

## *Desarrollo de la plataforma apertura de datos universitarios*

Teniendo en cuenta los requisitos no funcionales marcados por el escenario tecnológico de la Universidad de Alicante, se determinó que la mejor opción era desarrollar una solución propia. Al estudiar las características del escenario tecnológico de la Universidad de Alicante y los requisitos no funcionales, se determinó la arquitectura de la plataforma. Estos requisitos no funcionales del escenario tecnológico de la Universidad de Alicante tenían dos puntos diferenciadores: (i) el front-end de la plataforma para la apertura de datos debía usar el sistema de gestión de contenidos o CMS (Content Management System) desarrollado en la Universidad de Alicante, llamado Vualà (Vualà), en aras de la homogeneidad, la facilidad de uso y la integración con el resto de contenidos web de la universidad; (ii) además, la plataforma debía utilizar el sistema gestor de bases de datos o DBMS (Database Management System) Oracle, establecido en la universidad y gestionado por el personal responsable y cualificado del Servicio de Informática de la Universidad de Alicante, con el fin de cumplir con todos los requisitos de seguridad.

Además se debe tener en cuenta dos funcionalidades como (i) el versionado de los datos a lo largo del tiempo, es decir, el ecosistema debería permitir la fácil actualización de los datos y (ii) la recolección de los datos de fuentes de datos heterogéneas, tanto para el portal de transparencia como para el portal de datos abiertos.

Por ello, se diseñó una arquitectura genérica aplicable a cualquier escenario, que cubriera todas las necesidades detectadas y facilitara la apertura de datos. Esta arquitectura es detallada a continuación.

## *Arquitectura tecnológica del ecosistema de datos abiertos*

En lugar de desplegar un DMS de código abierto ya desarrollado como puede ser CKAN, DKAN o Socrata, se decidió desarrollar una plataforma propia (Carcel, y otros, 2014) que requería nuestro sistema y no encajaba en el escenario tecnológico ya definido de la Universidad de Alicante. La arquitectura de esta plataforma de apertura de datos de la Universidad de Alicante se basa en cuatro módulos, como se muestra en la Figura 1.

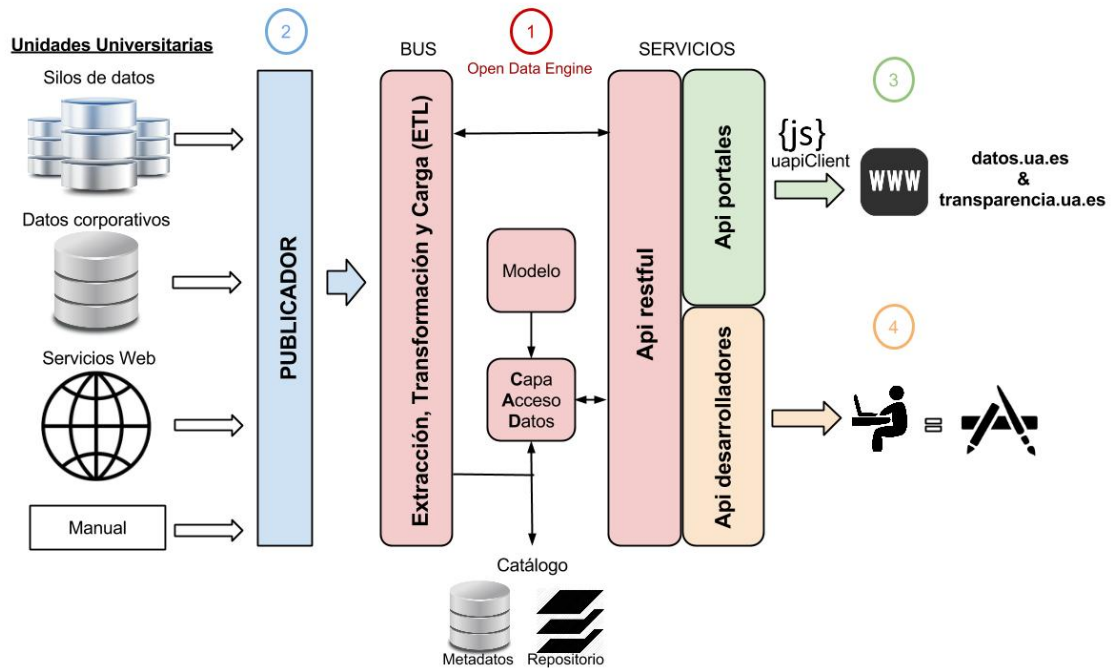


Figura 1. Plataforma tecnológica para el ecosistema de datos abiertos de la Universidad de Alicante

**Módulo 1: Open Data Engine.** Es el núcleo de la plataforma para la apertura de datos. Utiliza tecnologías open source, ha sido desarrollado en lenguaje Java y servido por Tomcat. Los componentes de este núcleo son las siguientes:

1. Catálogo, formado por el repositorio donde son almacenados los ficheros de datos y la base de datos donde se almacenan los metadatos, incluidos los enlaces a los recursos del repositorio.
2. Modelo de datos, encargado de instanciar las tablas de la base de datos. El modelo de datos está mapeado con las tablas del catálogo, facilitando así las operaciones.
3. Capa de acceso a datos (CAD), utiliza el modelo para el acceso a los datos persistentes en el catálogo y permite realizar operaciones CRUD (create, read, update, delete) sobre ellos.
4. API RESTful Web: interfaz que permite la comunicación del resto de módulos con la capa de acceso a datos (CAD) para realizar operaciones CRUD sobre el catálogo.
5. BUS de procesos ETL, es el encargado de almacenar los procesos de extracción de información de las diferentes fuentes de la universidad para extraer, transformar y cargar los datos en el catálogo. Como se puede observar en la Figura 2, los procesos definidos mediante la herramienta Spoon de Pentaho Data Integration están programados para ejecutarse según los metadatos de cada recurso, de esta manera estos procesos consumen los datos con la temporalidad definida en los metadatos del recurso y actualiza los datos del catálogo. También permitirá que las personas que posean permisos publiquen los conjuntos de datos en el portal (estableciendo un filtro de control

de calidad de los datos que realizará el grupo de trabajo del proyecto). Este bus de ETL automatizará el sistema para que este pueda ofrecer toda la información de la universidad y será encargado de mantener todos los datos de los portales actualizados.

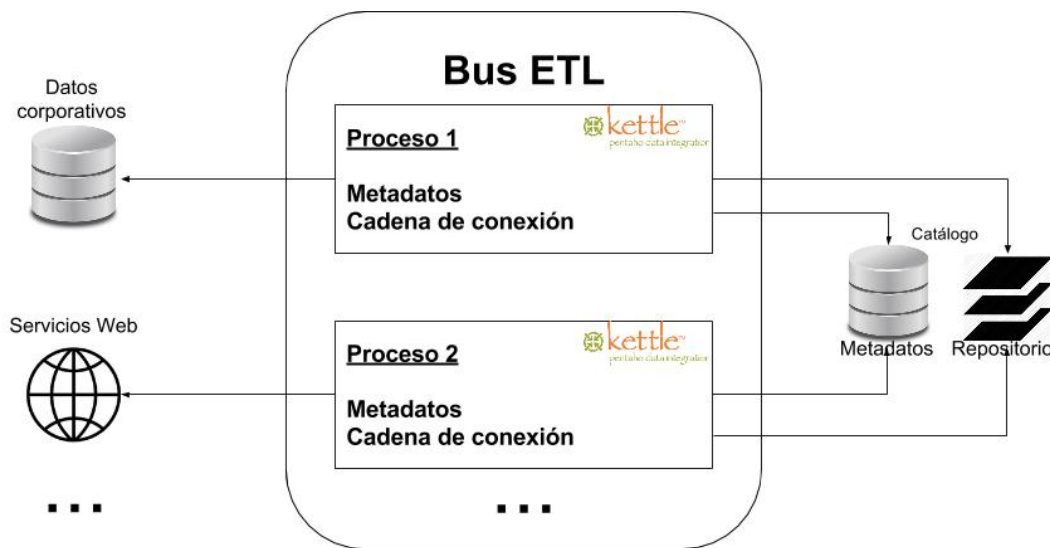


Figura 2. Detalle del bus ETL del módulo 1 'Open Data Engine'

**Módulo 2: El publicador** será encargado de introducir los datos en el catálogo para que puedan ser consumidos por ambos portales y por los desarrolladores. Además, el publicador permitirá modificar los datos y metadatos del catálogo, siendo la herramienta que se encargará de la gestión de los conjuntos de datos. Este módulo está diseñado para permitir la inserción de datos por parte de aquellas personas que tengan permisos, facilitando la participación de los responsables de los datos en el proceso de apertura. La inserción y actualización de datos por parte de los actores externos al desarrollo del proyecto contará con unos controles de calidad que realizarán los diferentes expertos (mencionados en el capítulo anterior) antes de que los datos sean publicados.

**Módulo 3:** incluye los **portales** de transparencia y de datos abiertos que sirven la información proporcionada por el Open Data Engine (módulo 1). La comunicación entre ambos portales y el API Restful se establece mediante llamadas HTTP a servicios web que devuelven los datos del catálogo. Para facilitar estas llamadas se usa una pequeña API Javascript (ver Figura 3) que tiene definidas las funciones parametrizadas que se encargan de hacer las llamadas al API Restful del módulo 1. En la Universidad de Alicante, ambos portales (UA - datos abiertos, 2014) (UA - Transparencia, 2015) están alojados en el CMS de la universidad (Vualà), pero podrían implantarse en cualquier otro CMS o contenedor.

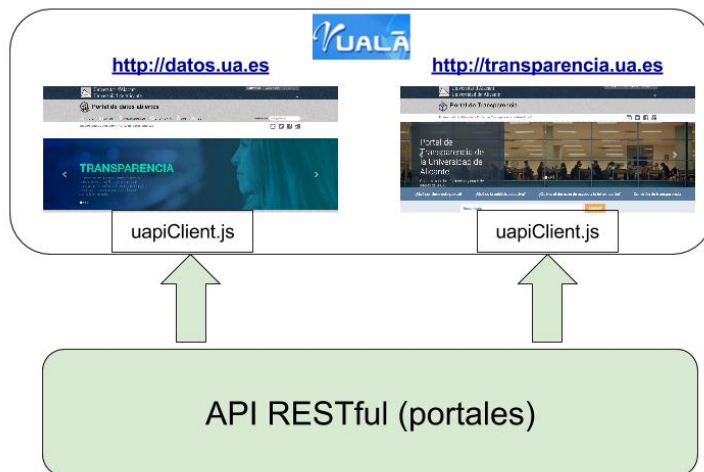


Figura 3. Arquitectura del módulo 3. Portales

**Módulo 4: El API de desarrolladores** (UA - DEV, 2014) está diseñada para proporcionar los datos abiertos en formato reutilizable, transformando aquellos formatos estructurados en JSON, facilitando así su reutilización de manera automática (ver Figura 4). Los desarrolladores pueden hacer uso de este servicio solicitando un clave de acceso que le proporcionará el sistema. Esta restricción servirá para mantener un control de acceso a los datos y poder restringir el mismo en el caso que sea necesario.

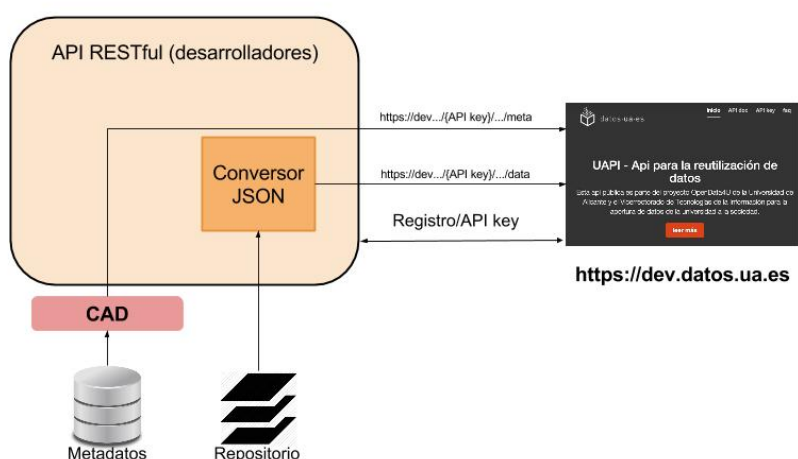


Figura 4. Arquitectura del módulo 4. API desarrolladores

Todos los módulos están desarrollados en lenguaje Java. El módulo 1 (Open Data Engine) es el núcleo del proyecto y utiliza tecnología Hibernate ORM (mapeo objeto relacional) para el mapeado del catálogo. Los servicios web están desarrollados sobre el framework de Jersey usando el API JAX-RS. El resto de módulos, utilizan las funcionalidades del núcleo para la gestión del catálogo, estos también están implementados en Java. El conjunto de todos estos módulos es un DMS que encaja perfectamente en cualquier escenario universitario, gracias a la independencia y desacoplamiento que ofrecen las capas que lo forman. El DMS ofrece facilidades de instalación de su parte open data engine a cualquier escenario tecnológico, teniendo en cuenta la clasificación de los datos universitarios.

Es interesante señalar que nuestra solución dispone de un solo catálogo que almacena tanto los recursos del portal de transparencia como los del portal de datos abiertos y que establece una relación (no obligatoria) entre ambos. Esta peculiaridad está diseñada para alentar la apertura en formato reutilizable de aquellos datos que se encuentren en el portal de transparencia (debido a las leyes vigentes y a la normativa establecida en la propia universidad) pudiendo ser ofrecidos también en formatos estructurados con el fin de que en un futuro todos los datos expuestos en el portal de transparencia tengan una relación y una versión estructurada en el portal de datos abiertos. De esta manera no solo se mejora la transparencia y la rendición de cuentas sino que también crecen las oportunidades de negocio basadas en la reutilización de la información del sector público.

### *Repositorio de código abierto OpenData4U*

Desde la Universidad de Alicante con el proyecto OpenData4U se defiende la apertura de los recursos que generan las administraciones e instituciones públicas. No se trata sólo de apertura, se trata de compartir los conocimientos en beneficio de la sociedad, usando como base los datos y como herramienta las tecnologías de la información y la comunicación.

Por este motivo, la Universidad de Alicante pone a disposición de toda la comunidad el código de las herramientas utilizadas para la apertura de datos abiertos y transparencia. El código está disponible en el repositorio de la Universidad de Alicante RUA y en la plataforma de desarrollo colaborativo GitHub (<https://github.com/UAdatos>) en el repositorio. La apertura del código desarrollado bajo el paraguas de la Universidad de Alicante tiene como objetivo alentar a poder enriquecer la solución actual por medio de contribuciones de la comunidad, además de que otras universidades e instituciones públicas puedan utilizar la plataforma para abrir sus datos. De esta manera pretendemos impulsar una comunidad para el desarrollo de herramientas para la apertura de datos universitarios que de manera colaborativa crezca a la vez que lo hace la transparencia, mejorando la formación en datos abiertos y el emprendimiento basado en reutilización de la información del sector público dentro de las universidades.