WEE-Net ALFA Project

Summer School at Trento University - Italy

Web Quality Measurement and Evaluation based on Metrics and Indicators

Contents

Dr. Luis Olsina

GIDIS_Web, Informatics Dept., Engineering School, UNLPam, Argentina

E-mail olsinal@ing.unlpam.edu.ar

Argentina

Motivation

- A recurrent challenge faced by many software and Web organizations is to have a **clear establishment** of a **measurement** and **evaluation framework** for **quality assurance** (QA) **processes and programs**.
  - For instance Measurement & Analysis process area in CMMI
- A well-established measurement and evaluation framework should rely on a **sound conceptual base**.
  - Ontology of Metrics and Indicators
- Organizations could succeed if resulting measurements and evaluations are tailored to their **information needs** for specific **purposes**, **contexts**, and **user viewpoints**.

Motivation

- Therefore, we argue that at least three columns are necessary to build, i.e. to design and to implement a robust and sound measurement and evaluation program, namely:
- A **process** for measurement and evaluation,
  - i.e. the main managerial and technical activities that have to be planned and performed;
- A **goal-oriented measurement and evaluation framework** that must rely on a sound conceptual (ontological) base; and
- Specific model-based **methods and techniques** in order to carry out the specific project's activities.

Motivation

- The present tutorial focuses mainly on discussing our measurement and evaluation framework so-called **INCAMI** (*Information Need*, *Concept model*, *Attribute*, *Metric* and *Indicator*), which is based on a metrics and indicators ontology.
- Without appropriate definitions (meta-data) of metrics and indicators it is difficult to ensure values are repeatable and comparable among organization's projects for datasets analyses.
- Moreover, inter and intra-project analyses and comparisons could be performed in an inconsistent way.

Motivation

- The tutorial aims to bring the attention of you about the usefulness of the INCAMI framework and strategy for measurement, evaluation and analysis process areas,
- Besides, we will discuss why this framework can be a more robust and well-established than the GQM (Goal/Question/Metric) paradigm for measurement and evaluation purposes, among others
- Ultimately, strengths and weaknesses of our framework are analysed as well

Outline

- **Introduction to Quality and Quality in Use**
  - **to Software and Web Entities**
- Conceptual Base for Metrics & Indicators
- Evaluation Process
- Goal-oriented Measurement and Evaluation Frameworks
  - GQM Paradigm
  - INCAMI Framework
    - INCAMI Components & Tool
- Conclusions

ISO Stds. about Quality, Measurement …

- Many ISO Stds deal with these concepts, e.g.:
  - **Quality (Sw Product): internal and external quality models, and quality in use model** for sw. (ISO 9126-1:2001)
  - **Quality (Process): process assessment and capability determination** for software organizations (ISO 15504:2003)
  - **Evaluation:** The **evaluation process** (ISO 14598:1998)
  - **Measurement:** The **measurement process** (ISO 15939:2002)

Very often, we have observed a lack of consensus in the terminology (same terms different meaning, different terms with similar meaning, absent terms, etc.)

What is Quality?

- **Quality**
  - **Quality** of an entity is hard to define and assess, but it is easy to recognize
  - Generally, the **perceived quality** of an entity is transparent when present, but noticeable when absent

What is Quality?

- **Quality usually has different views** (as analyzed by David Garvin, 87):
  - Transcendent View
  - User View
  - Product View
  - Producer View
  - Value-based View
    - quality/cost trade-off

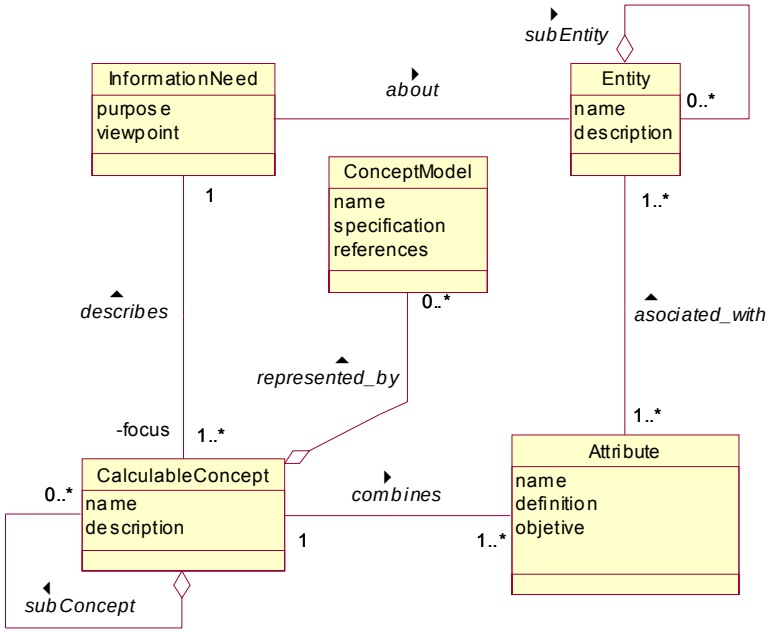
What is Quality?

- The meaning of the **quality term** is not simple and atomic, but a multidimensional and abstract concept.
- Quality can not be measured and evaluated directly,
  - at least in a not very trivial way
- Common practice assesses quality by means of the quantification of lower abstraction concepts, such as attributes of entities
- Given the inner complexity that a quality concept involves, it is necessary generally a **model** in order to specify the quality requirements.

What is Quality?

- Quality depends on a specific project/organizational **information need**, i.e., for a specific **purpose**, **user viewpoint**, and **context**
- Quality is an **abstract relationship** between **attributes** of an **entity** (a product, process, …) and a specific **information need** for a project, or organization.

NFR (Quality...) Framework



Quality vs. Project Variables

- Scope
  - Functionality / Services / Contents to deliver
- Time (Schedule)
  - Effort (persons per days)
  - Calendar (working and not-working days)
    - Time-to-Market
- Quality
  - Product
  - Process
    - Capability
  - Resource
    - Human Skills,
    - Methods, Tools, ...
- Cost
  - Budget

Non-Quality Cost and Impact

- Waste of:
  - effort (persons-hours)
  - materials
- Loss of time
  - to be the product available
- Re-work
  - For repairing / fixing defects
  - Impact of changes
- Impact wrt the customer
  - loss of the enterprise image
  - loss in the product trustfulness
    - likely lower sales

Define Quality is a hard job ...

Define, Specify the Quality depends on the:

- Entity to be applied
  - Project (Development, Maintenance, ...)
    - Process
    - Product
      - Product in Use
    - Resource
    - Service
- Perspective (User Viewpoint/Profile)
  - Developer, Manager, Final User, ...
    - Often, for the same user profile (to different –or similar, projects) there are different needs, priorities ...
- Domain
- Lifecycle Stages
  - Early, Late, Operative …

Evaluate Quality is a hard job ...

Evaluate Quality is a "human-intensive business"

- It is not an easy job:
  - Define, Specify,
  - Measure, Evaluate, Predict,

  for instance, the quality of a software or Web application
- It is not a simple task:
  - Define, Select, and
  - Use Systematically

  Porgrams, Strategies, Techniques and Methods for measurement and evaluation to different entities ... and quality perspectives
- Sometimes, a method or technique is not enough for an evaluation problem

Very often, one size does not fit all needs and preferences …

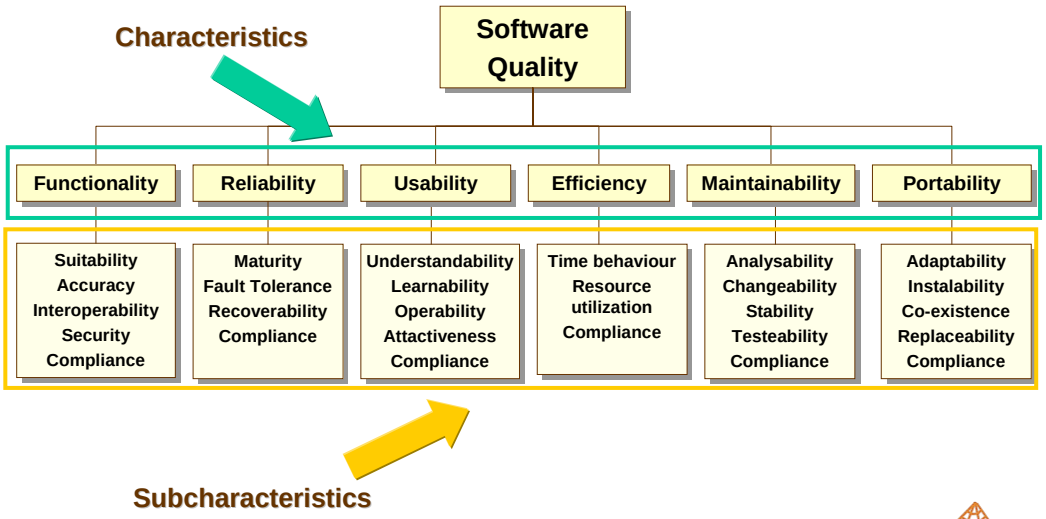
What is Quality?

Quality of a Software Product (ISO/IEC 9126-1: 2001)

Three Views for Quality:

- **Internal Quality – Def.**
  - The totality of attributes of a product that determines its ability to satisfy stated and implied needs when used under specified conditions
- **External Quality – Def.**
  - The extent to which a product satisfies stated and implied needs when used under specified conditions
- **Quality in Use – Def.**
  - The capability of software product to enable specified users to achieve specified goals with effectiveness, productivity, safety and satisfaction in specified context of use.

The ISO 9126-1 Quality Model



ISO 9126-1: Usability

The capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions

Subcharacteristic	Definition
Understandability	The capability of the software product to enable the user to understand whether the software is suitable, and how it can be used for particular tasks and conditions of use.
Learnability	The capability of the software product to enable the user to learn its application.
Operability	The capability of the software product to enable the user to operate and control it.
Attractiveness	The capability of the software product to be attractive to the user.
Compliance	The capability of the software product to adhere to standards, conventions, style guides or regulations relating to usability.

Perspectives of Quality: ISO 9126-1

- **Internal Quality** is specified by a **quality model** (the six characteristics shown)
- It can be measured and evaluated by static **attributes** of documents such as specification of requirements, architecture, or design; pieces of source code, and so forth.
- In **early phases** of a software or Web lifecycle, we can evaluate and control the internal quality of these early products.
- But assuring **internal quality** is not usually sufficient to assure **external quality**.

Perspectives of Quality: ISO 9126-1

- **External Quality** is specified by a quality model (the six characteristics shown)
- It can be measured and evaluated by dynamic **properties** of the running code in a computer system, i.e. when the module or full application is executed in a computer or network simulating as close as possible the actual environment.
- In **late phases** of a software lifecycle (e.g. in different kinds of testing, or even in the operational state of a software or Webapp), we can measure, evaluate and control the external quality of these late products,
- But assuring **external quality** is not usually sufficient to assure **quality in use**.

Perspectives of Quality: ISO 9126-1

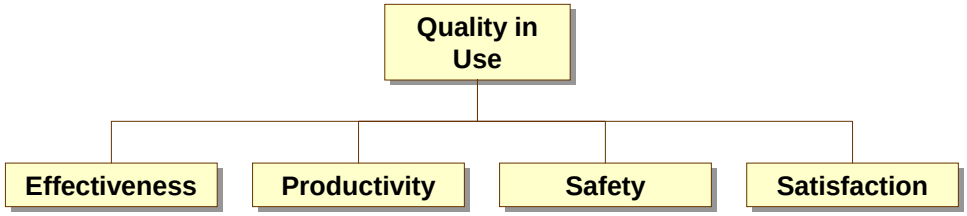
- **Quality in Use** is specified by a quality model (four characteristics),
- It can be measured and evaluated by the extent to which the software or Web application meets specific user's needs in the actual, real, specific context of use.
- Regarding the spirit of this standard, quality in use is the end user's view of the quality of a running system containing software, and is measured and evaluated in terms of the result of using the software, rather than by properties of the software itself.

Perspectives of Quality: ISO 9126-1

- Attributes of internal and external quality of a software product are rather **the cause**, attributes of quality in use rather **the effect**.
- QinU evaluates the degree of excellence, and can be used to validate the extent to which the software or Web meets specific user needs.
- Considering appropriate attributes of the software (or Web) for internal quality is a prerequisite to achieve the required external behavior, and considering appropriate attributes of the software to external behavior is a prerequisite to achieve quality in use

Quality in Use Model

- **Quality in use** is the final user's view of quality
  - similar to the definition of Usability in ISO 9241-11
- *The capability of software product to enable specified users to achieve specified goals with effectiveness, productivity, safety and satisfaction in specified context of use.* ISO/IEC 9126-1:2001.



Quality in Use Characteristics

- **Effectiveness**

  The capability of software product to enable users to achieve specified goals with accuracy and completeness in a specified context of use.

- **Productivity**

  The capability of software product to enable users to expend appropriate amounts of resources in relation to the effectiveness achieved in a specified context of use.

- **Satisfaction**

  The capability of software product to satisfy users in a specified context of use.

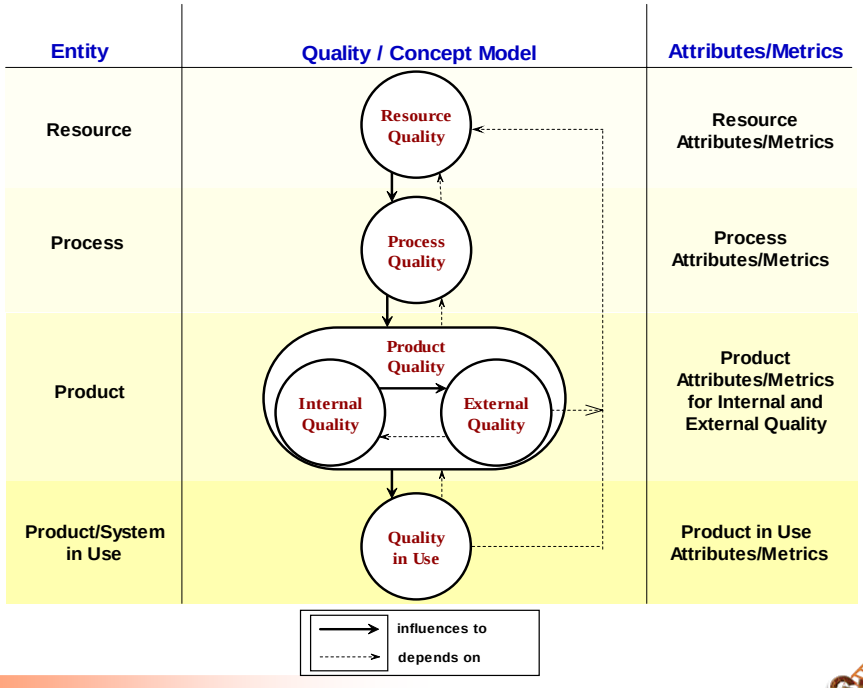
  - Satisfaction is the user's response to the interaction with the product (e.g. a website), and include attitudes towards use of the product.

Quality in Use Model

- **Instance of QinU MODEL with associated Attr.**

1. Quality in Use
    - 1.1 Effectiveness
        - *1.1.1 Task Effectiveness (TE)*
        - *1.1.2 Task Completeness (TC)*
    - 1.2 Productivity
        - *1.2.1 Efficiency related to Task Effectiveness (ETE)*
        - *1.2.2 Efficiency related to Task Completeness (ETC)*
    - 1.3 Satisfaction

Dependencies



What is Web Quality?

WebApps "*involve a mixture between print publishing and software development, between marketing and computing, between internal communications and external relations, and between art and technology*" [Powell 97]

- We argue the three ISO views (and quality models) are also applicable to a great extent to intermediate and final life-cycle Web products.
- Like any software line production, the Web lifecycle involves different stages of its products whether in early phases as inception and development, or late phases as deployment, operation and evolution.

What is Web Quality?

- Thus, to the general question if we can apply the same ISO internal and external quality, and quality in use models, the natural answer is yes
- However, to the more specific question whether we can use the same six-prescribed quality characteristics for internal and external quality, and the four characteristics for quality in use, our answer is yes for the latter, but some other considerations might be taken into account for the former.

What is Web Quality?

- The very nature of WebApps is a mixture of information (media) contents, functionalities and services.
- We argue that the six quality characteristics (i.e., Usability, Functionality, Reliability, Efficiency, Portability, and Maintainability) are not well suited (or they were not intended) to specify requirements for **information quality**.
- A new Characteristic related with information CONTENTS is needed

What is Web Quality?

Content characteristic has four major subconcepts not covered by the six-prescribed ISO characteristics

- **Information Accuracy** is the extent to which information is correct, unambiguous, authoritative (reputable), objective, and verifiable.
- **Information Suitability** is the extent to which information is appropriate (appropriate coverage for the target audience), complete (relevant amount), concise (shorter is better), and current.
- **Accessibility** emphasizes the importance of technical aspects of WebApps in order to make Web contents more accessible for users with various disabilities
- **Legal Compliance**. The capability of the information product to adhere to standards, conventions, and legal norms related to contents and intellectual property rights.

To Remark

- **Quality Assurance**, as a support process, should be a priority to main processes of Software and Web production lines
- Sw / Web **Quality Assurance** implies a set of planned and systematic activities in order to guarantee **products** (processes,...) will meet explicit and implicit **Quality** requeriments
- **Quality Models** can be core pieces for Quality and Metric Plans in QA.

To Remark

- The meaning of **quality** is not simple and atomic, but a multi-dimensional and abstract concept.
  - Not absolute but rather contextual
- Common practice assesses quality by means of the quantification of lower abstraction concepts, such as **attributes** of entities;
- The measurement of attributes can be made by means of **metrics**
- **Quality** and its **attributes** can be interpreted by means of **indicators**

To Remark

- For the shake of clarity and handling, the ISO general-purpose quality model contains a minimum amount of characteristics by which every kind of software can be evaluated;
  - And the Content side of WebApps?
- Define and instantiate a **model** depend on various considerations ...
- Product Quality is the **means**, Quality in Use the objective, the **ultimate goal**.

References

Olsina, L; Covella, G. Rossi, G; 2006, *Web Quality*; (Book Chapter of Web Engineering) Springer, E. Mendes & N. Mosley (Eds). pp 109-142. ISBN 3-540-28196-7

Olsina, L.; Rossi, 2002, *Measuring Web Application Quality with WebQEM*, In IEEE Multimedia Magazine, ISSN 1070-986X, Vol. 9, Nº 4, pp. 20-29

Covella, G., Olsina, L; 2006, *Assessing Quality in Use in a Consistent Way*, To appear in proceed. of ACM. Int'l Congress on Web Engineering, (ICWE'06), Stanford, USA

Outline

- Introduction to Quality and Quality in Use
  - For Software and Web
- **Conceptual Base for Metrics & Indicators**
- Evaluation Process
- Goal-oriented Measurement and Evaluation Frameworks
  - GQM Paradigm
  - INCAMI Framework
    - INCAMI Components & Tool
- Conclusions

Main Conceptual Base

Concepts related to:

- Model-centred Non-functional Requirements (NFR)
- Metrics and Measurement (M&M)
- Indicators and Evaluation (I&E)

Ontology: Building Process

An ontology is an explicit specification of a conceptualisation [Gruber, 95]

Ontology = Concepts + Properties + Relationships + Axioms

- We followed the main steps of METHONTOLOGY:
  - ***Specification*:** The ontology's goal, scope, granularity are specified, as well as the sources of knowledge
  - ***Conceptualisation*:** helps to organize and structure the acquired knowledge using an external representation language -independent of implementation languages.
  - ***Implementation***: It consists in implementing the conceptual model into a formal language like RDF/S (*Resource Description Framework/Schema*), OWL ...
  - ***Evaluation:*** A technical judgment of the ontology.

Main Sources of Knowledge

- Many ISO Standards deal with these concepts, e.g.:
  - **Quality: internal and external quality models, and quality in use model** for sw. (ISO 9126-1:2001),
  - **Evaluation:** The **evaluation process** (ISO 14598:1998),
  - **Measurement:** The **measurement process** (ISO 15939:2002)

Very often, we have observed a lack of consensus in the terminology (same terms different meaning, different terms with similar meaning, absent terms, etc.)

Main Sources of Knowledge

- We present the main terms and their meanings coming from an ontological study we made for this domain [Olsina et al 2002/04]
- We explicitly and formally specified the main concepts, properties and relationships. Some terms are:
  - **Information Need, Calculable Concept, Entity, Attribute, Metric, Scale, Unit, Measurement Method, Software Tool, Indicator, Elementary Indicator, Elementary Model** and **Decision Criteria,** among others

Quality-in-Use Case Study

- For illustration purposes, we will use a quality-in-use example for the E-learning domain.
- Quality in use is the combined effect of the internal and external quality sub-concepts (e.g., usability, functionality, reliability, and efficiency characteristics) for the end user.
- It can be measured and evaluated by the extent to which specified users can achieve specified goals with effectiveness, productivity, safety, and satisfaction in specified contexts of use.

Quality-in-Use Case Study

- When designing and documenting quality in use requirement, measurement and evaluation processes, at least the following information is needed
  - Descriptions of the components of the context of use including user type, equipment, environment, and application tasks
    - i.e., tasks are the sub-goals undertaken to reach an intended goal by a user group type
  - Quality in use **metrics** and **indicators** for the intended purpose and information need.

Quality-in-Use Case Study

- The "*QPlus Virtual Campus*" Web application (www.qplus.com.ar/) is being employed as support to a math preparatory course in the Engineering School at UNLPam since 2003
- Four tasks and six pre-enrolled students were chosen for testing purposes (in early 2004).
- We next use this case study as Proof of Concepts

Covella, G., Olsina, L; 2006, *Assessing Quality in Use in a Consistent Way*, To appear in proceed. of ACM. Int'l Congress on Web Engineering, (ICWE'06), Stanford, USA

Main Conceptual Base

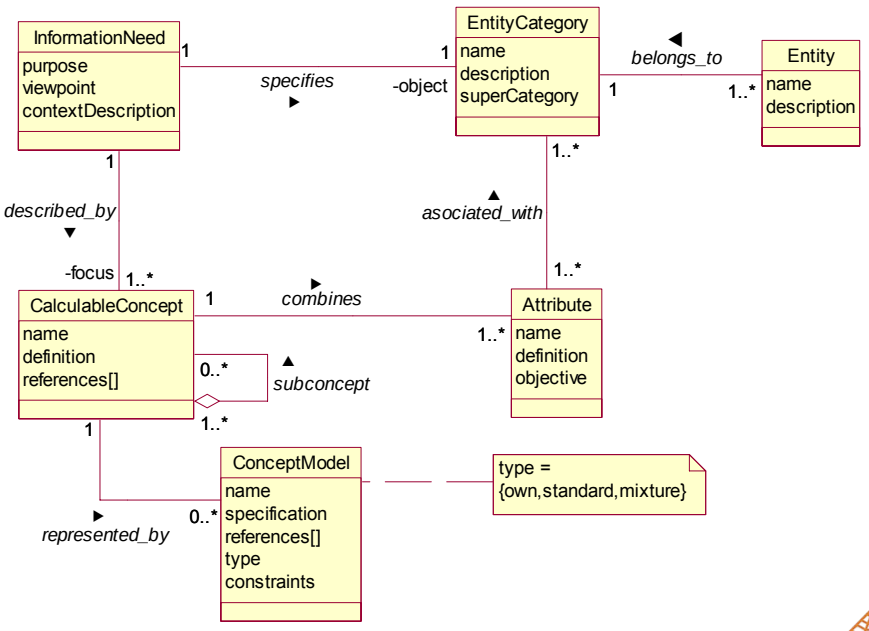
Concepts related to:

- Model-centred Non-functional Requirements (NFR)
- Metrics and Measurement (M&M)
- Indicators and Evaluation (I&E)

Concepts for NFR

- **Information Need**
- **Entity Category/Entity**
- **Quality, Quality in Use**
  - **CALCULABLE CONCEPT**
- **Quality Model, Quality in Use Model**
  - **CONCEPT MODEL**
- **Attribute**

Model for NFR



Term: Information Need

- **INFORMATION NEED**
  - Insight necessary to manage objectives, goals, risks, and problems [ISO-15939].
  - An information need is described by one or more Calculable Concepts (Quality, Quality in Use, etc.)
  - For example, for an academic organization, a basic information need may be "understand the quality in use of the YY e-learning application that supports courses tasks for pre-enrolled students".
    - Purpose = Understand
    - Viewpoint = pre-enrolled students
    - Calculable Concept = Quality in Use
  - So, an *entity category*, which is the *object* under analysis, and the *calculable concept*, which is the *focus* of *the information need* have to be defined.

Terms: Entity Category and Entity

- **ENTITY CATEGORY**
  - Object category that is to be characterized by measuring its attributes
  - High Level Categories: Product, Process, Resource, Project, Service ...
- **ENTITY** (syno Object)
  - A concrete object that belongs to an entity category.
  - Example: given the *entity category* (i.e., an e-learning application, which its *superCategory* is a product) a concrete object that belongs to this category is the "QPlus Virtual Campus" Web application.

Term: Calculable Concept

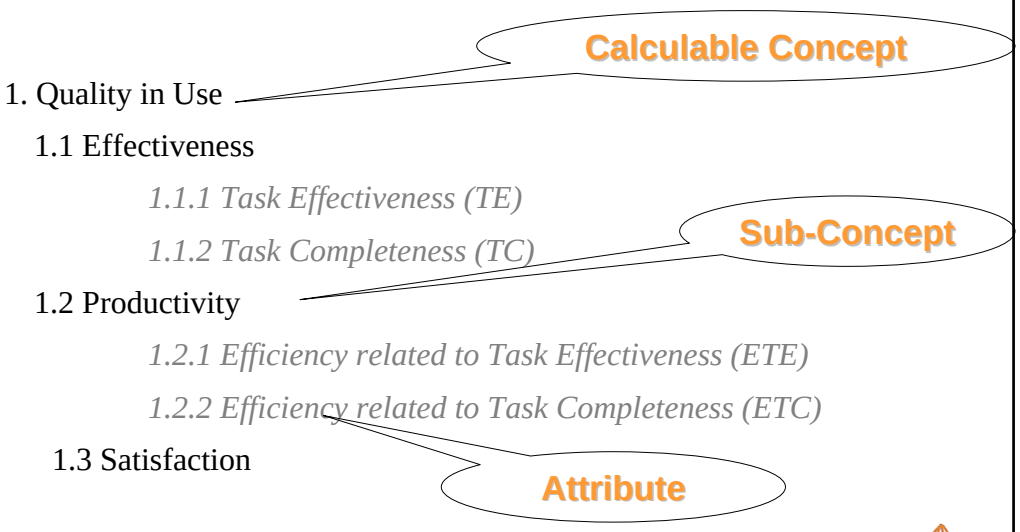
- **CALCULABLE CONCEPT** (syno Measurable Concept)
  - Abstract relationship between attributes of entities categories and information needs.
  - To our example, the *calculable concept* is "quality in use" and can have *sub-concepts* such as "effectiveness", "productivity", "safety", and "satisfaction".
  - For instance, the "effectiveness" sub-concept is defined as "the capability of the software product to enable users to achieve specified goals with accuracy and completeness in a specified context of use".
  - The calculable concept can be *represented by* a *concept model*.

Term: Concept Model

- **CONCEPT MODEL**
  - The set of sub-concepts and the relationships between them, which provide the basis for specifying the concept requirement and its further evaluation or estimation.
  - the concept model *type* can be either
    - a standard-based model (ISO, etc.)
    - an organization own-defined model, or
    - a mixture of both.
  - The concept model used in the example is of "standard" *type* that is based on the ISO quality-in-use model, and the *specification* is shown in the next slide
    - note the model shows also *attributes* combined to the *sub-concepts*.

Quality in Use Model

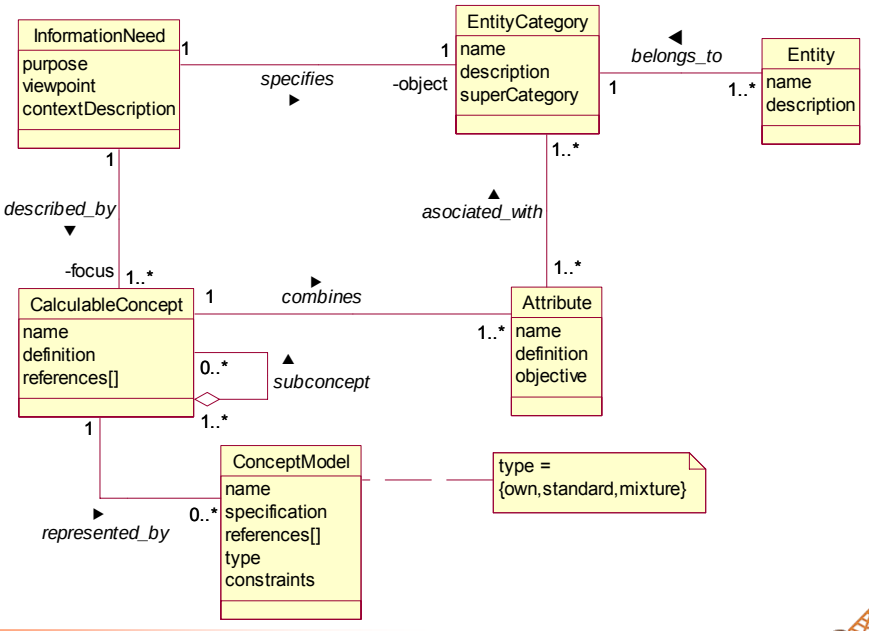
- **Instance of QinU MODEL with associated Attributes**



Term: Attribute

- **ATTRIBUTE** (syno Property, Feature)
  - A measurable physical or abstract property of an entity category.
  - Note that the selected attributes are those relevant properties for the defined information need.
  - The previous slide shows attribute *names* such as “Task Effectiveness”, “Task Completeness”, among others.
  - An attribute can be quantified (measured) by one or more direct or indirect metrics.

Model for NFR



Model for External Quality

- **1. Usability**
- 1.1 Global Site Understandability
  - 1.1.1 Book-domain Organization Scheme
    - 1.1.1.1 *Table of Contents*
    - 1.1.1.2 *Alphabetical Subject Index*
  - 1.1.2 Quality of Labeling System
  - 1.1.3 Guided Tour for First Time Visitors
- 1.2 Operability
  - 1.2.1 Presentation Permanence and Stability of Main Controls
    - *1.2.1.1 Direct Controls Permanence*
    - *1.2.1.2 Indirect Controls Permanence*
    - *1.2.1.3 Stability*

Model for External Quality

- **2. Functionality**
  - 2.1 Searching Issues
    - 2.1.1 Search Type
      - 2.1.1.1 Quick Search
        - (by author, title, ISSN, ISBN, etc.)
      - 2.1.1.2 Advanced Search
    - 2.1.2 Search Tolerancy
      - 2.1.2.1 Writing Error Tolerancy
        - Spell and Grammar Error Recognition
        - Shows Synonyme Dictionary
      - 2.1.2.2 Writing Variation Tolerancy
        - Composed Last Names
        - Hyphenized Descriptors
        - Synonym Recognition
        - Plural / Singular Recognition

Model for External Quality

- **3. Site Reliability**
  - 3.1 Link Maturity
    - 3.1.1 Link Errors
      - 3.1.1.1 *Broken Links*
      - 3.1.1.2 *Invalid Links*
      - 3.1.1.3 *Unimplemented Links*
      - .....

Main Conceptual Base

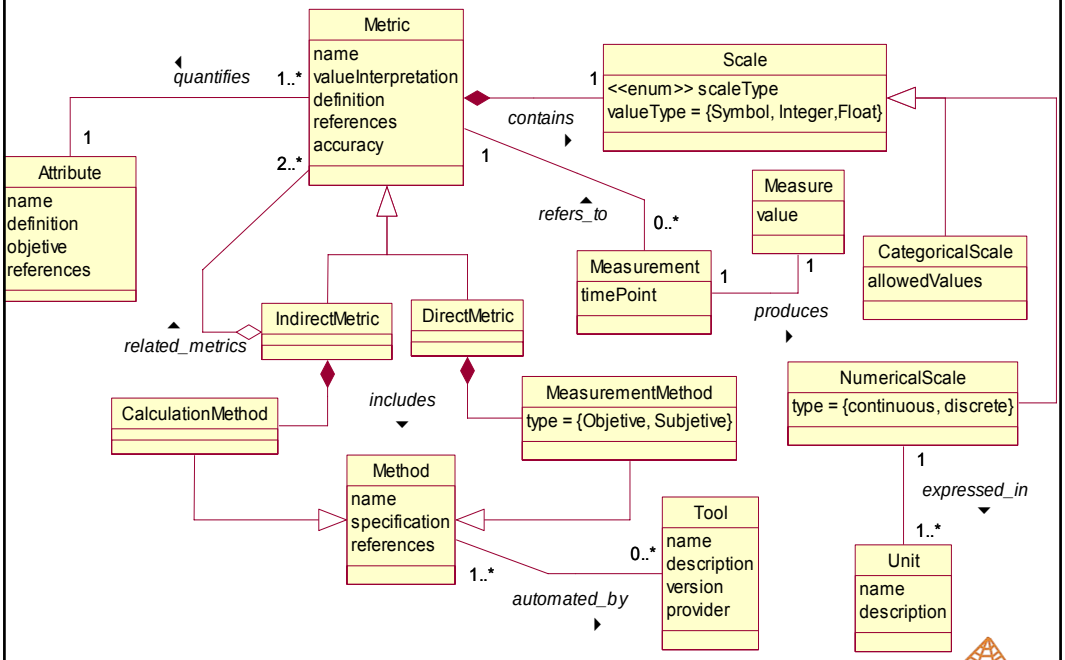
Concepts related to:

- Model-centred Non-functional Requirements (NFR)
- Metrics and Measurement (M&M)
- Indicators and Evaluation (I&E)

Concepts for Metrics

- **Attribute**
- **Metric**
  - Direct
  - Indirect (Formula)
- **Scale**
  - Scale Type
  - Categorical, Numerical (Unit)
- **Method**
  - Of Measurement, of Calculation (Sw Instrument)
- **Measurement**
- **Measure**

Model for Metric



Terms: Metric, Direct Metric

- **METRIC**
  - The defined measurement or calculation method and the measurement scale
    - Similar to the [ISO 14598-1] definition.
- **DIRECT METRIC** (syno Single, Base Metric)
  - a metric of an attribute that does not depend upon a metric of any other attribute

Terms: Indirect Metric, Function

- **INDIRECT METRIC** (syno Hybrid, Derived Metric)
  - a metric of an attribute that is derived from metrics of one or more other attributes.
- **FUNCTION** (syno Formula, Algorithm, Equation)
  - algorithm or formula performed to combine two or more metrics.

Example of Metric for TC

1. Quality in Use
   - 1.1 Effectiveness
     - *1.1.1 Task Effectiveness (TE)*
     - *1.1.2 Task Completeness (TC)*
   - 1.2 Productivity
     - *1.2.1 Efficiency related to Task Effectiveness (ETE)*
     - *1.2.2 Efficiency related to Task Completeness (ETC)*
   - 1.3 Satisfaction

Example of Metric for TC

- For the "Task Completeness" *attribute* we designed an *indirect metric* that specifies what proportion of the tasks is completed by a given user.
- The metric *name* is "Task Completeness Ratio"; the *formula specification* is TCR= #CT / #PT
- where both #CT ("Number of Completed Tasks"), and #PT ("Number of Proposed Tasks") are *direct metrics*
  - Note that the TCR metric specifies what proportion of the proposed tasks is fully completed by a user; the final metric we used is the average for the six selected users.

Again ...

- **METRIC**
  - The defined measurement or calculation **method** and the measurement **scale**
  - There are two key terms in the above definition: Method and *Scale*. For the latter, two types of scales have been identified, viz. *Categorical* and *Numerical Scales*

Terms: Scale, Scale Type

- **SCALE**
  - a set of values with defined properties [ISO 14598-1].
- **Scale Type**
  - The type of scales depends on the nature of the relationship between values of the scale.
  - The types of scales are commonly classified into *nominal*, *ordinal*, *interval*, *ratio*, and *absolute*.
  - The scale type of measured values affect
    - the sort of arithmetical and statistical operations that can be applied to values (e.g. we can't add numbers in an ordinal scale)
    - the admissible transformations (e.g. M' = a M for a ratio scale)

Terms: Scale, Scale Type

Scale type	Is ranking meaningful?	Are distances between scales the same?	Does the scale include an absolute zero?
Nominal	No	No	No
Ordinal	Yes	No	No
Interval	Yes	Yes	No
Ratio	Yes	Yes	Yes
Absolute	Yes	Yes	Yes

Terms: Scale, Scale Type

Scale type	Examples of suitable statistics	Suitable statistical tests
Nominal	Mode Frequency	Non-parametric
Ordinal	Median Percentile	Non-parametric
Interval	Mean Standard deviation	Non-parametric and parametric
Ratio	Mean Geometric mean Standard deviation	Non-parametric and parametric
Absolute	Mean Geometric mean Standard deviation	Non-parametric and parametric

Terms: Categorical and Num. Scales

- **Categorical Scale**
  - a scale where the measured or calculated values are categories, and cannot be expressed in units, in a strict sense.
- **Numerical Scale**
  - a scale where the measured or calculated values are numbers that can be expressed in units, in a strict sense.

Term: Unit

- **UNIT (for Numerical Scales)**
  - Particular quantity defined and adopted by convention, with which other quantities of the same kind are compared in order to express their magnitude relative to that quantity [ISO-15939]
    - Examples of Unit: LOC, bytes, words, links, tasks ...

Example of Scale for "Task Completeness Ratio" TCR= #CT / #PT

- The *scale type* of the TCR indirect metric is "ratio" represented by a *numerical scale* with a "real" *value type* and in a "continuous" representation form.
- The unit *description* is "completed tasks per proposed tasks by a user".
- In the *formula* intervenes two direct metrics, i.e. #CT, #PT respectively
  - note we can further specify thoroughly the metadata for each direct metric.

Terms: Method, Measurement Method

- **METHOD**
  - logical sequence of operations and possible heuristics, specified generically, for allowing the realisation of an activity description.
- **CALCULATION METHOD**
  - the particular logical sequence of operations specified for allowing the realisation of a formula or indicator description by a calculation.

Term: Measurement Method

- **MEASUREMENT METHOD** (syno Counting Rule, Protocol)
  - the particular logical sequence of operations and possible heuristics specified for allowing the realisation of a metric description by a measurement.
  - The **type of a measurement method** can be either
    - *subjective* i.e. where the quantification involves human judgement, or
    - *objective* i.e. where the quantification is based on numerical rules.
  - Usually an objective measurement method type can be automated or semi-automated by a software tool.

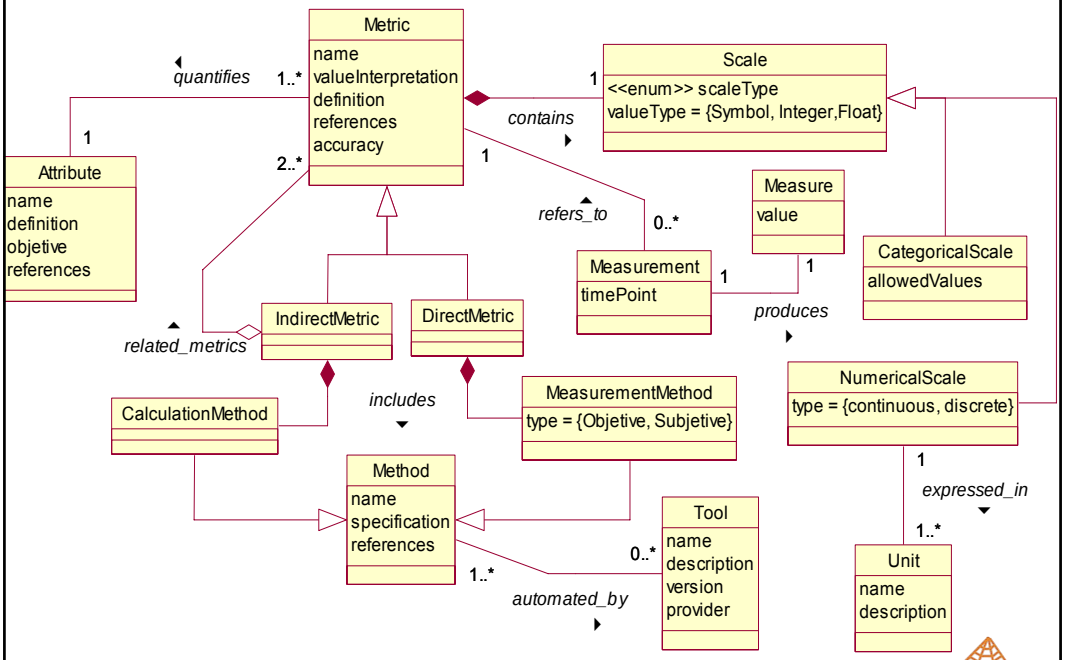
Term: Software Tool

- **SOFTWARE TOOL** (syno Software Instrument)
  - it is a tool that automates partially or totally a measurement or calculation method.
  - **Doctor HTML**: Imagiware [http://www2.imagiware.com/RxHTML];
  - **LIFT**: UsableNet.com [http://www.usablenet.com/];
  - **LinkBot:** WatchFire [http://www.watchfire.com/solutions/linkbot.asp];
  - .....

Terms: Measurement & Measure

- **MEASUREMENT**
  - activity that uses a metric definition in order to produce a measure's value.
- **MEASURE**
  - the number or category assigned to an attribute of an entity by making a measurement [ISO 14598-1]
  - A *measurement* activity must be performed for each metric that intervenes in the project.
  - It allows recording the *date/time stamp*, the *collector information* in charge of the measurement activity, and for the *measure,* the "actual" or "estimated" value *type* and the yielded *value* itself.

Model for Metric



To Remark

Metrics are welcome when they are clearly needed and easy to collect and understand
Pfleeger

- A Metric specifies in the numerical (formal) world a specific mapping of an entity's attribute of the empirical world
- A Metric **can not interpret** itself a calculable concept

Need of INDICATORS in order to get contextual information

Indicators are ultimately the foundation for interpretation of information needs and decision-making.

Main Conceptual Base

Concepts related to:

- Model-centred Non-functional Requirements (NFR)
- Metrics and Measurement (M&M)
- Indicators and Evaluation (I&E)

Concepts for Evaluation

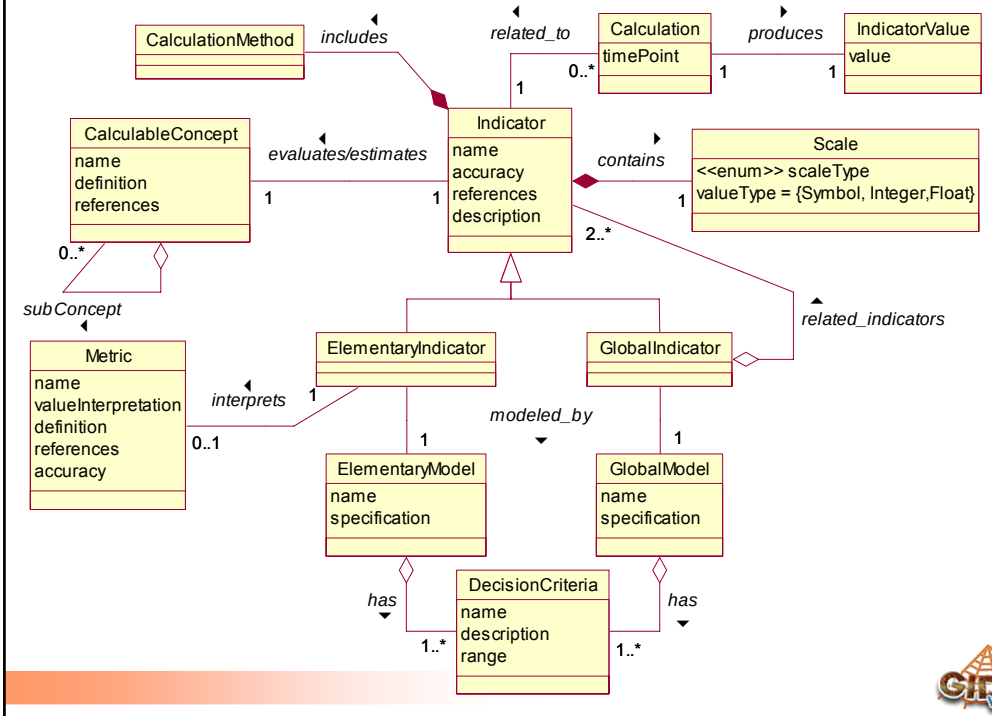
- **Information Need**
- **Concept Model**
- **Calculable Concept**
- **INDICATOR**
  - **Elementary (interprets Metric's measure)**
  - **Global (calculates Concept Model)**
- **ELEMENTARY and GLOBAL MODEL**
- **DECISION CRITERIA**
- **CALCULATION, INDICADOR VALUE**

Metric & Indicator

- The metric **m** represents the mapping **m: A -> X**, where **A** is an empirical attribute of an entity category (**the empirical world**), **X** the variable to which categorical or numerical values can be assigned (**the formal world**), and the arrow denotes a mapping.

- The indicator represents a **new mapping** coming from the interpretation of the metric's value (**formal world**) into the new variable to which categorical or numerical values can be assigned (**the new formal world**).
  - In order to do this mapping a **model** and **decision criteria** for a specific user information need is considered.

Indicators



Terms: Indicator, Elementary ...

- **INDICATOR** (syno Criterion)
  - the defined calculation method and scale in addition to the model and decision criteria in order to provide an estimate or evaluation of a calculable concept with respect to defined information needs.
- **Elementary Indicator** (syno Elementary Criterion)
  - an indicator that does not depend upon other indicators to evaluate or estimate a calculable concept.

Example of Elementary Indicator

- Considering the e-learning case study, an elementary indicator for each attribute of the concept model, i.e. for each leaf of the requirement tree can be defined.
- For instance, for the *Task Completeness* attribute the *name* of the elementary indicator is "*Task Completeness Performance Level"* (TC_PL).
- The elementary indicator interprets the metric's value of the attribute.
- For this end, an *elementary model* is needed.

Term: Elementary Model

- **ELEMENTARY MODEL**
  - algorithm or function with associated decision criteria that model an elementary indicator.

The *specification* of the elementary model can look like this:

TC_PL = 100% if TCR = 1; TC_PL = 0% if TCR <= $X_{min;}$

where X_{min} is some agreed lower threshold as 0.45;

otherwise TC_PL = TCR * 100 if X_{min} < TCR < 1

- Notice that, like a metric, an indicator has a *Scale.*
- To this case, we considered a *numerical scale* where the *Unit* can be a normalized "percentage" unit.

Terms: Decision Criteria, Range

- **DECISION CRITERIA** (syno Acceptability Levels)
  - thresholds, targets, or patterns used to determine the need for action or further investigation, or to describe the level of confidence in a given result [ISO 15939]
- **RANGE**
  - threshold or limit values that determine the acceptability levels.

Example: Decision Criteria, Range

- The decision criteria that a model of an indicator may have are the agreed **acceptability levels** in given **ranges** of the **scale**;
- E.g., it is "unsatisfactory" if the *range* (regarding *lower_threshold* and *upper_threshold*) is "0 to 45" respectively; "marginal" if it is "greater than 45 and less or equal than 70"; otherwise, "satisfactory".
  - A *description* or interpretation for "marginal" is that a score within this range indicates a need for improvement actions.
  - An "unsatisfactory" rating means change actions must take high priority.

Terms: Global Indicator, Model

- **GLOBAL INDICATOR** (syno Global Criterion)
  - an indicator that is derived from other indicators to evaluate or estimate a calculable concept.
- **GLOBAL MODEL** (syno Aggregation Model, Scoring Model or Function)
  - algorithm or function with associated decision criteria that model a global indicator.

Terms: Global Indicator, Model

- Regarding partial and global indicators, an *aggregation model* and *decision criteria* must be selected.
- The quantitative aggregation and scoring models aim at making the evaluation process well structured, objective, and comprehensible to evaluators.
- E.g., if our procedure is based on a "linear additive scoring model", the aggregation and computing of partial/global indicators (P/GI), considering relatives *weights* (W) is based on the following *specification*:

Global Evaluation: Scoring Models

- Linear Additive Scoring Model (T.Gilb)

  Partial/Global Indicator = ∑ (Weight x Elementary Indicator)

  $P/GI = W_1 EI_1 + ...+ W_n EI_n$

  where $W_1 + ...+ W_n = 1$;

- Non-linear Multi-criteria Scoring Model (LSP)

  (Weighted Power Mean Model - J. Dujmovic)

  $P/GI(r) = (W_1 EI^r_1 + W_2 EI^r_2 + ... + W_m EI^r_m)^{1/r}$

  - ✓ Simultaneity
  - ✓ Replaceability,
  - ✓ Neutrality,
  - ✓ Symmetric and Asymmetric Relationships

Example of Indicators

Code	Global/Partial Indicator Name	Elementary Indicator Name	Weight	Actual Value
1.	Quality in Use Level			57.43
1.1	Effectiveness Level		0.33	59.67
1.1.1		*Task Effectiveness Performance Level*	0.5	54.17
1.1.2		*Task Completeness Performance Level*	0.5	65.58
1.2	Productivity Level		0.33	51.87
1.2.1		*Efficiency Level related to Task Effectiveness*	0.5	49.76
1.2.2		*Efficiency Level related to Task Completeness*	0.5	54.04
1.3	Satisfaction Level		0.33	87.08
1.3.1		*Calculated Satisfaction Level*	1	87.08

In the case study we used the LSP model for calculation,
but if we'd use the additive model to calculate $PI_{1.1}$

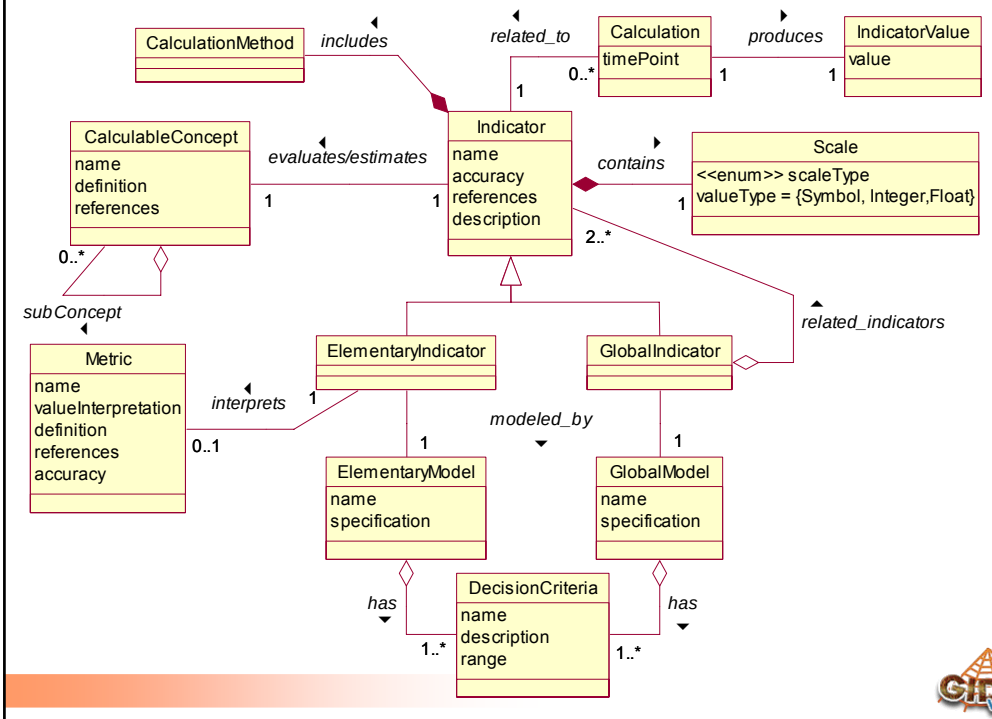
$PI_{1.1} = W_{1.1.1} EI_{1.1.1} + W_{1.1.2} EI_{1.1.2}$

gives 60.29 instead of 59.67

Terms: Calculation, Indicator Value

- **CALCULATION** (syno Computation)
  - activity that uses an indicator definition in order to produce an indicator's value.
- **INDICATOR VALUE**
  - The number or category assigned to a calculable concept by making a calculation.

Indicators



To Remark

Metrics are welcome when they are clearly needed and easy to collect and understand

Usefulness of Metrics

- Data coming from a measurement (objective, subjective)
- Mapping between an empirical world (entity attribute) to a numerical, formal world
- Heuristic operationalisation
- To serve as a "base" to Quantitative Methods for Evaluation and Prediction.
- A metric (and its measures) CAN NOT interpret by itself a calculable concept (Need of INDICATORS)

To Remark

Indicators are ultimately the foundation for interpretation of information needs and decision-making.

Usefulness of Indicators

- Mapping from a numerical world to another
- To serve as a base to quantify Calculable Concepts for an Information Need
- Indicators give contextual Information/Knowledge
- Indicators give contextual information for decision-making (Analyses and Recommendations)

References

Olsina, L.; Martín, M., 2004, *Ontology for Software Metrics and Indicators*, In Journal of Web Engineering, Rinton Press, US, Vol 2 Nº 4, pp. 262-281, ISSN 1540-9589

Molina, H; Papa, F.; Martin, M.; Olsina, L; 2004; *Semantic Capabilities for the Metrics and Indicators Cataloging Web System*. In: *Engineering Advanced Web Applications*, Matera M. Comai S. (Eds.), Rinton Press Inc., US, pp. 97-109, ISBN 1-58949-046-0

Martín, M.; Olsina, L., 2003, *Towards an Ontology for Software Metrics and Indicators as the Foundation for a Cataloging Web System*, In IEEE Computer Society (1st LA-WEB) Sant. de Chile, pp 103-113, ISBN 0-7695-2058-8.

Outline

- Introduction to Quality and Quality in Use
  - For Software and Web
- Conceptual Base for Metrics & Indicators
- **Evaluation Process**
- Goal-oriented Measurement and Evaluation Frameworks
  - GQM Paradigm
  - INCAMI Framework
    - INCAMI Components & Tool
- Conclusions

Evaluation Process

- An Evaluation Process (e.g. ISO 14598) is a generic and abstract specification of processes and activities, inputs and outputs, and check points.
  - customizable to different needs given a concrete evaluation process of software and Web quality products
- An Evaluation Process does NOT prescribe nor recommend specific procedures, methods and tools to perform the activities
  - It represents a generic framework.

Evaluation Process

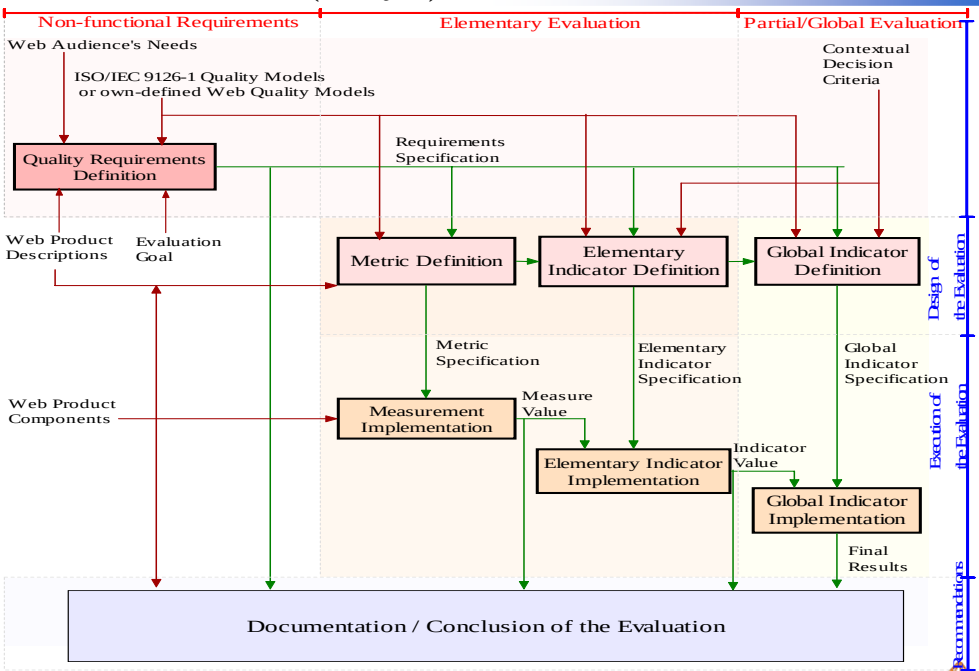
- **Main Processes defined in the ISO 14598:1998 Std.**
  - Establishment of Evaluation Requirements
  - Specification of the Evaluation
  - Design of the Evaluation
  - Execution of the Evaluation, and
  - Conclusion of the Evaluation

Web QEM´s main Steps

WebQEM: Web Quality Evaluation Method [Olsina 1999]

- Quality Requirements Definition
  - Evaluation Goal
  - User Viewpoint (manager, developer, visitor)
  - Nonfunctional Requirements Definition and Specification
- Measurement and Elementary Evaluation
  - Regarding Design and Execution stages
- Partial/Global Evaluation
  - Regarding Design and Execution stages
- Conclusion of the Evaluation
  - Regarding Recommendations

valuation Process (WebQEM)



Outline

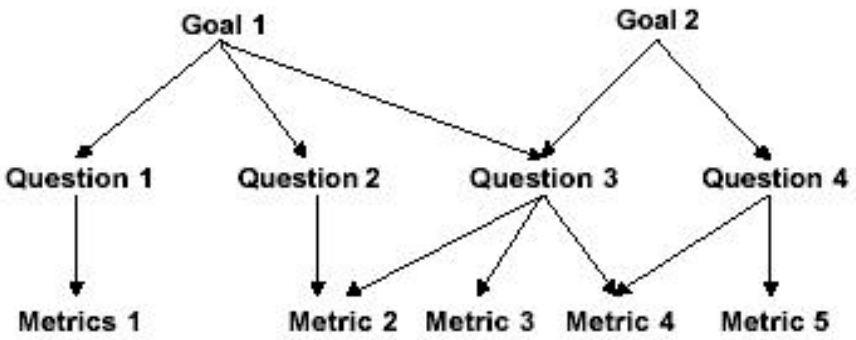
- Introduction to Quality and Quality in Use
  - For Software and Web
- Conceptual Base for Metrics & Indicators
- Evaluation Process
- **Goal-oriented Measurement and Evaluation Frameworks**
  - **GQM Paradigm**
  - INCAMI Framework
    - INCAMI Components & Tool
- Conclusions

What is GQM: Model, Approach?

- Measurement approach aimed to reach measurement goals in a sw organization
  - something more than collecting useful data...
- Top-down approach useful to determine:
  - Measurement objectives (for products, processes, etc.)
  - Framework to decide exactly what to measure
- Bottom-up approach “useful” to interpret data from metrics and objectives
- Brief History:
  - 1984: created at Universidad de Maryland (by Basili et al)
  - 1992: Kaiserslautern University (Rombach)
  - 1996: Fraunhofer Institute for Experimental Software Engineering
  - Strong adoption in the industry for measurement programs

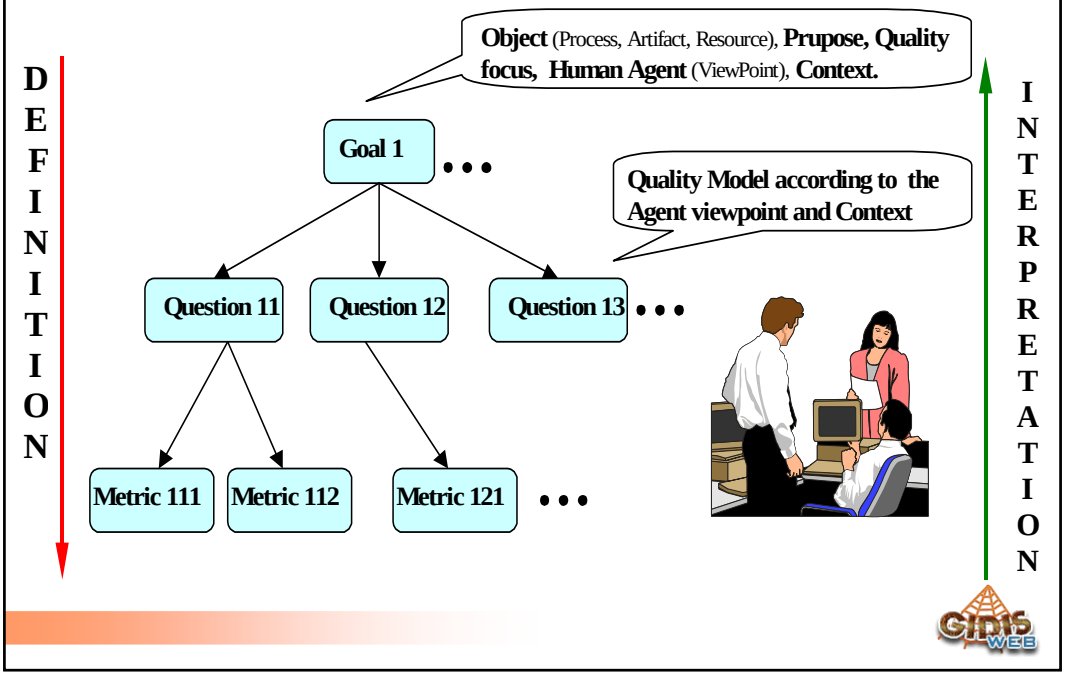
GQM Model: Structure

Goal / Question / Metric



[Basili-88]

GQM Model: Structure



GQM Model: Structure

- Conceptual Level (Goal)
  - According to a quality model, a goal is defined for an object (entity), for a variety of reasons, from one or many points of view, and relative to a particular context (organization, business or project goal).
- Operational Level (Question)
  - A set of questions is used to characterize the way the assessment/achievement of a specific goal is going to be performed based on some quality model.
  - Questions try to characterize the object of measurement with respect to a selected quality issue and to determine its quality from the selected viewpoint.
- Quantitative Level (Metric):
  - A set of metrics is associated with every question in order to answer it in a measurable way.

Goal Template: Examples

Goal for a Product:

- Analyze the **Web site** (entity, object)
- With the purpose of **understand** (objective, purpose)
- With regard to the **link reliability** (quality focus)
- From the **final user** viewpoint (viewpoint)
- In the context of the **X project** (context, environment)

Goal for a Process:

- Analyze the **testing process** (object)
- With the purpose of **improve** (objective)
- With regard to the **effectiveness** (quality focus)
- From the **tester** viewpoint (viewpoint)
- In the context of the **Y project** (context)

GQM: Example

Purpose **Characteristic**	*Understand* *Link Reliability*
Entity **Viewpoint**	*Static Pages of a Web Site* *Final User*
Question 1.1	*What is the level of internal and external broken links (physical error)?*
Metric 1.1.1	*Percentage of Internal Broken Links*
Metric 1.1.2	*Percentage of External Broken Links* *Frequency of Broken Links per Hit Pages*
Question 1.2 **Metric 1.1.3**	*What is the level of invalid links (logical error)?*
Metric 1.2.1	*Percentage of Invalid Links*
.....	.....

Basic Steps for GQM

- Develop objectives at organization, department or project level for quality, quality in use, etc.
- Generate questions (based on templates, models, previous experiences) in order to define objectives in an operationalisable way
- Specify useful metrics for answer the questions
- Develop mechanisms (procedures, tools) to data collection
- Collect, validate, analyze data in order to get feedback for further corrective, improvement actions
- Analyze post-mortem data for assessing goals compliance, etc.

To Remark about GQM

- GQM is a useful approach to decide what to measure.
- Measurement must be oriented to goals
  - Allows decision-makers to choose those metrics related to the most important objectives of the more urgent problems
  - Goal gives context for the analyses and interpretation of data
  - People should be strongly involved in the definition and interpretation
- Data collection should be based on documented or justified reasons
  - Useful and relevant metrics

To Remark about GQM

- GQM is a flexible but a generic approach
- Some weaknesses could be stressed:
  - It lacks an ontological base of metrics and indicators
    - GQM could not assure that measure values (and the associated metadata like scale, unit, measurement method, and so forth) are trustworthy and consistent for ulterior analysis among projects
  - It is not necessarily concept model-oriented
    - quality, quality in use models, etc.
  - Measure interpretation is not well defined for evaluation purposes
    - by means of elementary (and global) indicators
  - When many metrics intervene, it can be hard to perform analyses, interpretations, and recommendations
    - No aggregation model

Outline

- Introduction to Quality and Quality in Use
  - For Software and Web
- Conceptual Base for Metrics & Indicators
- Evaluation Process
- **Goal-oriented Measurement and Evaluation Frameworks**
  - GQM Paradigm
  - **INCAMI Framework**
    - INCAMI Components & Tool
- Conclusions

INCAMI Approach: Introduction

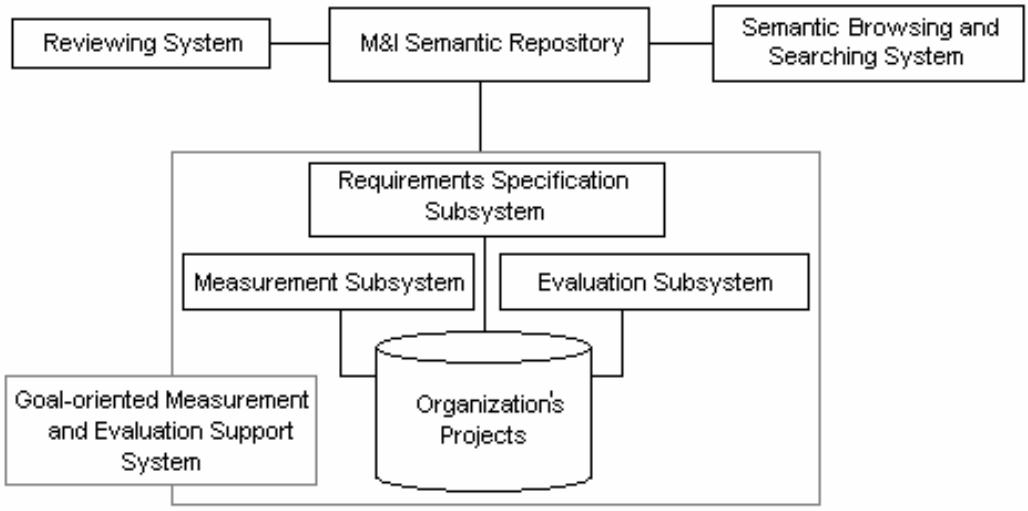
INCAMI (Information Need, Concept model, Attribute, Metric, and Indicator) [Olsina, Molina, Papa]

- INCAMI is a conceptual framework useful for NFR, measurement and evaluation processes
  - INCAMI_Tool is the in-progress supporting tool
- It is based on the Ontology of Metrics and Indicators introduced previously [Olsina, Martín, 03, 04]

INCAMI Approach: Introduction

- INCAMI is an approach (similar to GQM) useful for deciding **what** to measure and evaluate.
  - Measurements and evaluations must be oriented to specific goals (information needs) in the context of an organization, project/s, etc.
  - Concept Model-centered for Requirements
    - Quality, Quality in Use, Cost, etc.
  - Useful attributes, metrics and indicators must be selected for a concrete information need
  - Aggregation model-centered for Evaluation
    - Facilitates information needs interpretation and recommendations

M&E Environment



Outline

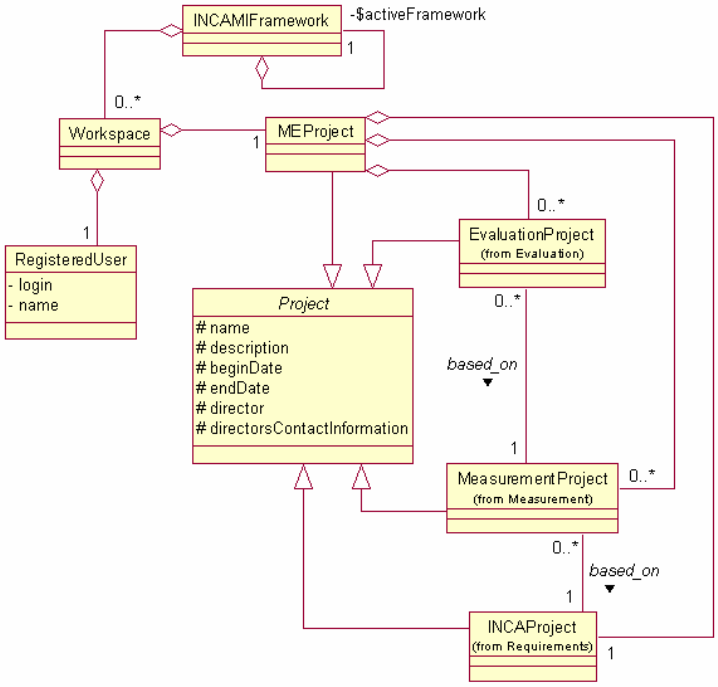
- Introduction to Quality and Quality in Use
  - For Software and Web
- Conceptual Base for Metrics & Indicators
- Evaluation Process
- **Goal-oriented Measurement and Evaluation Frameworks**
  - GQM Paradigm
  - **INCAMI Framework**
    - **INCAMI's Main Components & Tool**
- Conclusions

INCAMI Components

Main Components:

- Definition of Users and Projects
  - requirement, measurement and evaluation projects;
- Non-functional Requirements Definition and Specification
- Measurement Design and Execution
- Evaluation Design and Execution
- Analyses and Recommendations

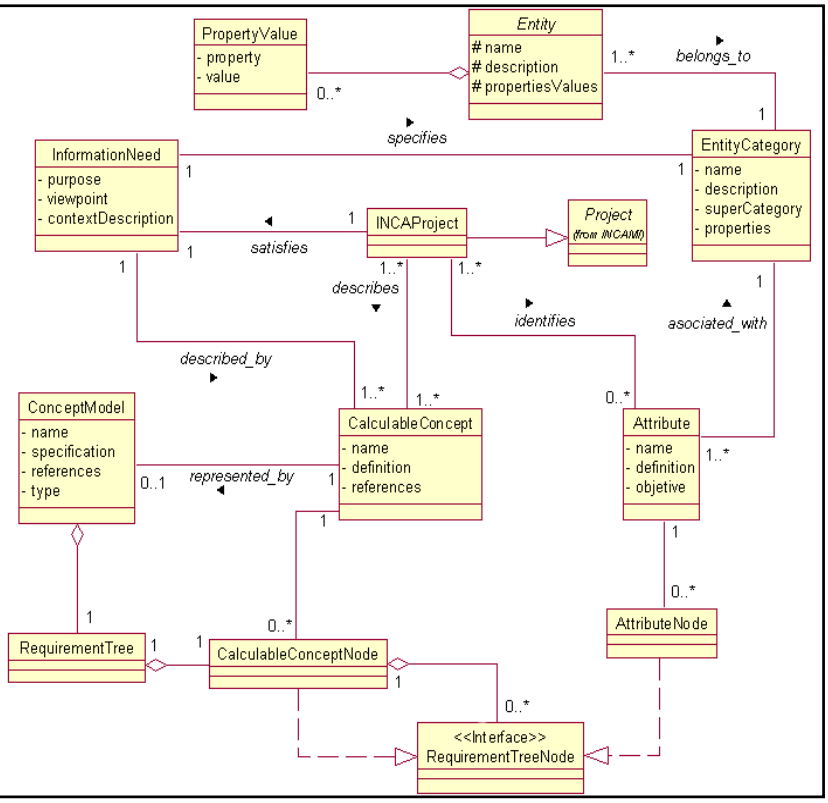
ponent: M&E Projects



ponent: M&E Projects

- **Requirement Project**
  - it is a project that allows specifying non-functional requirements for measurement and evaluation activities.
    - To our example, the project *name* is "QualityInUse_ESchool_04";
    - the *description* is "requirements for evaluating quality in use for a pre-enrolled student group in the Engineering School";
    - with starting date "2004/02/16" and ending date "2004/02/19", and
    - in charge of "Guillermo Covella" with the "covellag@ing.unlpam.edu.ar" *contact* email.

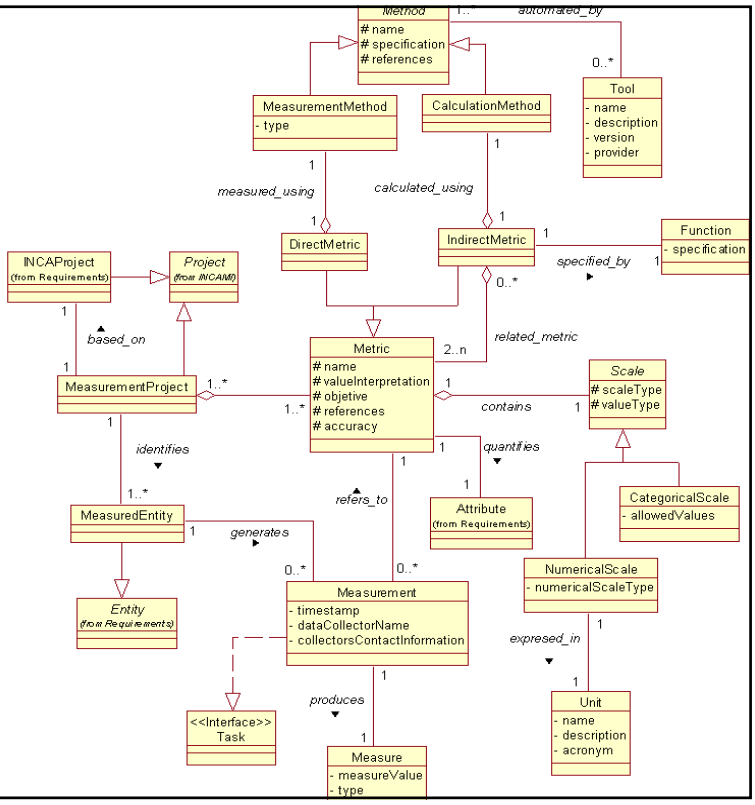
equirements
omponent



nponent: M&E Projects

- **Measurement Project**
  - it is a project that allows, starting from a requirement project, selecting the metrics and recording the values in a measurement process.
    - Once created, with similar information as that of a requirement project, the attributes in the requirement tree can be quantified by *direct* or *indirect metrics*.
    - To a specific measurement project just one metric should be selected for each attribute of the concept model.
    - Many measurement projects can rely on the same requirements,
      - for instance, in a longitudinal analysis. In this case, the starting and ending dates change for each project as likely the person in charge of.

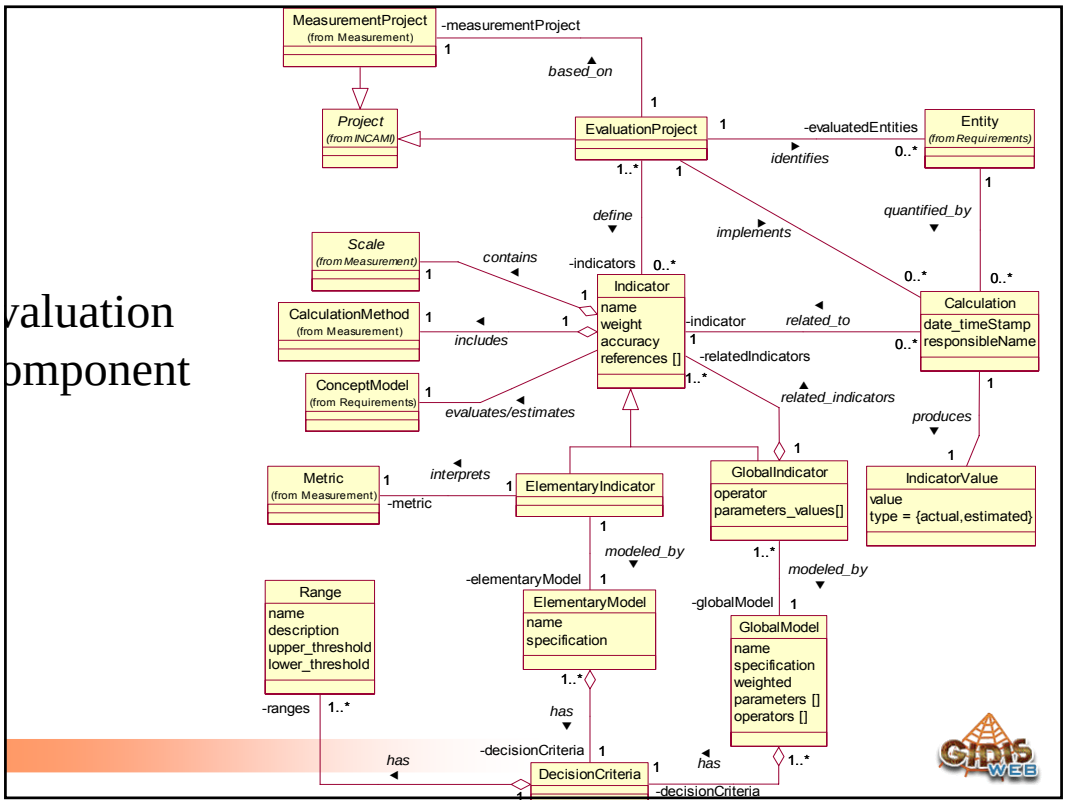
Measurement Component



Component: M&E Projects

- **Evaluation Project**
  - it is a project that allows, starting from a measurement project and a concept model of a requirement project, selecting the indicators and performing the calculation in an evaluation process.
    - Once a measurement project has been created and enacted one o more evaluation projects can be created relying on the recorded measurement data and metadata, and adding information related with *indicators*

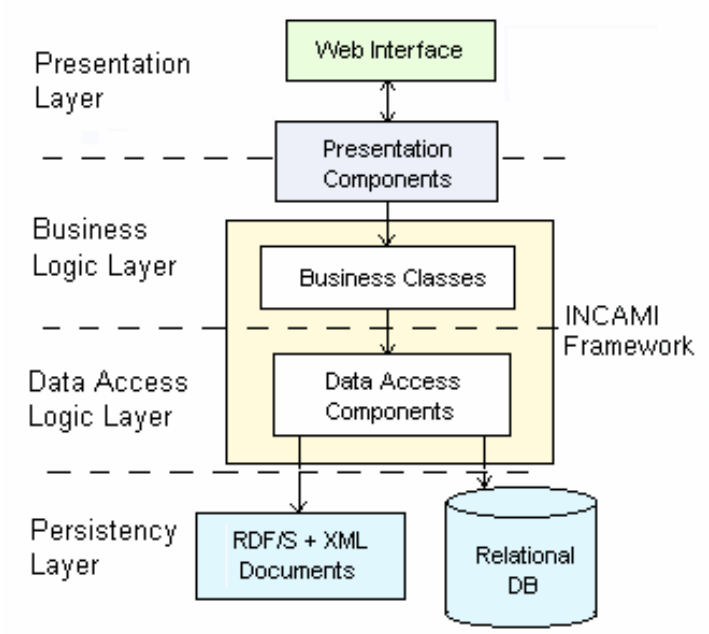
valuation
omponent



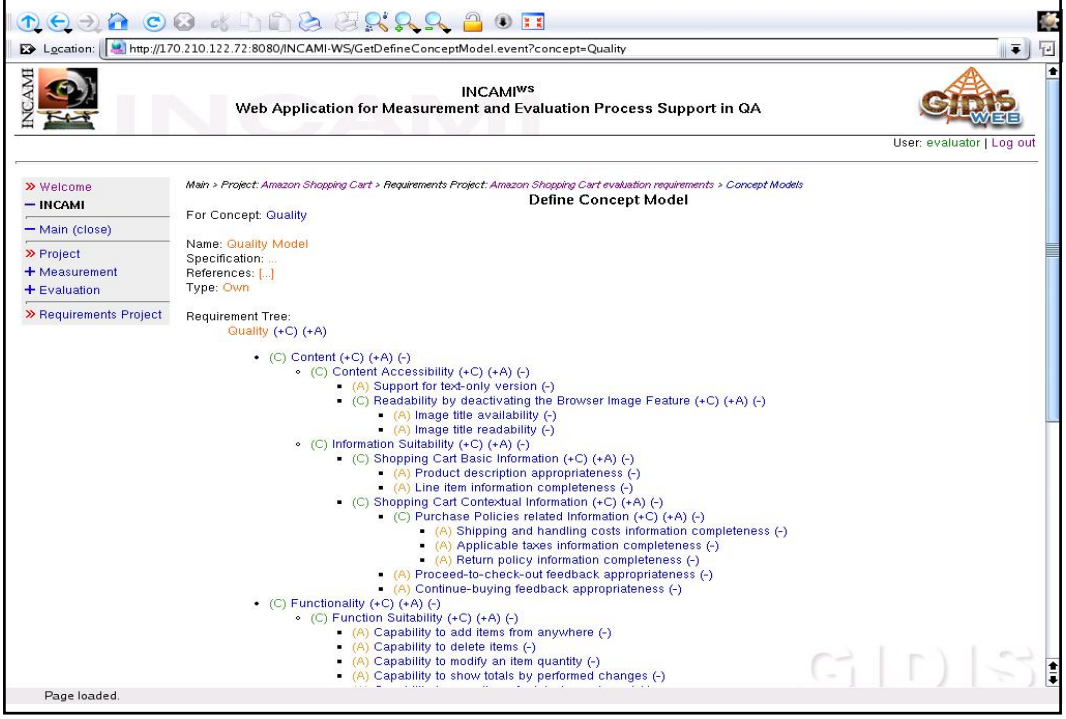
INCAMI_Tool

- INCAMI aims to give technological support to QA processes in organizational projects
- INCAMI_Tool is the prototype tool for the INCAMI framework [Papa, Molina, 05]
  - It takes metadata from the M&I ontology from a catalog (Sematic Web)
  - It saves M&I metadata and values for specific M&E projects

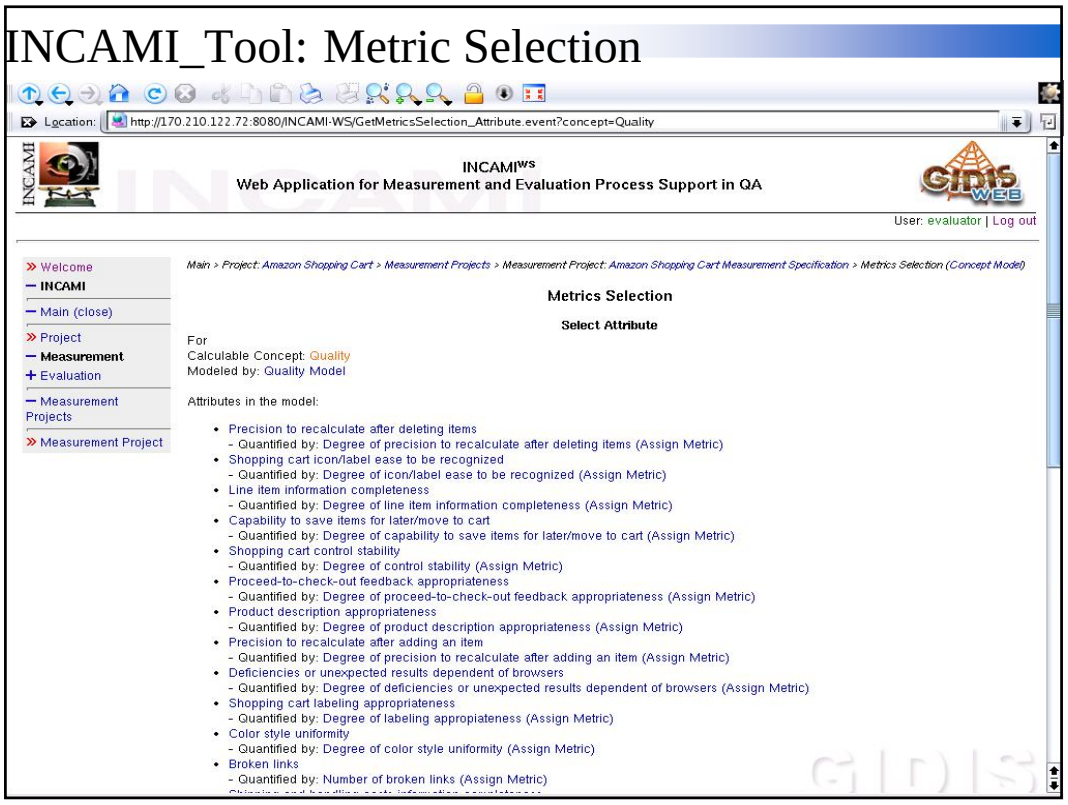
INCAMI_Tool Architecture



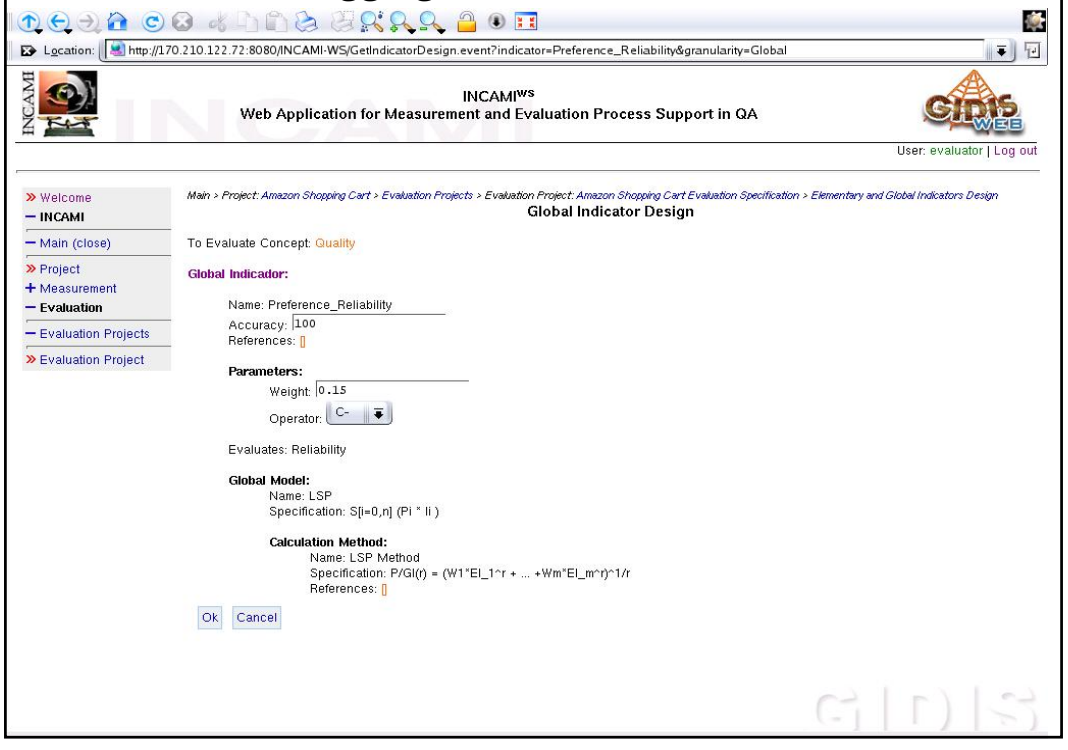
INCAMI_Tool: Model Definition



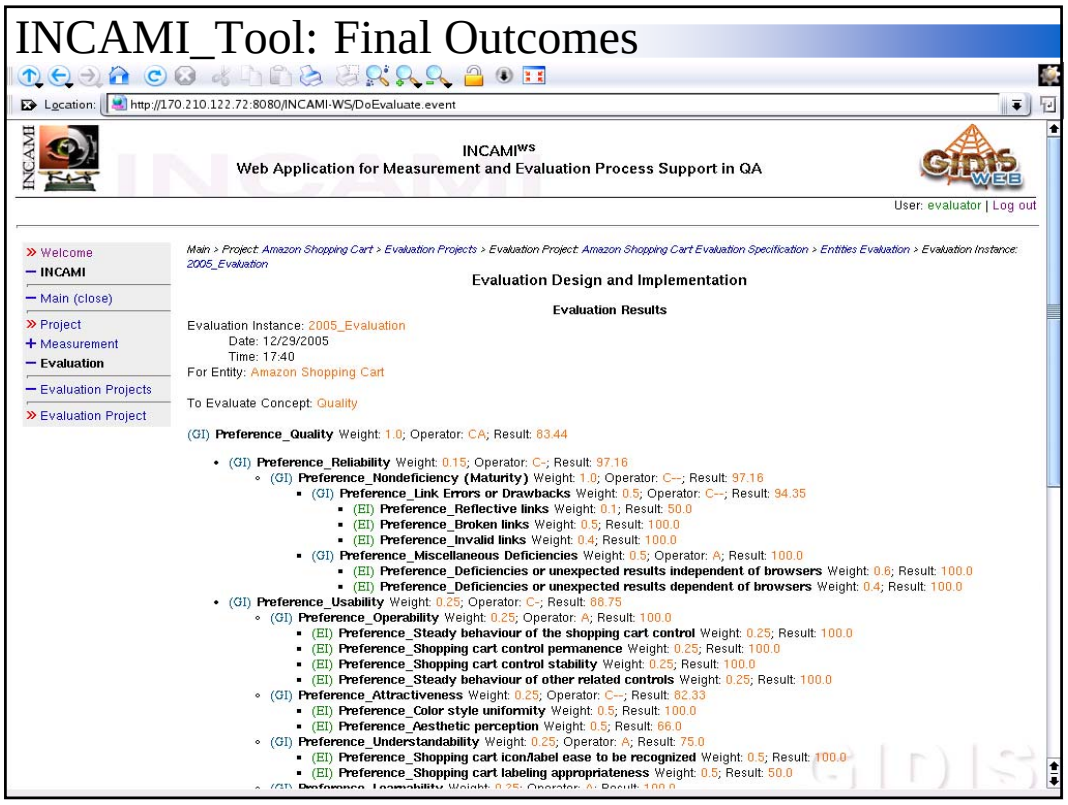
INCAMI_Tool: Metric Selection



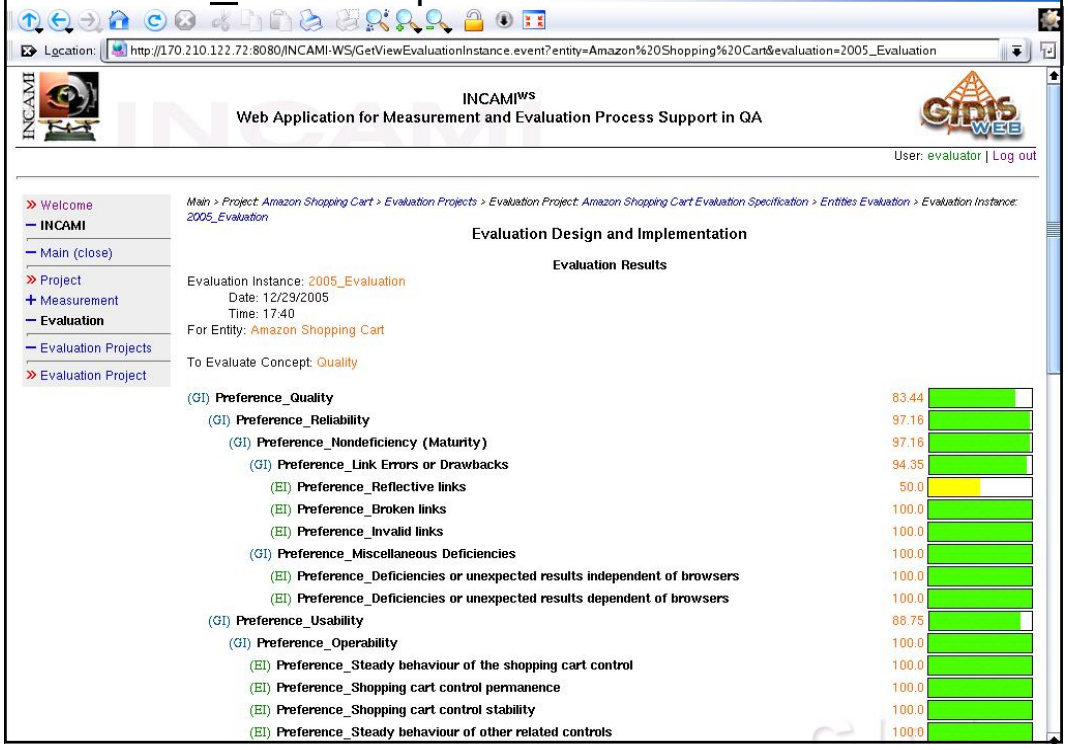
NCAMI_Tool: Aggregation Model Selection



INCAMI_Tool: Final Outcomes



INCAMI_Tool: Reports



To Remark

- Organizations could succeed in a **measurement and evaluation program** if resulting measurements and evaluations are tailored to their **information needs** for specific **purposes**, **contexts**, and **user viewpoints**.

- **INCAMI** is a framework which allows the definition and specification of NFR, in addition to the specification and implementation of measurement and evaluation processes driven by the **Information Needs** of an **organization** or project.

To Remark

- The INCAMI framework is based upon the assumption that for an organization to measure and evaluate in a purpose-oriented way it must first
  - specify nonfunctional requirements starting from information needs, then
  - it must design and select the specific set of useful metrics for measurement purpose, and lastly
  - interpret the metrics values by means of contextual indicators with the aim of evaluating or estimating the degree the stated requirements have been met.

To Remark

- Without appropiate definitions (meta-data) of metrics and indicators it is difficult to ensure values are repeatable and comparable among organization's projects for datasets analyses.
- Moreover, inter and intra-project analyses and comparisons could be performed in an inconsistent way.

To Remark

- GQM is a simple, flexible, goal-oriented approach with strong adoption in the industry for measurement programs
- GQM is not based on a sound conceptualization of metrics and indicadotors
  - there is no ontological base of metrics and indicators, so it can not assure that measure values (and the associated metadata like scale, unit, measurement method, and so forth) are trustworthy and consistent for ulterior analysis among projects

To Remark

- GQM is not necessarily concept model-oriented
  - quality, quality in use models, etc.
- Measure interpretation is not well defined for evaluation purposes
  - by means of elementary (and global) indicators
- When many metrics intervene, it can be hard to perform analyses, interpretations, and recommendations
  - No aggregation models
- GQ(I)M is an enhanced paradigm issued in the end of 2003 (SEI)
  - Strenghts and weaknesses

To Remark

- Kitchenham *et al.* worked in the definition of a framework (based on the ER model) to specify entities, attributes and relationships for measuring and instantiating projects,
  - with the purpose of analysing datasets in a consistent way.
- This is the closest framework to our research
- we tried to strengthen not only from the conceptual modeling viewpoint (using O-O models), but also from the ontological viewpoint including a broader set of concepts.
  - Particularly, we deal with evaluation concepts that Kitchenham *et al.* did not.

References

Olsina, L; Molina, H; Papa, F.; 2005; *Organization Oriented Measurement and Evaluation Framework for Software and Web Engineering Projects*, In LNCS 3579 of Springer, Int'l Congress on Web Engineering, (ICWE'05), Australia, pp. 42-52

Basili V., Rombach H.D., 1989, *The TAME Project: Towards Improvement-Oriented Software Environments*, IEEE Trans. on Software Engineering, 14(6), pp. 758-773.

Kitchenham B.A.; Hughes R.T.; Linkman S.G.; 2001, Modeling Software Measurement Data, IEEE Transaction on Software Engineering, 27 (9), pp.788, 804.

Outline

- Introduction to Quality and Quality in Use
  - For Software and Web
- Conceptual Base for Metrics & Indicators
- Evaluation Process
- Goal-oriented Measurement and Evaluation Frameworks
  - GQM Paradigm
  - INCAMI Framework
    - INCAMI Components & Tool
- **Conclusions**

Final Remarks

- To make QA a useful support process to Sw and Web development and maintenance projects, organizations must have sound specifications of M&I metadata associated consistently with data sets, as well as a clear establishment of frameworks and programs for measurement and evaluation projects.
- Organizations will not willingly waste their resources if resulting measurements and evaluations are not tailored to their information needs for specific purposes, contexts, and user viewpoints

Final Remarks

- Therefore, without sound specifications of M&I metadata, and engineered establishment of measurement and evaluation frameworks, organization's projects are less repeatable and controllable, and hence more prone to fail.
- This tutorial highlighted why the INCAMI framework can make a contribution in this direction

Further Issues

- Importance of managing the acquired organizational knowledge during quality assurance projects,
  - a semantic infrastructure that embraces organizational memory is being considered in our research.
  - recommender system
- Importance of M&I and the INCAMI framework for supporting CMMI (*Capability Maturity Model Integration*) upper levels
  - recommender system

Thank you, ANY QUESTION?

Dr. Luis Olsina

E-mail: **olsinal@ing.unlpam.edu.ar**

URL: http://gidis.ing.unlpam.edu.ar

Grupo de Investigación y Desarrollo en Ingeniería de Software y Web (GIDIS_Web)

Departamento de Informática – Facultad de Ingeniería – Universidad Nacional de La Pampa

General Pico – La Pampa

Argentina