



Departamento de Lenguajes y  
Sistemas Informáticos



Universitat d'Alacant  
Universidad de Alicante

## Un método global basado en UML para el diseño de Almacenes de Datos

Sergio Luján-Mora  
Juan Trujillo

JISBD 2003

Un método global basado en UML para el diseño de Almacenes de Datos

### Contenido

- **Motivación**
- Mecanismos de extensión en UML
- Esquemas de modelado de AD
- Aplicación de los modelos
- Conclusiones
- Trabajos futuros

## Motivación

- Almacenes de Datos (AD) son sistemas de información complejos
- Proporcionan soporte para:
  - OLAP
  - Data mining
  - Decision Support Systems
  - ...
- Construcción AD:
  - Consume mucho tiempo
  - Costoso
  - Propenso a fallos

## Motivación

- Aproximaciones parciales:
  - Procesos ETL
  - Diseño lógico y conceptual de AD
  - Derivación del esquema del AD a partir de esquemas ER de las fuentes de datos
  - ...
- Diferentes métodos para AD, pero no un método (con su modelo) que abarque las distintas fases de diseño

## Motivación

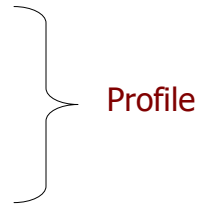
- Objetivo: **Un método completo de diseño de Almacenes de Datos**
- Principios de nuestra aproximación:
  - Notación estándar → UML
  - Completo → Incluye las principales fases de diseño
  - Potente pero fácil de entender
    - Diferentes niveles de detalle para diferentes usuarios (técnicos y usuarios finales) → Empleo de paquetes
  - Método flexible → Punto de inicio, pero no un esquema estricto
  - Aplicable → Soportado por herramientas CASE

## Contenido

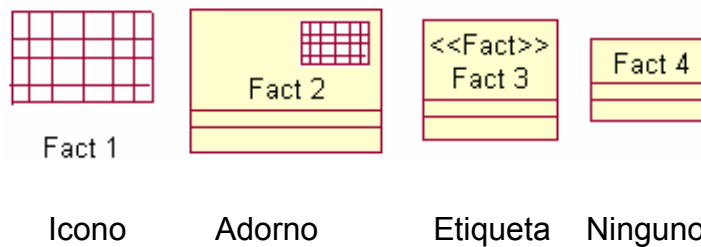
- Motivación
- **Mecanismos de extensión en UML**
- Esquemas de modelado de AD
- Aplicación de los modelos
- Conclusiones
- Trabajos futuros

## Mecanismos extensión UML

- UML es un lenguaje de modelado visual de **propósito general**
- Mecanismos de extensión permiten adaptarlo a dominios específicos
- Mecanismos:
  - Stereotypes → Nuevos elementos de construcción
  - Tagged values → Nuevas propiedades
  - Constraints → Nuevas semánticas

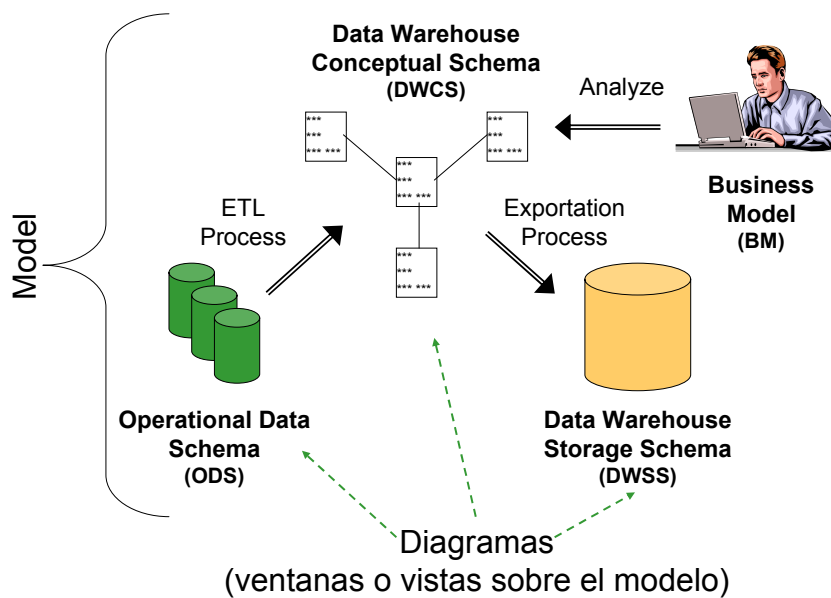


## Mecanismos extensión UML



## Contenido

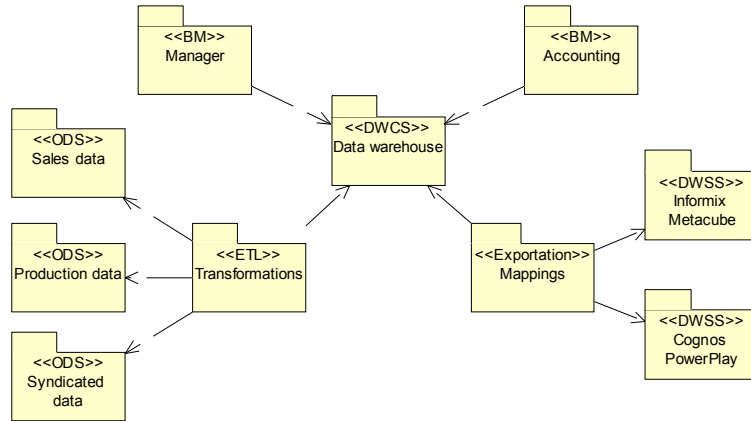
- Motivación
- Mecanismos de extensión en UML
- **Esquemas de modelado de AD**
- Aplicación de los modelos
- Conclusiones
- Trabajos futuros



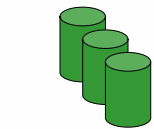
**Un método global basado en UML para el diseño de Almacenes de Datos**

**Diagrama general (nivel 0)**

<<ODS>>, <<DWCS>>, <<DWSS>>, <<BM>>, <<ETL>>, <<Exportation>>



**Un método global basado en UML para el diseño de Almacenes de Datos**



**Operational Data  
Schema  
(ODS)**

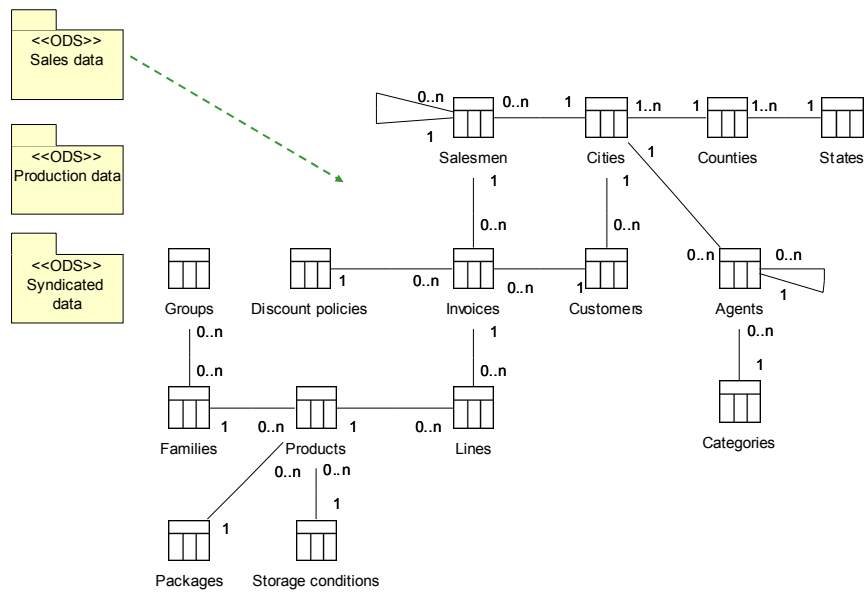
## ODS

- *Operational Data Schema*
- Representa:
  - Sistemas de procesamiento transaccional (OLTP)
  - Fuentes de datos externas (datos censales, datos económicos, datos sobre los competidores, etc.)
- No existe una extensión de UML para modelar distintos tipos de fuentes de datos

## ODS

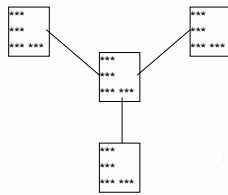
- RDBMS → *Rational's UML Profile for Database Design*: <<Database>>, <<Schema>>, <<Table>>, ...
- ORDBMS → *Marcos et al. UML Profile for Object-Relational Database Design*: <<array>>, <<row>>, <<ref>>, ...
- XML → *Rational's XML-DTD UML Profile*: <<DTDElement>>, <<DTDElementEmpty>>, <<DTDEntity>>,
- ...

### Un método global basado en UML para el diseño de Almacenes de Datos



### Un método global basado en UML para el diseño de Almacenes de Datos

#### Data Warehouse Conceptual Schema (DWCS)

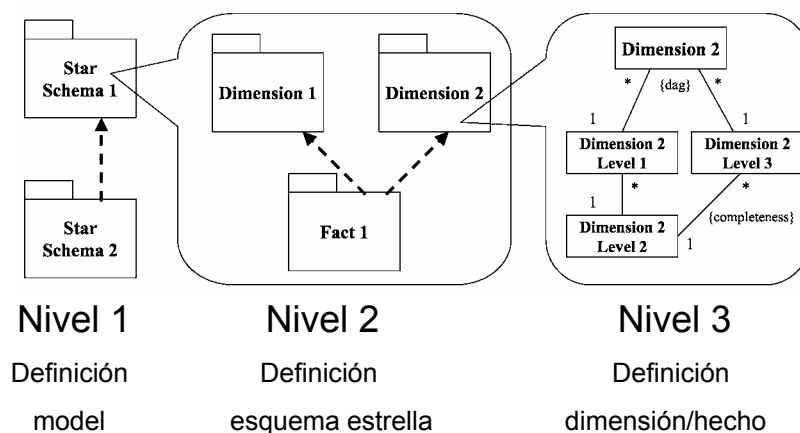




## DWCS

- *Data Warehouse Conceptual Schema*
- Propuesta: *UML Profile for Multidimensional Modeling*
- Componentes básicos:
  - Hechos: las transacciones o valores a analizar
  - Dimensiones: información descriptiva de los hechos
- Propiedades capturadas:
  - Dimensiones compartidas
  - Dimensiones heterogéneas
  - Hechos y dimensiones degerados
  - Jerarquías de clasificación múltiples y alternativas
  - ...

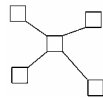
## DWCS



Un método global basado en UML para el diseño de Almacenes de Datos

## DWCS

### Estereotipos de paquete



StarPackage  
(Nivel 1)

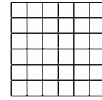


FactPackage  
(Nivel 2)



DimensionPackage  
(Nivel 2)

### Estereotipos de clase



Fact  
(Nivel 3)



Dimension  
(Nivel 3)

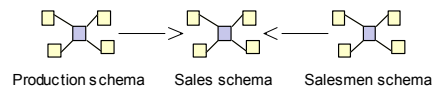


Base  
(Nivel 3)

Un método global basado en UML para el diseño de Almacenes de Datos

### Definición del modelo (nivel 1)

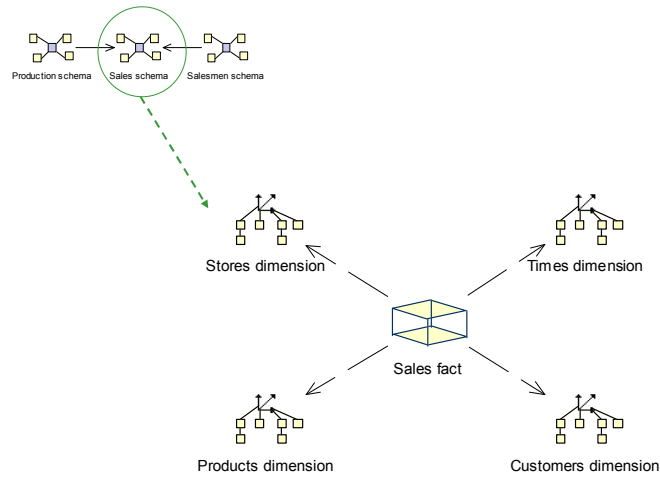
<<StarPackage>>



**Un método global basado en UML para el diseño de Almacenes de Datos**

**Definición del esquema estrella (nivel 2)**

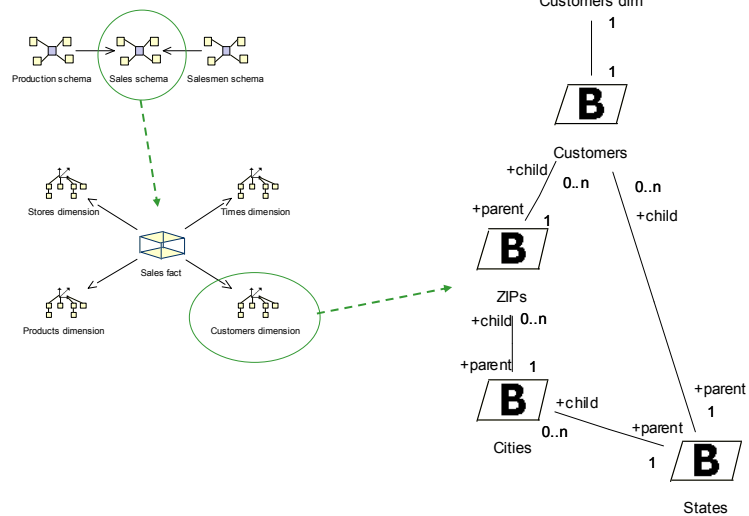
<<FactPackage>>, <<DimensionPackage>>



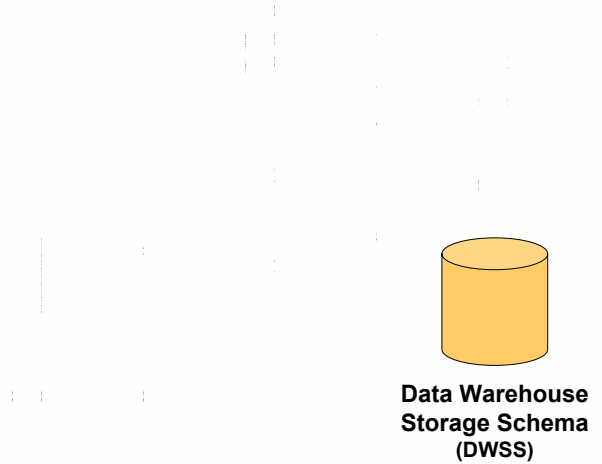
**Un método global basado en UML para el diseño de Almacenes de Datos**

**Definición dimensión/hecho (nivel 3)**

<<Fact>>, <<Dimension>>, <<Base>>



Un método global basado en UML para el diseño de Almacenes de Datos



Un método global basado en UML para el diseño de Almacenes de Datos

## DWSS

- *Data Warehouse Storage Schema*
- Representación en función de la implementación (RDMS, ORDBMS, MD, ...) → Similar a ODS
- Dos posibilidades: manual o automática

Un método global basado en UML para el diseño de Almacenes de Datos



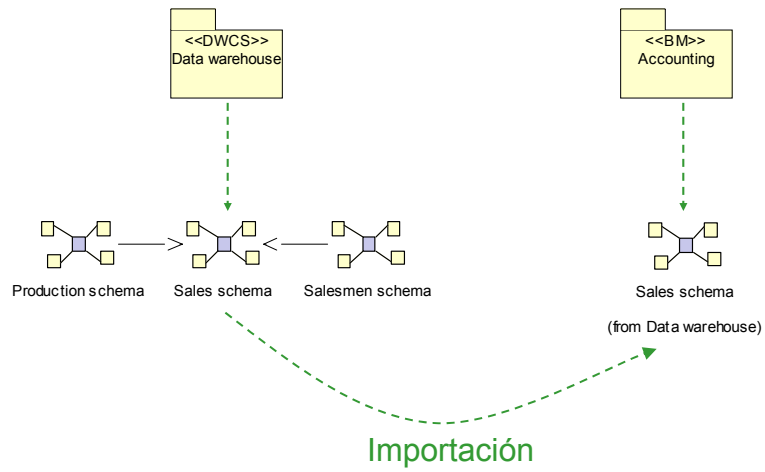
**Business  
Model  
(BM)**

Un método global basado en UML para el diseño de Almacenes de Datos

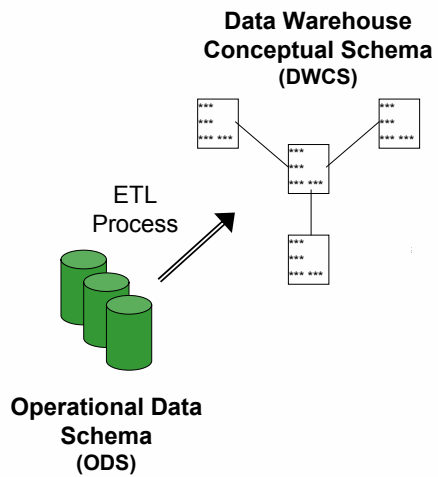
## BM

- *Business Model*
- Adaptar el AD a los diferentes usuarios finales:
  - Más fácil de entender
  - Incluye temas de seguridad
  - ...
- Mecanismo de importación de UML →  
Diferentes submodelos a partir del  
DWCS

**Un método global basado en UML para el diseño de Almacenes de Datos**



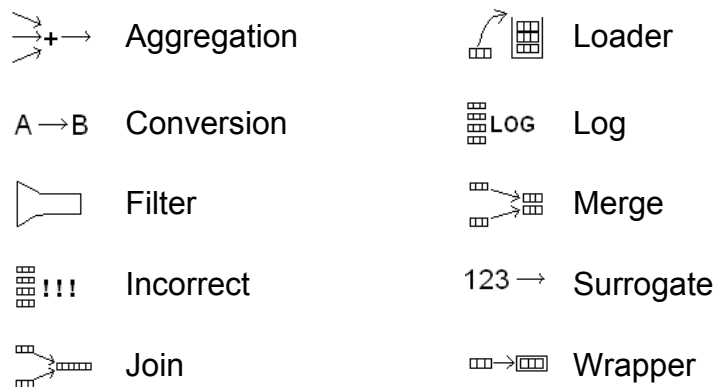
**Un método global basado en UML para el diseño de Almacenes de Datos**



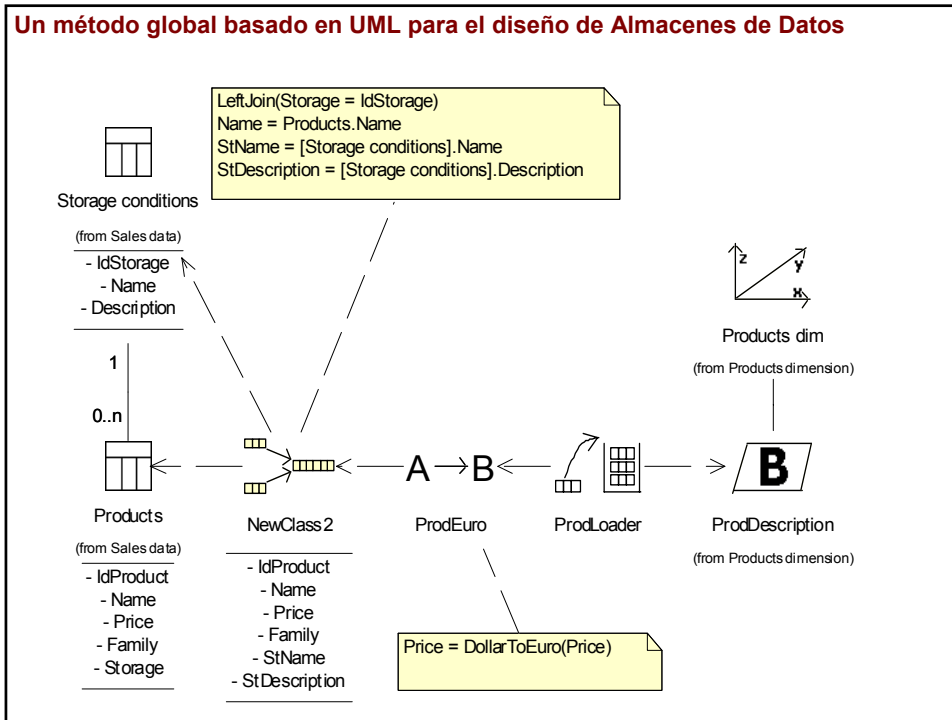
## ETL Process

- *Extraction-Transformation-Loading*
- Mapeo entre ODS y DWCS
- Propuesta: *UML Profile for Modeling ETL Processes*
- Paleta de mecanismos comunes:
  - Integración de diferentes fuentes de datos
  - Transformaciones
  - Generación de claves sustitutas (*surrogate keys*)
  - ...

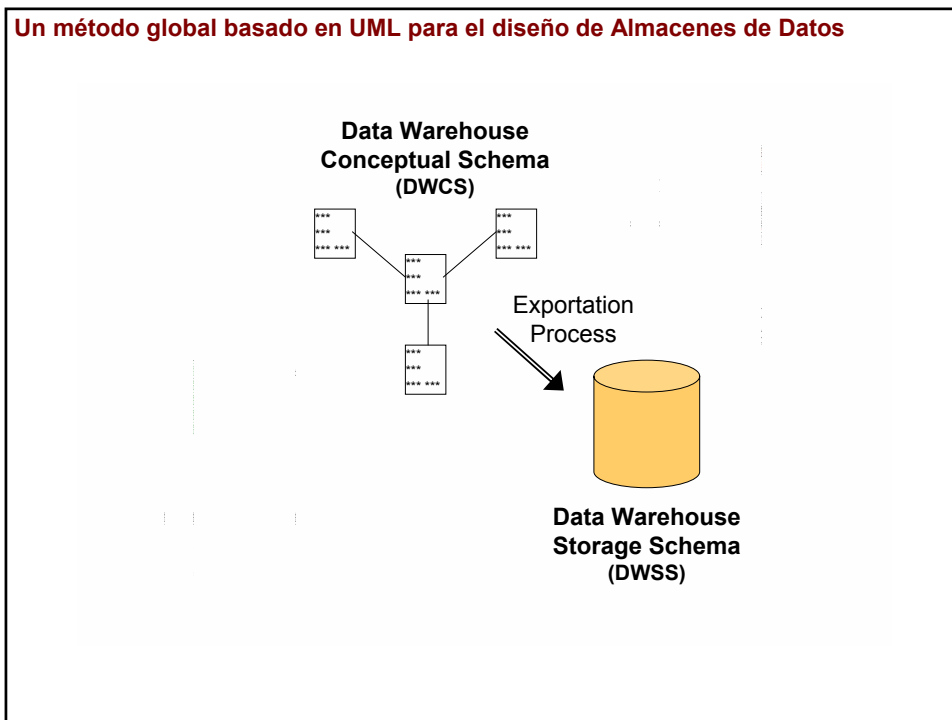
## ETL Process



**Un método global basado en UML para el diseño de Almacenes de Datos**



**Un método global basado en UML para el diseño de Almacenes de Datos**





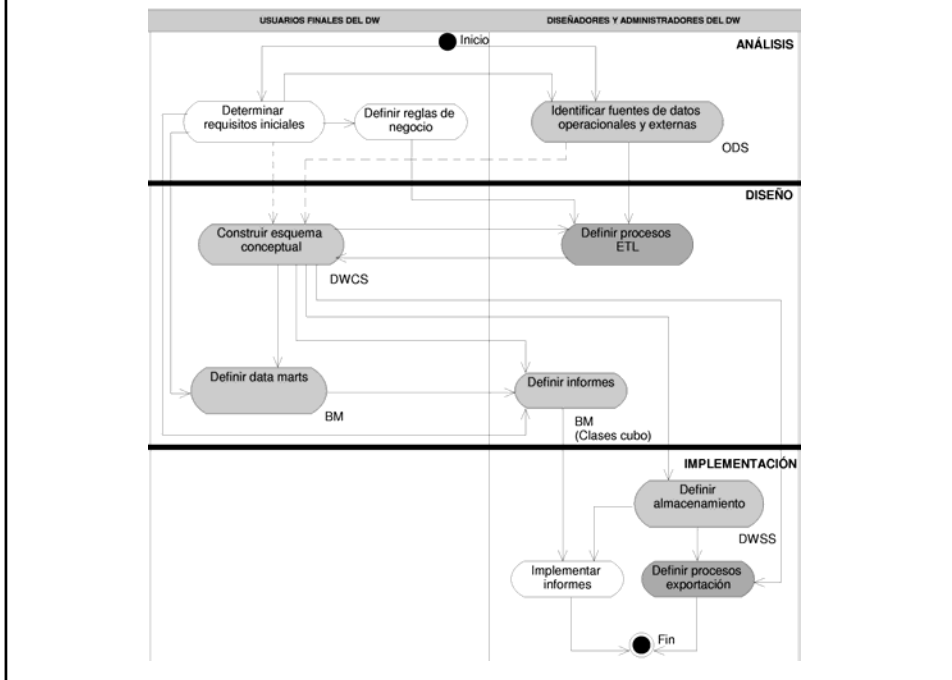
## Exportation Process

- Mapeo entre DWCS y DWSS
- Dos posibilidades: manual o automático (depende del modo de creación del DWSS)

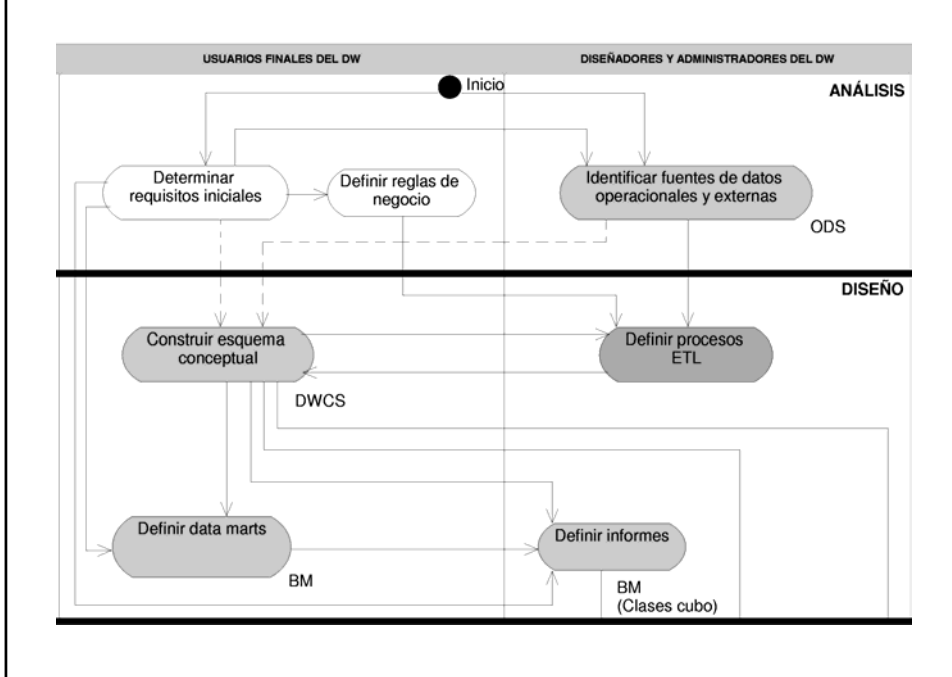
## Contenido

- Motivación
- Mecanismos de extensión en UML
- Esquemas de modelado de AD
- **Aplicación de los modelos**
- Conclusiones
- Trabajos futuros

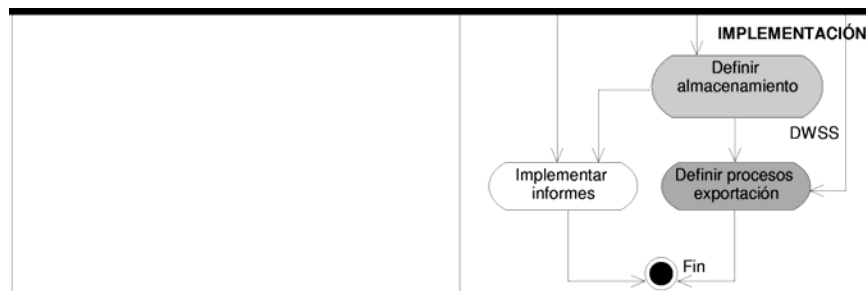
### Un método global basado en UML para el diseño de Almacenes de Datos



### Un método global basado en UML para el diseño de Almacenes de Datos



## Un método global basado en UML para el diseño de Almacenes de Datos



## Un método global basado en UML para el diseño de Almacenes de Datos

### Contenido

- Motivación
- Mecanismos de extensión en UML
- Esquemas de modelado de AD
- Aplicación de los modelos
- **Conclusiones**
- Trabajos futuros

## Conclusiones

- Método de diseño global
- Principales ventajas:
  - Misma notación estándar (UML) para todos los modelos
  - Perfil → Definición formal (OCL)
  - Integración de diferentes fases de diseño en un marco único
  - Se adapta a ADs grandes y complejos
- Soporte en la herramienta CASE Rational Rose → Add-in

## Contenido

- Motivación
- Mecanismos de extensión en UML
- Esquemas de modelado de AD
- Aplicación de los modelos
- Conclusiones
- **Trabajos futuros**

## Trabajos futuros

- Mapeo de atributos en el nivel de atributo
- Guías de estilo y de modelado para crear diagramas correctos
- Modelado de fases de diseño adicionales del ciclo de vida de un AD (implantación, procesos de refresco, etc.)

## Un método global basado en UML para el diseño de Almacenes de Datos

Sergio Luján-Mora  
Juan Trujillo

JISBD 2003