

## **Introducción a la tecnología donantonio**



## **Introducción a la tecnología donantonio**

Publicado \$Id: libro-blanco-tecnologias-doc.sgml,v 1.6 2001/06/29 15:29:07 olea Exp \$



## Tabla de contenidos

Prefacio .....	vii
.....	ix
Requisitos .....	11
Especificaciones técnicas .....	11
¿Especificación funcional? .....	12
Acciones que lleva a cabo un servidor Donantonio: .....	12
Acciones que lleva a cabo un cliente Donantonio: .....	12
Acciones del interfaz de usuario Donantonio .....	13
Sistema de consultas .....	13
Búsqueda de patrones .....	13
Navegación taxonómica .....	13
Ontología bibliográfica .....	14
Sistema de replicación de recursos .....	14
Esquemas bibliográficos .....	15
RDF .....	15
Esquemas estándar .....	15
Ampliaciones Donantonio de esquemas estándar .....	15
Consecuencias prácticas de la arquitectura Donantonio .....	16
Herramientas a desarrollar .....	16
Terminología Donantonio .....	17
Elementos .....	17
Tipos de agentes donantonio .....	17
La metáfora Donantonio .....	17
La metáfora astral .....	17
Equivalencia entre la metáfora Donantonio y la metáfora astral .....	18
Quién es Don Antonio .....	18
Bibliografía .....	18
.....	21
.....	25
Dorsales .....	27
Replicación de recursos .....	27
Tolerancia a caídas de la red .....	27



## Prefacio

**Aviso para esta edición:** Este libro no ha sido revisado a fondo. Si alguna información aquí expuesta entra en conflicto con las especificaciones donantonio (publicadas en documentos aparte) siempre tendrán prioridades dichas especificaciones.

Este libro recopila algunos textos para ilustrar el uso práctico de las tecnologías donantonio.

*Prefacio*



# **Introducción a Donantonio**

**Juan José Amor**

**David Escorial**

**Ismael Olea**

Este artículo ya está obsoleto por otros posteriores pero sigue siendo de interés hasta que no se extraiga y ordene la información útil.

Donantonio es un sistema de publicación distribuido de nueva generación. En primera instancia se desarrolla para la publicación de documentos libres pero es inmediatamente generalizable para la publicación, distribución y localización de toda clase de recursos electrónicos o como sistema de catalogación genérico.

## **Tabla de contenidos**

Requisitos .....	11
Especificaciones técnicas .....	11
¿Especificación funcional? .....	12
Sistema de consultas .....	13
Sistema de replicación de recursos.....	14
Esquemas bibliográficos.....	15
Consecuencias prácticas de la arquitectura Donantonio .....	16
Herramientas a desarrollar.....	16
Terminología Donantonio .....	17
La metáfora Donantonio.....	17
Quién es Don Antonio .....	18
Bibliografía .....	18



## Requisitos

- Publicación distribuida: no importa dónde se encuentra el recurso, este puede ser localizado desde cualquier punto de la red donantonio.
- Responsabilidad de publicación distribuida: no hay jerarquías en la red de servidores; todos los servidores se tratan de igual a igual; además cualquier sistema en Internet que acepte los protocolos donantonio y que esté permanentemente conectado puede ser un servidor donantonio.
- Uso y reutilización de protocolos Internet: siempre que se sea posible, dentro de los protocolos donantonio se usarán las tecnologías internet estándar.
- Implementación libre: licencia GPL.
- Un sistema de consulta entre aplicaciones Donantonio.
- Técnicas para la recuperación de información por parte de usuarios finales.
- Facilidades de replicación automática en base a consultas Donantonio.
- Sencillez de diseño e implementación, independiente de lenguajes y plataformas a fin de que pueda ser fácilmente implementado y rápidamente difundido.

## Especificaciones técnicas

El sistema Donantonio funcionará conforme a las siguientes reglas.

- Uso de descripciones bibliográficas (metadatos); cada recurso publicado necesita de un único documento que lo describa. Para ello se usarán esquemas bibliográficos estándar (bien ya existentes, bien ligeramente modificados bien de nueva factura).
- La descripción no tiene por qué estar alojada en el mismo servidor que el recurso.
- Cada servidor mantendrá sus propias descripciones, ofreciéndolas a quien esté dispuesto a copiarlas.
- Existirá una diferencia formal entre una descripción original y su copia. Esta diferencia se plasmará en una marca de originalidad-o-no. La marca de originalidad sólo la tendrá la descripción original, situada en el servidor de publicación. Cada vez que se copia una descripción original la marca de originalidad debe desaparecer.
- El hecho de publicar descripciones originales o copias se hace de forma conjunta y sin distinción.
- Las descripciones replicadas (de diferentes fuentes) pueden ser sometidas a un proceso de filtrado para eliminar duplicados idénticos. Queda por definir el algoritmo que distinga, por ejemplo, versiones diferentes, si es que es preciso (tal vez sea un filtro que deba tener disponible sólo el usuario final o el mecanismo de consulta). Lo más ideal tal vez sea disponer de «módulos de filtros» adaptables, intercambi-

ables y modificables a placer. *NOTA: Este párrafo no se puede quedar así por nada del mundo.*

- Un servidor donantonio puede ser a su vez cliente donantonio.
- Cada cliente replicará las descripciones de los servidores que le plazcan, sin mantener ninguna topología o norma concreta fuera del libre albedrío del administrador del servicio donantonio.
- por ahora no se decide cuál debe ser el método de almacenamiento interno. Probablemente no haya que hacerlo nunca.
- Es obligatorio importar/exportar descripciones en el formato definido por la norma donantonio.
- Sistema de consulta entre aplicaciones.
- Técnicas de usuario final para recuperar información:
  - búsqueda de patrones en descriptores;
  - navegación por taxonomías dinámicas;
  - ontología bibliográfica (váyase usted a saber qué es eso).
- Procesos de replicación/filtro/consulta (POR DEFINIR).

## **¿Especificación funcional?**

### **Acciones que lleva a cabo un servidor Donantonio:**

- registrar descripciones
- publicar almacén de descripciones visible a través de
  - mecanismo de consulta
  - protocolos internet de transferencia de ficheros
- (opcional) publicar descripciones (o sea, publicar la dirección a través de una url como un fichero convencional más)
- atender a consultas Donantonio

## Acciones que lleva a cabo un cliente Donantonio:

- replicar almacenes de descripciones mediante
  - protocolos internet de transferencia de ficheros
  - consultas Donantonio
- replicar almacenes de recursos en función de una consulta (con la que se genera previamente el almacén de descripciones).

## Acciones del interfaz de usuario Donantonio

(el interfaz de usuario es una herramienta interactiva de consultas)

- Diseño de consultas (¿a estilo QBE?)
- Representación del resultado de las mismas
- Acceso a los recursos referenciados en las consultas

## Sistema de consultas

Las consultas son el mecanismo mediante el cual el usuario final (humano o cibernético) accede al conocimiento y a los recursos.

Se proponen tres tipos de consulta:

*(Nota: ahora que lo pienso puede esta división sólo tenga sentido a nivel de interfaz de usuario y que puedan codificarse internamente con el mismo mecanismo... ¿tal vez SQL? Es probable)*

- búsqueda de patrones en descriptores bibliográficos;
- navegación taxonómica (qué nombre tan grandilocuente, espero no haber metido la gamba).
- ontología bibliográfica (chúpate esa, grep).

## Búsqueda de patrones

La búsqueda de patrones es una búsqueda clásica. Si es necesario intentamos hacer una descripción más formal. El interfaz de usuario tal vez podría ser tipo QBE.

## **Navegación taxonómica**

La navegación taxonómica creo que es una herramienta potentísima. No he comprobado si sitios como Yahoo lo usan pero es ideal.

¿En qué consiste? Fácil. Las descripciones bibliográficas están hechas en base a un esquema. Ese esquema tiene una sintaxis y contenidos predeterminados. Si ordenamos un conjunto de descripciones en función de un orden determinado de los campos que forman parte del esquema, obtenemos una jerarquía. Lo ideal es que el orden de la jerarquía se adapte a cada consulta en función del usuario. Así, una consulta podría empezar ordenando descripciones de mamíferos por terrestres/anfibios/marinos/o voladores y hacer que la siguiente jerarquía esté en función de otro campo como podría ser... joer, no me acuerdo cómo se clasifican los mamíferos.

La cuestión es que otra consulta puede tener un orden diferente.

En cada consulta el usuario va navegando más profundamente refinando la consulta.

Para que este refinamiento tenga más sentido, en cada incremento de la profundidad puede añadir restricciones en el sentido de campos que cumplan o no determinadas propiedades.

La consulta finalmente debe tener la posibilidad de ser guardada, a fin de reutilizarla en otro momento (como una vista al almacén de descripciones).

Como es de suponer, a lo largo de la navegación van apareciendo referencias a las descripciones que cumplen las condiciones de la búsqueda. Finalmente las descripciones nos llevarán al recurso. Como Freshmeat, pero bien hecho.

Sé que esto es un infierno de incongruencias, pero para eso estamos, para limpiar, fijar y dar esplendor.

## **Ontología bibliográfica**

Ilumínanos, David.

## **Sistema de replicación de recursos**

Si construimos una consulta Donantonio se supone que es porque tenemos alguna clase de interés en los recursos referidos. En ese caso es interesante poder disponer de ellos rápidamente en modo local.

Para eso nada más fácil que replicar los mismos manteniendo una copia sincronizada de cada uno de ellos.

Disponiendo de las descripciones de nuestro gusto y dado que los recursos son accesibles vía internet es casi trivial diseñar una herramienta que los replique.

Los usos son muchos y variados:

- creación de bibliotecas-e departamentales
- confección de cdrom recopilatorios

- mantenimiento de distribuciones Linux.

## Esquemas bibliográficos

Los esquemas bibliográficos son conjuntos de descriptores de metadatos normalizados que se usan para clasificación bibliográfica. *NOTA:* apuesto a que tiene una definición más formal.

## RDF

Las siglas RDF responden (SI NO ME EQUIVOCO) a Resource Description Framework, que es un estándar del W3C para la codificación de descripciones bibliográficas usando tecnologías XML.

Una de las ventajas del diseño de RDF es que, gracias a los *name spaces* de XML, una descripción puede estar almacenada en un documento XML de forma completamente transparente. Eso facilita enormemente el mantenimiento de ambos.

RDF está desarrollándose rápidamente en la industria del software (comercial y libre) y las nuevas versiones de los navegadores más importantes contemplan implementarlo.

## Esquemas estándar

Siempre que sea posible se procurará usar esquemas ya estandarizados en RDF, XML o en otras tecnologías adaptándolos, si es preciso, al nuevo medio.

Algunos esquemas interesantes son

- OMF, para documentación libre;
- lsm, para software libre;
- rdf/rpm (no sé su nombre exacto) para catalogar paquetes de software RPM;
- UDDI, nueva norma para la catalogación y publicación de servicios B2B en internet;
- donantonio-biblo (pendiente de definir), para la publicación y difusión de servidores.

## Ampliaciones Donantonio de esquemas estándar

No es el propósito de esta tecnología contaminar estándares abiertos, pero que esta tecnología funcione como se espera es preciso codificar cierta información en cada descripción.

Está pendiente de resolver si esta información puede codificarse transparentemente mediante los *name spaces* o es preciso crear variantes mínimas de los esquemas a usar.

## **Consecuencias prácticas de la arquitectura Donantonio**

- Los recursos y las descripciones tienen puntos fijos de publicación y son accesibles a través de una URL.
- Es posible la existencia de recursos y descripciones idénticos con diferentes URL. Aunque precisamente se ha diseñado Donantonio con la idea de erradicar esta costumbre, la funcionalidad del mismo sólo empeora ligeramente si se usa.
- Las copias de recursos y descripciones (copias de los mismos) no tienen el mismo interés bibliográfico y por tanto tienen un tratamiento claramente separado.
- El almacén de descripciones de cada servidor donantonio es visible en cualquier parte de la red.
- Los servidores pueden ser genéricos (replican todo) o especializados (recopilan sólo cierto tipo de información de interés local).
- Las facilidades de replicación a través de consultas diseñadas convenientemente permiten redistribuir la carga de servidores a medida de las necesidades
- Cualquier recurso localizado a través de donantonio es accesible vía internet.
- No se puede garantizar la localización de un recurso en la red donantonio, pero sí en su espina dorsal (subred de servidores genéricos).
- Los clientes donantonio podrían ser otros servidores, interfaces web o un programa cliente en cualquier lenguaje clásico.
- pueden diseñarse redundancias entre servidores (un servidor puede proveerse de dos servidores idénticos filtrando duplicados) que fortalecen la red contra caídas de algunos nodos.

## **Herramientas a desarrollar**

- editor de descripciones
- creador semi/automático de descripciones
- publicador de descripciones
- replicador de descripciones
- gestor de descripciones
- motor de consultas;
- herramienta de usuario para navegación y recuperación de información;



- replicador de ficheros (sobre una consulta prefijada) (organiza una consulta/replicación-de-metadatos en una jerarquía de recursos que son copia de los originales)

## Terminología Donantonio

### Elementos

#### Recurso

fichero o documento accesible a través de una URL.

#### Descripción

fichero de metadatos de un recurso, accesible a través de una URL.

#### Repositorio

colección de recursos o de copias de recursos.

#### Biblioteca

colección de descripciones o de copias de descripciones.

## Tipos de agentes donantonio

#### Servidor

nodo actuando como servidor donantonio.

#### Cliente

nodo actuando como cliente donantonio o cliente donantonio puro.

#### Representante

nodo que actúa a la vez como cliente y como servidor donantonio. *NOTA: tal vez este agente sea formalmente prescindible.*

## La metáfora Donantonio

Para explicar, aunque de forma completa, lo que es Donantonio puede usarse la «metáfora astral».

## **La metáfora astral**

Si creemos a algunos astrólogos o a ciertas religiones, el cuerpo físico de las personas está unido a un cuerpo astral: el alma. En ciertas circunstancias el alma puede viajar fuera de su correspondiente cuerpo, que duerme en algún lugar de la Tierra, permaneciendo ambos unidos gracias a una cinta plateada. Si la cinta se rompe por algún motivo, es la muerte para esa persona, cuyos cuerpo y alma quedan irremisiblemente separados para siempre.

## **Equivalencia entre la metáfora Donantonio y la metáfora astral**

Según el esquema Donantonio, los recursos son los cuerpos, que permanecen fijos en algún sitio, dormidos, mientras que sus almas, que son los metadatos que los describen, viajan libremente por la red Donantonio, unidos ambos por una URL.

Si el enlace se pierde, el recurso desaparece. El documento ha muerto.

La idea principal es que mientras los recursos permanecen tranquilos, durmiendo en una URL de referencia, sus descripciones se mueven con libertad y son procesadas con libertad sin peligro a perder el nexo entre ambos que hacen que el mecanismo de la comunicación tenga lugar.

## **Quién es Don Antonio**

Don Antonio es el nombre del bibliotecario donde estudié el bachillerato.

Elegí el nombre porque me pareció sonoro, expresivo y como una nota de color y alegría frente a los fríos nombres de los protocolos y aplicaciones de Internet.

## **Bibliografía**

- XML
- DC
- RFC 2413 (DC)
- RDF
- Los de la clasificación automática.
- URL de LuCAS
- del borrador del libro de estilo de Insflug
- al glosario de LuCAS
- ldp-core.dtd: <http://casbah.org/cgi-bin/cvsweb.cgi/XML/DTDs/ldpCoreMetaML.dtd>  
Creada por Kendall Clark en XML. Se supone un paso previo a la versión RDF.

DC	<a href="ftp://sunsite.rediris.es/docs/rfc/24xx/2413">ftp://sunsite.rediris.es/docs/rfc/24xx/2413</a> <sup>1</sup> <a href="http://purl.org/DC/">http://purl.org/DC/</a> <sup>2</sup>
DocBook	<a href="http://www.docbook.org">http://www.docbook.org</a> <sup>3</sup>
DTD	TACATACA
GNU	<a href="http://www.gnu.org">http://www.gnu.org</a> <sup>4</sup> <a href="http://www.fsf.org">http://www.fsf.org</a> <sup>5</sup>
Glosario de LuCAS	<a href="http://LuCAS.HispaLinux.ES/Otros/glosario/Glosario/glosario.html">http://LuCAS.HispaLinux.ES/Otros/glosario/Glosario/glosario.html</a> <sup>6</sup>
GPL	TACATACA
ISO	International Standards Organization
LDP	<a href="http://www.linuxdoc.org">http://www.linuxdoc.org</a> <sup>7</sup>
LDP-es	<a href="http://LuCAS.HispaLinux.ES">http://LuCAS.HispaLinux.ES</a> <sup>8</sup>
LuCAS	<a href="http://LuCAS.HispaLinux.ES">http://LuCAS.HispaLinux.ES</a> <sup>9</sup>
Metalab	<a href="http://metalab.unc.edu">http://metalab.unc.edu</a> <sup>10</sup>
RDF	<a href="http://www.w3.org/Metadata/RDF">http://www.w3.org/Metadata/RDF</a> <sup>11</sup>
RFC	<a href="ftp://sunsite.rediris.es/docs/rfc/">ftp://sunsite.rediris.es/docs/rfc/</a> <sup>12</sup>
SQL	TACATACA
URL	Uniform Resource Locators <a href="ftp://sunsite.rediris.es/docs/rfc/17xx/1738">ftp://sunsite.rediris.es/docs/rfc/17xx/1738</a> <sup>13</sup>

SGML

TACATACA

XML

<http://www.w3.org/XML/><sup>14</sup>

W3C

<http://w3.org/><sup>15</sup>

## **Notas**

1. <ftp://sunsite.rediris.es/docs/rfc/24xx/2413>
2. <http://purl.org/DC/>
3. <http://www.docbook.org>
4. <http://www.gnu.org>
5. <http://www.fsf.org>
6. <http://LuCAS.HispaLinux.ES/Otros/glosario/Glosario/glosario.html>
7. <http://www.linuxdoc.org>
8. <http://LuCAS.HispaLinux.ES>
9. <http://LuCAS.HispaLinux.ES>
10. <http://metalab.unc.edu>
11. <http://www.w3.org/Metadata/RDF>
12. <ftp://sunsite.rediris.es/docs/rfc/>
13. <ftp://sunsite.rediris.es/docs/rfc/17xx/1738>
14. <http://www.w3.org/XML/>
15. <http://w3.org/>

# **Un día en la vida de un recurso en la red Donantonio**

**Ismael Olea**

Nota: esto es sólo una visión parcial que pretende ser ilustrativa. Casi nada para estos tiempos.



1. Un recurso es publicado en un servidor inet (con una URL)
2. Creación de la descripción bibliográfica del recurso
  - creación manual
  - asistida por un editor de metadatos
  - semiautomática, mediante un programa de extracción de palabras clave.
3. Publicación de la descripción en inet (con una URL)
4. Registro de la descripción en un servidor Donantonio (Sugerencia: probablemente el servidor debe distinguir entre:
  - descripciones revisadas
  - descripciones no revisadas
  - descripciones copiadas de otros servidores Donantonio)
5. El servidor Donantonio:
  - archiva una copia de la descripción en su almacén público
  - alimenta con la descripción su sistema de gestión de datos.
6. El servidor Donantonio atiende consultas Donantonio
7. El cliente Donantonio que lanzó la consulta recoge del servidor consultado como resultado una serie de descripciones que la cumplen.
8. El cliente Donantonio puede archivar el resultado de la consulta en su propio almacén o bien puede acceder al recurso/s vía URL.

Nota: ¿cómo lo véis? ¿os aclara algo? :-?

*Un día en la vida de un recurso en la red Donantonio*



# Aplicaciones de donantonio

Ismael Olea

Ejemplos de aplicaciones donantonio

## Tabla de contenidos

Dorsales .....	27
Replicación de recursos .....	27
Tolerancia a caídas de la red .....	27



## **Dorsales**

Pueden construirse redes dorsales que garanticen la disponibilidad de todas las descripciones registradas en una red donantonio:

### **Dorsal de Donantonio**

red de servidores en la que siempre es posible localizar cualquier recurso dado de alta en Donantonio.

### **Servidor genérico**

servidor donantonio que forma parte de la dorsal de la red.

### **Servidor especializado**

servidor donantonio que replica solo una parte de las descripciones que guarda la dorsal de la red donantonio, y por tanto no puede formar parte de dicha dorsal.

## **Replicación de recursos**

El sistema permitirá, mediante una selección previa (vía consulta), realizar una replicación local de los recursos seleccionados.

De esta forma, mediante protocolos estándar Internet y otros específicamente diseñados para este sistema, se podrá realizar una copia local de los recursos que resulten de nuestro interés (por ejemplo, cumplan un criterio de búsqueda taxonómico determinado) con el fin de crear una biblioteca-e local.

## **Tolerancia a caídas de la red**

El diseño de la red de servidores dará un matiz de fiabilidad al sistema. Es decir, una caída de un servidor no impedirá necesariamente la localización de un recurso, cuando su descripción se encuentre replicada en otros servidores.

*Aplicaciones de donantonio*