

Gestión de la Identidad en la UC3M

Enrique Bueno

P. David Escudero

Área de Desarrollo y Soporte - SDIC - UC3M

Objetivos del proyecto

- **Centralizar** la gestión de la información de personas relacionadas con la Universidad.
- Ser **independiente** de sistemas fuente y aplicaciones.
- **Punto único** de acceso: fácil y evitando copias de datos.
- La Universidad es imperfecta y los datos también, **aceptarlo**.

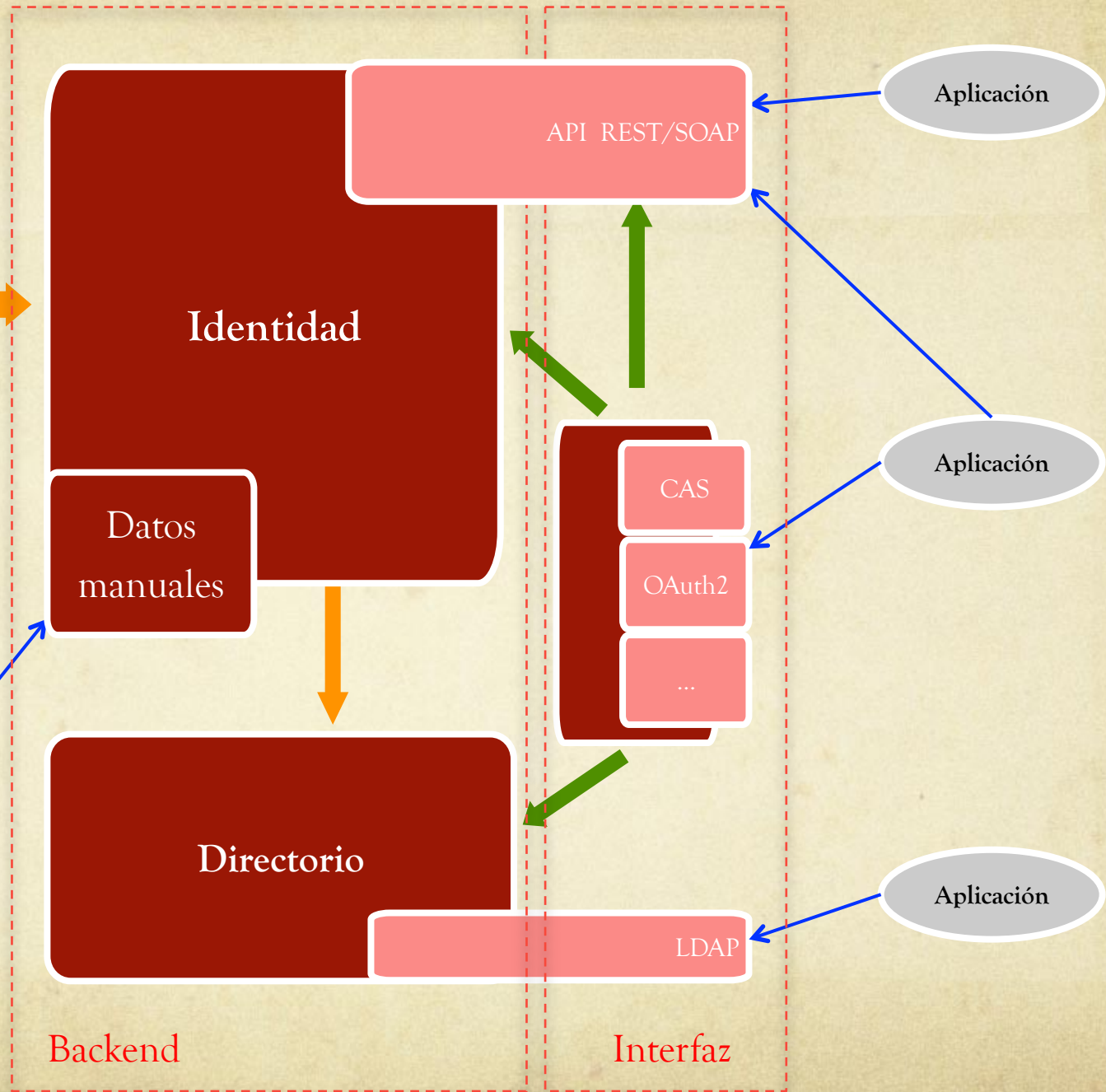
¿De dónde venimos?

- Evolución de un sistema previo: GENTE.
 - Información incompleta, sólo de personal.
- Datos de estudiantes gestionados por separado.
 - Cada aplicación gestionaba la mezcla.
- Consulta mediante acceso directo a la base de datos. Los sistemas accedían como podían.
- Dos identidades por usuario. Dos login.
- Intercambios de información complejos. Dos consultas a sistemas diferentes devolvían distinta información.

El Objetivo

Bases de datos corporativas

Cliente



Identidad

- Información manejada
 - Personas
 - Datos personales
 - Relaciones dentro la UC3M
 - Despacho y teléfono
 - Datos académicos
 - Tablas maestras
 - Grupos
 - Listas de distribución
 - Cargos
 - Permisos
- Arranque del proyecto
 - Reuniones con ‘productores’ y ‘consumidores’.
 - Puesta en común de la mayor cantidad de datos posible.
 - Resultado: modelo de datos desde el que empezar a trabajar.

La vida real. Los problemas

- Proceso de creación y asimilación muy lento. Poco vistoso.
- Afrontado por módulos. Algunos corrían prisa y “viven” por separado.
- Difícil encontrar soluciones que conserven el dato original pero con coherencia.
- Conceptos que cada área entiende de forma distinta.
- Datos históricos. Demasiados casos especiales.
- Limitar el ámbito: hay cosas que no tienen cabida.
- Al devolver información, hay mucha, ¿dónde cortas?

Puntos clave (1)

- El núcleo es una **base de datos relacional**.
- Mucha lógica del sistema en los **traspasos** de datos.
- Múltiples formas de acceso, pero siempre a la **misma información**.
- No todas las personas vienen de otros sistemas: Gestión “manual”.
- **Aplicaciones web** para consulta y gestión de datos.
- Traspaso de datos **sin parada de servicio**.
- Modelo de datos:
 - Información procedente de varios sistemas. **Consensuado** lo mejor posible
 - Tiene que reflejar **la realidad** (estructura de origen) ...
 - ... pero mantener la **coherencia**

Puntos clave (y 2)

- Acceso a la información:
 - Todo lo relativo a las **personas** de la Universidad debe salir de aquí.
 - Cada sistema que accede tiene unas características diferentes.
 - **Múltiples formas de acceso:** WS SOAP, REST, LDAP, SSO, ...
- Datos:
 - Sólo podemos recopilar la información, **no cambiarla**.
 - Buscar siempre la **coherencia**. No trayendo duplicidades ni problemas ajenos.
 - Crear abstracciones de datos que **simplifiquen**.
 - Parte de la información se introduce aquí ya que no existe en otros sistemas (despachos, teléfonos, etc.)
 - Aportar **estabilidad** a la información.

Modelo de datos

- Uso de capacidades del **modelo relacional**.
 - Integridad referencial, mantenimiento de claves únicas.
- Definición central de “tablas maestras”.
- El **IDU**:
 - Nuevo **identificador** (realmente) **único** para cada persona.
 - No servía ninguno de los existentes.
- Documentos: se guardan todos los válidos.
- Información de **auditoría**.
- Para datos de otros sistemas, guardamos la referencia.

Acceso a la información

- Cambio de mentalidad:
 - Lo ideal es **consultas puntuales** y guardar sólo el IDU sobre una persona, lo demás puede cambiar.
- **Cada sistema tiene sus propias limitaciones** para acceder a la información.
- Inicialmente la forma principal de acceso: WS SOAP vía HTTP: para consultas y traspasos. Pasando a REST.
 - En traspasos se consulta sólo los registros modificados.
- También hay sistemas que se llevan información con procesos ETL, aunque de forma coordinada.

Carga de datos

- El otro punto fuerte del sistema:
 - **Control de duplicados.**
 - Decisión de qué **información tiene prioridad.**
 - Otros: adivinar tipo documento, decidir si modificar.
- Resultado de la ejecución: correo-informe con
 - Errores: dato no 'digerido'.
 - Avisos: se ha tomado una decisión que conviene revisar.
 - Acciones: lo que se ha hecho sin problemas.
- Controla si hay que hacer cambios (no “sobrescribir” siempre).
- Se debe **respetar el dato de origen**, no se transforma.
- Posibilidad de bloquear los datos.

Carga de datos, control de duplicados

- Es más peligroso un “**solapado**”.
- Antes de dar un nuevo IDU nos aseguramos de que sea una **persona nueva**.
 - Se busca por Documento, Nombre y Apellidos.
 - Si hay varios resultados se trata como un error (para análisis humano).
 - Si se encuentra sólo una persona candidata, se intenta decidir por un sistema de puntuación que tiene en cuenta:
 - Coincidencia “nombre apellido1 apellido2”.
 - Evitando comparar apellidos más comunes.
 - Fecha de nacimiento.
 - ¿Qué viene cuando no se sabe un dato? ¿Y si no hay segundo apellido?
 - Formas diferentes de registrar documentos.
 - En **continua evolución**.

Detalles técnicos

- Base de datos Oracle
- Programación sobre la JVM: lenguajes Java y Groovy
 - Aplicaciones web en Java y Grails
 - Traspasos en Groovy
- Servicios web con JAX-WS en SOAP 1.1 vía HTTP diseñados “contract first”
- Definición común de interfaces de datos que reflejan modelo e implementan otros módulos
- Capa común de acceso a datos
 - Utilizada en traspasos, servicios web, aplicaciones gestión
 - Combina JPA (ORM) con JDBC Raw (eficiencia)

Consecuencias

- Somos el “canario de la mina”. **Detectamos los errores.**
- Te conviertes en **intermediario** entre áreas.
- Si los datos están mal, **coordinas los cambios.**
- Papel de consultor/clarificador/paño de lágrimas.
- ¿Quién gestiona ciertos datos ahora que el sistema es centralizado?
- Está funcionando y **está vivo**. Continua evolucionando.



Muchas gracias

Enrique Bueno <ebueno@di.uc3m.es>

P. David Escudero <barahona@di.uc3m.es>

Área de Desarrollo y Soporte - SDIC - UC3M