

Análisis Comparativo de Estrategias Integradas de Medición y Evaluación

María Fernanda Papa, Luis Olsina

GIDIS_Web, Facultad de Ingeniería, UNLPam, General Pico, LP, Argentina
{pmfer, olsinal}@ing.unlpam.edu.ar

Abstract. Este trabajo especifica e implementa un estudio para comprender y comparar estrategias integradas de medición y evaluación, considerando a una estrategia como a un recurso desde el punto de vista del ente a evaluar. El objetivo del estudio es comprender 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. El resultado del estudio comparativo permite un análisis de fortalezas y debilidades de las estrategias GQM⁺Strategies (*Goal-Question-Metric*) y GOCAME (*Goal-Oriented Context-Aware Measurement and Evaluation*) con el fin de mejora.

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

1 Introducción

Para llevar adelante programas de medición y evaluación (M&E), las organizaciones de software necesitan establecer un conjunto de actividades para especificar, recolectar, almacenar y utilizar métricas e indicadores. A su vez, para robustecer el análisis y la mejora en la toma de decisiones es necesario asegurar que los valores medidos y los indicadores calculados sean repetibles y comparables entre distintos proyectos de una organización. Por lo que se debe almacenar no sólo datos de mediciones y evaluaciones sino también metadatos de métricas e indicadores como escala, tipo de escala, unidad, método de medición, entre otros.

En [15] se propuso una estrategia de M&E que permite desarrollar programas con estas características. Tal estrategia integra los siguientes fundamentos o capacidades: 1) *Un marco conceptual*, flexible y terminológicamente consistente. Un marco bien establecido debe estar construido sobre una base terminológica robusta (diccionarios, ontologías, etc.), en donde se especifique de manera formal y explícita, los conceptos, relaciones y restricciones acordadas para un dominio particular, además de su agrupación en componentes. 2) *Un proceso de M&E*, que especifica qué hacer, es decir, describe las principales actividades que deben ser planificadas y ejecutadas, sus entradas y salidas, roles, interdependencias, etc. 3) *Métodos y herramientas*, que permiten llevar a cabo la descripción de las actividades.

Por lo que consideramos que una estrategia de M&E es integrada si define un proceso de M&E –que permita garantizar repetitividad y reproducibilidad en la

ejecución de las actividades; un modelo de dominio que defina explícitamente todos los conceptos necesarios y sus relaciones –para garantizar uniformidad terminológica y consistencia en los resultados; y, un soporte metodológico y tecnológico, para instanciar y automatizar actividades de un modo flexible.

Nuestra investigación se centra en el diseño e implementación de la evaluación con el fin de comprender y comparar estrategias integradas de M&E existentes. El objetivo es comprender la *calidad de las capacidades* de dichas estrategias bajo la premisa que una estrategia es apta o adecuada para realizar proyectos y programas de M&E repetibles si posee e integra las tres capacidades mencionadas. La evaluación valorará el nivel de satisfacción alcanzado para estos tres aspectos, y en definitiva, para la calidad de las capacidades. En un trabajo previo [17] se ilustró sólo el diseño del estudio comparativo de las estrategias integradas mientras que en el presente se discuten detalles de su ejecución y análisis, el cual permite comprender fortalezas y debilidades. En el estudio intervinieron GQM⁺Strategies [2, 3] y GOCAME [15] como estrategias seleccionadas, dado que son de dominio público (esto es, existentes en la literatura), con cierto impacto en la academia o industria, y con algún nivel de integración de los tres fundamentos mencionados.

En resumen, las principales contribuciones de esta investigación son: *i)* comprender y comparar estrategias integradas de M&E existentes, considerando a una estrategia como a un ente de tipo recurso; *ii)* diseñar los requerimientos no funcionales teniendo como foco la *calidad de las capacidades* de este recurso; *iii)* implementar un caso de estudio con el fin de analizar y brindar conclusiones/recomendaciones a partir de fortalezas y debilidades detectadas, siendo el objetivo ulterior establecer cursos de acción para la mejora de GOCAME.

El resto del artículo se organiza así: en la Sección 2 damos un panorama de las estrategias evaluadas. La Sección 3, se encuentra dividida en tres sub-secciones donde se ilustra el diseño, la implementación y el análisis de la evaluación respectivamente. En la Sección 4 analizamos la contribución de este trabajo con respecto a otros relacionados y, por último, se destacan las conclusiones y líneas de avance.

2 Panorama de dos Estrategias Integradas de M&E

A seguir, se resumen las dos estrategias seleccionadas que forman parte del estudio comparativo, a saber: GOCAME y GQM⁺Strategies. La selección se basó en los criterios mencionados previamente, siendo importante el criterio de nivel de integración de las tres capacidades (características) enunciadas, con cierto cumplimiento simultáneo: marco/base conceptual, proceso y métodos/técnicas. Con cumplimiento simultáneo queremos indicar que el valor de entrada de estas tres características a la función de agregación (cuya salida es el valor del indicador de la calidad de las capacidades) fuera distinto de cero, como se discutirá más adelante.

2.1 GOCAME

Es una estrategia integrada de M&E que sigue un enfoque orientado a objetivos, sensible al contexto y centrada en la necesidad de información organizacional.

Soporta de manera simultánea las tres capacidades indicadas, que pasamos a describir.

GOCAME posee su base terminológica definida en una ontología [14] a partir de la cual surge el marco conceptual C-INCAMI (*Contextual - InformationNeed, Concept model, Attribute, Metric and Indicator*) [15]. Dicha ontología provee un modelo de dominio que define todos los conceptos, propiedades y relaciones que a su vez ayuda a diseñar las actividades de M&E. El marco C-INCAMI permite especificar los datos y metadatos utilizados en las actividades y artefactos del proceso, como así también, definir catálogos que contengan información reusable entre proyectos e instanciar métodos y 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: permite la definición de una *necesidad de información*¹ la cual pretende satisfacer un requerimiento no funcional de alguna *entidad* para un *propósito* particular y un *punto de vista* de usuario determinado. Los requerimientos no funcionales son representados por un *modelo de concepto* que incluye *conceptos* (características), *subconceptos* (sub-características) y *atributos*. Los atributos son propiedades cuantificables de la entidad bajo análisis; 2) Especificación del Contexto: permite la descripción del *contexto* relevante para una necesidad de información por medio de *propiedades de contexto*, que son *atributos*; 3) Diseño y Ejecución de la Medición: permite la especificación de *métricas directas e indirectas* a utilizar en la *medición*, los *valores medidos*, el *método de medición o cálculo* y la *escala* de valores; 4) Diseño y Ejecución de la Evaluación: permite la definición de la *evaluación* mediante el uso de *indicadores*, que especifican cómo interpretar los valores obtenidos de los atributos y de los conceptos de más alto nivel del modelo de concepto, para conocer el grado de satisfacción de todos los requerimientos para una necesidad de información. Un *indicador elemental* interpreta el valor medido de un atributo mediante el uso de un *modelo elemental*, mientras que, un *indicador global* permite evaluar conceptos de más alto nivel de abstracción, y su valor se calcula a partir de otros indicadores por medio de un *modelo global*. A su vez la *escala* del indicador tiene *criterios de decisión*; 5) Análisis y Recomendación: soporta el análisis de datos y brinda recomendaciones para la toma de decisiones.

GOCAME tiene un proceso de M&E bien definido y modelado (ver Fig. 1) que consta de seis procesos principales: 1) *Definir los requerimientos no funcionales*; 2) *Diseñar la medición*; 3) *Implementar la medición*; 4) *Diseñar la evaluación*; 5) *Implementar la evaluación*; y 6) *Analizar y recomendar*. Cada uno de estos procesos se encuentra desglosado en actividades, sub-actividades y tareas modelados en SPDM y UML. La descripción textual de las actividades, utiliza los conceptos definidos en su base terminológica (al igual que los diagramas de actividad) y documenta el objetivo, precondiciones, post-condiciones, entradas y salidas, etc. (ver detalles en [4, 5]).

Respecto a la metodología y herramientas, GOCAME es soportada por WebQEM [16] y C-INCAMI_Tool [15]. La metodología aporta el 'cómo' implementar los procesos de medición, evaluación y análisis con el fin de dar recomendaciones. Comprende un conjunto de métodos, técnicas y herramientas para llevar a cabo las actividades del proceso introducido.

¹ Los conceptos en cursiva se refieren a términos definidos en la ontología [14].

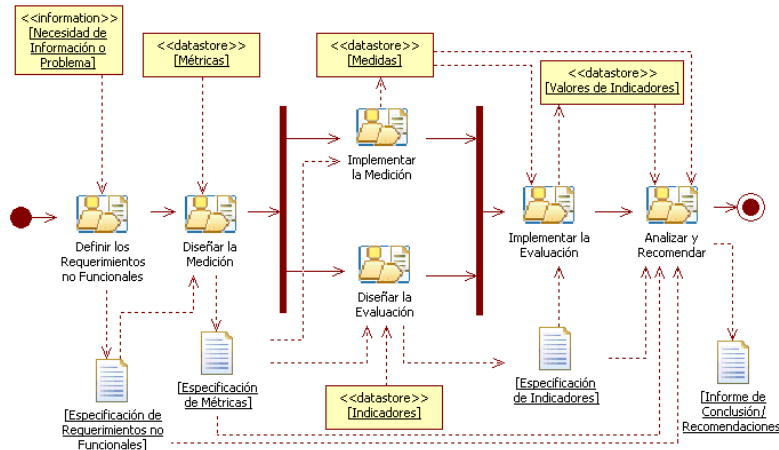


Fig. 1. Principales procesos de la estrategia de medición y evaluación GOCAME.

2.2 GQM⁺Strategies

Es un enfoque de medición construido sobre GQM [1] que permite planificar e implementar programas de medición de software orientados a objetivos. GQM⁺Strategies agrega a GQM un mecanismo para enlazar explícitamente los objetivos de medición de software con los objetivos de alto nivel de la organización, como así también, objetivos y tácticas a todos los niveles del negocio. Este enlace ayuda a justificar los esfuerzos realizados en la medición de software y, a su vez, los datos recolectados contribuyen a la toma de decisiones de alto nivel.

La base terminológica de GQM⁺Strategies está formada por un conjunto de términos agrupados en un glosario [2, 3]. Los términos forman parte de dos componentes (ver Fig. 2) denominados *Elementos de GQM⁺Strategies* y *Grafo GQM*. A estos términos se le deben agregar aquellos vocablos que pertenecen al enfoque GQM y que están presentes en la descripción del mismo.

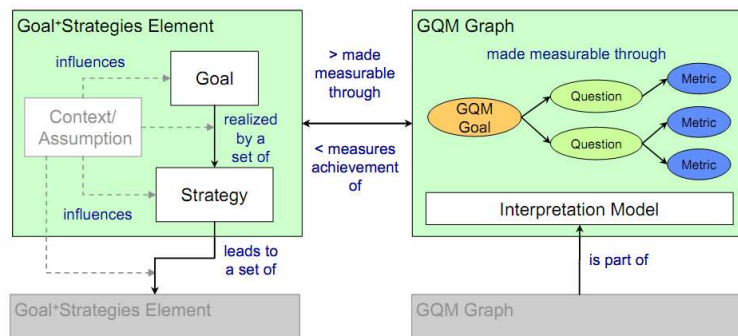


Fig. 2. Conceptos y relaciones de los componentes de GQM⁺Strategies (tomado de [3]).

El primer componente incluye un objetivo simple (*Goal*) y las tácticas (*Strategy*)

derivadas, tanto como toda la información de contexto y las suposiciones (*Context/assumption*) que explican cómo los objetivos y las tácticas se enlazan. En cambio, el segundo componente, refleja un objetivo GQM simple, las preguntas (*Question*), métricas (*Metric*) correspondientes y un *modelo de interpretación*. Este modelo permite la interpretación de los datos para determinar si los objetivos han sido alcanzados. Entre ambos componentes se puede ver una relación bidireccional, por lo que un *Elemento Goal⁺Strategies* es medible por medio de un *grafo GQM* y a su vez el grafo GQM mide el logro de un *Elemento Goal⁺Strategies*.

Con respecto a su proceso explícito, los autores especifican dos procesos principales: 1) *relacionar los objetivos de negocio de alto nivel a objetivos más operativos*, mediante el uso de escenarios y tácticas; 2) *relacionar objetivos de medición con preguntas, y éstas, con sus métricas*. El primero involucra actividades tales como: *Elicitar suposiciones y contexto en general*; *Definir objetivos de alto nivel*; *Tomar decisiones tácticas*; y *Definir objetivos*. El segundo proceso involucra actividades como: *Identificar objetivos de medición potenciales*; *Determinar las preguntas, métricas, criterios de evaluación y modelos de interpretación*; e, *Identificar relaciones entre los modelos de interpretación*. Luego resta *implementar la medición e interpretar los resultados*.

En la literatura de dominio público revisada no se menciona ninguna metodología para GQM⁺Strategies, mientras que el enfoque GQM define explícitamente una, que cubre las fases de planificación, definición, recolección de datos e interpretación. En [18] se describe las características de una herramienta que es configurable para cada programa de medición en una organización.

3 Diseño e Implementación del Estudio Comparativo

En esta sección se presenta el diseño e implementación del estudio comparativo de las estrategias integradas de M&E antes mencionadas. Se divide en tres sub-secciones que involucran: el diseño del estudio (3.1), la implementación (3.2) y el análisis de los resultados (3.3). Notar que, tal cual fue discutido en [17], GOCAME fue utilizada para la realización del diseño y ejecución del estudio comparativo.

3.1 Diseño

La primera actividad fue la definición de los requerimientos no funcionales. 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 super-categoría *recurso*). El concepto calculable raíz es la *Calidad de las Capacidades*, y dado que no existe un modelo de conceptos estándar ni tampoco en la literatura relacionada que lo especifique, fue necesario definir un modelo *de facto*.

Las sub-características asociadas a la *Calidad de las Capacidades* son los que representan a los tres fundamentos (ver 1^{ra} columna en tabla 1), y son: *Calidad de las Capacidades del Proceso* (codificado 1.1); *Calidad de las Capacidades del Marco Conceptual* (1.2), y *Calidad de las Capacidades de la Metodología* (1.3).

Tabla 1. Especificación del árbol de requerimientos en la 1^{ra} columna (en itálica, *atributos*). La leyenda (1) en la 2^{da} columna representa los valores de los indicadores elementales, parciales y global (en %) para GOCAME y la leyenda (2) en la 3^{ra} columna para GQM⁺Strategies.

	(1)	(2)
1. Calidad de las Capacidades (para una Estrategia de M&E)	66,48	45,89
1.1 Calidad de las Capacidades del Proceso	58,88	54,34
1.1.1 Adecuación de las Actividades	46,67	38,37
<i>1.1.1.1. Disponibilidad de la Descripción de las Actividades</i>	31,91	24,75
<i>1.1.1.2. Completitud de la Descripción de las Actividades</i>	15,47	14,40
<i>1.1.1.3. Granularidad del Proceso</i>	70	70
<i>1.1.1.4. Formalidad de la Descripción de las Actividades</i>	100	61,39
<i>1.1.1.5. Disponibilidad de Asignación de Roles a Actividades</i>	0	17,82
1.1.2 Adecuación de los Artefactos	3	31,58
<i>1.1.2.1. Disponibilidad de la Descripción de los Artefactos</i>	0	27,59
<i>1.1.2.2. Completitud de la Descripción de los Artefactos</i>	0	26,96
<i>1.1.2.3. Granularidad de los Artefactos</i>	30	70
1.1.3 Adecuación del Modelado del Proceso	83,56	70,70
1.1.3.1 Adecuación de la Vista Funcional	88	76,42
<i>1.1.3.1.1. Disponibilidad de la Vista Funcional</i>	100	100
<i>1.1.3.1.2. Completitud de la Vista Funcional</i>	100	61,39
<i>1.1.3.1.3. Granularidad de la Vista Funcional</i>	70	70
1.1.3.2 Adecuación de la Vista Informacional	82,13	74,97
<i>1.1.3.2.1. Disponibilidad de la Vista Informacional</i>	100	100
<i>1.1.3.2.2. Completitud de la Vista Informacional</i>	90,32	72,41
<i>1.1.3.2.3. Granularidad de la Vista Informacional</i>	30	30
1.1.3.3 Adecuación de la Vista de Comportamiento	88	60,42
<i>1.1.3.3.1. Disponibilidad de la Vista de Comportamiento</i>	100	100
<i>1.1.3.3.2. Completitud de la Vista de Comportamiento</i>	100	61,39
<i>1.1.3.3.3. Granularidad de la Vista de Comportamiento</i>	70	30
1.1.3.4 Adecuación de la Vista Organizacional	0	63,78
<i>1.1.3.4.1. Disponibilidad de la Vista Organizacional</i>	0	100
<i>1.1.3.4.2. Completitud de la Vista Organizacional</i>	0	44,44
<i>1.1.3.4.3. Granularidad de la Vista Organizacional</i>	0	30
1.1.4 Conformidad del Proceso	85,79	71,11
<i>1.1.4.1. Conformidad del Proceso a la Terminología de la Base Conceptual</i>	94,74	88,89
<i>1.1.4.2. Conformidad del Proceso a un Estándar de Procesos de M&E</i>	50	0
1.2 Calidad de las Capacidades del Marco Conceptual	75,09	35,82
1.2.1 Adecuación del Marco Conceptual	75	25
<i>1.2.1.1. Modularidad del Marco Conceptual</i>	50	0
<i>1.2.1.2. Formalidad del Modelado del Marco Conceptual</i>	100	50
1.2.2 Adecuación de la Base Conceptual	68,53	18,53
<i>1.2.2.1. Completitud de la Base Conceptual</i>	21,53	1,33
<i>1.2.2.2. Riqueza de la Estructura de la Base Conceptual</i>	100	30
1.2.3 Conformidad del Marco Conceptual	84,31	81,82
<i>1.2.3.1. Conformidad del Marco Conceptual a la Terminología de la Base Concep.</i>	84,31	81,82
1.3 Calidad de las Capacidades de la Metodología	77,43	57,35
1.3.1 Adecuación de la Metodología	83,19	51,88
<i>1.3.1.1. Disponibilidad de la Metodología</i>	100	100
<i>1.3.1.2. Completitud de la Asignación de Métodos a Actividades</i>	82,98	29,70
<i>1.3.1.3. Disponibilidad de Herramienta que Automatice a la Metodología</i>	50	0
1.3.2 Conformidad de la Metodología	73,68	61,11
<i>1.3.2.1. Conformidad de la Metodología a la Terminología de la Base Conceptual</i>	73,68	61,11

Los conceptos en el árbol de requerimientos deben estar definidos; por ejemplo, *Calidad de las Capacidades del Proceso (1.1)* representa “el grado en que el proceso es adecuado y apropiado para soportar y realizar las acciones especificadas”. Esta sub-característica a su vez está integrada por cuatro subconceptos que representan cuán adecuadas son las actividades, los artefactos, el modelado del proceso y sus vistas, y la conformidad del proceso a estándares y a su base terminológica. La sub-característica *Adecuación de las Actividades (1.1.1)* se define como “el grado en que las actividades proveen propiedades que satisfacen necesidades explícitas e implícitas para el propósito de usuario previsto”. Como se observa en la tabla 1, *Adecuación de las Actividades* está conformada por cinco atributos; por ejemplo, el atributo *1.1.1.2* se define como “el grado en que las actividades enunciadas están descritas” y su objetivo es “conocer el grado de completitud de las descripciones de las actividades enunciadas”. El total de atributos mensurables relacionados a sub-características son treinta y uno.

Una vez instanciado y especificado el árbol la siguiente actividad es seleccionar las métricas que cuantifiquen los atributos. Para el atributo de nuestro ejemplo se seleccionó desde el repositorio la métrica indirecta *Grado de Completitud de la Descripción de las Actividades (GCDA)* cuyo objetivo es “calcular el grado de completitud que presenta la descripción de las actividades enunciadas del proceso”, su escala es porcentaje (%) y la especificación de su fórmula es: (Ec. 1)

$$GCDA=0 \text{ si TAE} = 0, \text{ sino } GCDA=[(\#AMD*0,10+\#APD*0,35+\#ACD*0,55)/(\text{TAE}*0,55)]*100,$$

siendo TAE el Número Total de Actividades Enunciadas; #ACD el Número de Actividades Completamente Descriptas; #APD el Número de Actividades Parcialmente Descriptas; y #AMD el Número de Actividades Mínimamente Descriptas.

Una vez seleccionadas todas las métricas que cuantifican a los 31 atributos (esto es, 16 directas y 15 indirectas) se comenzó con el diseño de la evaluación. Para calcular indicadores parciales/globales se utilizó el modelo de agregación LSP (*Logic Scoring of Preference*) [7], el cual soporta pesos y operadores lógicos que permiten modelar relaciones de simultaneidad (operadores C) y reemplazabilidad (operadores D) entre características y sub-características. Por ejemplo, el operador de conjunción débil C-+ permite modelar el criterio de simultaneidad entre las sub-características 1.1, 1.2 y 1.3 (que era uno de los criterios de selección de las estrategias integradas como se indicó al inicio de la Sección 2). La tabla 2 muestra la especificación del modelo LSP.

Tabla 2. Especificación del modelo de agregación LSP.

$GI(r) = (W_1 * I_1^r + W_2 * I_2^r + \dots + W_m * I_m^r)^{1/r}$
Donde:- GI representa el indicador parcial o global a ser calculado.
- I_i son los valores de los indicadores del nivel más bajo, $0 \leq I_i \leq 100$.
- W_i representa los pesos, $(W_1 + W_2 + \dots + W_m) = 1$; $W_i > 0$; $i = 1 \dots m$.
- r es un coeficiente conjuntivo/disyuntivo para el modelo de agregación LSP.

Además, se definieron los niveles de aceptabilidad (criterio de decisión) en una escala porcentual utilizados para la interpretación de los valores de los indicadores obtenidos. Un valor entre *cero y cincuenta ([0-50])* corresponde a un *nivel insatisfactorio* que indica que *acciones de cambio deben tomarse con una alta prioridad*. Un valor entre *cincuenta y setenta y cinco ([50-75])* corresponde a un *nivel marginal* que indica que *se deberían tomar acciones de mejora*. Mientras que un valor entre *setenta y cinco y cien ([75-100])* corresponde a un *nivel satisfactorio*. Un

indicador elemental evalúa el grado de satisfacción o nivel de aceptabilidad alcanzado por un requerimiento elemental (atributo), en tanto que un indicador parcial/global evalúa el grado de satisfacción alcanzado por un requerimiento parcial (sub-característica) o global (característica).

Como resultado del diseño del estudio se obtuvieron los documentos de *Especificación de los requerimientos no funcionales*, *Especificación de métricas* y *Especificación de indicadores* (ver [17] para mayor detalle del diseño).

3.2 Implementación

La recolección de los datos de la medición se realizó sobre material publicado y accesible (libros, artículos científicos, tesis de posgrado, etc.) referente a ambas estrategias de M&E. Se tomaron aquellos documentos más relevantes luego de una lectura exhaustiva de los mismos, descartando aquellos en que no participara al menos algún miembro del grupo de autores originales de la investigación. Se le dio además mayor prioridad a los documentos más actuales en tanto representaran una contribución respecto de los previos.

Tabla 3. Valores medidos de las métricas directas para el cálculo de la métrica GCDA (Ec. 1).

	GOCAME	GQM ⁺ Strategies
Número Total de Actividades Enunciadas (TAE)	47	101
Número de Actividades Mínimamente Descriptas (#AMD)	5	3
Número de Actividades Parcialmente Descriptas (#APD)	10	22
Número de Actividades Completamente Descriptas (#ACD)	0	0

Siguiendo con el ejemplo anterior, la tabla 3 muestra los valores medidos para las métricas directas a partir de las cuales se calcula el valor de la métrica indirecta *GCDA* (según la Ecuación 1), resultando para GOCAME el valor de 15,47% y 14,40% para GQM⁺Strategies. Así se fueron midiendo todos los atributos a partir de la definición de sus métricas correspondientes.

En la implementación de la evaluación, se calcularon los indicadores elementales a partir de los valores medidos de cada atributo conforme al modelo elemental de indicador correspondiente (ver [17] para un ejemplo de modelo elemental). Luego se aplicó el modelo parcial/global especificado en la tabla 2 para obtener los valores de los indicadores parciales y globales. La tabla 1 muestra todos los valores de indicadores (esto es, elementales, parciales y global) para nuestro estudio comparativo.

3.3 Análisis de Resultados

Para justificar los resultados obtenidos se realizó el análisis de la información surgida de los indicadores y a su vez de los datos de las métricas. Como producto se obtuvo informes de conclusión y recomendación, los cuales están conformados por tablas, gráficos, cuadros comparativos, por ejemplo, el que resume debilidades/fortalezas y acciones de mejora para facilitar la ulterior toma de decisiones. Por cuestiones de espacio, resaltaremos algunos aspectos del estudio comparativo de un modo textual.

Analizando los resultados de la tabla 1, se observa que el nivel de satisfacción alcanzado para el indicador global de la *Calidad de las Capacidades* para las estrategias evaluadas es *marginal* para GOCAME (66,48%, lo que ya implica como conclusión que se *deberían tomar acciones de mejora*) e *insatisfactorio* para GQM⁺Strategies (45,89%, lo que implica que *urgen cambios*). Teniendo en cuenta el criterio de simultaneidad de las tres sub-características que conforman al concepto foco (a saber: proceso, marco conceptual y método), observamos que para *Calidad de las Capacidades del Proceso* (1.1) ambas estrategias están en el nivel marginal (GOCAME: 58,88 / GQM⁺Strategies: 54,34); que para *Calidad de las Capacidades del Marco Conceptual* (1.2) el valor del indicador en GOCAME es 75,09 por lo que está en el nivel satisfactorio, en tanto que GQM⁺Strategies cae al 35,82% de satisfacción de este requerimiento; y que para *Calidad de las Capacidades de la Metodología* (1.3) el valor del indicador en GOCAME también es satisfactorio (77,43), en tanto que GQM⁺Strategies alcanza el nivel marginal (57,35).

Realizando un refinamiento en el análisis comparativo (a partir de la tabla 1), para la sub-característica 1.1 (la cual está conformada por 1.1.1 *Adecuación de las Actividades*, 1.1.2 *Adecuación de los Artefactos*, 1.1.3 *Adecuación del Modelado del Proceso*, y 1.1.4 *Conformidad del Proceso*), observamos que para 1.1.1 ambas estrategias caen en el nivel insatisfactorio. Además, *Adecuación de las Actividades*, combina cinco atributos. Uno de sus atributos ilustrado previamente en la sub-sección 3.1, el 1.1.1.2 *Complejidad de la Descripción de las Actividades*, da 15,47 para GOCAME y 14,40 para GQM⁺Strategies. El diseño de la métrica (Ec. 1) y su valor medido (y a su vez el de sus métricas directas mostradas en tabla 3), nos ayuda a comprender las razones por la que el indicador elemental ranquea insatisfactoriamente, y nos permite realizar recomendaciones para su ulterior mejora. Por ejemplo, la *Complejidad de la Descripción de las Actividades* está en función de las actividades descriptas (ya sean mínima, parcial o completamente descriptas) versus las enunciadas. Se considera (por definición del atributo) que una actividad está descripta completamente cuando está enunciada y presenta de forma textual su objetivo específico, descripción, pre-condiciones, post-condiciones, entradas y salidas. Y una actividad es enunciada cuando pertenece al proceso bajo análisis y su etiqueta tiene un único significado. De los datos arrojados en tabla 3 para la métrica directa *Número de Actividades Completamente Descriptas* se concluye que ninguna estrategia posee la descripción textual de los campos arriba indicados. Por lo que una recomendación de acción de mejora es definir una plantilla con dichos campos, para luego completar todas las descripciones.

Por otra parte, para 1.1.3 *Adecuación del Modelado del Proceso*, observamos que para GOCAME la vistas de proceso funcional, informacional y de comportamiento se encuentran en un nivel satisfactorio (ver en tabla 1, los atributos evaluados). Sin embargo, la vista organizacional está completamente ausente en esta estrategia, en tanto que para GQM⁺Strategies da 63,78%. Notar que en términos generales, una vista organizacional específica qué agentes intervienen en la realización de qué actividades en cumplimiento de roles. Por lo tanto analizando las métricas de los tres atributos empleados, y las fortalezas observadas en GQM⁺Strategies, nos puede guiar a la mejora de estos atributos para GOCAME.

Por último, y como indicado previamente, para *Calidad de las Capacidades del Marco Conceptual* (1.2), GOCAME está en el nivel satisfactorio (75,09), en tanto que

GQM⁺Strategies cae al nivel insatisfactorio (35,82). En la tabla 1 se observa que intervienen 5 atributos combinados a sub-características. Por ejemplo, el atributo *Riqueza de la Estructura de la Base Conceptual (1.2.2.2)* se satisface el 100% para GOCAME. Teniendo en cuenta que una base conceptual puede ser, desde el punto de vista de su estructuración, una ontología, una taxonomía, un diccionario o glosario, GOCAME emplea una ontología, en tanto que GQM⁺Strategies usa un glosario de términos.

En resumen, observando los valores de los indicadores elementales de las tres capacidades evaluadas en ambas estrategias, se puede concluir que existen fortalezas como así también varias oportunidades de mejora.

4 Trabajos Relacionados

En esta sección presentamos trabajos relacionados, a saber, estrategias, enfoques, modelos y estándares que presentan algunas de las tres capacidades discutidas (modelo conceptual, proceso, método/herramienta) justificando el porqué no fueron incluidas en el estudio comparativo.

En [17] se discutieron distintas propuestas publicadas en el área de M&E entre las que se citaron a Kitchenham *et al.* [12], los estándares ISO 14598 [9] y 15939 [10], y el estándar de facto CMMI (*Capability Maturity Model Integration*) [6]. Estos trabajos se centran en algunos aspectos de M&E como lo son la definición de términos y modelos conceptuales, procesos de evaluación y medición, etc., pero sin considerar la integración de las capacidades mencionadas. Además el estándar ISO 25000 [11], denominado *Requerimientos y Evaluación de Calidad de Productos de Software* (SQuaRE), provee una guía para el uso de la nueva serie de estándares SQuaRE, de los cuales algunos están basados en normas previas como ISO 14598, entre otras. El estándar ISO 25000 tiene como objetivo dar un panorama de los contenidos de los documentos SQuaRE, definiciones y modelos de referencia, entre otros aspectos, como la arquitectura SQuaRE. Si bien se observa en el objetivo de la serie, cierto nivel de integración de las capacidades mencionadas, en definitiva no participó del estudio dado que al momento de la recolección de los datos (de setiembre a diciembre de 2010) no se habían publicado algunos documentos claves de la serie.

GQM es un enfoque de medición propuesto por Basili *et al.* [1] utilizado ampliamente en la academia e industria, que ha sido ampliado por ejemplo con la definición de su proceso [18] y complementado con otras investigaciones. Este enfoque si bien no fue seleccionado directamente como una entidad a evaluar, fue considerado como parte de la estrategia integrada GQM⁺Strategies [3].

Otro aporte respecto a enfoques de M&E es CQA-Meth (*Continuous Quality Assessment*) [8], una metodología flexible que permite la evaluación de la calidad de cualquier tipo de modelo software. La metodología junto con la herramienta que la soporta forma el entorno integrado CQA que puede ser utilizado por empresas para realizar evaluaciones de la calidad de sus propios productos o de terceros. CQA-Meth define los procesos necesarios para llevar a cabo la evaluación de los modelos UML, y facilitar la comunicación entre el cliente (patrocinador de la evaluación) y el equipo

de evaluación. CQA-Meth no se seleccionó dentro de las estrategias a evaluar dado que no satisfacía los tres criterios simultáneamente. En particular, carece de un marco conceptual explícito a partir de una base terminológica. Si bien CQA-Meth surge del mismo grupo de investigación que el enfoque FMESP [13] el cual posee un marco conceptual basado en una ontología, no queda constancia explícita de su relación en la literatura ni en las referencias proporcionadas por sus autores.

5 Conclusiones y Trabajos Futuros

A seguir, enfatizaremos algunas conclusiones a partir de las contribuciones indicadas en la sección 1, a saber: *i*) el diseño de un estudio comparativo de estrategias integradas de M&E existentes, considerando a una estrategia como a un ente de tipo recurso, desde el punto de vista de la categoría de entidad, y teniendo como foco de evaluación a *la calidad de las capacidades*; *ii*) implementar el estudio con el fin de analizar y brindar conclusiones/recomendaciones a partir de fortalezas y debilidades detectadas, siendo el objetivo ulterior establecer cursos de acción para la mejora de GOCAME.

Respecto de la primera contribución, además del diseño del estudio, que abarca la especificación de requerimientos no funcionales, el diseño de la medición y evaluación (ya documentados en [17]), en el presente artículo se seleccionaron las dos estrategias integradas (GOCAME y GQM⁺Strategies) como entes concretos a evaluar. Se estableció claramente qué se entiende por estrategias integradas de M&E que satisfagan simultáneamente los tres criterios o capacidades (modelo conceptual, proceso, método/herramienta), entre otros criterios de selección; se dio un panorama en la sección 2 de ambas estrategias y se justificó en trabajos relacionados, el por qué otros métodos, enfoques o estrategias no fueron seleccionadas para el presente estudio.

Considerando la segunda contribución, se llevó a cabo, entre setiembre y diciembre del 2010, la recolección de datos a partir de la especificación de las métricas teniendo en cuenta el material publicado y accesible referente a ambas estrategias de M&E, el cálculo de los indicadores y el análisis de los resultados. En la sección 3.3 se resumieron algunas conclusiones y recomendaciones a partir de fortalezas y debilidades detectadas.

El nivel de satisfacción alcanzado para el indicador global de la calidad de las capacidades ha resultado *marginal* para GOCAME (66,48%) e *insatisfactorio* para GQM⁺Strategies (45,89%). Observando con mayor detalle los valores de los indicadores parciales y elementales de las tres capacidades evaluadas en ambas estrategias, se pudo concluir que existen varias fortalezas como así también debilidades. Estas oportunidades de mejora detectadas para GOCAME (ya surgidas tanto del análisis de las fortalezas específicas de GQM⁺Strategies como de los requerimientos no-funcionales insatisfechos para GOCAME) es motivo de una línea actual de investigación para robustecer a nuestra estrategia (que era el objetivo final del presente trabajo).

Cabe destacar que GOCAME ha servido como estrategia general por la cual se ha diseñado una estrategia particular denominada SIQinU (*Strategy for Improving*

Quality in Use) que se alinea con las tres capacidades mencionadas [5]. SIQinU ha sido empleada en un caso de estudio real en una empresa de testing. Otra línea futura de trabajo es ampliar el alcance de GOCAME para que soporte objetivos de M&E a distintos niveles organizacionales, tal cual lo hace GQM⁺Strategies.

Referencias

1. Basili R., Caldiera G., Rombach H. D.: The goal question metric approach. In *Encyclopedia of Software Engineering*. Vol. 1. pp 528-532. (1994).
2. Basili V.R., Heidrich J., Lindvall M., Münch J., Regardie M., Rombach H.D., Seaman C.B., Trendowicz A.: GQM strategies®: A comprehensive methodology for aligning business strategies with software. pp. 253-266. (2007).
3. Basili V.R., Lindvall M., Regardie M., Seaman C., Heidrich J., Jurgen M., Rombach D., Trendowicz A.: Linking Software Development and Business Strategy through Measurement. *IEEE Computer*, 43(4), 57–65. (2010).
4. Becker P., Molina H., Olsina L.: Measurement and Evaluation as quality driver. In: *Journal ISI (Ingénierie des Systèmes d'Information)*, Special Issue "Quality of Information Systems", Lavoisier, Paris, France, (15) 6, pp. 33-62. (2010).
5. Becker P., Lew P., Olsina, L.: Strategy to Improve Quality for Software Applications: A Process View, To appear in ACM proceedings of ICSE, Int'l Conference of Software and System Process (ICSSP), Honolulu, Hawaii, USA. (2011).
6. CMMI Product Team. CMMI for Development, Ver.1.2. SEI, USA, (2006).
7. Dujmovic, J.J.: A Method for Evaluation and Selection of Complex Hardware and Software Systems, 22nd Int'l Conference for the Resource Management and Performance Evaluation of Enterprise CS. CMG 96 Proceedings, Vol. 1, pp. 368-378. (1996).
8. Genero M., Torre D., Blasco B., Piattini M., Rodríguez M.: CQA-Meth: una Metodología para la Evaluación de la Calidad de los Modelos Software. ASSE'10, Simposio Argentino de Ingeniería de Software (39 JAIIO), pp. 309-322. (2010).
9. ISO/IEC 14598. International Standard, Information technology - Software product evaluation - Part 5: Process for evaluators, Geneva, Switzerland. (1999).
10. ISO/IEC 15939. International Standard, Information technology - Software Engineering: Software Measurement Process, Geneva, Switzerland. (2002).
11. ISO/IEC 25000. Software Engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Guide to SquaRE. (2005).
12. Kitchenham B.A., Hughes R.T., Linkman S.G.: Modeling Software Measurement Data. *IEEE Transactions on Software Engineering*. 27(9), pp. 788-804. (2001).
13. Mora B., García F., Ruiz F., Piattini M., Boronat A., Gómez A., Carsi J. Á., Ramos I.: Software generic measurement framework based on MDA. *IEEE Latin America Transactions*, 6 (4), pp. 363-370. (2008).
14. Olsina L., Martín M.: Ontology for Software Metrics and Indicators. *Journal of Web Engineering*. Santiago de Chile. Rinton Press. US. 4 Vol. 2. pp. 262-281. (2004).
15. 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).
16. Olsina L., Rossi G.: Measuring Web Application Quality with WebQEM. In *IEEE Multimedia Magazine*. Vol. 9. N° 4. pp. 20-29. (2002).
17. Papa F., Becker P., Olsina L.: Estrategias de Medición y Evaluación: Diseño de un Estudio Comparativo. ASSE'10, Simposio Argentino de Ingeniería de Software (39 JAIIO), pp. 309-322. (2010).
18. Solingen R.V., Berghout E.: *The Goal/Question/Metric Method: a practical guide for quality improvement of software development*. McGraw-Hill Co. (1999).