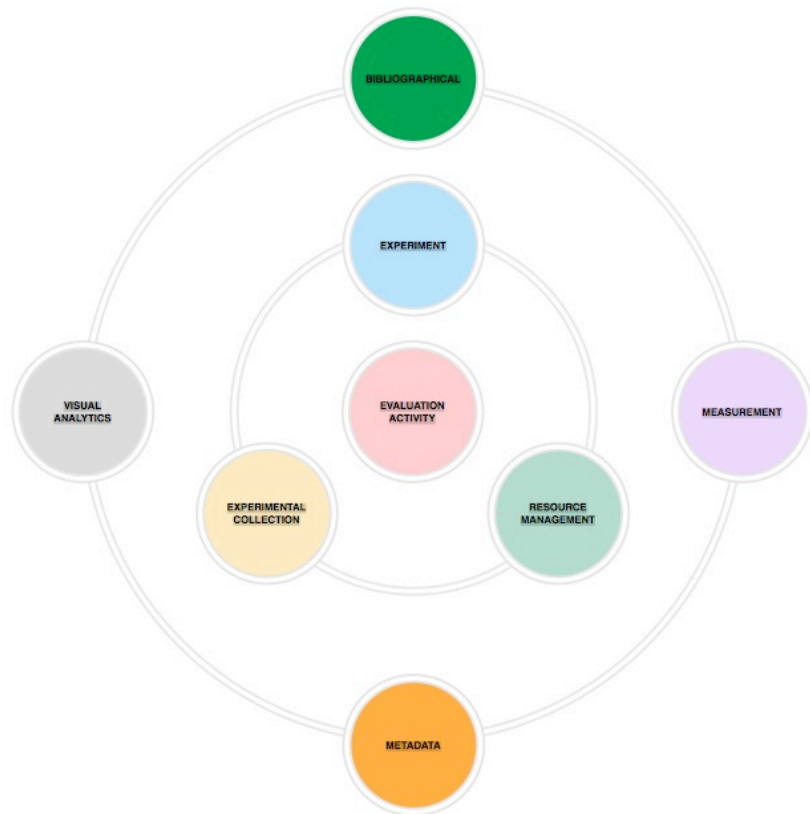
```
F/tim:{nzdl2000_30September 1996-1December1996}
F/tim:{nzdl2000_April1996- July1997}
```

Promise

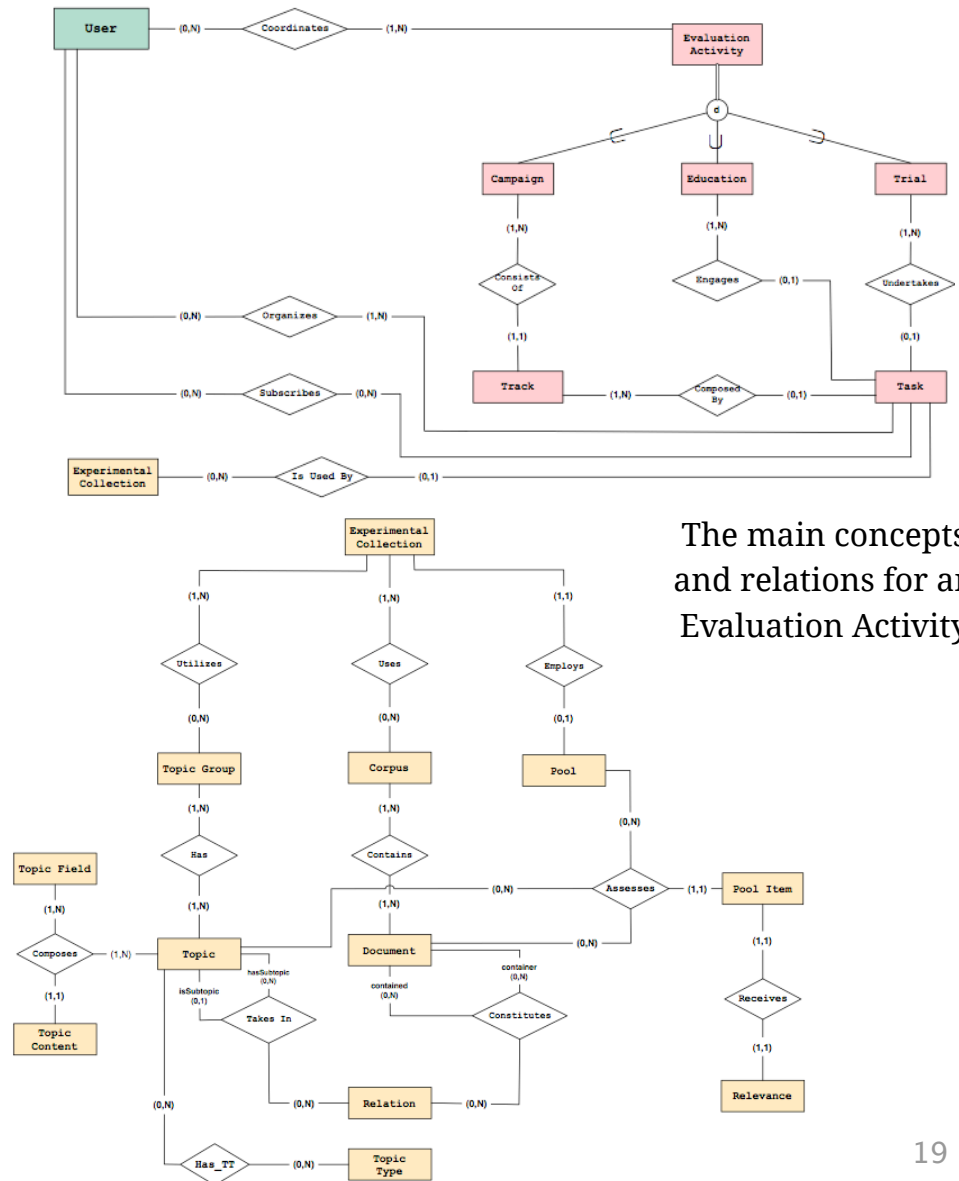


- Open Evaluation Infrastructure
 - guided by example; the DIRECT case
- Use Cases as a Bridge between Benchmarking and Validation
 - widening the basis of evaluation data and systems
 - introducing concerns about real use cases; some uncontrolled parameters
- Continuous Experimentation
 - reusing evaluation data; making them accessible and comprehensible

Promise conceptual model

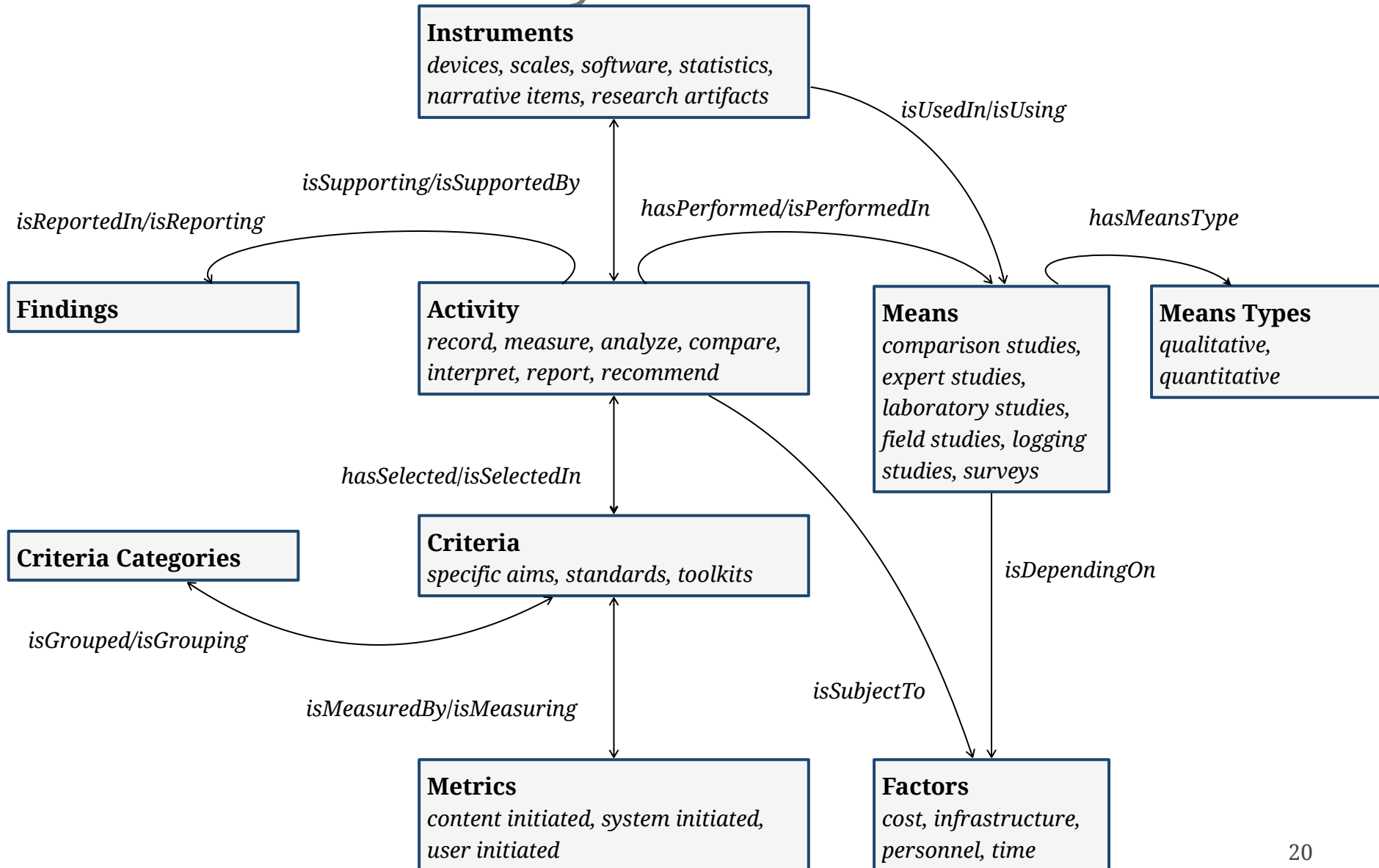


Entities of the PROMISE infrastructure and their relationships



The main concepts and relations for an Evaluation Activity

the DiLEO Activity



...can DiLEO address Promise issues?

- Use Cases as a Bridge between Benchmarking and Validation
 - DiLEO tries to model the context in which information access services live
 - covers several perspectives of evaluation (system/user-centered), application layers (system, content, user, communities, etc.) and objects (data, metadata, types of users, etc.)
 - can offer extensibility on concepts that appear in CH evaluation constructs, such as the Europeana Personas
- Continuous Experimentation
 - instances are reusable through a poly-prismatic view, which can forward reiteration

the challenge

- PROMISE conceptual model is oriented to information retrieval evaluation
- DiLEO is a generic model for digital libraries evaluation
- a rough comparison:
 - DiLEO does not cover the concepts of ‘Evaluation Campaigns’, ‘Collections’ while ‘Visualization’ is not an explicit subclass in DiLEO, but it is implicit in the ‘Activities’ subclass ‘Report’
 - DiLEO could define the context of work through the concepts ‘Dimensions’ and ‘Levels’ and to link them with the main evaluation activities, incorporating concepts such as ‘Criteria’ and ‘Means’ expressing explicitly the parameters for the design of large scale experiments
- the issue: could the alignment of the two models provide a concrete model for evaluating information access to CH sources?

DiLEO resources

<http://www.ionio.gr/labs/dbis/dileo.html>

Thank you for your attention
questions?

