Gestión Sistemática de la Calidad de la Información en los Procesos de Selección de Componentes de Software

Claudia Ayala[1,2], Xavier Franch[1]

[1] Technical University of Catalunya. UPC-Campus Nord. 08034, Barcelona, Spain
[2] Norwegian University of Science and Technology (NTNU). Trondheim, Norway
{cayala, franch}@lsi.upc.edu

Resumen. La selección de componentes de software desarrollados por terceros juega un papel primordial en el desarrollo de sistemas de software basados en componentes. Uno de los problemas que incrementa potencialmente el riesgo de elegir componentes no apropiados es basar las decisiones en información de poca calidad. La información relacionada con los componentes se caracteriza por estar disgregada, crecer de forma continuada, cambiar con frecuencia, provenir de múltiples y variadas fuentes (no siempre confiables), estar frecuentemente desestructurada; y muchas veces incompleta. En este contexto, el análisis y reuso de la información relativa a los componentes y su dominio, así como la gestión de su calidad, representan un factor importante en el riesgo asociado al proceso de selección de componentes. En este artículo, se propone un modelo de referencia para facilitar la colección, almacenamiento, gestión, análisis y reuso de esta información bajo un entorno de aseguramiento de calidad.

Palabras Clave: Calidad de la información, componentes de software, selección.

1 Introducción

Actualmente, el Desarrollo de Software Basado en Componentes pre-existentes (DSBC) se ha convertido en una necesidad económica y estratégica en una gran variedad de áreas de aplicación. La existencia de un enorme y creciente mercado de componentes desarrollados por terceros, tanto de naturaleza comercial como open source (denominados Componentes Off-The-Shelf, OTS) han hecho de esta tecnología la forma "estándar" de desarrollo de software [1]. Sin embargo, a pesar de sus beneficios potenciales (ej., reducción de costes y tiempo de desarrollo), también implica nuevos riesgos y retos para la Ingeniería de Software (IS).

Uno de los procesos más críticos del DSBC es el proceso de selección de componentes a ser integrados: si un componente es seleccionado erróneamente, el riesgo de fracaso de un proyecto se incrementa dramáticamente [2], [3]. Los factores que afectan al proceso de selección son cuantiosos y muy variados y van desde factores funcionales y técnicos a factores políticos y legales [1], [4]. En la última década, un importante sector de la comunidad de IS ha dedicado sus esfuerzos a lidiar con el proceso de selección. Algunos ejemplos son las metodologías CAP, CARE,

CEP, CRE, OTSO, PECA, PORE, QESTA, Scarlet, STACE, etc. (ver estado del arte en [5]). Sin embargo, aún son muchos los factores que deben ser abordados para mejorar el DSBC en la práctica. Uno de ellos es el aseguramiento de la *calidad de la información* (CI) que describe los componentes y su entorno.

La problemática abordada en este artículo reside en que para seleccionar componentes OTS, el *seleccionador* (es decir, el equipo de personas a cargo de la selección) necesita obtener información sobre los componentes con el fin de poder compararlos con los requisitos del sistema a desarrollar, así como con otros componentes candidatos para decidir qué componente debe ser integrado en el sistema. Sin embargo, sin tomar en cuenta que la búsqueda de los componentes candidatos que puedan satisfacer sus necesidades en el exponencialmente creciente mercado de componentes es de por sí una tarea ardua; el seleccionador se encuentra con el hecho de que las descripciones de los componentes están muy lejos de ser estandarizadas u homogéneas. Esto es que generalmente, para cada uno de los componentes existe una cantidad ingente de información disgregada, heterogénea y difícil de procesar, provista por diversos agentes y no siempre fiable, completa o útil [4], [6], [7]. Así pues, el proceso de toma de decisiones sobre qué componente elegir se ve claramente afectado por la CI sobre la que se basan las decisiones, y esto repercute enormemente en la calidad del sistema de software y su desarrollo [8], [9]. Por tanto, asegurar que la información que describe los componentes y su dominio posee niveles adecuados de calidad, permitiría incrementar los factores de éxito del DSBC. Además, el reuso de la información acerca de la CI de los componentes en diversos procesos de selección puede representar un impacto incrementalmente positivo en la eficiencia y efectividad de futuros procesos de selección.

Es importante destacar que en este artículo, nos referimos a la CI o calidad de los datos que describen al componente OTS y su entorno, y no a la calidad del componente en sí. Así pues, por *información del componente* nos referimos a cualquier tipo de dato o información (ej. oral o escrita) que describe propiedades del componente y/o su entorno de uso o dominio. Por ejemplo, documentación provista por el proveedor del componente, estándares relacionados, experiencias de uso, análisis de mercado, etc., que son transmitidas por diversos *agentes* (individuos u organizaciones) a través de diversos medios tales como charlas, blogs, fórums, artículos, libros, etc. También usamos el término *fuentes de información* para referirnos de manera general al conjunto de datos o información relacionada que puede ser referida individuamente (ej. un artículo sobre un componente publicado en una revista, un portal web que contiene información sobre componentes, etc.)

El objetivo principal de la investigación presentada en este artículo es aminorar el riesgo de usar información de poca calidad en los procesos de selección. El contenido se estructura la manera siguiente. La sección 2 describe la motivación del estudio. La sección 3 detalla los procesos, métodos y técnicas usados para capturar las necesidades de CI en la práctica de selección de componentes OTS y su óptima representación. La sección 4 se refiere a la obtención de un esquema de CI mesurable basado en el estándar ISO/IEC 9126-1 y enfocado a procesos de selección. La sección 5 resume los resultados anteriores en un modelo conceptual y subsecuentemente en una herramienta de software que provee un soporte sistemático para evaluar la CI y disminuir el riesgo en la toma de decisiones en la selección de componentes OTS.

2 Motivación

En esta denominada "era de la información" caracterizada por el hecho de que Internet pone al alcance de todos una cantidad ilimitada de información, la investigación en el campo de la CI sobresale como un área emergente y transversal a un sinfín de áreas en que la información juega un papel primordial [10]. Desde años atrás, en el área de las ciencias de la información se han propuesto un amplio conjunto de criterios para evaluar la CI basándose en una cuidadosa exploración de la fuente y su contenido, por ejemplo: reputación de la fuente, reputación del autor, formato de presentación, objetivo, alcance, etc. Sin embargo, el concepto de CI es inherentemente subjetivo y solo puede ser determinado en el contexto específico de su uso [10]. Así pues, estos criterios resultan demasiado genéricos y reconocidamente ineficaces para casos específicos como el de evaluar la CI asociada a la toma de decisiones en procesos de selección de componentes OTS y para procesar la gran cantidad de fuentes de información disponibles en Internet [11], [12]. Por otro lado, en el campo de la selección de componentes OTS, la mayoría de los trabajos existentes se centran en el uso de herramientas de búsqueda semi-automática para identificar componentes (ej. [13], [14], *Google Code Search*); o en el estudio de las propiedades de los componentes para asegurar su correcta integración y satisfacción de los requerimientos [9], [15], [17] (ver survey en [16]). En este contexto, aunque la CI que describe a los componentes ha sido considerada un aspecto crucial [4], [6], [7], [9], el problema de dónde y cómo obtener información acerca de los componentes y asegurar su calidad no ha sido abordado en profundidad [18].

Durante los últimos años, los autores de este artículo hemos desarrollado investigación, y participado en proyectos industriales referentes a la selección de componentes OTS. Nuestra experiencia en el ámbito industrial [5], [16], [17] reveló la dificultad reportada por los ingenieros de software involucrados en procesos de DSBC para recopilar, procesar y analizar la vasta cantidad de fuentes de información para obtener información "confiable" sobre los componentes que contribuya al éxito de las decisiones a tomar. En general, las consecuencias de una pobre CI en la toma de decisiones incluyen problemas de integración, sobre-costo o incluso abandono del proyecto [1], [3]. Así pues, asegurar la CI de los componentes en procesos de selección se considera un problema crítico todavía a resolver [4], [6], [8], [19].

3 El Estudio

El estudio consistió en 2 fases. El objetivo de la primera fase fue identificar empíricamente los problemas relacionados con la CI de los componentes en la práctica industrial de DSBC. La segunda fase está basada en propuestas relevantes en el área de la CI [10], [20], [21], [22] para definir las dimensiones e indicadores de CI de los componentes desde el punto de vista de la práctica de los seleccionadores de componentes OTS.

3.1 Fase 1: Identificando Problemas de la CI en Procesos de Selección

Para identificar los principales dilemas que las compañías afrontan para buscar componentes OTS e información acerca de estos durante los procesos de selección, realizamos un estudio explorativo en 16 compañías que realizan DSBC en Noruega. El estudio consistió en entrevistas semi-estructuradas diseñadas por un equipo de 5

investigadores y 1 estudiante de maestría de la Universidad Noruega de Ciencia y Tecnología (NTNU, Trondheim). El carácter semi-estructurado de la entrevista permitió explorar las necesidades de los seleccionadores de componentes OTS, así como ahondar en cuestiones relevantes que surgieron durante las charlas. Diferentes roles involucrados en el proceso de selección fueron entrevistados: managers de proyecto, arquitectos de software y desarrolladores. Las entrevistas fueron grabadas y posteriormente transcritas para su análisis en equipo. El reporte detallado está disponible en [23]. En general, el proceso de selección de componentes en la industria está orientado a alcanzar los plazos de entrega asociados a los proyectos. En casi todos los casos, los entrevistados manifestaron no tener tiempo suficiente para inspeccionar detalladamente la información relacionada con el componente, ni de reflexionar adecuadamente acerca de su calidad. Así pues, la toma de decisiones raras veces es documentada (aún en los casos de organizaciones que seleccionan componentes periódicamente) y usualmente está basada en factores como recomendaciones de colegas, o incluso intuición -esto coincide con lo reportado en [25]-. Todo esto no solo contribuye a la pérdida continua de conocimiento tácito de miembros del equipo de desarrollo cuando son reemplazados, sino también a incrementar los riesgos de fallo en la toma de decisiones. Sin embargo, la mayoría de los entrevistados también afirmaron que disponer de información con mayores niveles de calidad disminuiría los riesgos de fracaso en la toma de decisiones en proyectos de DSBC.

Para comprender los inconvenientes que los entrevistados dijeron afrontar al buscar y procesar información asociada a los componentes, examinamos en profundidad las fuentes de información que durante las entrevistas, los seleccionadores mencionaron conocer o usar para buscar componentes OTS y/o información relacionada con estos. La Tabla 1 resume algunas de estas fuentes, así como los mecanismos de búsqueda-retorno y los utilizados para caracterizar y representar la información o datos; también se incluye cierta información adicional comúnmente hallada.

Tabla 1. Algunas fuentes que proveen información acerca de los componentes OTS (en algunos casos también proveen el componente en sí)

Fuente de información	Caracterización	Esquema de Búsqueda-Retorno	Presentación de Información	Información adicional
eCots.org	Categorías de Software (CS)	Browsing (B), Búsqueda por clave (BC)	No Estructurada	*Newsletter*
SourceForge.net	CS	B, BC	No Estructurada	*Newsletter*
ComponentSource.com	CS; Plataformas	B, BC	No Estructurada	*Newsletter*, Forums, Demos
KnowledgeStorm.com	Soluciones IT (SIT)	B, BC	No Estructurada	Estudios de caso, *Análisis de Mercados, white papers*, Demos.
Forrester.com	SIT	B, BC	No Estructurada	*IT Support*, de pago
Gartner.com	SIT	B, BC	No Estructurada	*IT Support*, de pago
Incose.org	CS	B, BC	Semi-estructurada	*Tendencias*
Google.com/codesearch	-	BC	-	Búsqueda pública de código fuente por la web
Tucows.com	Plataformas	B, BC	No Estructurada	*Freeware/Shareware* Forums
Proveedores de Componentes OTS	Nombre	B, BC	No Estructurada	Documentación, forums, *blogs*
CMSmatrix.org	Nombre	Listado de nombres	Semi-estructurada	Tablas comparativas
Internet Mail Consortium	-	-	No Estructurada	Estándares y tendencias
CeBASE	-	B, BC	No Estructurada	*Experiencias* y enlaces

Para el análisis de los resultados de las entrevistas y de las fuentes de información, se llevaron a cabo varias rondas de discusión con miembros del equipo a fin de detectar y comprender la problemática existente. Los resultados fueron completados con experiencias en diversos casos de estudio en los que miembros del equipo habían participado anteriormente, así como los reportados en la literatura. Subsecuentemente, basándonos en el paradigma grounded-theory [24], articulamos una lista extensiva de los problemas detectados. La Tabla 2 muestra un resumen de los problemas más comunes agrupados por tipos de información y algunos ejemplos representativos.

Tabla 2. Tipos de información relacionada con componentes OTS y sus problemas más comunes

Tipos de Información	Descripción	Problemática	Ejemplos de recursos
Jerarquías o Taxonomías existentes	Diversas categorizaciones que con diferentes objetivos indexan y ofrecen descripciones de varios componentes	• Proliferación de iniciativas de catalogación de componentes que contienen descripciones usualmente breves y desestructuradas de los componentes. • Diferentes descripciones del mismo componente en diferentes catálogos. • La confiabilidad de la información es muchas veces cuestionable. • El uso de las categorizaciones propuestas es muchas veces confuso.	ComponentSource.com KnowledgeStorm.com SourceForge.net
Información del Proveedor	Información provista por los proveedores de los componentes, tales como: características, documentación, comparativas con versiones previas, etc.	• Información muchas veces incompleta y desestructurada. • En algunos casos se tiende a destacar aspectos favorables a los componentes y encubrir puntos débiles. • La veracidad de la información algunas veces es cuestionable o subjetiva.	Cualquier proveedor de componentes OTS: Comunidades open source o vendedores de componentes
Estándares relacionados	Algunas organizaciones pertenecientes a diferentes dominios industriales se han agrupado para pactar cierta información como estándar.	• No existe un estándar plenamente reconocido.	Internet Mail Consortium (IMC), Workflow Management Coalition (WfMC), Enterprise Content Management Association (AIIM)
Reportes Independientes	Organizaciones independientes ya sean del campo de la investigación o consultoría que proveen soporte en la selección.	• El análisis que proveen es generalmente a corto plazo, ad-hoc, y en algunos casos, muy caros.	Científicos: Journals Divulgación: Websites Técnicos: Gartner, Forrester,
Experiencias	Información o conocimiento práctico provisto por expertos y/o practicantes	• Las experiencias provistas son dependientes del contexto específico de quien las aporta.	Forums, Charlas, Reuniones, Recomendaciones, Experiencias, repositorio CeBASE,
Test de Pruebas de componentes o sistemas	Descripciones de Test de pruebas que informan del comportamiento del componente.	• Dependen del contexto y requerimientos específicos del proyecto reportado.	Resultados de pruebas: Tucows.com CMSmatrix.org
Otros	Algún otro recurso que provea funcionalidad específica para soportar la selección de OTS, ej., herramientas de búsqueda especializadas, de código, etc.	• Permiten encontrar grandes conjuntos de información. El problema principal es que comúnmente, la mayor parte de esta no es relevante y debe ser escrudiñada cuidadosamente.	Herramientas: Koders Tool Google.com/codesearch

3.2 Fase 2: Identificación de Dimensiones de CI de Componentes OTS y sus Indicadores en Procesos de Selección

Investigadores en el área de la CI concuerdan que su significado es subjetivo y debe ser determinado por el contexto específico del usuario de dicha información [21]. Para definir lo que la CI significaba en el contexto de selección de componentes OTS, nos basamos en el trabajo de [20]. Dicho trabajo propone un marco de 4 dimensiones de alto nivel (que a su vez son descritas por un conjunto de indicadores) que definen los diferentes "significados" de CI dependiendo de las diversas necesidades de los usuarios de la información. Cabe destacar que este marco es bastante reconocido y

ampliamente utilizado en el área de la CI, además existen varias evidencias de su uso satisfactorio en diversos contextos [21], [22], [26]. La Tabla 3 muestra las dimensiones e indicadores de calidad propuestos por dicho marco. Para una mejor compresión de dichas dimensiones, se sugiere también consultar [10] y [21]. Por otro lado, tal como sugiere [22], usamos la noción de información como producto.

Tabla 3. Dimensiones e Indicadores de calidad sugeridas en [20] para describir la CI en diversos contextos

Dimensión de CI	Definición	Indicadores de Calidad
Intrínseca	Denota la calidad interna de la información contenida.	Credibilidad, Precisión, Objetividad, Reputación.
Representacional	Incluye aspectos relacionados con el formato y el significado de los datos.	Representación Concisa, Consistencia en la Representación, Facilidad de Comprensión, Interpretación.
Accesibilidad	Enfatiza la importancia de los sistemas para acceder a la información de forma segura.	Accesibilidad, Seguridad en el Acceso
Contextual	Remarca los requerimientos de calidad necesarios para el contexto de uso de la información.	Relevancia, Fecha, Completitud, Cantidad Apropiada de Datos, Valor Añadido.

Dada la naturaleza de la información relativa a los componentes y del proceso de selección, decidimos replicar una parte del estudio reportado en [20] para verificar si las dimensiones e indicadores propuestos eran adecuados para este contexto. Esto se hizo a través de un cuestionario on-line en el que invitamos a participar a investigadores y practicantes del área, además de rondas de aplicación del cuestionario in-situ en el transcurso de varias conferencias internacionales relacionadas con DSBC, tales como OSS'07 y ICCBSS'08, en esta última además presentamos un póster para exponer los objetivos de nuestro estudio e invitar a la comunidad a participar [27]. El cuestionario se encuentra aún disponible con el fin de refinar los indicadores de CI iterativamente y obtener el mayor feedback posible.

El siguiente paso consistió en analizar las respuestas obtenidas en el cuestionario y las entrevistas previas, para su integración en las dimensiones propuestas por [20]. Como resultado de este proceso, algunos de los indicadores propuestos fueron redefinidos, abstraídos o eliminados para obtener un esquema que reflejara las respuestas obtenidas de manera óptima. La Tabla 4 muestra una síntesis de los resultados. Los cambios más significativos son:

- El indicador *Seguridad en el Acceso* perteneciente a la dimensión *Accesibilidad* (ver Tabla 3) fue eliminado dado que las respuestas indicaron que los seleccionadores de componentes OTS actúan como usuarios de la información por lo que los aspectos de seguridad en el acceso a la información no son relevantes desde su perspectiva.
- El nombre del indicador *Accesibilidad* fue cambiado por *Disponibilidad* para evitar confusiones con el nombre de la dimensión a la cual pertenece.
- Se añadió una sub-dimensión especial llamada *Necesidades de Información* como parte de la dimensión *Contextual* para representar las necesidades de calidad de la información específica de cada proyecto de selección. Los indicadores de esta sub-dimensión son *Requisitos de CI del Proyecto* y *Recursos Disponibles*. Esta sub-dimensión complementa a la dimensión *Contextual* propuesta en [20], y sus indicadores son la base para interpretar en qué medida las propiedades de la información de los componentes (representadas por los indicadores asociados a las dimensiones *Intrínseca*, *Representacional* y *Accesibilidad*) cubren las necesidades contextuales de un proyecto de selección específico. Cabe señalar que el concepto

detrás de esta sub-dimensión es similar al usado en el estándar ISO/IEC 25012 (Software Engineering- SQuaRE – Data Quality Model) recientemente publicado y que se refiere al aseguramiento de la calidad de los datos.

Tabla 4. Resumen de necesidades de la CI en procesos de selección de componentes OTS

Dimensión de la CI/ Indicadores	Algunas necesidades de la CI en la selección de componentes OTS	Algunas respuestas representativas
1.-Intrinseca		
Credibilidad	Información verosímil	Información sobre componentes comerciales tiende a ser parcial.
Precisión	Información precisa	La información de los componentes se vuelve obsoleta muy rápidamente dada la alta volatilidad del mercado tecnológico.
Objetividad	Punto de vista imparcial	Los recursos informacionales provenientes de fuentes comerciales o patrocinadas tienen a ofrecer información parcial.
Reputación	Proveniente de fuentes de reconocida reputación	El uso de fuentes de información provistas por autores reconocidos en el área u organizaciones de prestigio dan buen indicio de su probable alta calidad.
2.-Representacional		
Representación Concisa	Bien estructurada	La información acerca de los componentes OTS, en la mayoría de los casos se encuentra desestructurada y es difícil de ser localizada en un solo lugar, especialmente en lo referente a información no técnica y aspectos de su calidad.
Consistencia en la Representación	Homogeneidad	La ausencia de estándares para documentar los componentes OTS genera una gran heterogeneidad en la información que a su vez dificulta su comparación.
Facilidad de Comprensión	Fácil de entender	La información dirigida al público en general es en la mayoría de los casos fácil de comprender, sin embargo la información dirigida a expertos requiere un sólido conocimiento del área para ser interpretada.
Interpretación	Fácil de interpretar	Existen diferentes tipos de representación (ej. modelos Entidad-Relación, UML, lenguaje natural, etc.).
3.-Accesibilidad		
Disponibilidad	Preferencia por información gratuita	Algunos recursos de información sólo están disponibles a través de pago.
4.-Contextual		
Relevancia	Útil y apropiada a las necesidades	Comprender y procesar la información debe ser una tarea factible a realizar con los recursos disponibles al proyecto. Por ejemplo, para interpretar la información, los miembros del equipo deben estar familiarizados o capacitados para lidiar con la forma en que la información esta codificada (ej. modelos).
Fecha	Suficientemente actualizada	La fecha de creación o actualización de la fuente es el indicador más frecuentemente usado para reconocer si una fuente es actual.
Completitud	Información que cubra los aspectos requeridos por el proyecto	La información a usar debe preferiblemente cubrir las áreas involucradas en el proyecto con el fin de asegurar un entendimiento global.
Cantidad Apropiada de Datos	Volumen de información procesable	La forma de obtención de la información (ej., descarga de internet, previa subscripción, etc.) y los recursos requeridos para procesarla deben ser compatibles con los recursos del proyecto.
Valor Añadido	Provisión de un valor añadido	La información desestructurada y heterogénea es difícil de ser procesada. Su facilidad de procesamiento se considera un valor añadido.
Necesidades de Información		
Requisitos de CI del Proyecto	Describe la CI esperada en el proyecto	Los objetivos de alto nivel del proyecto determinan en gran medida sus requisitos de CI.
Recursos Disponibles	Aspectos relacionados a los recursos disponibles para el análisis de la información	Los recursos asociados al proyecto son básicos para determinar si un producto de información es adecuado o no. Es decir, para decidir si cierta información es apropiada se deben considerar, entre otras cosas, características de la fuente de información y las habilidades de los miembros del equipo y otros recursos disponibles para procesarla.

Es evidente que existe una alta sinergia entre las dimensiones e indicadores de calidad, lo cual es también soportado en [12]. Por ejemplo, el indicador *Valor Añadido* está íntimamente relacionado con *Representación Concisa* y *Consistencia en la Representación* para denotar en qué grado estos últimos ofrecen un valor añadido a las necesidades del proyecto denotadas por el indicador *Requisitos de CI del Proyecto* perteneciente a la sub-dimensión *Necesidades de Información*.

4 Esquema Mesurable para el Análisis de CI en DSBC

Nuestro siguiente objetivo consistió en el desarrollo de un modelo de referencia que integrara no solo los indicadores de CI de los componentes OTS, sino también las métricas que permiten reusar los indicadores de calidad y lidiar con las diferentes perspectivas de CI esperadas en distintos proyectos de selección. Consideramos que los modelos calidad eran un tipo de artefacto adecuado para obtener este modelo de referencia integral, pues nos permiten definir y consolidar las diferentes perspectivas de calidad y sus métricas asociadas. Específicamente, usamos como base la estructura del estándar de calidad ISO/IEC 9126-1 [28]. Su estructura se define como jerárquica multi-nivel dividida en características, refinadas en diferentes niveles de sub-características, las cuales son refinadas a su vez en atributos. El último nivel de la jerarquía la conforman los atributos mesurables cuyos valores pueden ser calculados usando métricas. Las razones primordiales que motivaron el uso de este estándar son: i) El estándar ISO/IEC 9126 es uno de los más reconocidos por la comunidad de IS, por lo tanto, la mayoría de los seleccionadores están familiarizados con él. ii) Es compatible con los modelos de evaluación de calidad de componentes propuestos en nuestros trabajos anteriores [5], [16], [17], para los cuales también hemos desarrollado herramientas (ver sistema integral DesCOTS [29]). De esta manera, procuramos dar soporte a varias fases del proceso de selección de componentes OTS, incluyendo el soporte a la evaluación de la CI sobre la que las decisiones son basadas y la elección de las fuentes de información que más se adecuen al proyecto específico, ofreciendo los mejores niveles de calidad posible. En general, el hecho de concebir nuestro modelo de referencia de CI de componentes OTS en el mismo formato que nuestros modelos anteriores nos permite considerar los aspectos de CI como requisitos, desde el inicio del proceso de selección, de la misma manera que consideramos y evaluamos los requisitos técnicos y no técnicos. Además, logramos un nivel de reusabilidad óptima del análisis de la CI, así como una mejora continua en la toma de decisiones dado que el modelo de referencia nos permite estimar el grado de "confiabilidad" o calidad de la información en cada proceso de selección.

4.2 Elaboración del modelo de la CI de Componentes OTS

Para elaborar el modelo de referencia a partir de las dimensiones e indicadores presentados en la Tabla 4, utilizamos la técnica Goal-Question-Metric (GQM) [30]. En GQM, primero se determinan los objetivos de cada elemento a ser elaborado y posteriormente se formulan algunas preguntas que ahondan en el estudio de dicho elemento; por último, se asocia un conjunto de métricas a cada pregunta con el fin de darles respuesta. El resultado final de la aplicación de GQM es una estructura jerárquica en forma de grafo en la que las métricas definidas pueden influenciar a más de una pregunta y las preguntas pueden estar relacionadas a más de un objetivo. Los

objetivos están compuestos por 4 elementos: propósito, referente, objeto y punto de vista. En nuestro esquema, los elementos fueron los siguientes:

- *Propósito*: Análisis y refinamiento de los indicadores que describen la CI de los componentes OTS en procesos de selección de componentes OTS.
- *Referente*: Cada uno de los indicadores de CI de componentes identificados en la Tabla 4 (por supuesto, tomando en cuenta la dimensión a la que pertenecen).
- *Objeto*: Información del componentes (ver definición en Sección 1).
- *Punto de Vista*: Necesidades de los seleccionadores de componentes OTS.

La Tabla 5 muestra un resumen del resultado de este proceso.

Tabla 5. Un extracto del resultado de aplicar GQM para refinar el indicador *Credibilidad*

Goal	Question	Metric	Valor	Tipo de métrica
Propósito: Análisis y *Refinamiento de indicadores de calidad* **Referente**: *Credibilidad* (perteneciente a la dimensión *Intrínseca*) **Objeto:** Información del componente OTS **Punto de Vista:** Seleccionador de componentes *OTS*	¿Quién es el autor(es) del producto?	Nombre del Autor(es)	ANom = Set (String)	Objetiva, parte de las características intrínsecas
	¿Cuál es la "credibilidad" del autor?	"Credibilidad" de cada autor	AutorCred = Function(String→TScore) TScore: {Muy Alto, Alto, Bajo, Muy Bajo}	Subjetiva, propiedad externa
	¿Cuál es la "credibilidad" global de todos los autores?	Promedio de la "credibilidad" de todos los autores	GlobAutorCred = Mean(AuthorsBel)	Subjetiva, propiedad externa
	¿Cómo se obtiene la "credibilidad" del Autor?	Promedio de todas las evaluaciones que ha recibido	IndAutorCred = Mean(AuthorBelMarks)	Subjetiva, propiedad externa
	¿Cuáles son las evaluaciones que el autor ha recibido?	Evaluaciones disponibles (Evaluador debe existir)	AutorCredEval= Function(String → TScore)	Subjetiva, propiedad externa
	¿Cuál es la "credibilidad" de estas evaluaciones al autor?	Para cada evaluador, calcular su "credibilidad"	CredEval = Function(String → TScore)	Subjetiva, propiedad externa
	…			
	¿Cuál es la organización proveedora del producto?	Nombre de la Organización	OrgNom= String	Objetiva, parte de las características intrínsecas
	¿Cuál es la "credibilidad" de la organización?	"Credibilidad" de la organización	OrgCred = Mean(ProvBelMarks)	Subjetiva, propiedad externa
	¿Cuáles son las evaluaciones que el proveedor ha recibido?	Evaluaciones disponibles (Evaluador debe existir)	ProvCredEval = Function(String → TScore)	Subjetiva, propiedad externa
	¿Cuál es la "credibilidad" de estas evaluaciones?	Para cada Evaluador, calcular su "credibilidad"	CredEval = Function(String → TScore)	Subjetiva, propiedad externa
	…			

Para definir las métricas asociadas al modelo, usamos [31] como base conceptual. Usamos varios tipos de métricas: Objetivas, Subjetivas, Internas y Externas. En general, éstas pueden ser tan simples como valores Enteros o Booleanos o más complejas como Conjuntos o Funciones. Un ejemplo de una Función predefinida es la función *Mean* usada en la Tabla 5; esta función retorna la media de todos los valores de las funciones que incluye, por ejemplo *AutorCred*, *AutorCredEval*, *ProvCredEval*. Para determinar algunos atributos de calidad que fomentasen la reusabilidad, consideramos ciertas evaluaciones externas, por ejemplo: determinar la credibilidad de un autor por medio de las evaluaciones que otras personas han hecho de él/ella; o determinar la credibilidad de las evaluaciones por medio de la credibilidad asociada a

las personas que hicieron las evaluaciones. Además, se asociaron condiciones de aplicabilidad a ciertas métricas; por ejemplo: para que existan evaluaciones para el autor de algún producto (fuente de información), debe existir al menos un evaluador.

Siguiendo los principios establecidos en [32] para construir modelos basados en el estándar ISO/IEC 9126 e integrando los resultados de la aplicación de GQM, obtuvimos un modelo de referencia de CI altamente reusable en diversos procesos de selección de componentes OTS. La Tabla 6 muestra un extracto del modelo (ver [5] para una versión más completa). A continuación se describe la estructura resultante:

Tabla 6. Ejemplo del refinamiento de la sub-característica *Credibilidad*

Sub-característica/Atributo	Descripción	Métrica	Comentarios
Credibilidad			
Credibilidad Basada en el Autor	Describe la "credibilidad" del producto en base a su(s) autor(es)		
1 Nombre del Autor	Describe el nombre del/los autor(es) del producto	ANom = Set (String)	ANom ≠ ∅ Los nombres son obtenidos directamente.
2 Credibilidad Global de los Autores	Describe la "credibilidad" global como promedio de la "credibilidad" de los autores	GlobAutorCred = Mean(AutorCredI)	Debería existir más de 1 autor Atributo derivado
1 Credibilidad Individual del Autor	Describe la "credibilidad" individual del/los autor(es)	AutorCred = Function(String→ TScore)	1) Dom(AutorCredI) = ANom 2) ∀ s ∈ ANom: AutorCred (s) = Mean(AutorCredEval)
1 Evaluación Externa de la Credibilidad del Autor	Describe las evaluaciones existentes acerca del autor	AutorCredEval = Function(String→ TScore)	Estas evaluaciones son obtenidas directamente.
…			
Credibilidad Basada en el Proveedor	Describe la "credibilidad" del producto en base a la organización que lo provee		
1 Nombre del Proveedor	Describe el nombre de la organización que provee.	OrgNom = String	OrgNom ≠ ∅
2 Credibilidad de la Organización	Describe la "credibilidad" de la organización que provee el producto	OrgCred = Function(String → TScore)	1) Dom(OrgCred) = OrgNom2) ∀ p ∈ OrgNom: OrgCredI(p) = Mean(ProvCredEval)
1 Evaluación Externa acerca de la Organización	Describe las evaluaciones existentes acerca de la organización proveedora	ProvCredEval = Function(String→ TScore)	Estas evaluaciones son obtenidas directamente.
…			
Credibilidad Basada en los Datos	Describe la "credibilidad" de los datos en base a sus propiedades inherentes		
…			

- **Características:** 4 características correspondientes a las 4 dimensiones descritas en la Tabla 4. De acuerdo a su naturaleza, las primeras tres características agrupan propiedades generales que describen la CI de los componentes y pueden ser reusadas en todos los proyectos de selección de componentes OTS. La característica denominada *Contextual* sirve para registrar el grado en el que las propiedades de CI de los componentes satisfacen los requisitos específicos de la CI para un proyecto determinado. Esta estructura permite una alta reusabilidad no sólo de la CI sino también del conocimiento adquirido en cada uso de la información en diversos proyectos de selección.

- **Primer nivel de Sub-características**: Las 4 características denotadas anteriormente, se refinaron en 15 sub-características correspondientes a los 14 indicadores descritos en la Tabla 4, más la sub-dimensión *Necesidades de Información* perteneciente a la dimensión *Contextual*.
- **Sub-características intermedias**: Se obtuvieron más de 28 sub-características intermedias, incluyendo los 2 indicadores pertenecientes a la sub-dimensión *Necesidades de Información*, es decir *Requisitos de CI del Proyecto* y *Recursos Disponibles*. Este refinamiento en sub-características se debió principalmente a que las primeras 15 sub-características de primer nivel descritas anteriormente eran aún muy abstractas para ser mesurables y además, varias de las necesidades de CI detectadas en nuestro estudio aún requerían refinamiento para ser lo suficientemente expresivas. Así pues, siempre que consideramos necesario, añadimos un nivel adicional de sub-características. Éstas sirvieron principalmente como medio de clasificación, por ejemplo: la sub-característica *Credibilidad* perteneciente a la característica *Intrínseca* fue descompuesta en otras sub-características: *Credibilidad Basada en el Autor*, *Credibilidad Basada en el Proveedor,* y *Credibilidad Basada en los Datos*.
- **Atributos**: Para medir los atributos del modelo de calidad resultante, añadimos las métricas identificadas a través de GQM. Se obtuvieron más de 75 atributos mesurables. En general, hay 2 tipos de atributos: atributos básicos que implican propiedades objetivamente evaluables (por ejemplo: el atributo *Nombre del Autor* perteneciente la sub-característica *Credibilidad*); y atributos derivados, que para poder ser evaluados requieren su descomposición en otros atributos (tal como: *Credibilidad Global de los Autores*).

Tal como se puntualizó anteriormente, algunos atributos influencian a otros atributos o sub-características en el modelo [12] (tal como es mostrado en la última columna de la Tabla 6), por este motivo, nuestra propuesta soporta el solapamiento jerárquico, considerando la jerarquía obtenida como un grafo. Además, permite integrar tipos elaborados de relaciones entre características, así como intensidades de las mismas, por medio de representaciones tabulares como las propuestas en [33].

5 Gestión y Reuso Sistemático de Información relacionada con Componentes OTS

El modelo de referencia de la CI obtenido en las secciones anteriores describe satisfactoriamente los requisitos de CI en diversos proyectos de selección. Sin embargo, dado el volumen creciente de la información relacionada con los componentes OTS, es crucial auxiliar a los selectores mediante una gestión sistematizada de las características de la CI que ayuden a reusarla y reconocer su probable utilidad para sus proyectos sin tener que inspeccionarla detalladamente cada vez. Con este objetivo, implementamos nuestros resultados en un modelo conceptual que sirvió de base para el desarrollo de una herramienta de software [34]. Esta herramienta da soporte a los selectores para recoger y almacenar información (meta-información) acerca de la información de los componentes, basándose en los atributos identificados en el modelo introducido en la sección anterior. Así, la herramienta facilita el análisis y uso óptimo de esta meta-información en diversos

proyectos de selección para asegurar niveles convenientes de calidad en la toma de decisiones, además de su gestión y reuso.

La Fig. 1 presenta un extracto del modelo conceptual obtenido (el modelo completo puede consultarse en [34]). La información acerca de la información de los componentes es descrita por medio de características, sub-características y atributos de calidad definidos en la sección anterior. De esta manera, los seleccionadores pueden consultar la calidad de la información y decidir si ésta es adecuada para sus objetivos. Por razones de espacio y para dar una idea general del modelo conceptual, sólo describiremos algunas particularidades de la elaboración de la sub-característica *Credibilidad*. Así pues, como parte central del modelo tenemos la clase *FuenteDeInformación*, descrita por un conjunto de atributos y relaciones basadas en el modelo de la Tabla 6. Existen diversos tipos de fuentes de información (clase *TipoDeFuenteDeInformación*), como también se denota en la Tabla 2. Heurísticas son provistas para auxiliar esta caracterización (ver [5]).

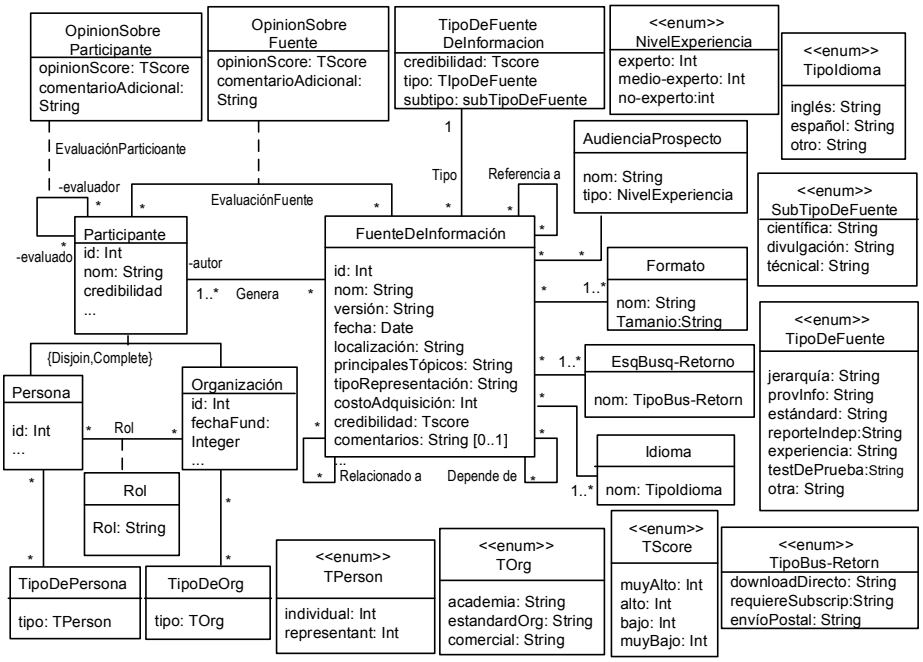

Fig. 1. Extracto del modelo conceptual para la gestión y reuso de la CI de componentes OTS

Para asegurar el reuso y la gestión, se definió la clase *Participante*, descrita por un conjunto de atributos objetivos (ej., *nom*) y uno subjetivo llamado *credibilidad*. Esta clase hace referencia a personas (*Persona*) u organizaciones (*Organización*). Una persona puede poseer diferentes roles en una organización. Los participantes juegan el rol de autores (*autor*) o evaluadores (*evaluador*). *Autor* se refiere al creador de la fuente de información (*FuenteDeInformación*). *Evaluador* se refiere a las personas que dan una opinión o valoración acerca de la fuente de información basada en su análisis o experiencia previa. Esta opinión es denotada por la clase asociativa *OpinionSobreFuente*. De la misma manera, los participantes pueden proveer

evaluaciones o comentarios acerca de otros participantes, denotado por *OpinionSobreParticipante*. Algunas relaciones que aparecen en el modelo conceptual de la Fig. 1 son fácilmente inferibles; por ejemplo, algunos indicadores de la dimensión *Accesibilidad* (ej. Costo de Adquisición se reflejan en el modelo conceptual como el atributo *costoAdquisición* de la clase *FuenteDeInformación*); o de la dimensión *Representacional* (ej. Tipo de Representación denotado también como el atributo *tipoRepresentación* de la clase *FuenteDeInformación*). Igualmente, algunas relaciones especiales entre las fuentes denotadas por: *Relacionado A*, *Depende De*, o *Referencia A*, que se refieren a atributos intrínsecos que denotan las fuentes de información que están relacionadas, que dependen de, o son referenciadas por una fuente de información, respectivamente. Por último, también se definieron ciertas restricciones de integridad en el modelo, ej.: los autores no pueden proveer evaluaciones sobre sí mismos ni sobre las fuentes de información creadas por ellos mismos. La herramienta de software resultante de este modelo se llama IQ-Tool [34].

6 Conclusiones y Trabajo Futuro

Este trabajo propone un esquema que representación y captura de los aspectos de CI más importantes para los seleccionadores de componentes OTS. Los aspectos más relevantes de esta investigación son que las dimensiones, indicadores y métricas que definen la calidad de la información de los componentes OTS en los procesos de selección fueron obtenidos empíricamente basados en la práctica industrial de DSBC. Todo esto se integró en un modelo de referencia de la calidad de la información usada en procesos de selección, basado en el estándar ISO/IEC 9126-1. Además, se elaboró un modelo conceptual como referencia para diseñar la colección, almacenamiento, gestión, y análisis de fuentes de información y su calidad en procesos de selección de componentes OTS. Este modelo conceptual se utilizó como base para diseñar una herramienta de soporte al trabajo teórico llamada IQ-Tool [34]. En general, este esquema asesora a los seleccionadores de componentes OTS en la toma de decisiones y evaluación de la información a ser usada en la toma de decisiones de acuerdo a los requisitos específicos de sus proyectos.

Nuestro trabajo actual se centra en: *i)* la integración de la herramienta IQ-Tool [34] con el sistema DesCOTS [29]; *ii)* integración de los resultados de esta investigación y la herramienta IQ Tool en [35]; *iii)* hacer más eficiente el esquema de gestión de las fuentes, poblando automáticamente la base de datos con ciertos atributos objetivos como nombre de autores, organizaciones y referencias, etc.

Referencias

1. Reifer, D., Basili, V. R., Boehm, B.W. and Clark, B.: Eight Lessons Learned during COTS-Based Systems Maintenance (2003), in: *IEEE Software*, 20:5(94-96)
2. Vitharana, P., et al. “Design, Retrieval, and Assembly in Component-Based Software Development”. *Communications of ACM*. November 2003/Vol.46, No. 11
3. Bhuta, J., Boehm, B.: “Attribute-Based COTS Product Interoperability Assessment” Int. Conference on COTS-Based Software Systems (ICCBSS 2007).
4. Cechich, A., Réquilé-Romanczuk, A., et al. “Trends on COTS Component Identification and Retrieval” Int. Conference on COTS-Based Software Systems (ICCBSS 2006).
5. Ayala, C. “Systematic Construction of Goal-Oriented COTS Taxonomies”. PhD Thesis. Technical University of Catalunya (UPC). GESSI Group. 2008.

6. Bertoa, M.F., Troya, J.M., Vallecillo, A. “A Survey on the Quality Information Provided by Software Component Vendors”. In *Proceedings of* QAOOSE, 2003.
7. Astudillo, H., Pereira, J., López, C.: “Evaluating Alternative COTS Assemblies from Imperfect Component Information”. QoSA 2006: 27-42
8. Ankolekar, A., Herbsleb, J., Sycara, K. Addressing Challenges to Open Source Collaboration with Semantic Web. 3rd Wshop on Open Source Software Engineering, (ICSE 2003).
9. Bertoa M., Troya, J.M, Vallecillo, A. ”Measuring the Usability of Software Components” The Journal of Systems and Software 79 (2006) 427–439.
10. Batini, C; Scannapieco, M.: *Data Quality: Concepts, Methodologies and Techniques. Data-Centric Systems and Applications*. Springer-Verlang. Berlin-Heidelberg, 2006.
11. Shankaranayanan, G.; Cai, Y.: Supporting Data Quality Management in Decision-Making. *Decision Support Systems*, 2006. 42(1): p. 302-317)
12. Gackowski, Z.: “Logical Interdependence of Data/Information Quality Dimensions—A Purpose Focused View on IQ”. Proceedings of the 9th International Conference on Information Quality (ICIQ-04), Cambridge, Massachusetts, 2004
13. Seacord, R., Hissam, A., Wallnau, K. “Agora: A Search Engine for Software Components”. Internet Computing, Vol. 2 No. 6, December 1998, pp. 62-70
14. Yanes, N., Sassi, S.B., Jilani, L. “ MoReCOTS: a Specialized Search Engine for COTS Components on the Web”. Int.Conference on COTS-Based Software Systems, IEEE, 2006.
15. Ayala, C.; Franch, X.: "Domain Analysis for Supporting Commercial Off-The-Shelf Components Selection". Intnal. Conf. on Conceptual Modelling (ER 2006). LNCS, Vol: 4215/2006, pp. 354-370.
16. Carvallo, J.P.: "Systematic Construction of Quality Models for COTS-Based Systems”. PhD Thesis. Technical University of Catalunya (UPC). GESSI Group. 2005.
17. Carvallo, J.P., Franch, X., Quer, C.: Determining Criteria for Selecting Software Components: Lessons Learned. *IEEE Software*. March/April 2007.
18. Caballero, I., et al.: “A Data Quality Measurement Information Model based on ISO/IEC 15939”. Proceedings of 12th International Conference on Information Quality. 2007. MIT, Cambridge.
19. Simmons, G.L., Dillon, T.S. “Towards an Ontology for Open Source Software Development”. In IFIP Intl. Federation for Information Processing, Springer, pp.65-75., 2006.
20. Wang, R.Y., Strong, D.M. “Beyond Accuracy: What Data Quality Means to Data Consumers” *Journal of Management of Information Systems*. Vol. 12, No.4, pp5-34. 1996.
21. Pipino, L., Y. Lee, and R. Wang, Data Quality Assessment. *Communications of the ACM*, 2002. 45(4): p. 211-218.
22. Wang, R., Allen, T., Harris, W., Madnick, S. “An Information Product Approach for Total Information Awareness” Proceedings of IEEE Aerospace Conference. Vol, 6 pp. 6_3005 - 6_3020, 2003.
23. Gerea, M.: "Selection of Open Source Components: A Qualitative Survey in Norwegian IT Industry". Master Thesis. NTNU. http://www.idi.ntnu.no/grupper/su/su-diploma-2007/dipl07-gerea.pdf
24. Charmaz, K. Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis. Thousand Oaks, CA: Sage Publications, 2006.
25. Torchiano, M., Morisio, M. Overlooked Aspects of COTS-Based Development. *IEEE Software* 21(2): 88-93 (2004).
26. Shankaranarayan, G., Ziad, M., Wang, R.Y. “Managing Data Quality in Dynamic Decision Environments” Journal of Database Management, 14(4), 14-32, Oct-Dec 2003.
27. Ayala, C., Franch, X.: “Assessing What Information Quality Means in OTS Selection Processes”. ICCBSS 2008. IEEE Society Press.
28. International Organization for Standardization. ISO/IEC Standard 9126: Software Engineering. Product Quality 2001.
29. DesCOTS web page: http://www.lsi.upc.edu/~gessi/QMTool/QMTool.html
30. Basili, V., Caldiera, G., Rombach, D. Goal/question/metric paradigm. In Encyclopedia of Software Engineering. Vol. 1, 1994. J. C. Marciniak, Ed. John Wiley and Sons, New York.
31. Fenton, N.E., Pfleeger, S.L. *Software Metrics: A Rigorous and Practical Approach*. 2nd Edition, 1997.
32. Carvallo, J.P., Franch, X. “Extending the ISO/IEC 9126-1 Quality Model with Non-Technical Factors for COTS Components Selection” Proceedings WQSQ 2006. ACM
33. Chung, L., Nison, B.A., Yu, E., Mylopoulos, J. *Non-Functional Requirements in Software Engineering*. Kuwler Academic Publishers, 2000.
34. Messeguer, F. “Eina de suport per a l’anàlisi de dominis”. UPC. 2007. Texto en catalán http://www.lsi.upc.edu/~cayala/Papers/IQToolDocumentation.pdf.
35. Ayala, C., Sørensen, C.F., Conradi, R., Franch, X., Li, J. "Open Source Collaboration for Fostering Off-The-Shelf Components Selection" 3rd Int. Conference on Open Source Systems (OSS 2007).