

Título: "La gestión de objetos digitales en FEDORA ."

Autor:

**Luis Zorita Vicente**

## **Resumen.**

La e-ciencia se percibe , cada vez más, dentro de un entorno con dos elementos clave : la **red** y las posibilidades de **colaboración** que ella ofrece.

En ese contexto se está generando un nuevo tipo de unidad de información de naturaleza compleja cuya gestión, acceso, almacenamiento y reutilización requieren un sistema capaz de tratar esa complejidad.

En esta presentación se hacen 3 propuestas:

- 1- Un nuevo modelo de Objeto Digital que estructura la naturaleza compleja de la unidad de información en red y
- 2.-una arquitectura, FEDORA, capaz de integrar ese modelo de Objeto Digital.
- 3.-Utilizar Fez como cliente web para acceso y gestión de los DO de Fedora

Explicaré estas propuestas y presentaré algunas aplicaciones que utilizan Fedora y/o Fez como contenedor y gestor de la información digital:

Partiendo de estos conceptos de arquitectura-aplicación explicaré como se pueden asociar diferentes esquemas de metadatos a un mismo objeto digital utilizando Fez y someramente algunas aplicaciones desarrolladas utilizando Fedora: creación de un repositorio de objetos digitales, un sistema para la creación de revistas digitales mediante query RDF y transformaciones xslt, un servicio web de búsqueda a texto completo (datos y assets textuales) mediante indexación con estándar Lucene, un servicio de generación de RSS en la búsqueda.

## **Software libre utilizado**

Practicamente el conjunto de desarrollos que aquí se presentan utilizan software libre y/o open source.

Tomcat como contenedor de aplicaciones

Apache como servidor Web

Mysql como base de datos

Fedora como backend de información digital

Fez como front-end web de Fedora

Jhove como extractor automático de metainformación digital.

ImageMagick para presentación de imágenes

Saxon como procesador de documentos xml

...

## El modelo de objeto digital (DO)

De forma cada vez más frecuente, los contenidos a gestionar tienen información digital incorporada : El pdf de un artículo de revista, las imágenes de una colección, los resultados de una investigación de laboratorio, páginas web, etc. Además, en muchos casos se desea que esta información no sea estática sino dinámica en el sentido de que se pueda incorporar nuevos elementos a ella (hacer anotaciones, relacionarla con otra información digital, aumentar su visibilidad en la red, etc)

Es decir, las unidades de información (que denominaremos **objetos digitales -DO**), tienen en la red una naturaleza compleja y dinámica, son objetos "vivos" que tienen diferentes ubicaciones en la red, se agregan y mantienen relaciones entre sí, y cuyas necesidades de almacenamiento, gestión, acceso, difusión y reutilización, en un entorno de trabajo en colaboración, requieren un sistema escalable y flexible, capaz de tratar y representar esa complejidad y adaptarse a los más que previsibles cambios tecnológicos.

En este contexto, se propone un nuevo modelo de Objeto Digital que tiene una uri asociada, capaz de representar cualquier data, soportando la agregación de varios data en un único Objeto digital, y capaz de expresar relaciones semánticas entre ellos. Sus características más relevantes son:

- Es un contenedor capaz de "estructurar" los datos que contiene. Se trata, por tanto, de un agregador de componentes. Característica clave para funcionar en la red es que tanto el contenedor como sus componentes sean direccionables mediante URI's.
- Se basa en la idea de que las unidades de información pueden ser:

- a) heterogéneas
- b) complejas
- c) generadas dinámicamente
- d) y mantienen relaciones entre sí

a) DO Heterogéneo:

Ha de poder representar muchos tipos de **unidades de información**:

- objetos textuales,
- imágenes,
- libros electrónicos,
- objetos multimedia,
- datasets,
- metadatos
- y muchas otras entidades

b) DO Complejo:

Ha de soportar la agregación en único DO de más de un componente de cualquiera de los tipos especificados anteriormente. Esos componentes pueden estar incluidos en el objeto o bien estar referenciados mediante URL.

c) DO dinámico:

Debe disponer de métodos asociados (otro tipo de DO) capaces de actuar sobre él. Por ejemplo, ofrecer la vista ampliada de una imagen o la tabla de contenidos del documento libro obtenida mediante su generación en tiempo de ejecución.

e) Soporta relaciones semánticas

Ha de ser capaz de expresar esas tripletas (sujeto, verbo y predicado). Por ejemplo :

```
<rdf:description   rdf:about="info:fedora/bibliuned:ETFSerieV-55C14806-F9E2-237B-99D3-045AD53B1069">
  <rel:isMemberOf rdf:resource="info:fedora/bibliuned:ETFSerieV2004"/>
  <rel:isMemberOf rdf:resource="info:fedora/bibliuned:Setarticulo"/>
</rdf:description>
```

## **Estandares de metainformación**

Frente a la creación de esquemas de metadatos que se adapten a nuestros casos

particulares, estimamos que es más provechoso la combinación de estándares ya definidos y vigentes en la red aumentando de esta manera la interoperabilidad. Así utilizaremos DublinCore con carácter general, LOM para representar Objetos de aprendizaje, RDF para representar relaciones semánticas, y estaremos atentos a las nuevas estructuras que defina el W3C.

Otro tipo de estándares como OAI-PMH ( y posiblemente ORE en el futuro) serán referencia cara a la difusión e interoperabilidad de los DO's.

Creemos que esta apuesta por los estándares es imprescindible cara a las posibilidades futuras de conexión a esa red de la siguiente generación, dónde los contornos de la información digital se difuminarán . Quién no esté en condiciones de conectarse al resto estará condenado a desaparecer, por muy buenos que sean sus valores específicos.

## **Fedora. Una arquitectura para DO's**

FEDORA<sup>1</sup> (no confundir con el sistema operativo) es el middleware que hemos escogido como arquitectura capaz de integrar este modelo de objeto digital.

Sus características principales son:

- Es un middleware con arquitectura SOA desarrollado en Java
- Todas sus funciones (acceso ,creación , modificación y borrado) están expuestas como servicios web.

<http://www.fedora.info/definitions/1/0/api/Fedora-API-A-LITE.wsdl>

<http://www.fedora.info/definitions/1/0/api/Fedora-API-A.wsdl>

<http://www.fedora.info/definitions/1/0/api/Fedora-API-M.wsdl>

<http://www.fedora.info/definitions/1/0/api/Fedora-API-M-LITE.wsdl>

- Montado sobre Tomcat como servlet
- Serializa los objetos digitales en XML con arreglo al siguiente esquema:  
<http://www.fedora.info/definitions/1/0/foxml1-0.xsd>
- Múltiples vistas de los DO mediante su asociación con métodos definidos en servicios web.
- Pueden gestionar el(los) asset(s) en forma local o remota

---

<sup>1</sup> <http://www.fedora-commons.org/>

- Relaciones entre objetos digitales : oArquitectura basada en RDF  
<http://www.fedora.info/definitions/1/0/fedora-relsext-ontology.rdfs>
- Metadatos sobre relaciones entre DO's basados en RDF  
<http://62.204.194.45:8080/fedora/get/bibliuned:ETFSerie1-F444A3E3-4230-C5FF-C492-BA3D011D7738/RELS-EXT>
- Posibilidad de buscar en el repositorio como un grafo , esto es mediante las relaciones definidas en RELS-EXT(relaciones con otros objetos)
- Control de acceso utilizando el estándar XACML con diferentes niveles de granularidad:
  - de repositorio,
  - del objeto
  - de componentes dentro de un objeto (datastreams).
- Permite incorporar LDAP y Shibboleth como sistemas de autenticación
- Control y mantenimiento de las diferentes versiones de un objeto digital
- Auditoría
- Preservación

## Fez: Un cliente Web para Fedora

Cuando se habla de crear un repositorio digital, se suelen barajar 3 posibilidades: Dspace, Eprints y Fedora. Por inmediatez y facilidad de instalación los 2 primeros tienen una clara preponderancia desde el punto de vista de usuario final. ¿Por qué entonces hemos elegido Fedora?. Porque creemos que el alcance y sus posibilidades de desarrollo en un entorno de red son mucho mayores. Esperamos que empiecen a aparecer aplicaciones no sólo orientadas al concepto de repositorio digital sino a otras aplicaciones :gestión del workflows de la publicación científica, portal específico de objetos de aprendizaje o de contenidos multimedia. Fedora , como tal, no ofrece el interfaz web para acceso y gestión de sus contenidos digitales. Este problema se ha resuelto con la aparición del cliente Fez cuyas características principales son:

a) gestión de información digital estructurada jerárquicamente (comunidad, colección, registro)

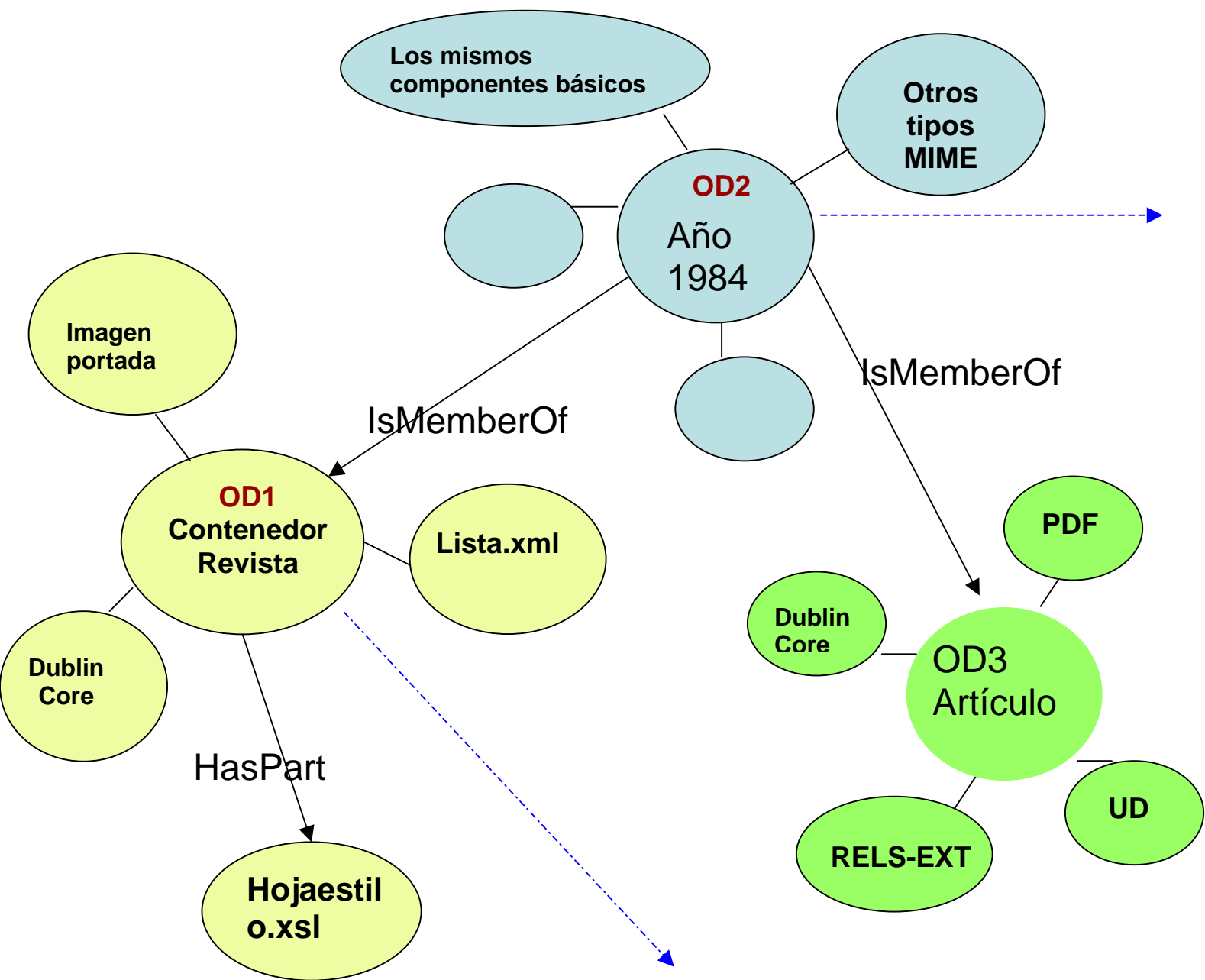
- b) servicios de histórico del registro, estadísticas de acceso individuales y generales
- c) importación y exportación en diferentes formatos (csv, rss, foxml )
- d) otros servicios (los más visitados, los más descargados)
- e) Autenticación -usuario, ldap, Shibboleth- y accesibilidad ACML.
- f) creación, en forma interactiva, de metainformación asociada a cierto tipo de contenidos digitales. Esta funcionalidad posibilita la creación de DOs con metainformación adaptada a sus tipos de contenidos. Así podré crear el objeto digital video con metadatos correspondientes a dublicore (para su información bibliográfica) lom (para su "aspecto" de objeto de aprendizaje) y VMD (para su "aspecto" técnico). La utilización de los estándares nos permitirá conseguir una mayor interoperabilidad con otros objetos digitales (locales o remotos)

## **Algunos servicios desarrollados en torno a los objetos digitales**

Partiendo de estos conceptos y arquitectura , vamos a indicar algunas aplicaciones desarrolladas utilizando Fedora :

-Creación de un repositorio conteniendo objetos digitales (locales y remotos):  
<http://e-spacio.uned.es> . En este momento disponemos de unos 11.500 DO's .procedentes mayoritariamente de la migración de datos ya existentes en BBDD

-Creación mediante query RDF y transformaciones xslt de 4 revistas digitales basadas en la existencia de las bases de datos de tripletes y en transformaciones xslt:  
<http://62.204.194.45:8080/fedora/get/bibliuned:revistaETF/demo:Collection/view/>



-Servicio web de búsqueda a texto completo (datos y assets textuales) mediante indexación con estándar Lucene.

<http://62.204.194.45:8080/fedoragsearch/rest> incluyendo acceso directo a texto ,imagen y video:

<http://62.204.194.45:8080/fedoragsearch/rest?operation=gfindObjects&query=maso&hitPageSize=>

-Servicio de generación de RSS en la búsqueda, por ejemplo:

<http://62.204.194.45:8080/fedoragsearch/rest?operation=gfindObjects&query=casa&hitPageStart=1&hitPageSize=10&indexName=DemoOnRepoured&restXslt=salidarss>

-Servicio web de diseminación mediante protocolo OAI-PMH.

<http://www.madrimasd.org/informacionidi/e-ciencia/buscar-documentos/default.asp>

## **Cara al futuro:**

- Aprovechar la posibilidad de intercambiar información en formato XML para crear y desarrollar proyectos de colaboración en red
- Estudiar y, si es posible, participar en la en el proyecto ORE (Object Reuse and Exchange ) orientado a modificar la forma en que se realiza el proceso de la comunicación y publicación científica.

## **Conclusiones**

- Creemos que esta elección tecnológica nos permite un alto grado de independencia de nuestros contenidos respecto de la representación elegida.
- Está orientada a su despliegue en la web
- Permite crear relaciones internas y externas entre unidades de información
- Permite asociar data con información de una forma sencilla .
- Utiliza esquemas estándar y publicados en la red para representar sus contenidos lo cual facilita la posibilidad de federación de contenidos

- Los DOs creados pueden ser vistos y utilizados por otros repositorios independientemente de los clientes de acceso utilizados