

Estrategias de Medición y Evaluación: Diseño de un Estudio Comparativo

María Fernanda Papa, Pablo Becker y Luis Olsina

Facultad de Ingeniería, UNLPam, General Pico, La Pampa, Argentina
{pmfer, beckerp, olsinal}@ing.unlpam.edu.ar

Abstract. Este trabajo especifica un diseño de evaluación para comprender y comparar estrategias integradas de medición y evaluación, considerando a una estrategia como a un recurso de un proyecto, desde el punto de vista del ente a valorar. Nuestro objetivo es evaluar la calidad de las capacidades de una estrategia de medición y evaluación teniendo en cuenta tres fundamentos: 1) el marco conceptual centrado en una base terminológica, 2) el soporte explícito de proceso, y 3) el soporte metodológico/tecnológico. A su vez, para diseñar el estudio tuvimos en cuenta para una estrategia aspectos de requerimientos no funcionales, de medición, de evaluación, y de análisis y recomendación. Como resultado de esta investigación hemos especificado el diseño del estudio; esto es, el árbol de requerimientos en función de características y atributos, el diseño de las métricas que cuantifican a estos atributos y su interpretación por medio del diseño de indicadores. Como consecuencia, documentaremos en otro artículo la implementación del estudio comparativo que permitirá un análisis riguroso de debilidades y fortalezas de estrategias como GQM (*Goal-Question-Metric*), GOCAME (*Goal-Oriented Context-Aware Measurement and Evaluation*), etc., con el fin de definir cursos de acción para mejorarlas.

Keywords: Estrategia de medición y evaluación, base terminológica, C-INCAMI, marco conceptual, proceso.

1 Introducción

Las organizaciones de software que deciden realizar programas de medición y evaluación (M&E), necesitan establecer un conjunto de actividades para especificar, recolectar, almacenar y utilizar métricas e indicadores. Además, deberían asegurar que los valores medidos y los indicadores calculados sean repetibles y comparables entre distintos proyectos, con el objetivo de realizar análisis más robustos y consecuentemente tomar mejores decisiones. Por lo que no sólo se deben almacenar datos de mediciones y evaluaciones sino también metadatos de métricas e indicadores como escala, tipo de escala, unidad, método de medición, especificación del modelo de indicador elemental, entre otros.

En [19] propusimos una estrategia de M&E que permite desarrollar programas con esas características. Tal enfoque o estrategia integra los siguientes tres fundamentos o pilares:

1. *Un proceso para medición y evaluación*, en el cual se especifican las

principales actividades que deben ser planificadas y ejecutadas. Contar con un modelo de proceso permite garantizar repetitividad y reproducibilidad en la ejecución de las actividades y consistencia en los resultados. El modelo no sólo debe prescribir un conjunto de actividades, sino también sus entradas y salidas, roles, interdependencias, etc.

2. *Un marco conceptual*, flexible y terminológicamente consistente. Un marco bien establecido debería estar construido sobre una base robusta (diccionarios, ontologías, etc.), en donde se especifique de manera formal y explícita, los componentes, conceptos, relaciones y restricciones acordadas para un dominio particular. Este pilar ayuda a las organizaciones al momento de comenzar un programa de medición a que no tengan que empezar nuevamente desde cero sino que posean una base desde donde comenzar.
3. *Métodos y herramientas* concretos para llevar a cabo las actividades que se derivan del proceso y marco.

Es decir, una estrategia de M&E integrada debe estar conformada al menos por la definición de un proceso de M&E de requerimientos no funcionales, un modelo de dominio que define todos los conceptos necesarios y sus relaciones, y un soporte tecnológico y metodológico para realizar actividades usando dicha conceptualización.

Como discutiremos, las estrategias de M&E que logren la integración de estos tres fundamentos permitirán una implementación repetible y robusta de programas y proyectos de M&E en una organización. En el presente trabajo, nos centramos en especificar el diseño de la evaluación que permita comprender y comparar estrategias de M&E existentes. El objetivo del estudio es evaluar la *calidad de las capacidades* de dichas estrategias partiendo de la idea que una estrategia es apta o adecuada para llevar a cabo programas y proyectos de M&E repetibles si posee: un proceso explícitamente definido y modelado, una base conceptual robusta (como lo es una ontología) sustentada a su vez en un marco conceptual, y un soporte metodológico y/o tecnológico. Bajo esta premisa podremos evaluar y comparar las distintas estrategias de M&E según el nivel de satisfacción alcanzado para estos tres aspectos, y en definitiva, para la calidad de las capacidades.

El diseño propuesto en este trabajo es el primer paso en la ejecución del estudio comparativo que permitirá un análisis riguroso de debilidades y fortalezas de las estrategias seleccionadas. Intervendrán en el estudio enfoques de medición, evaluación y análisis de dominio público existentes en la literatura, con cierto impacto en la academia o industria, y con cierto nivel de integración de los tres aspectos arriba mencionados. En resumen, la contribución general de esta investigación consiste en: i) comprender y comparar enfoques integrados de M&E existentes; ii) diseñar los requerimientos no funcionales teniendo como foco la calidad de las capacidades de una estrategia integrada de M&E; y, iii) brindar conclusiones/recomendaciones a partir del análisis de fortalezas y debilidades de los enfoques seleccionados y, en definitiva, establecer cursos de acción para mejorar y robustecer las estrategias integradas de M&E; y, en particular, nuestra estrategia denominada GOCAME (*Goal-Oriented Context-Aware Measurement and Evaluation strategy*).

El resto del artículo se organiza de la siguiente manera. En la Sección 2 analizamos la contribución de este trabajo con respecto a otros relacionados. En la Sección 3 damos un panorama de la estrategia integrada GOCAME que fue tomada como base para la realización del diseño del estudio comparativo. En la Sección 4,

detallamos el diseño de la evaluación ilustrando especificaciones de requerimientos no funcionales, métricas e indicadores. Finalmente, se destacan las conclusiones y líneas de avance.

2 Trabajos Relacionados

Existen muchas propuestas publicadas en el área de M&E. Por ejemplo, algunos autores como Kan [14], Fenton *et al.* [6] y Zuse [22] brindan definiciones acerca de métricas, modelos, métodos y herramientas de medición, mientras que otros como Kitchenham *et al.* [15] y Abran *et al.* [1] se focalizan en la definición y/o modelado de los datos y metadatos de la medición.

Por otro lado, como estrategia de medición relevante podemos mencionar el enfoque GQM (*Goal Question Metric*) propuesto por Basili *et al.* [2]. Además, existen muchos trabajos relacionados [4, 9, 10, 21] que extienden y mejoran el propuesto por Basili *et al.* definiendo el proceso, el modelo de información o completándolo con otros enfoques existentes.

También, existen un conjunto de estándares ISO como lo son: ISO 14598 [11] que define el proceso de evaluación de productos software e ISO 15939 [12] que define el proceso de medición de software. Más recientemente, se comenzó a desarrollar una nueva generación de estándares en el marco del proyecto SQuaRE (*Software Product Quality Requirements and Evaluation*) [13] que intenta reunir los dos estándares antes mencionados, entre otras contribuciones específicas para el área de requerimientos de calidad. Por otra parte, otro trabajo relevante es el estándar *de facto* CMMI (*Capability Maturity Model Integration*) [5] el cual reúne las mejores prácticas de una organización, agrupándolas en diferentes áreas de procesos una de las cuales es medición y análisis.

Otro aporte respecto a enfoques de M&E es FMESP (*Framework for the Modeling and Evaluation of Software Processes*) [17]. Su marco conceptual cuenta con una base ontológica (presentada en [8] y luego refinada en [7]), que si bien es similar a nuestra propuesta ontológica, en [18] hemos modelado algunos otros conceptos (por ej. indicador elemental, indicador global, etc.) y relaciones (por ej. entre medición y medida, entre métrica e indicador, etc.) que difieren semánticamente de las propuestas en [7, 8]; además hemos incluido nuevos conceptos como contexto, propiedad de contexto, etc. no consideradas en tales trabajos. En lo que respecta al soporte explícito del proceso de M&E en FMESP, los autores más que centrarse en la modelización del mismo, hacen uso de su ontología, del paradigma MDE (*Model-Driven Engineering*) y MDA (*Model-Driven Architecture*) para la modelización arquitectónica de la medición, empleando transformaciones de modelos para instanciar herramientas.

Sin embargo, ninguno de los trabajos citados considera la necesidad de una estrategia integrada de M&E basada en los tres principios mencionados en la sección introductoria. Por otra parte, conforme a las búsquedas exhaustivas de literatura que hemos efectuado, no existen al presente trabajos relacionados que traten con la evaluación de la calidad de las capacidades de estrategias integradas de M&E, basadas en esos tres fundamentos.

3 Panorama de GOCAME

Como se mencionó en la Introducción el diseño del estudio comparativo se llevó a cabo siguiendo la estrategia GOCAME, realizando el diseño de los requerimientos no funcionales, el diseño de la medición y el diseño de la evaluación. A seguir se presenta un panorama de esta estrategia.

GOCAME es una estrategia integrada de M&E que sigue el enfoque orientado a objetivos, sensible al contexto y centrada en la organización. Está fundamentada en los tres pilares mencionados, a saber: 1) un marco conceptual centrado en una base terminológica, 2) el soporte explícito de proceso, y 3) el soporte metodológico/tecnológico. Está diseñada para permitir la definición de proyectos de M&E incluyendo descripciones de contexto, las que proveen interpretaciones más sólidas entre resultados de evaluaciones intra e inter-organizacionales [16].

GOCAME tiene su base terminológica definida a partir del marco conceptual C-INCAMI (*Contextual - Information Need, Concept model, Attribute, Metric and Indicator*) [19] y su base ontológica [18]. Esta base provee un modelo de dominio que define todos los conceptos, propiedades y relaciones necesarios para implementar el proceso de M&E (ver Fig. 1). C-INCAMI permite especificar y estructurar los datos y metadatos utilizados en las actividades y artefactos del proceso. También, siguiendo las especificaciones de este marco se pueden definir catálogos que contengan información reusable entre proyectos, y construir herramientas de soporte para automatizar todo o parte del proceso. C-INCAMI está estructurado en cinco módulos principales: 1) Definición y Especificación de Requerimientos no Funcionales, 2) Especificación del Contexto, 3) Diseño y Ejecución de la Medición, 4) Diseño y Ejecución de la Evaluación; y, 5) Análisis y Recomendación.

El primer componente (Definición y Especificación de Requerimientos no Funcionales) permite definir la Necesidad de Información (*InformationNeed*)¹, la cual pretende satisfacer un requerimiento no funcional de alguna entidad (*Entity*) para un propósito particular (*purpose*) y un punto de vista determinado (*userViewpoint*). Los requerimientos no funcionales son representados por un modelo de concepto (*ConceptModel*) que incluye conceptos (*CalculableConcept*), subconceptos y atributos (*Attribute*). Los atributos son propiedades cuantificables de la entidad bajo análisis. Por otro lado, el componente Especificación del Contexto, permite describir de forma estructurada el contexto (*Context*) relevante para una necesidad de información por medio de propiedades (*ContextProperty*) –ver detalles en [16]. Respecto al componente Diseño y Ejecución de la Medición, este permite especificar métricas directas e indirectas (*Direct e Indirect Metric*) a utilizar en la medición (*Measurement*) como así también los valores medidos (*Measure*), el método de medición o cálculo (*Measurement o Calculation Method*) y la escala (*Scale*) de los valores. El componente Diseño y Ejecución de la Evaluación permite definir la evaluación (*Evaluation*) mediante el uso de indicadores (*Indicator*), que especifican cómo interpretar los valores de los atributos y de los conceptos de más alto nivel, permitiendo conocer de este modo el grado de satisfacción de todos los requerimientos para una necesidad de información. Un indicador elemental

¹ Los conceptos en cursiva se refieren a términos definidos en la ontología de métricas e indicadores [18].

(*ElementaryIndicator*) interpreta el valor de un atributo mediante el uso de un modelo elemental (*ElementaryModel*). Además, un indicador global (*GlobalIndicator*) permite evaluar un concepto de alto nivel de abstracción (como son los conceptos calculables), y su valor se deriva de otros indicadores haciendo uso de un modelo global (*GlobalModel*). A su vez, los modelos tienen asociados criterios de decisión (*DecisionCriteria*). Por último, el componente Análisis y Recomendación (no mostrado los conceptos en Fig. 1) tiene la finalidad de soportar el análisis de datos y brindar recomendaciones para la toma de decisiones.

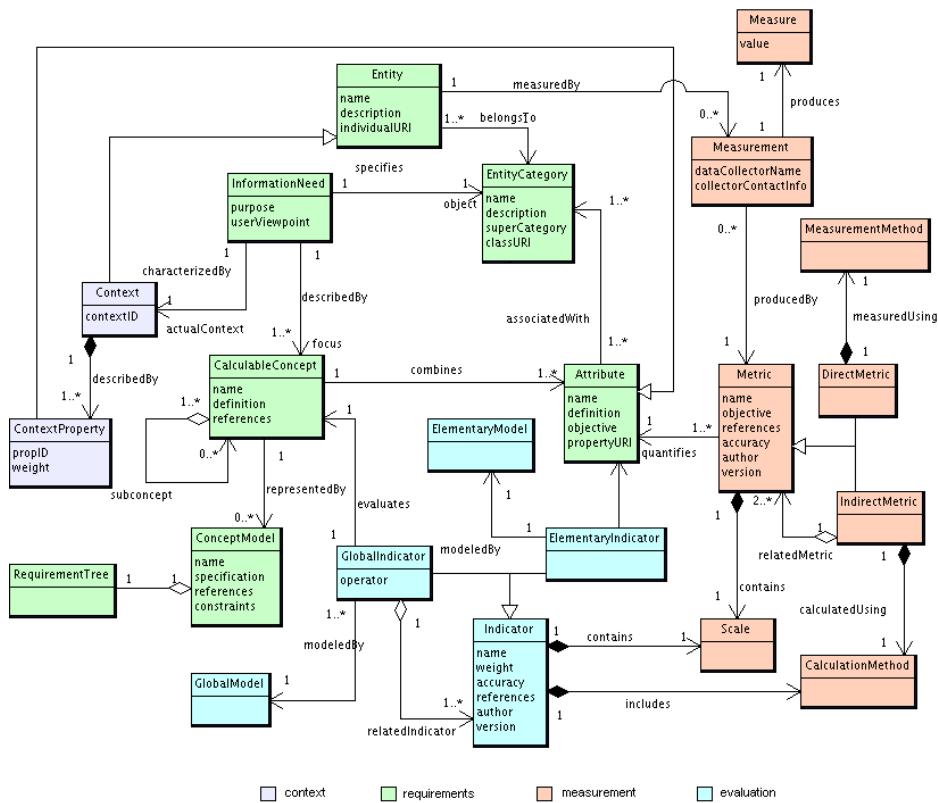


Fig. 1. Principales conceptos y relaciones de los componentes del marco C-INCAMI.

Por otro lado, desde el punto de vista del soporte explícito al proceso de M&E [3], GOCAME cuenta con la definición y modelado de seis actividades principales (ver Fig. 2), las cuales comentamos a continuación².

La primera actividad técnica a ejecutar en un proyecto de M&E es *Definir los Requerimientos no Funcionales*. Su objetivo es servir como guía al resto de las actividades. Como primer paso (ver Fig. 3), se debe *Establecer la Necesidad de Información*, es decir, se debe definir el propósito de la evaluación, el tipo de usuario,

² Se utiliza cursiva al hacer mención a nombres de actividades del proceso. Notar que hay una correspondencia entre nombres de actividades y términos de C-INCAMI.

el objeto o categoría de entidad bajo análisis, la entidad concreta a evaluar y el foco o concepto de más alto nivel. Luego, se debe *Especificar el Contexto* con la finalidad de poder realizar análisis más consistentes de los resultados, principalmente, si estos provienen de distintos proyectos y organizaciones. La especificación del contexto se lleva a cabo mediante *Seleccionar las propiedades relevantes* que caracterizan el contexto y *Cuantificar la propiedad de contexto* en base a la métrica asociada.

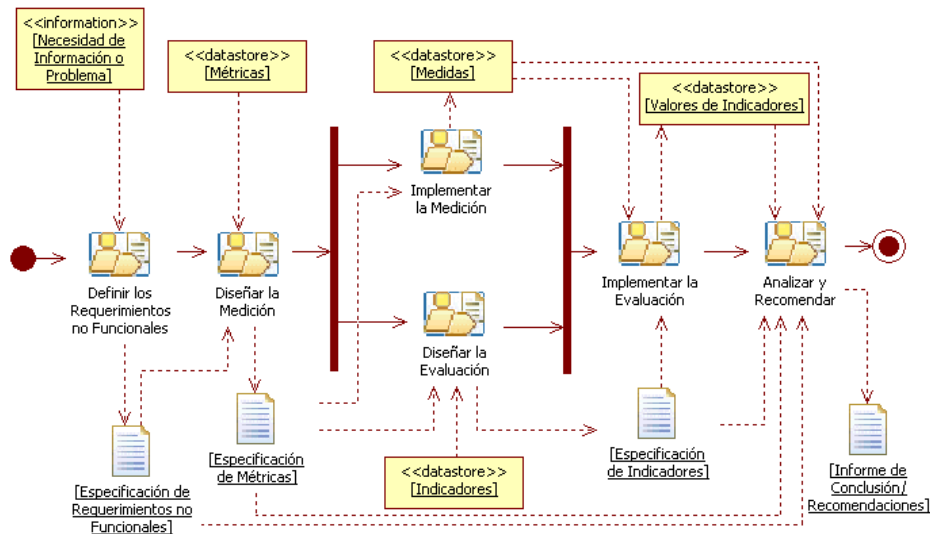


Fig. 2. Principales actividades de la estrategia de medición y evaluación GOCAME.

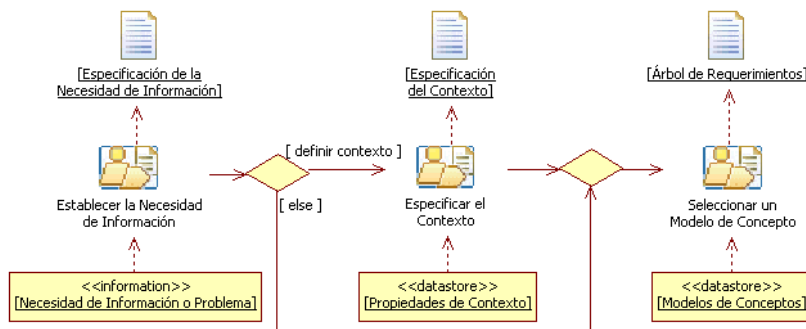


Fig. 3. Flujo de actividades al Definir los Requerimientos no Funcionales.

Seleccionar un Modelo de Conceptos, es otra de las actividades necesarias para definir los requerimientos no funcionales. Tiene como objetivo agregar conceptos, subconceptos y atributos, como así también, especificar la forma en la cual se relacionan los mismos. Se puede seleccionar un modelo estándar o *de facto* (definido por la propia organización) teniendo en cuenta el punto de vista del usuario, como así también, la especificación del contexto. En caso que el modelo elegido no sea

totalmente adecuado a la necesidad, ya sea por relaciones inapropiadas o carencia de atributos, se tiene la posibilidad de editar el mismo. El árbol de requerimientos tendrá como raíz al concepto foco y como hojas a los atributos. Al finalizar las actividades descriptas anteriormente se obtiene como salida un documento estructurado con la especificación de los requerimientos no funcionales.

Una vez definidos los requerimientos no funcionales, se conoce con claridad cuál es la finalidad del proceso de M&E, por lo que se puede identificar de manera objetiva cuales serán las métricas que se utilizarán en la medición. En la actividad *Diseñar la Medición* se debe establecer una métrica para cada atributo del árbol. En el caso que la métrica sea indirecta también se deben identificar las métricas relacionadas y los atributos que son cuantificados por éstas. Además si el método puede automatizarse se debe especificar una herramienta. La salida de la actividad *Diseñar la Medición* es la especificación de las métricas asociadas a los atributos del árbol de requerimientos.

La próxima actividad principal es *Diseñar la Evaluación*. Su objetivo consiste en definir para cada atributo y concepto calculable del árbol de requerimientos un indicador que lo evaluará. En primer lugar se deben *Identificar Indicadores Elementales*, es decir, especificar un indicador para cada atributo del árbol de requerimientos, el cual especificará la manera de realizar el mapeo del valor medido del atributo a un valor que indique su nivel de aceptación según los objetivos y requerimientos establecidos. Luego, se deben *Identificar Indicadores Parciales/Globales*. Para ello se debe proceder de la misma forma que para los indicadores elementales, y además asociar pesos (si son requeridos por el modelo de agregación) a los sub-conceptos. Como resultado se obtiene un documento con la especificación de todos los indicadores.

La siguiente actividad es *Implementar la Medición*, y su objetivo es obtener un valor o medida (numérica o categórica) para cada uno de los atributos de una o más entidades a partir de las métricas diseñadas y las herramientas que automatizan la medición (si las hubiera). Como resultado se obtienen los valores asociados a cada uno de los atributos que caracterizan a las entidades bajo análisis quedando almacenados en un repositorio para su posterior utilización en la siguiente actividad.

A continuación, se debe *Implementar la Evaluación*. El objetivo de esta actividad es obtener los valores de los distintos indicadores y conocer en qué proporción se satisfacen los requerimientos establecidos en la necesidad de información. Para llevar a cabo esta actividad se necesita por un lado, *Calcular los Indicadores Elementales* para lo cual se realiza un mapeo de los valores heterogéneos obtenidos en la medición a un conjunto de valores normalizados. Una vez realizado esto, se puede *Calcular los Indicadores Parciales/Globales*. El cálculo de los indicadores puede realizarse en forma manual o por medio de herramientas que faciliten los cálculos como puede ser C-INCAMI Tool [19]. Al finalizar esta actividad se obtienen los valores de los diferentes indicadores, tanto elementales como parciales y globales, que servirán como entrada a la próxima actividad.

La última actividad, *Analizar y Recomendar*, consiste en la realización de actividades orientadas al análisis de los datos y la justificación de los resultados obtenidos. Esto implica diseñar e implementar el análisis, producir el informe de los resultados y realizar recomendaciones si fuese pertinente, por ejemplo, especificando cursos de acción a seguir, fortalezas y debilidades, etc.

Respecto al tercer pilar (métodos, técnicas y herramientas), se cuenta con la metodología WebQEM [20] y la herramienta citada. Esta metodología aporta una estrategia eficaz, centrada en expertos o en usuarios finales, para evaluar y analizar la calidad de aplicaciones web o de productos software en general. Comprende un conjunto de métodos, técnicas y herramientas para llevar a cabo las actividades del proceso introducido.

4 Diseño del Estudio Comparativo de Estrategias de M&E

En la sección introductoria se justificó la necesidad de contar con una estrategia integrada de M&E de soporte a procesos de aseguramiento de calidad en las organizaciones de software, y además, se indicó que para comprender el estado actual de este recurso –y potencialmente mejorarlo- es necesario diseñar y ejecutar su evaluación. En esta sección presentamos el diseño realizado que permite comprender y comparar distintas estrategias integradas de M&E tomando como guía a GOCAME.

El foco principal de este estudio es evaluar la *calidad de las capacidades* de una estrategia de M&E teniendo en cuenta los tres principios o fundamentos ya enunciados e introducidos. Definimos al concepto calculable *calidad de las capacidades (Capability Quality)* como “the degree to which a resource is suitable and appropriate for supporting and performing the actions when used under specified conditions”.

Siguiendo el proceso definido, la primera actividad fue *Definir Requerimientos no Funcionales* (recordar figuras 1 y 3). Por lo que el propósito de la Necesidad de Información es *comprender y comparar* desde el punto de vista del *líder de aseguramiento de la calidad* una *Estrategia Integrada de Medición y Evaluación* como categoría de entidad (la que a su vez pertenece a la supercategoría *Recurso*). El concepto calculable de más alto nivel (foco) es *Capability Quality*, y dado que no existe un modelo de conceptos estándar ni tampoco en la literatura relacionada que evalúe dicho foco, fue necesario definir uno *de facto*.

Los subconceptos de más alto nivel asociados a *Capability Quality* según se observa en la Fig. 4 son los que representan a los tres pilares, a saber: *Process Capability Quality* (codificado 1.1), *Conceptual Framework Capability Quality* (1.2), y *Methodology Capability Quality* (1.3) respectivamente.

A su vez 1.1 se subdivide en sub-características que representan cuán apropiadas son las actividades (*Activities Suitability*), los artefactos (*Artifacts Suitability*), el modelado del proceso y sus vistas (*Process Modeling Suitability*) y la conformidad del proceso a estándares y su base terminológica (*Process Compliance*). Por ejemplo, *Activities Suitability* se define como “the degree to which the activities provide properties that meet stated and implied needs for the intended user purpose”.

Por otra parte, el concepto calculable *Conceptual Framework Capability Quality* está especificado en función de sub-características que representan cuán apropiada es la base conceptual terminológica (*Conceptual Base Suitability*) y su marco conceptual (*Conceptual Framework Suitability*), como así también el nivel de conformidad entre ambos (*Conceptual Framework Compliance*).

Por último, el concepto *Methodology Capability Quality* está definido en función

de sub-características que representan cuán apropiada es la metodología (*Methodology Suitability*), además del nivel de conformidad de la metodología con la base conceptual terminológica (*Methodology Compliance*).

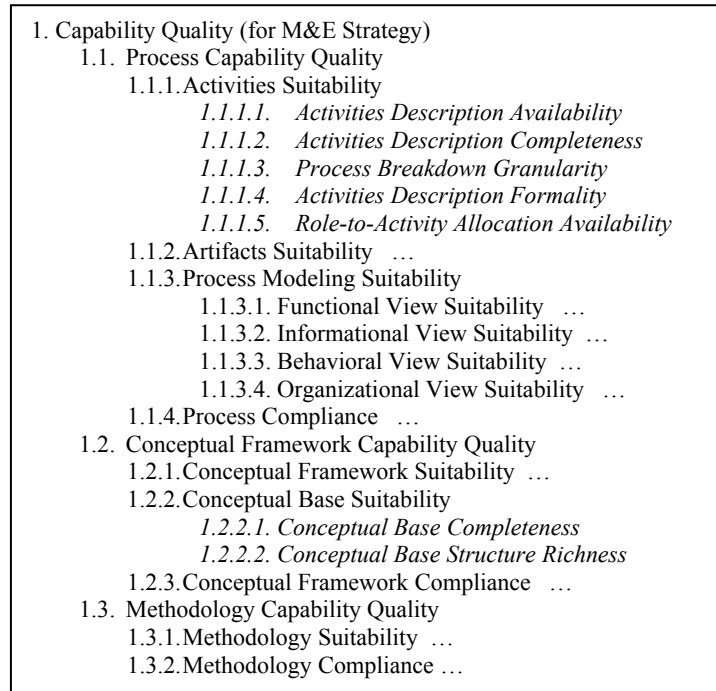


Fig. 4. Fragmento del árbol de requerimientos para Capability Quality. En *italica* se muestran algunos atributos cuantificables.

Tabla 1. Especificación de algunos atributos del árbol de requerimientos (ver Fig. 4)

Nombre	Definición	Objetivo
<i>Activities Description Availability</i> (A.1.1.1.1)	It represents the explicit specification of the process activities' description.	To find out the extent to which the process activities' descriptions are available
<i>Process Breakdown Structure Granularity</i> (A.1.1.1.3)	It represents the granularity level of the process breakdown structure, i.e. a process can be fine- or coarse-grained with regard to its hierarchical structure division.	To find out the degree of process granularity according to its breakdown structure.
<i>Conceptual Base Structure Richness</i> (A.1.2.2.2)	It represents the kind of the conceptual base structuredness.	To find out the extent to which of conceptual base is rich, considering that it can be an ontology, taxonomy, dictionary, etc.

Attribute Name: Activities Description Availability		
Indirect Metric		
<u>Name:</u> Activities Description Availability Degree (ADAD)		
<u>Objective:</u> To quantify how many activities are described with regard to the total enunciated activities.		
<u>Author:</u> Fernanda Papa	<u>Version:</u> 0.1	
Calculation Method		
<u>Name:</u> ADAD Determination	<u>Specification:</u> To apply the related function.	
Function		
<u>Name:</u> ADAD function	<u>Specification:</u> $ADAD = (\#DA / TEA) \times 100$	
Numerical Scale:		
<u>Representation:</u> Continue	<u>Value Type:</u> real	<u>Scale Type:</u> ratio
<u>Unit:</u>		
<u>Name:</u> Percentage	<u>Acronym:</u> %	
<u>Description:</u> Value that represents a proportion of a whole		
Related Metric:		
Total number of Enunciated Activities (TEA)		
Number of Described Activities (#DA)		
Attribute Name: Enunciated Activities		
Direct Metric:		
<u>Name:</u> Total number of Enunciated Activities (TEA)		
<u>Objective:</u> To quantify how many process' activities are enunciated.		
...		
Measurement Method:		
<u>Name:</u> TEA Determination		
<u>Specification:</u> Add 1 for each different enunciated activity (see attribute definition) found in the documentation. The name/label for each enunciated activity should be different semantically.		
<u>Type:</u> Objective		
Numerical Scale:		
<u>Representation:</u> discrete	<u>Value Type:</u> integer	<u>Scale Type:</u> absolute
<u>Unit:</u>		
<u>Name:</u> Activity	<u>Acronym:</u> Ac.	
<u>Description:</u> An activity is an atomic or compound operation that is performed for one or more agents into a process.		
Attribute Name: Described Activities		
Direct Metric:		
<u>Name:</u> Number of Described Activities (#DA)		
<u>Objective:</u> To quantify how many activities in the process under analysis are described.		
...		
Measurement Method:		
<u>Name:</u> #DA Determination		
<u>Specification:</u> Add 1 for each different described activity (see attribute definition) found in the documentation. The description could be together or dispersed in the documentation. In the last case the relation must be explicit.		
<u>Type:</u> Objective		
...		

Fig. 5. Especificación de la métrica *Activities Description Availability Degree*.

<p><u>Attribute Name:</u> Conceptual Base Structure Richness</p> <p><u>Direct Metric:</u> <u>Name:</u> Conceptual Base Structure Richness Degree (CBSRD) <u>Objective:</u> To determine the extent to which the conceptual base is rich, considering that it can be an ontology, taxonomy, dictionary, etc.</p> <p>...</p> <p><u>Measurement Method:</u> <u>Name:</u> CBSRD Determination <u>Specification:</u></p> <p>CBSRD { High → If the conceptual base is represented as ontology. Medium → If the conceptual base is represented as taxonomy. Low → If the conceptual base is represented as dictionary or list of terms. None → If the conceptual base does not exist.</p> <p><u>Type:</u> Objective</p> <p><u>Categorical Scale:</u> <u>Value Type:</u> symbol <u>Scale Type:</u> ordinal <u>Allowed Values:</u></p> <table border="1"> <thead> <tr> <th><u>Name</u></th> <th><u>Value</u></th> <th><u>Description</u></th> </tr> </thead> <tbody> <tr> <td>None</td> <td>0</td> <td>There is no conceptual base.</td> </tr> <tr> <td>Low</td> <td>1</td> <td>The conceptual base is represented as dictionary or list of terms.</td> </tr> <tr> <td>Medium</td> <td>2</td> <td>The conceptual base is represented as taxonomy.</td> </tr> <tr> <td>High</td> <td>3</td> <td>The conceptual base is represented as ontology.</td> </tr> </tbody> </table>	<u>Name</u>	<u>Value</u>	<u>Description</u>	None	0	There is no conceptual base.	Low	1	The conceptual base is represented as dictionary or list of terms.	Medium	2	The conceptual base is represented as taxonomy.	High	3	The conceptual base is represented as ontology.
<u>Name</u>	<u>Value</u>	<u>Description</u>													
None	0	There is no conceptual base.													
Low	1	The conceptual base is represented as dictionary or list of terms.													
Medium	2	The conceptual base is represented as taxonomy.													
High	3	The conceptual base is represented as ontology.													

Fig. 6. Especificación de la métrica *Conceptual Base Structure Richness Degree*.

Los atributos combinados a los distintos conceptos calculables se refieren a propiedades cuantificables por métricas tales como: *disponibilidad, granularidad, completitud, nivel de formalidad, conformidad y nivel de riqueza*, entre otros. En la tabla 1, se muestra la definición y el objetivo de los siguientes atributos: *Activities Description Availability, Process Breakdown Structure Granularity, Conceptual Base Completeness y Conceptual Base Structure Richness*.

Al finalizar esta actividad de diseño se obtiene como salida un documento estructurado con la especificación de los requerimientos no funcionales, es decir, el modelo instanciado mostrado parcialmente en la Fig. 4, con todas las definiciones. (Notar que en el presente artículo, no incluimos la especificación de las propiedades de contexto, por razones de espacio).

Siguiendo el proceso definido en GOCAME, la segunda actividad fue *Diseñar la Medición*, por lo que para cada atributo del árbol se diseñó una métrica que la cuantificara. La métrica –según discutido en la sección 3- puede ser directa o indirecta. El documento de especificación de métricas quedó conformado por 55 métricas de las cuales 40 son métricas directas y el resto indirectas. En las figuras 5 y 6, se muestran la especificación de las métricas *Activities Description Availability Degree* y *Conceptual Base Structure Richness Degree*, respectivamente. La primera (Fig. 5), es una métrica indirecta, y cuantifica al atributo *Activities Description Availability*. Al ser indirecta se tuvieron que definir las métricas directas *Total number of Enunciated Activities* y *Number of Described Activities*. La segunda (Fig. 6), cuyo nombre es *Conceptual Base Structure Richness Degree (CBSRD)* es una métrica directa y tiene por objetivo “to determine the extent to which the conceptual

base is rich, considering that it can be an ontology, taxonomy, dictionary, etc.”. El lector puede observar que la plantilla de especificación de la métrica, contiene todos los metadatos que deben ser almacenados, además de valores concretos (datos) que a la postre se registrarán al momento de la implementación de la medición. Estos metadatos provienen de la base conceptual C-INCAMI introducida en la sección 3 (y detallada en [19]). Por lo que para una métrica directa se almacenan (en un catálogo) además del nombre y objetivo, el método de medición, tipo, especificación de las reglas de conteo, potencial herramienta que automatiza al método, escala, tipo de escala, unidad si corresponde (recordar que sólo las escalas numéricas soportan el concepto de unidad), etc. A modo de ejemplo, para la métrica CBSRD de la Fig. 6, la escala es categórica y el tipo de escala es ordinal, cuyos valores o categorías permitidas son: “None” (que es un símbolo y se mapea al valor entero 0) y su descripción es: “There is no conceptual base”; “Low”, (1), “The conceptual base is represented as dictionary or list of terms”; “Medium”, (2), “The conceptual base is represented as taxonomy”; y “High”, (3), “The conceptual base is represented as ontology”. En definitiva, la salida de la actividad *Diseñar la Medición* es la especificación de las métricas asociadas a los atributos del árbol de requerimientos.

<p>Attribute: <i>Activities Description Availability</i></p> <p>Elemental Indicator: <u>Name:</u> Preference of Activities Description Availability (P_ADA) <u>Author:</u> Fernanda Papa <u>Version:</u> 0.1 <u>Weight:</u> 0.20</p> <p>Numerical Scale: <u>Scale Type:</u> absolute <u>Unit name:</u> Percentage <u>Acronym:</u> %</p> <p>Function (Elementary Model): <u>Name:</u> P_ADA function <u>Specification:</u> P_ADA = ADAD</p> <p>Decision Criteria/Acceptability Levels: if $0 \leq X \leq 45$: “unsatisfactory” indicates change actions must take high priority. if $45 < X \leq 70$: “marginal” indicates a need for improvement actions. if $70 < X \leq 100$: “satisfactory” indicates satisfactory quality of the analyzed feature.</p>
<p>Attribute: <i>Conceptual Base Structure Richness</i></p> <p>Elemental Indicator: <u>Name:</u> Preference of <i>Conceptual Base Structure Richness</i> (P_CBSR) <u>Author:</u> Fernanda Papa <u>Version:</u> 0.1 <u>Weight:</u> 0.50</p> <p>Numerical Scale: idem as before (percentage)</p> <p>Function: <u>Name:</u> P_CBSR function <u>Specification:</u></p> $P_CBSR = \begin{cases} \text{High} & \rightarrow 100\% \\ \text{Medium} & \rightarrow 70\% \\ \text{Low} & \rightarrow 30\% \\ \text{None} & \rightarrow 0\% \end{cases}$ <p>Decision Criteria: idem as before to the P_ADA indicator.</p>

Fig. 7. Especificación de los Indicadores Elementales que evalúan los atributos *Activities Description Availability* y *Conceptual Base Structure Richness*

Una vez especificadas todas las métricas –y considerando el proceso discutido en la sección 3- se continuó con la actividad *Diseñar la Evaluación*. Para cada atributo (o requerimiento no-funcional elemental) se especificó un indicador elemental, como se observa en la plantilla de la Fig. 7. Aquí se definen metadatos como su nombre, peso, modelo elemental, el cual especifica la manera de realizar el mapeo del valor medido del atributo (a partir de la métrica) a un valor que indica el nivel de satisfacción alcanzado según los objetivos establecidos en la necesidad de información. Respecto a la escala y criterios de decisión (o niveles de aceptabilidad), se optó por utilizar los mismos para todos los indicadores, tanto elementales como parciales y globales. A modo de ejemplo, en la Fig. 7 se detallan los indicadores elementales que evalúan los atributos *Activities Description Availability* y *Conceptual Base Structure Richness*.

Por otra parte, al momento de establecer los indicadores parciales/globales, también se decidió emplear un mismo modelo global para todos los indicadores parciales y globales con el fin de calcular y evaluar el modelo de *Capability Quality*.

<u>Global (Aggregation) Model:</u>		
<u>Function:</u>		
<u>Name:</u> LSP		
<u>Specification:</u> $GI(r) = (W_1 * I_1^r + W_2 * I_2^r + \dots + W_m * I_m^r)^{1/r}$		
donde:		
- GI represents the partial or global indicator to be calculated		
- I _i are the lower-level indicators values, $0 \leq I_i \leq 100$		
- W _i represents weights, $(W_1 + W_2 + \dots + W_m) = 1$; $W_i > 0$; $i = 1 \dots m$		
- r is a conjunctive/disjunctive coefficient for the LSP aggregation model		
<u>Numerical Scale:</u>		
<u>Scale Type:</u> absolute	<u>Unit name:</u> Percentage	<u>Acronym:</u> %
<u>Decision Criteria/Acceptability Levels:</u>		
if $0 \leq X \leq 45$: “unsatisfactory” indicates change actions must take high priority.		
if $45 < X \leq 70$: “marginal” indicates a need for improvement actions.		
if $70 < X \leq 100$: “satisfactory” indicates satisfactory quality of the analyzed feature.		

Fig. 8. Especificación del modelo global, escala y criterios de decisión.

La Fig. 8 muestra la especificación del modelo global, junto a los criterios de decisión y escala establecidos. Como modelo global se decidió utilizar un modelo de agregación lógica de preferencias, particularmente el modelo LSP (*Logic Scoring of Preference*), el cual está basado en pesos para modelar importancias relativas, y permite utilizar operadores lógicos para modelar distintos niveles de intensidad de conjunción/disyunción (“y/o”) entre características y sub-características del modelo de concepto. Este modelo fue seleccionado debido a que se consideró más apropiado por su sensibilidad un modelo de agregación lógico que uno meramente lineal y aditivo. Básicamente esta decisión está justificada en el hecho de que un modelo de agregación lógico multi-criterio permite manejar consistentemente relaciones de simultaneidad, reemplazabilidad y neutralidad entre elementos de un árbol, además de soportar el problema de modelar requerimientos obligatorios (mandatorios), del que un modelo lineal y aditivo carece. Debido al modelo de agregación elegido, se tuvo que especificar para cada indicador su peso, y para cada indicador global/parcial el operador que transforma entradas en salidas. Finalmente, el documento de

especificación de la evaluación quedó conformado por 48 indicadores de los cuales 17 son indicadores globales y parciales, y 31 indicadores elementales.

Si nos remitimos a la Fig. 2, podemos observar que como resultado del diseño de la M&E ahora contamos con documentos de especificación de requerimientos no-funcionales, de especificación de métricas y de indicadores. El siguiente paso es implementar la medición y evaluación con el fin de obtener las medidas y los cálculos que permitan (a partir del proceso *Analizar y Recomendar*) comprender y comparar las estrategias integradas de M&E según la necesidad de información establecida. El estudio permitirá realizar un análisis riguroso de debilidades y fortalezas de estrategias como GQM, GOCAME, FMESP y recomendar posibles mejoras.

4 Conclusiones y Trabajos Futuros

En el presente trabajo hemos argumentado que una estrategia integrada de M&E debe estar basada en tres principios o fundamentos, a saber: una especificación explícita y formal de un proceso de M&E; un modelo conceptual que defina todos los términos, propiedades y relaciones del dominio; y un soporte metodológico y tecnológico, en el que métodos y herramientas se puedan instanciar a partir del proceso y componentes del marco conceptual. La integración de estos tres pilares permitirá una robusta implantación de programas y proyectos de M&E en una organización facilitando el análisis, comparación y justificación de los resultados de un modo consistente, y por ende la mejora en la toma de decisiones.

Existen muchas propuestas publicadas en el área de M&E, como se discutió en la sección de trabajos relacionados, sin embargo ninguno de ellos considera la necesidad de una estrategia integrada basada en esos tres pilares. Con el fin de comprender y ulteriormente mejorar este recurso es necesario realizar un estudio comparativo que permita un análisis de debilidades y fortalezas de las estrategias existentes teniendo en cuenta la calidad de sus capacidades. Conforme a la búsqueda sistemática de literatura que hemos efectuado, no existen al presente trabajos relacionados que traten con la evaluación de la calidad de las capacidades de estrategias integradas de M&E, basadas en esos tres fundamentos.

En este artículo, hemos ilustrado por razones prácticas la fase de diseño del estudio comparativo, incluyendo los requerimientos no funcionales, la medición y la evaluación para el foco que hemos denominado calidad de las capacidades. Por lo que en otro manuscrito documentaremos la implementación del estudio comparativo y una discusión rigurosa de los hallazgos encontrados.

En definitiva, la contribución en general de esta investigación es brindar conclusiones/recomendaciones a partir del análisis de fortalezas y debilidades de los enfoques seleccionados para luego establecer cursos de acción con el fin de mejorar las estrategias integradas de M&E. Como línea de avance futura y contribución particular, nuestro último objetivo es considerar como factor clave de mejora a incorporar a nuestra estrategia GOCAME las mejores características encontradas en el estudio de evaluación comparativa de las distintas estrategias.

Reconocimientos. Esta investigación está soportada parcialmente por el proyecto PAE-PICT 2188 de la Agencia de Ciencia y Tecnología, Argentina.

References

1. Abran, A., Sellami, A., Suryn, W.: Metrology, measurement and metrics in software engineering. IEEE International Symposium on Software Metrics (2003).
2. Basili, R., Caldiera, G., Rombach, H. D.: The goal question metric approach. In Encyclopedia of Software Engineering. Vol. 1. pp 528-532. (1994).
3. Becker, P., Olsina, L. Towards Support Processes for Web Projects, To appear in LNCS, Springer, 1st Int'l Workshop QWE at ICWE2010, Vienne, Austria. (2010).
4. Briand, L.C., Morasca, S., Basili, V.R.: An operational process for goal-driven definition of measures. 28(12), pp. 1106-1125. (2002).
5. CMMI Product Team. CMMI for Development, Ver.1.2. SEI, USA, (2006).
6. Fenton, N., Pfleeger, S.: Software metrics: a rigorous and practical approach. PWS (1997).
7. García, F., Bertoa, M.F., Calero, C., Vallecillo, A., Ruiz, F., Piattini, M., Genero, M.: Towards a consistent terminology for software measurement. Information & Software Technology. 48(8):631–644 (2006).
8. García, F., Ruiz, F., Bertoa, M.F., Calero, C., Genero, M., Olsina, L., Martín, M., Quer, C., Condori, N., Abrahao, S., Vallecillo, A., Piattini, M.: An ontology for software measurement. University of Castilla-La Mancha, Spain, TR UCLM DIAB-04-02-2. (2004).
9. Goethert, W., Fisher, M.: Deriving Enterprise-Based Measures Using the Balanced Scorecard and Goal-Driven Measurement Techniques, Software Engineering Measurement and Analysis Initiative, CMU/SEI-2003-TN-024 (2003).
10. Gresse C., Hoisl B., Wüst J.: A process model for GQM-based measuremet. TR STTIKL (1995).
11. ISO/IEC 14598-5. International Standard, Information technology - Software product evaluation - Part 5: Process for evaluators, Geneva, Switzerland. (1999).
12. ISO/IEC 15939. International Standard, Information technology - Software Engineering: Software Measurement Process, Geneva, Switzerland. (2002).
13. ISO/IEC 25000. Software Engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Guide to SquaRE. (2005).
14. Kan, S. H.: Metrics and Models in Software Quality Engineering. Addison Wesley. (2002).
15. Kitchenham, B.A., Hughes, R.T., Linkman, S.G.: Modeling Software Measurement Data. IEEE Transactions on Software Engineering. 27(9), pp. 788-804. (2001).
16. Molina H., Olsina L.: Assessing Web Applications Consistently: A Context Information Approach. In proceedings of IEEE CS. 8th ICWE08. pp. 224-230. (2008).
17. Mora, B., García, F., Ruiz, F., Piattini, M., Boronat, A., Gómez, A., Carsí, J. Á., Ramos, I.: Software generic measurement framework based on MDA. IEEE Latin America Transactions, 6 (4), pp. 363-370. (2008).
18. Olsina L., Martín, M.: Ontology for Software Metrics and Indicators. Journal of Web Engineering. Rinton Press. US. 4 Vol. 2. pp. 262-281. (2004).
19. Olsina, L., Papa, F., Molina, H.: How to Measure and Evaluate Web Applications in a Consistent Way. Ch. 13 in Springer book: Web Engineering: Modeling and Implementing Web Applications, Rossi, Pastor, Schwabe & Olsina (Eds), pp. 385–420, (2008).
20. Olsina, L., Rossi, G.: Measuring Web Application Quality with WebQEM. In IEEE Multimedia Magazine. Vol. 9. N° 4. pp. 20-29. (2002).
21. Solingen R., Berghout E.: Improvement by goal-oriented measurement. In: European Software Engineering Process Group conference (E-SEPG). The Netherlands. (1997).
22. Zuse, H.: A Framework of Software Measurement. Walter de Gruyter. Berlin-NY. (1998).