

Desarrollo de almacenes de datos dirigido por modelos

Juan Trujillo

Dpto. de Lenguajes
y Sistemas Informáticos
Universidad de Alicante
jtrujillo@dlsi.ua.es

Emilio Soler

Dpto. de Informática
Universidad de Matanzas
Matanzas, Cuba
emilio.soler@umcc.cu

Jose Zubcoff

Dpto. de Ciencias del Mar
y Biología Aplicada
Universidad de Alicante
Jose.Zubcoff@ua.es

Jose-Norberto Mazón

Dpto. de Lenguajes
y Sistemas Informáticos
Universidad de Alicante
jnmazon@dlsi.ua.es

Octavio Glorio

Dpto. de Lenguajes
y Sistemas Informáticos
Universidad de Alicante
oglorio@dlsi.ua.es

Jesús Pardillo

Dpto. de Lenguajes
y Sistemas Informáticos
Universidad de Alicante
jesuspv@dlsi.ua.es

Resumen

Esta contribución presenta una descripción de las líneas de investigación que el grupo IWAD (*Ingeniería Web y Almacenes de Datos*) de la Universidad de Alicante, está realizando en el campo de los almacenes de datos (ADs). La investigación se basa en el modelado conceptual como piedra angular para el desarrollo del AD. Actualmente, se estudian diversos estándares del OMG (*Object Management Group*) y su aplicabilidad al diseño del repositorio del AD, las aplicaciones para la explotación del AD como las herramientas OLAP (*On-Line Analytical Processing*) o de minería de datos, la incorporación de seguridad y métricas de calidad a los modelos conceptuales, así como la transformación e implementación de estos modelos en distintas plataformas.

1. Introducción

Un *almacén de datos* (AD) es una colección de datos históricos que sirven de apoyo a la toma de decisiones para mejorar un proceso de negocio. Su arquitectura se compone de varias capas: las fuentes de datos, los procesos de *extracción, transformación y carga* (ETL),

el repositorio, los cubos y las aplicaciones de explotación. Asimismo, el modelado *multidimensional* (MD) es considerado como el paradigma para estructurar el AD.

El grupo IWAD (*Ingeniería Web y Almacenes de Datos*) trabaja en el desarrollo de una metodología que cubra completamente el diseño del AD mediante una aproximación dirigida por modelos. Para abordar este objetivo se aboga por el uso del marco de trabajo MDA (*Model Driven Architecture*) [20] propuesto por el OMG (*Object Management Group*). MDA no sólo permite el desarrollo integral y formal de modelos usando una notación estándar, también permite especificar transformaciones para obtener un AD ajustado a los criterios de diseño, aportando así un entorno de desarrollo para ADs integrado, bien estructurado y completo, permitiendo decrementar el esfuerzo y tiempo de su desarrollo.

Esquema de la contribución. La siguiente sección introduce la investigación previa sobre ADs realizada por el grupo IWAD. Luego, se describen las líneas actuales: la aplicación de MDA para el diseño de ADs (sección 3), la transformación de reglas de seguridad con MDA (sección 4) y el modelado de técnicas de minería de datos sobre el AD (sección 5).

2. Investigación

Esta contribución presenta una descripción de las investigaciones que se están realizando actualmente en el grupo IWAD de la Universidad de Alicante. La investigación actual del grupo IWAD sobre ADs estudia su desarrollo dirigido por modelos apoyándose en estándares del OMG. Así, investigamos el diseño de los modelos de ADs del repositorio, aplicaciones para la explotación del AD como las herramientas OLAP (*On-Line Analytical Processing*) o de minería de datos, la incorporación de seguridad y métricas de calidad, así como el desarrollo de nuestros sistemas desde el modelado conceptual hasta su implementación.

Las investigaciones realizadas se reflejan en cerca de 50 artículos de revista y 50 ponencias en congresos nacionales e internacionales. A continuación, se resumen brevemente los trabajos previos a las líneas de investigación que se siguen en la actualidad.

2.1. Antecedentes

Las investigaciones anteriores en ADs se han centrado en la propuesta de una metodología de modelado basada en UML (*Unified Modeling Language*) [21] y UP (*Unified Process*). En esta propuesta se emplean los perfiles de UML como mecanismo para especializar este lenguaje al dominio de los ADs. Contemplando cada una de las fases de su desarrollo, se han diseñado perfiles para el modelado MD del repositorio del AD y los procesos ETL. El repositorio del AD se modela a nivel conceptual mediante el perfil propuesto en [4]. De esta forma se acomete el modelado MD del AD a nivel conceptual, siendo la base fundamental sobre la que se articula nuestra metodología de modelado actual. Complementando al perfil [4], se incorporaron también aspectos de seguridad sobre el repositorio del AD mediante un perfil para el modelado conceptual de reglas de seguridad siguiendo el modelo ACA (*Access Control and Audit*) [2, 3, 37]. Respecto a los procesos ETL, propusimos un perfil para su modelado a nivel conceptual y lógico [34]. Sin embargo, debido a las diferentes granularidades con las que pueden presentarse los datos a capturar, también

se propuso un perfil para habilitar el diseño de procesos ETL con datos atómicos [5]. Asimismo, dando soporte a las diferentes necesidades de desarrollo de ADs, se propuso un perfil para el modelado físico del AD y la integración con el resto de sistemas que lo componen [6], así como el diseño con UML de las aplicaciones OLAP para su explotación [35].

La propuesta de modelado a través de perfiles UML ha sido integrada en un proceso de desarrollo de ADs basado en UP [7]. Los beneficios del uso de UML para soportar este proceso de desarrollo han sido analizados en [8]. Conforme a esta propuesta, se ha estudiado también el uso de CWM (*Common Warehouse Metamodel*) [22] para la representación de los elementos MDs [18]. El estándar XML (*eXtensible Markup Language*) para la distribución de datos entre sistemas, se estudió como medio para la serialización de los modelos UML de ADs [19] e intercambio de datos con las aplicaciones OLAP [36].

Adicionalmente, las investigaciones en el modelado conceptual del AD se han complementado con el estudio de la calidad en los modelos propuestos. En [24, 1], se propusieron un conjunto de indicadores y métricas para la medición de la calidad del modelado del repositorio del AD conforme al perfil para su modelado MD [4]. Las métricas propuestas fueron además validadas empíricamente en [25].

3. MDA para el diseño de ADs

El diseño MD de un AD se asemeja al método tradicional de diseño de bases de datos, ya que se estructura en una serie de pasos que incluyen fases de modelado conceptual y lógico, y transformaciones entre ellas hasta la obtención de la implementación final. Las propuestas actuales tratan de representar las principales propiedades MD a nivel conceptual mediante la abstracción de detalles de la plataforma donde el AD será implementado. Sin embargo, no definen mecanismos formales para la obtención automática de representaciones lógicas del modelo conceptual como base de la implementación final. Estas propuestas sólo describen de manera informal algunas guías

de diseño para llevar a cabo esta tarea manualmente, por lo que el tiempo y coste de desarrollo del AD se incrementa considerablemente, llegando a ser una tarea tediosa para el diseñador y muy susceptible a fallos.

Por otro lado, a diferencia de las bases de datos tradicionales, debido a que el AD integra diversas fuentes de datos, el desarrollo de un modelo MD conceptual debe basarse, no sólo en un análisis de requisitos de información del usuario, sino también en un análisis de las fuentes operacionales que poblarán el AD, con el fin de implementar un AD que cumpla con las necesidades de los usuarios y que sea a la vez consistente con las fuentes de datos.

Teniendo en cuenta estas consideraciones, se propone el uso de MDA para el desarrollo MD del AD de una manera estructurada que facilite la tarea del diseñador mediante la automatización del proceso de obtención de un modelo lógico a partir de un modelo MD conceptual derivado tanto de los requisitos de información como de las fuentes de datos [9, 10]. En nuestra propuesta (ver figura 1) a partir de los requisitos de los usuarios del AD se construye un CIM (*Computation Independent Model*) [11, 12]. Después, se deriva del CIM un modelo conceptual MD o PIM (*Platform Independent Model*) inicial [13, 12] que contiene los elementos MDs necesarios para cumplir con los requisitos sin tener en cuenta ninguna tecnología de bases de datos. Se ha definido un conjunto de relaciones QVT (*Query/View/Transformation*) [23] para reconciliar este modelo con las fuentes operacionales [13, 14], obteniéndose un PIM enriquecido que será consistente con las fuentes de datos. El siguiente paso consiste en el desarrollo de diferentes PSMs (*Platform Specific Models*) a partir del PIM (mediante la definición de transformaciones QVT), dependiendo de la tecnología en la que se base la implementación del AD [9, 15, 10].

Por otro lado, también existen situaciones (p.e., prototipado) en las cuales es conveniente derivar directamente el modelo MD de las fuentes de datos. Esto también se puede hacer mediante nuestra aproximación, una vez marcado el modelo [16].

El uso de MDA permite un desarrollo del

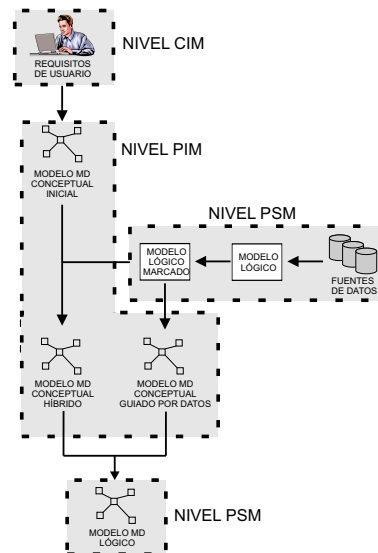


Figura 1: Desarrollo de ADs con MDA

AD sistemático y automatizado, además de aportar las siguientes ventajas [9]:

Productividad: la implementación del AD puede automatizarse una vez definido el PIM mediante las transformaciones de modelos. Por tanto, la productividad se mejora al decrementarse el tiempo y coste de desarrollo.

Perspectiva de negocio: al empezar el desarrollo mediante la especificación de un PIM, los desarrolladores del AD prestan más atención al desarrollo de modelos conceptuales que cumplan con los requisitos de usuario, en lugar de centrar los esfuerzos en detalles técnicos de una plataforma concreta. Este conocimiento técnico se embebe en las transformaciones. El AD implementado cumplirá con las expectativas del usuario.

Portabilidad: el mismo PIM puede transformarse a varios PSMs (dependiendo de la plataforma de implementación) según las necesidades de cada momento.

Integración e interoperabilidad: el desarrollo de ADs, debido a que se forman a

partir de sistemas heterogéneos, necesita de una gestión de los metadatos que asegure integridad e interoperabilidad entre las diferentes capas. MDA suministra un entorno para el desarrollo de manera integrada del AD.

Reusabilidad: las mejores prácticas del diseño de ADs pueden incluirse en las transformaciones, creando un repositorio de transformaciones que el diseñador puede aplicar en cada proyecto de AD según sus necesidades, asegurando la calidad final del AD.

Adaptabilidad: si aparecen nuevas tecnologías para el desarrollo del AD, no es necesario comenzar el proyecto desde cero, sino que el PIM, al ser independiente de la tecnología subyacente, no debe ser modificado y los diseñadores sólo deben preocuparse de las transformaciones del PIM al PSM relacionado con la nueva tecnología.

A continuación se detallan brevemente los principales puntos de nuestra propuesta desarrollados hasta la fecha.

3.1. Especificación de requisitos

La especificación de requisitos permite a los diseñadores desarrollar un AD que cumpla con las necesidades reales de los usuarios, incrementando de esta manera el éxito del proyecto de AD. Esta tarea no es trivial, ya que el análisis de requisitos en ADs está relacionado con la comprensión, por parte del diseñador, de los procesos de toma de decisiones y los objetivos que el usuario pretende alcanzar con la información suministrada por el AD. Así pues, obtener directamente del usuario requisitos de información es una tarea compleja, por lo que se requiere de técnicas especiales.

Teniendo en cuenta esta complejidad, se propone una aproximación basada en objetivos (mediante el uso de i^*) para definir requisitos de información [11, 12]. Para especificar los requisitos en un CIM, se ha definido un perfil de UML que permita obtener los diferentes tipos de objetivos que un usuario de

AD tiene. Primero se definen objetivos estratégicos como los principales objetivos del proceso de negocio. Estos objetivos corresponden con cambios desde una situación actual hacia otra mejor; por ejemplo, “*incrementar ventas*” o “*decrementar costes*”. Una vez que se definen los objetivos estratégicos, se deben especificar objetivos decisionales, que permiten definir las acciones que se deben tomar para cumplir con un objetivo estratégico. Por ejemplo, “*determinar una promoción*” o “*abrir nuevas tiendas*”. El cumplimiento de un objetivo decisional sólo es beneficioso para la organización si ayuda a alcanzar un objetivo estratégico. Finalmente, se deben determinar los objetivos informacionales que están relacionados con la información necesaria para alcanzar un objetivo decisional, como pudiera ser “*analizar compras de clientes*”. De esta manera, los requisitos de información se obtienen fácilmente de estos objetivos informacionales, y se evita la complejidad de intentar especificarlos de manera directa. Estos requisitos están relacionados con medidas de interés del proceso de negocio y con el contexto de análisis de esas medidas, por lo que los elementos MDs pueden descubrirse a partir de ellos. Así, el análisis de requisitos en ADs necesita añadir algunos conceptos MDs. En concreto, con el proceso de negocio con el que se relaciona el AD, con sus medidas y contexto de análisis. Así, los siguientes conceptos se deben tener en cuenta en el CIM: procesos de negocio relacionados con los objetivos de los usuarios del AD, medidas relevantes relacionadas con los requisitos de información y contexto de análisis de esas medidas.

3.2. Modelado MD conceptual

Un PIM describe un sistema abstrayendo los detalles relacionados con una plataforma o tecnología determinada. Este punto de vista corresponde con un nivel conceptual en el desarrollo del AD. El objetivo más importante a este nivel es la representación de las principales propiedades MDs sin tener en cuenta detalles específicos de tecnología alguna de bases de datos, posibilitando así la independencia del modelo MD del AD respecto de la plataforma en la cual sea implementado. Este PIM

para el modelado MD del AD se desarrollará siguiendo nuestro perfil MD para ADs [4].

Una vez que los requisitos de información se han especificado en un CIM, debe derivarse un modelo MD conceptual inicial como un PIM. Para ello se han definido varias reglas de transformación QVT [12], de tal manera que se obtiene un PIM inicial que contiene los elementos MD necesarios para suministrar la información requerida por los usuarios del AD.

3.3. Inclusión de las fuentes de datos

El AD integra diversas fuentes de datos operacionales, por lo que además de tener en cuenta los requisitos de información del usuario, se deben considerar estas fuentes de datos en etapas tempranas del desarrollo, con el fin de obtener un PIM consistente que cumpla con las siguientes propiedades [14]: completitud (el potencial de análisis que permiten las fuentes de datos se captura en el PIM) y fidelidad (el PIM representa un modelo MD que podrá ser poblado correctamente con los datos disponibles en las fuentes operacionales).

Para ello, se han desarrollado varias relaciones QVT basadas en formas normales MDs [14, 13] que fuerzan estas propiedades en el PIM mediante la eliminación, adición o modificación de elementos del esquema. Estas relaciones se aplican sobre un modelo en tercera forma normal de las fuentes de datos (realizado con el metamodelo relacional de CWM) y sobre el PIM inicial. Cabe destacar que el modelo de las fuentes de datos debe ser previamente marcado con conceptos MDs (p.e., hechos o dimensiones). El marcado de modelos es una técnica que permite extender elementos para capturar información adicional. Las marcas representan conceptos de un modelo que pueden aplicarse a elementos de otros modelos con el fin de guiar su emparejamiento.

Actualmente se está trabajando en el marcado automático de las fuentes de datos [16] y en la situación de este marcado en un contexto de modernización del *software* [17].

3.4. Modelado MD lógico

Un PSM representa el mismo sistema descrito por su correspondiente PIM, pero teniendo en cuenta cómo el sistema hace uso de una determinada plataforma o tecnología. En el modelado MD, *específico de plataforma* significa que el modelo se diseña a nivel lógico; es decir, está especialmente diseñado para un tipo concreto de tecnología de bases de datos: relacional (representación de estructuras de datos MD mediante el uso de bases de datos relacionales) o MD (representación directa de los datos en estructuras MD).

En nuestra propuesta cada PSM se modela mediante el uso de la capa *Resource* de CWM, ya que los metamodelos que contiene dicha capa permiten representar diversos tipos de estructuras de datos. En concreto, se usan los metamodelos relacional y MD.

Mediante el lenguaje de relaciones de QVT se han desarrollado transformaciones (i) entre nuestro PIM y un PSM para una plataforma relacional [9, 10] y (ii) entre nuestro PIM y un PSM para una plataforma MD [15].

4. Diseño de ADs seguros

Es común que en los proyectos de ADs, los aspectos de seguridad se implementan en las fases finales de su desarrollo. Sin embargo, la seguridad de los datos es un serio requisito que debe ser considerado cuidadosamente, no como un aspecto aislado, sino como un elemento que esté presente en todas las etapas del ciclo de vida de desarrollo: desde el análisis de requisitos, hasta la implementación y el mantenimiento. Lo anterior justifica que es crucial especificar medidas de confidencialidad en el diseño de ADs y hacerlas cumplir.

A partir del perfil UML [4] en [2, 3, 37], se propone una nueva extensión que permite representar requisitos de seguridad y auditoría en el modelado conceptual de ADs. Para ello, se propone el modelo ACA: un modelo de control de acceso y auditoría para el modelado MD seguro de los ADs. Para garantizar la integración de la seguridad en todas las etapas del ciclo de vida de desarrollo de un AD (aná-

lisis de requisitos, modelo conceptual, modelo lógico y modelo físico) utilizamos la propuesta metodológica basada en MDA.

La propuesta MDA permite transformar todos los requisitos de seguridad y auditoría capturados desde fases tempranas de diseño hasta el nivel físico; es decir, hasta su implementación en un sistema gestor de bases de datos. Debido a la frecuente complejidad y dimensiones de los modelos de ADs, ampliando la propuesta [4], hemos agrupado los elementos para el modelado de la seguridad en tres niveles de abstracción [38, 39]: definición del modelo, definición del esquema estrella y la definición de hechos y dimensiones.

Según la figura 1 (sección 3), para aplicar el marco de MDA, necesitamos construir un PSM que permita modelar todos los requisitos de seguridad y auditoría capturados durante la etapa del modelado conceptual. En [9], ha sido empleado MDA para el desarrollo de ADs, eligiendo el metamodelo relacional de CWM como PSM. El metamodelo relacional de CWM permite el intercambio entre bases de datos relacionales para la mayoría de los sistemas comerciales. En [26], hemos mostrado las posibilidades que ofrece CWM para representar reglas de seguridad y auditoría para ADs a nivel lógico. Esta propuesta ha sido completada y formalizada utilizando los mecanismos de extensión que ofrece CWM [27, 28].

Hemos presentado un marco general basado en MDA y QVT para el desarrollo de ADs seguros que abarca todo su ciclo de vida [29]. En nuestro caso (ver figura 1), el CIM se corresponde con el *Secure Multidimensional* (SMD) CIM que modela los requisitos de seguridad del negocio. Transformándolo, obtenemos el *Secure Multidimensional* (SMD) PIM que modela los requisitos a nivel conceptual (correspondiendo con el nivel PIM de la figura 1). A partir del SMD PIM, podemos obtener el SMD PSM (nivel PSM, figura 1). En [30, 31], hemos definido una transformación entre los niveles SMD PIM y SMD PSM. El SMD PSM ha sido definido mediante la extensión de CWM propuesta en [27, 28]. Ésta permite la representación lógica del repositorio del AD con toda la información relacionada con la seguridad y

la auditoría capturada a nivel conceptual; es decir, contenida en el SMD PIM. Esta transformación entre el SMD PIM y el SMD PSM ha sido definida utilizando el estándar QVT. Finalmente, con una transformación del SMD PSM a un sistema de gestión de bases de datos específico. Con el objetivo de validar los metamodelos SMD PIM, SMD PSM, así como la correspondiente transformación QVT, hemos desarrollado también un caso de estudio que ha sido presentado en [32, 33].

5. Modelos de minería de datos

También hemos hecho aportes en el campo del modelado conceptual de las técnicas de minería de datos. Dado que la minería de datos permite obtener conocimiento a partir de grandes volúmenes de datos, hemos propuesto modelos conceptuales [40, 41, 42, 43] y perfiles de UML [44, 45, 46, 47] para modelar distintas técnicas sobre el modelo MD de un AD. Por tanto, partimos en todos los casos de los datos estructurados de los modelos MDs para diseñar la minería de datos. Esta decisión está soportada por la ventaja de trabajar con datos orientados a tema que provienen de diversas fuentes (incluso cuando ya están integradas en el AD), evitando la duplicidad de tareas en el preprocesado de datos, dado que no hay que limpiar y unificar los datos de distintas fuentes si se quieren analizar con herramientas OLAP, de minería de datos, informes, etc. De este modo, los modelos de minería pueden aprovechar tanto el esfuerzo realizado en las etapas previas, como el conocimiento adquirido y reflejado en el diseño MD del AD, que representa de una manera clara los hechos, dimensiones, jerarquías de los datos. Esto contrasta con el método habitual en minería de datos de utilizar datos no estructurados, generalmente ficheros planos, con gran cantidad de columnas y por supuesto gran cantidad de casos (extensional e intensionalmente). Mayor dificultad presenta la integración de esos datos y su posterior limpieza (valores erróneos, validez de los datos nulos, etc.), todo ello sin una metodología que facilite el proceso. Por todo lo anterior, nuestras propuestas han integrado el uso de mo-

delos de minería en el desarrollo global de un AD desde las primeras etapas del diseño del mismo. En [40, 45, 44], propusimos extender UML para poder diseñar modelos de reglas de asociación. Esta técnica de minería de datos se utiliza ampliamente para obtener conocimiento en forma de reglas “*si A entonces B con una confianza de x% y un soporte de y%*” que expresan probabilidades de ocurrencia conjunta. El uso de nuestros modelos permite, por primera vez, diseñar modelos independientes de la plataforma de implementación.

En cuanto a técnicas de clasificación, presentamos el modelo conceptual en UML utilizando perfiles [41, 46]. Con ellas, podremos extraer conocimiento en forma de modelos (p.e., árboles de decisión) que describen las clases de datos. Los principales objetivos de estos trabajos son, en primer lugar proponer un modelo conceptual que describa completamente el proceso de minería de datos usando la técnica de clasificación, y en segundo lugar, la definición de las primitivas y semántica necesarias para diseñar modelos de minería de clasificación integrados en un proyecto de ADs en un lenguaje de modelado visual. Hemos usado UML dado que: es un lenguaje de modelado universal, ampliamente conocido y extensible en su metamodelo y semántica para adaptarse a un dominio específico.

Respecto a la minería de datos con *clustering* (agrupando poblaciones de datos en función de su semejanza y diferencias), presentamos modelos conceptuales basados en los datos MDs de un AD [47, 42, 48]. De esta forma, para obtener conocimiento usando técnicas de *clustering*, sólo es necesario identificar cuáles son los atributos del AD que participan en el proceso de minería y los ajustes del modelo de *clustering*. Nuevamente el modelado conceptual facilita la tarea, permitiendo un nivel de abstracción razonable para esta etapa del diseño de modelos de minería de datos.

También hemos propuesto modelos conceptuales para técnicas de análisis de series temporales [43], en el mismo sentido que en los trabajos anteriores, proponiendo una metodología para el diseño de modelos de análisis de series temporales integrada en ADs.

6. Conclusiones

Este trabajo resume las líneas de investigación respecto a ADs que actualmente se están llevando a cabo en el grupo IWAD de la Universidad de Alicante. Dichas líneas de investigación están relacionadas con el desarrollo de varias tesis, a saber: “*Metodología de desarrollo de almacenes de datos basada en MDA*” (Jose-Norberto Mazón), “*Desarrollo de almacenes de datos seguros con métricas de calidad*” (Emilio Soler) y “*Modelado conceptual en minería de datos*” (Jose Zubcoff).

Parte de la investigación se ha desarrollado en colaboración con otros grupos de investigación de importancia contrastada en el campo de las bases de datos, como son: *Database Group* de la Universidad de Drexel, grupo Alarcos de la Universidad de Castilla-La Mancha, *Bias Research Group* de la Universidad de Bolonia, *KDSL* de la Universidad de Atenas, *Database Research Group* de la Universidad de Münster, etc. Asimismo, estas investigaciones han posibilitado la organización de varios eventos como el *Workshop on Best Practices of UML (BP-UML)*, dentro de la conferencia internacional ER (*International Conference on the Entity Relationship Approach*), en los años 2005, 2006 y 2007; del *DOLAP (International Workshop on Data Warehousing and OLAP)* en 2005; y del congreso *DaWaK (Data Warehousing and Knowledge Discovery)* durante los años 2005 y 2006.

Finalmente, cabe destacar que los trabajos descritos se han realizado bajo el marco de varios proyectos de investigación, los cuales han estado vigentes durante el último año. Durante el período 2005–2007, se está trabajando en el proyecto *METASIGN (Metodologías y Herramientas Avanzadas para la Producción de Software Industrial de Nueva Generación)*, financiado por el Ministerio de Educación y Ciencia. En el ámbito regional, por un lado, se trabaja sobre el proyecto *DADASMECA (Desarrollo de Almacenes de DATos Seguros con MÉtricas de CALidad)* de la Conselleria de Empresa, Universidad y Ciencia de la Comunidad Valenciana; y por otro lado, en el proyecto *DADS (Diseño de Almacenes de*

Datos Seguros) de la Consejería de Educación y Ciencia de Castilla-La Mancha. Asimismo, el grupo IWAD cuenta con dos becarios FPU (Jose-Norberto Mazón –AP2005-1360– y Jesús Pardillo –AP2006-00332–) cuyos trabajos son financiados por el Ministerio de Educación y Ciencia.

Referencias

- [1] Berenguer, G., Romero, R., Trujillo, J., Serrano, M., Piattini, M.: A Set of Quality Indicators and Their Corresponding Metrics for Conceptual Models of Data Warehouses. In: DaWaK. (2005) 95–104
- [2] Fernández-Medina, E., Trujillo, J., Villarroel, R., Piattini, M.: Developing secure data warehouses with a UML extension. *Inf. Syst.* **32**(6) (2007) 826–856
- [3] Fernández-Medina, E., Trujillo, J., Villarroel, R., Piattini, M.: Access control and audit model for the multidimensional modeling of data warehouses. *Decision Support Systems* **42**(3) (2006) 1270–1289
- [4] Luján-Mora, S., Trujillo, J., Song, I.Y.: A UML profile for multidimensional modeling in data warehouses. *Data Knowl. Eng.* **59**(3) (2006) 725–769
- [5] Luján-Mora, S., Vassiliadis, P., Trujillo, J.: Data Mapping Diagrams for Data Warehouse Design with UML. In: ER. (2004) 191–204
- [6] Luján-Mora, S., Trujillo, J.: Modeling of Data Warehouse Using UML Component and Deployment Diagrams: Design and Implementation Issues. *Journal of Database Management (JDM)* **17**(2) (2006) 12–42
- [7] Luján-Mora, S., Trujillo, J.: Applying the UML and the Unified Process to the design of Data Warehouses. *Journal of Computer Information Systems (JCIS)* **46**(5) (2006) 30–58
- [8] Luján-Mora, S., Trujillo, J., Vassiliadis, P.: Advantages of UML for Multidimensional Modeling. In: ICEIS (1). (2004) 298–305
- [9] Mazón, J.N., Trujillo, J.: An MDA approach for the development of data warehouses. *Decis. Support Syst.* **In Press** (2007) doi:10.1016/j.dss.2006.12.003
- [10] Mazón, J.N., Trujillo, J., Serrano, M., Piattini, M.: Applying MDA to the development of data warehouses. In: DOLAP. (2005) 57–66
- [11] Mazón, J.N., Trujillo, J., Serrano, M., Piattini, M.: Designing data warehouses: from business requirement analysis to multidimensional modeling. In: RE Workshops (REBNITA). (2005) 44–53
- [12] Mazón, J.N., Pardillo, J., Trujillo, J.: A Model-Driven Goal-Oriented Requirement Engineering Approach for Data Warehouses. In: ER Workshops (RIGiM). (2007) In Press
- [13] Mazón, J.N., Trujillo, J., Lechtenböcker, J.: Reconciling requirement-driven data warehouses with data sources via multidimensional normal forms. *Data Knowl. Eng.* **In Press** (2007) doi:10.1016/j.datak.2007.04.004
- [14] Mazón, J.N., Trujillo, J., Lechtenböcker, J.: A Set of QVT Relations to Assure the Correctness of Data Warehouses by Using Multidimensional Normal Forms. In: ER. (2006) 385–398
- [15] Mazón, J.N., Pardillo, J., Trujillo, J.: Applying Transformations to Model Driven Data Warehouses. In: DaWaK. (2006) 13–22
- [16] Mazón, J.N., Ortega, E., Trujillo, J.: Ingeniería inversa dirigida por modelos para el diseño de almacenes de datos. In: JISBD. (2007) In Press
- [17] Mazón, J.N., Trujillo, J.: A Model Driven Modernization Approach for Automatically Deriving Multidimensional Models in Data Warehouses. In: ER. (2007) In Press

- [18] Medina, E., Trujillo, J.: A Standard for Representing Multidimensional Properties: The Common Warehouse Metamodel (CWM). In: ADBIS. (2002) 232–247
- [19] Medina, E., Luján-Mora, S., Trujillo, J.: Handling Conceptual Multidimensional Models Using XML through DTDs. In: BNCOD. (2002) 66–69
- [20] Object Management Group: Model-Driven Architecture (MDA), version 1.0.1. <http://www.omg.org/cgi-bin/doc?omg/03-06-01> (June 2003)
- [21] Object Management Group: Unified Modeling Language (UML), version 2.1.1. <http://www.omg.org/technology/documents/formal/uml.htm> (February 2007)
- [22] Object Management Group: Common Warehouse Metamodel (CWM), version 1.1. <http://www.omg.org/technology/documents/formal/cwm.htm> (March 2003)
- [23] Object Management Group: Query/View/Transformation Language (QVT), version 2.0. <http://www.omg.org/cgi-bin/doc?ptc/2005-11-01> (November 2005)
- [24] Serrano, M., Calero, C., Trujillo, J., Luján-Mora, S., Piattini, M.: Towards a Metrics Suite for Conceptual Models of Datawarehouses. In: Software Audit and Metrics. (2004) 105–117
- [25] Serrano, M., Calero, C., Trujillo, J., Luján-Mora, S., Piattini, M.: Empirical Validation of Metrics for Conceptual Models of Data Warehouses. In: CAiSE. (2004) 506–520
- [26] Soler, E., Villarroel, R., Trujillo, J., Fernández-Medina, E., Piattini, M.: Representing Security and Audit Rules for Data Warehouses at the Logical Level by Using the Common Warehouse Metamodel. In: ARES. (2006) 914–921
- [27] Soler, E., Trujillo, J., Fernández-Medina, E., Piattini, M.: SECRDW: an extension of the Relational Package from CWM for representing Secure Data Warehouses at the Logical Level. In: ICEIS Workshops (WOSIS). (2007) 245–256
- [28] Soler, E., Trujillo, J., Fernández-Medina, E., Piattini, M.: Una extensión del metamodelo relacional de CWM para representar Almacenes de Datos Seguros a nivel lógico. In: JISBD. (2007) In Press
- [29] Soler, E., Trujillo, J., Fernández-Medina, E., Piattini, M.: A Framework for the Development of Secure Data Warehouses based on MDA and QVT. In: ARES. (2007) 294–300
- [30] Soler, E., Trujillo, J., Fernández-Medina, E., Piattini, M.: Un Conjunto de Transformaciones QVT para el Modelado De Almacenes de Datos Seguros. In: JISBD Workshops (DSDM), <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-227/> (2007)
- [31] Soler, E., Trujillo, J., Fernández-Medina, E., Piattini, M.: A set of QVT relations to transform PIM to PSM in the Design of Secure Data. In: ARES. (2007) 644–654
- [32] Soler, E., Trujillo, J., Fernández-Medina, E., Piattini, M.: Application of QVT for the Development of Secure Data Warehouses: A case study. In: ARES. (2007) 829–836
- [33] Soler, E., Trujillo, J., Fernández-Medina, E., Piattini, M.: Aplicación de QVT al Desarrollo de Almacenes de Datos Seguros: Un caso de Estudio. In: IDEAS. (2007) 209–222
- [34] Trujillo, J., Luján-Mora, S.: A UML Based Approach for Modeling ETL Processes in Data Warehouses. In: ER. (2003) 307–320
- [35] Trujillo, J., Luján-Mora, S., Song, I.Y.: Applying UML For Designing Multidi-

- mensional Databases And OLAP Applications. In: *Advanced Topics in Database Research*, Vol. 2. (2003) 13–36
- [36] Trujillo, J., Luján-Mora, S., Song, I.Y.: Applying UML and XML for designing and interchanging information for data warehouses and OLAP applications. *J. Database Manag.* **15**(1) (2004) 41–72
- [37] Villarroel, R., Fernández-Medina, E., Piattini, M., Trujillo, J.: A UML 2.0/OCL Extension for Designing Secure Data Warehouses. *Journal of Research and Practice in Information Technology (JRPIT)* **38**(1) (2006)
- [38] Villarroel, R., Soler, E., Fernández-Medina, E., Trujillo, J., Piattini, M.: Using UML Packages for Designing Secure Data Warehouses. In: *ICCSA* (3). (2006) 1024–1034
- [39] Villarroel, R., Soler, E., Fernández-Medina, E., Piattini, M., Trujillo, J.: Representing levels of abstraction to facilitate the Secure Multidimensional Modeling. In: *ARES*. (2006) 678–684
- [40] Zubcoff, J.J., Trujillo, J.: Extending the UML for Designing Association Rule Mining Models for Data Warehouses. In: *DaWaK*. (2005) 11–21
- [41] Zubcoff, J.J., Trujillo, J.: Conceptual Modeling for Classification Mining in Data Warehouses. In: *DaWaK*. (2006) 566–575
- [42] Zubcoff, J., Trujillo, J.: An approach for the conceptual modeling of clustering mining in the KDD process. In: *ECDM*. (2007) In Press
- [43] Zubcoff, J., Trujillo, J., Cuzzocrea, A.: On the suitability of time series analysis on data warehouses. In: *ECDM*. (2007) In Press
- [44] Zubcoff, J., Trujillo, J.: A UML 2.0 profile to design Association Rule mining models in the multidimensional conceptual modeling of data warehouses. *Data Knowl. Eng.* **63**(1) (2006) 44–62
- [45] Zubcoff, J., Trujillo, J.: Extensión de UML para el Diseño de modelos de Reglas de Asociación en Almacenes de Datos. In: *JISBD Workshops (Data Gadgets)*. (2005) 17–26
- [46] Zubcoff, J., Trujillo, J.: Diseño de Modelos de Minería de Datos de Clasificación en Almacenes de Datos. In: *JISBD*. (2006) 465–470
- [47] Zubcoff, J., Pardillo, J., Trujillo, J.: Integrating Clustering Data Mining into the Multidimensional Modeling of Data Warehouses with UML Profiles. In: *DaWaK*. (2007) In Press
- [48] Zubcoff, J., Pardillo, J., Trujillo, J.: Minería de datos con clustering en espacios multidimensionales. In: *JISBD*. (2007) In Press