

# METADATOS: Introducción e historia

Cristian Vásquez Paulus

Comentarios y sugerencias a:  
cvasquez@dcc.uchile.cl

## Índice

1	Introducción .....	2
1.1	¿Que son los metadatos?.....	2
2	Evolución de los metadatos en la Web.....	3
2.1	Raíces.....	3
2.2	MAchine Readable Cataloguing (MARC) .....	4
2.3	ISO 23950 (Z39.50) .....	4
2.4	Standard Generalized Markup Language (SGML) .....	5
2.5	Document Type Definition (DTD).....	5
2.6	The Warwick Framework.....	6
3	Metadatos extrínsecos e intrínsecos .....	6
4	Identificadores.....	7
4.1	Uniform Resource Identifier (URI).....	7
5	La recuperación de información con metadatos .....	9
6	Beneficios a la hora de utilizar metadatos .....	9
7	Estándares de metadatos.....	10
8	Clasificación de los estándares de metadatos .....	12
8.1.1	Administrativos .....	13
8.1.2	Descriptivos y de descubrimiento:.....	13
8.1.3	Técnicos y modelos .....	13
9	Los lenguajes de metadatos .....	15
10	Puntos importantes. ....	16
11	Bibliografía .....	18

# 1 Introducción

En estos días la mayoría del contenido de la Web esta diseñada por humanos para ser leída por ellos mismos, formando una caótica biblioteca de enormes dimensiones desde la cual resulta complicado y costoso extraer conocimiento alguno.

Para solucionar esto, numerosas comunidades desarrollan y promueven el uso de la "Web semántica" la cual tiene entre sus objetivos modificar la forma en que se presenta la información de la Web de un modo que facilite el procesamiento de la misma por parte de las máquinas y de esta forma establecer canales para un factible procesamiento, integración y reutilización de la información contenida en la Web, apostando así a la extracción de conocimiento de mayor utilidad a los humanos.

La Web semántica no sería sino a una extensión de la ya existente, en que la información adquiere significado mediante el uso de **metadatos** para proveer una categorización semántica de su contenido, permitiendo razonar automatizadamente sobre la información.

## 1.1 ¿Que son los metadatos?

Si bien podríamos definir (etimológicamente hablando) que un metadato es algo que va "más allá de los datos", lo cierto es que en la literatura actual no existe consenso sobre lo que realmente son estos metadatos. Una definición utilizada con frecuencia nos dice que los metadatos son "datos sobre datos", en general un objeto que describe o dice algo sobre otro objeto de información.

Aunque el uso de la palabra "metadato" se masificó en un contexto que se refiriere a la era de la información digital, la generación de metadatos data de siglos atrás. Los bibliotecarios han creado metadatos que han tomado la forma de catálogos de libros, catálogos de tarjetas y en la actualidad catálogos en línea. Hoy en día la generalización del concepto ha cubierto cualquier tipo de información descriptiva (estandarizada) sobre recursos, incluyendo los que no son digitales.

De manera formal podríamos decir que un metadato es un dato que se encarga de mantener un registro sobre el significado, contexto o propósito de un objeto informativo, de tal forma de poder descubrir, entender, extraer y administrar dicho objeto. En general estos registros son de menor envergadura que los objetos que describen, y son creados en un formato corto y conciso de tal forma que puedan intercambiarse.

Los metadatos pueden describir colecciones de objetos y también los procesos en los que están involucrados, describiendo cada uno de los eventos, sus componentes y cada una de las restricciones que se les aplican. Los metadatos definen las relaciones entre los objetos, como las tuplas en una base de datos o clases en orientación a objetos, generando estructuras.

A lo largo del documento se utilizará cierta terminología técnica sobre los lenguajes:

**Semántica:** Es el significado que asocia una comunidad a un elemento de metadatos o a los valores de ese elemento, organizado en un vocabulario de entendimiento común.

**Estructura:** Parte del lenguaje que impone un orden a las expresiones semánticas, genera coherencia en la codificación de las semánticas no-ambiguas. Permite una interpretación consistente para los que se comunican.

**Sintaxis:** Parte del lenguaje que provee de medios para representar las estructuras, entrega mecanismos para codificar, intercambiar, visualizar y procesar los metadatos. Un ejemplo es XML.

## 2 Evolución de los metadatos en la Web

### 2.1 Raíces

Los metadatos tienen sus raíces en el **catálogo**, probablemente inventado poco después del comienzo de la historia por parte de los Sumerios. A lo largo de los siglos las tabletas de arcilla utilizadas evolucionan hasta listas manuscritas y posteriormente a catálogos de libros después de la invención de la imprenta. Estos primeros catálogos de libros eran impresos y eran listas ordenadas alfabéticamente sin criterios de clasificación sofisticados. Un avance importante en cuanto a esquemas de clasificación se desarrolla alrededor del 1900 cuando los catálogos de libros son reemplazados completamente por tarjetas, las cuales entre otras cosas pueden ser actualizadas. En la década del sesenta los métodos de producción en masa (como los computadores) hacen necesario disponer de múltiples copias de los catálogos existentes, surgen masivas colecciones distribuidas de libros y los catálogos de tarjetas no logran satisfacer los nuevos requerimientos. Es necesario entonces desarrollar estándares de codificación, llamados hoy en día metadatos.

Los primeros metadatos (digitales) y sus bases se desarrollan a finales del siglo XX, cuando emergen múltiples estándares de codificación, lenguajes y protocolos que se utilizan en la generación y uso de catálogos.

## 2.2 MACHINE READABLE CATALOGUING (MARC)

MARC fue concebido para transmitir datos de un sistema a otro, y fue revolucionario al incorporar campos de largo variable. Contiene códigos "directorios", alfanuméricos de largo fijo que determinan el nombre, largo y donde empieza cada campo de descripción, campos de "control", utilizados para clasificar la información en términos de tiempo y lugar. Los campos de descripción variable son los que contienen datos de catalogación tradicionales, y son precedidos por un código definido que va desde 001 a 999, donde por ejemplo el código 650 es la materia por tópicos del recurso.

Desde la creación de MARC se generaron más de veinte estándares nacionales (DenMARC, AZMARC, CHMARC, UKMARC, CAN/MARC etc) los cuales tienden a 'armonizarse'. El más conocido de ellos es USMARC (*United States MARC*), también llamado LC-MARC que fue desarrollado en 1968 por la Biblioteca del Congreso de Estados Unidos y deriva de MARC. Otro estándar de renombre es MARC21 producto de la conjugación de CAN/MARC (de Canadá) y USMARC en 1999.

También desde 1977 existe una interlingua entre los distintos estándares MARC, creada gracias a un esfuerzo de cada agencia bibliográfica nacional, en donde se generaron traductores desde su estándar propio a un esquema UNIMARC (UNiversal MARC) y viceversa.

## 2.3 ISO 23950 (Z39.50)

Este es un protocolo para la generación de consultas a lo largo de múltiples catálogos online. De origen estadounidense, data de 1988 momento en que fue aprobado por la NISO (*National Information Standards Organization*) y permite a un usuario de un sistema buscar y recuperar la información sin saber la sintaxis utilizada por los otros sistemas. Posee un protocolo XML llamado XER y es portable a SQL.

Ambos, MARC y Z39.50, son utilizados ampliamente por las entidades bibliotecarias tradicionales y es probable que lo sigan siendo por un tiempo debido al alto costo en el que deben incurrir estas entidades para mutar de formato, además del poco financiamiento de que disponen para estos fines.

En el desarrollo de los lenguajes que son utilizados en el mercado de metadatos juega un papel fundamental *Standard Generalized Markup Language* (SGML), los *Document Type Definition* (DTD) y *The Warwick Framework*.

## 2.4 Standard Generalized Markup Language (SGML)

SGML es un lenguaje de marcado de documentos. Sus raíces se remontan a 1969 cuando en los laboratorios de IBM se desarrolla *Generalized Markup Language* (GML), lenguaje que evoluciona hasta 1974 donde pasa a llamarse SGML. La *International Organization for Standardization* (ISO) aprueba y publica este lenguaje en 1984 con el nombre de estándar ISO 8879.

Este estándar internacional consta de un conjunto de reglas para describir la estructura de un documento de tal forma que puedan ser intercambiados a través de las plataformas computacionales. SGML es extremadamente flexible y es la base de los lenguajes de marcado más utilizados hoy en día.

En SGML un documento está definido en función de la estructura de las entidades que lo conforman. Estas entidades se organizan en una estructura lógica de manera jerarquizada determinando la estructura de los elementos del documento. Las entidades pueden ser compartidas por distintos documentos.

El marcado se lleva a cabo mediante delimitadores y etiquetas de la forma ‘<etiqueta> elemento </etiqueta>’. Las etiquetas pueden estar anidadas y se representan mediante el conjunto de caracteres básicos de acuerdo al estándar ISO 8879.

En el contexto histórico de los metadatos, la introducción de SGML jugó un papel fundamental, pues estableció un nuevo paradigma, en que los datos dejan de ser sólo datos. Los documentos SGML contienen separadamente (en el sentido lógico) los contenidos, la estructura y el formato.

## 2.5 Document Type Definition (DTD)

Son aplicaciones de SGML y son las utilizadas para definir las estructuras de un tipo de documento en especial. Sus raíces se remontan a 1978 cuando en los laboratorios de IBM se publican los primeros DTD como parte del desarrollo de SGML.

Un DTD puede crearse para describir las estructuras de múltiples documentos o para alguno en particular. Estas estructuras se definen indicando reglas tales como nombres de los elementos permitidos, el contenido de cada tipo de elemento y el orden en el cual los elementos pueden aparecer.

Uno de los más conocidos es el DTD de HTML (*HyperText Markup Language*) que define las reglas que dan a luz este masivo lenguaje para marcado de páginas Web.

Particularmente en las bibliotecas se utilizan variados DTD, tales como EAD (*Encoded Archival Description*) para la descripción bibliográfica y TEI (*Text Encoding Initiative*) para el marcado de las versiones electrónicas de textos culturales.

## 2.6 The Warwick Framework

Esta iniciativa tiene sus orígenes en abril de 1996, cuando se lleva a cabo un *workshop* en la Universidad de Warwick. El encuentro contó con más de cincuenta representantes de bibliotecología, estándares de Internet, marcado de texto y proyectos de bibliotecas digitales.

El concepto detrás de esta iniciativa es mantener múltiples conjuntos de metadatos independientes unos de otros en un solo lugar o '*framework*', con el objeto de proveer de medios para administrar y acceder a conjuntos de metadatos por separado. Este marco permite la existencia de distintas sintaxis en cada conjunto de metadatos de acuerdo a los requerimientos semánticos, promoviendo la interoperatividad y extensibilidad al momento de manipular (en forma selectiva) dichos paquetes por parte de los agentes o sistemas que los utilizan.

Los *framework* pueden tener dos tipos de objetos: contenedores y paquetes. Un paquete está definido como un conjunto de metadatos tal como puede ser una descripción de *Dublin Core* y un contenedor está definido como el lugar donde se almacenan otros contenedores o paquetes. Un contenedor puede tener estados transientes y persistentes, en donde un contenedor transiente es un objeto de transporte entre repositorios, clientes y agentes; y a su vez los contenedores persistentes son los que perduran en el tiempo y son accesibles mediante un identificador universal.

Este *framework* resultó de un análisis de *Dublin Core* e influenció en gran medida la creación de *Resource Description Framework* (RDF).

## 3 Metadatos extrínsecos e intrínsecos

Dependiendo del contexto en que se generan los metadatos, pueden distinguirse dos grandes grupos de metadatos de muy distinta naturaleza, los metadatos extrínsecos y los metadatos intrínsecos<sup>1</sup>.

Los metadatos extrínsecos son los que disponen de un enlace persistente entre el registro de metadato y el objeto que describen, enlace que existe separadamente de estos objetos. Entre los principales beneficios de los metadatos extrínsecos está el que permiten el intercambio de la información sin necesidad de involucrar el intercambio de los recursos mismos en la transacción, disminuyendo los costos de distribución y administración de los objetos. Por esta razón, pueden efectuarse operaciones de una forma limpia. Un caso práctico es pedir autenticación antes de acceder a los datos, gracias a que estos datos están separados

---

<sup>1</sup> En la literatura, los términos intrínsecos y extrínsecos pueden referirse también a datos sobre las características esenciales de los objetos (intrínsecos) y al contexto en que son utilizados (extrínsecos).

(seguridad en la distribución). También los metadatos extrínsecos son flexibles y escalables a la hora de alterarlos, dado que las modificaciones y agregaciones se efectúan sobre el metadato sin re-codificar los datos principales.

Los metadatos intrínsecos son los que están sincronizados con el objeto que describen, para un uso de ellos que sea sensible al contexto en que se utilizan, siendo muy dificultoso el residir fuera del objeto que describen. Los casos más comunes son los metadatos que describen objetos generados en forma dinámica y los que describen objetos con características que varían frecuentemente a lo largo del tiempo, como las distribuciones espaciales de los objetos o la disponibilidad de un ítem en una casa comercial.

El problema principal de los metadatos intrínsecos es que la administración de los metadatos esta estrechamente interrelacionada con el objeto que describen, generando problemas técnicos a la hora de alterar alguno de los dos.

Estos dos tipos de metadatos nos son exclusivos, existiendo sistemas que son combinación de ambos. Algunas comunidades recomiendan el uso conjunto de metadatos extrínsecos e intrínsecos para sus descripciones (sin dejar de distinguirlos).

## **4 Identificadores**

Un identificador es un objeto que identifica alguna entidad en forma única y persistente, es decir identificando a dicha entidad durante toda su existencia.

Los identificadores son la base de la descripción con metadatos, dado que en este ámbito los objetos sólo existen cuando son identificables y distinguibles.

### **4.1 Uniform Resource Identifier (URI)**

Una URI es un objeto que identifica alguna entidad de forma única y persistente.

Una URI es un código que se manifiesta mediante una secuencia de caracteres de un alfabeto limitado. Estas URIs son totalmente independientes de las formas en que se manifiestan, ya sea en forma de letras impresas o símbolos en pantalla, lo realmente importante es que identifiquen un objeto en forma única y persistente.

Una URI siempre define o especifica una entidad, las cuales pueden ser cualquier cosa, pueden ser un documento, un dispositivo físico, una película, un concepto abstracto (por ej: "autor") o incluso un ciudadano de un país. Una analogía interesante de las URIs es el uso de números de carnet (RUT) de alguna persona, el cual es único para alguna entidad persona y persiste durante toda su existencia, esto sin importar donde se encuentre o cómo haya cambiado a lo largo del tiempo.

Uno de los casos más importantes son los identificadores URL (*Uniform Resource Locator*) que se refieren al subconjunto URIs las cuales identifican los recursos mediante la

representación de su mecanismo de acceso primario (lugar), más que identificar el recurso por un nombre o algún otro atributo de ese recurso.

El término URN (Uniform Resource Name) se refiere al subconjunto de URIs que se requiere para identificar un recurso en forma globalmente única y persistente incluso si el recurso deja de existir.

Existen de distintos esquemas de URIs, con distintos esquemas de representación de los que dependen. Estos distintos esquemas difieren por identificar distintos componentes de un recurso o representar estos recursos mediante un alfabeto distinto, por ejemplo el alfabeto de las URL considera que D y d son el mismo carácter, por lo que `<http://www.dcc.uchile.cl>` es equivalente a `<http://www.DCC.UChile.CL>`, a diferencia de los identificadores de archivos de Unix.

Existen muchos esquemas de identificadores, con distintos grados de aceptación y uso, algunos ejemplos de identificadores son:

El esquema <b>FTP</b> :	<code>ftp://ftp.is.co.za/rfc/rfc1808.txt</code>
El esquema <b>HTTP</b> :	<code>http://www.math.uio.no/faq/compression-faq/part1.html</code>
El esquema <b>MailTo</b> :	<code>mailto:mduerst@ifi.unizh.ch</code>
El esquema <b>USENET</b> :	<code>news:comp.infosystems.www.servers.unix</code>

Un identificador siempre se refiere a un recurso independientemente del contexto en que este identificador es utilizado, De ahí viene el término 'uniforme'. Distintos tipos de identificadores de recursos también deben poder ser utilizados en el mismo contexto, por ejemplo en una descripción con metadatos. Es importante notar que los objetos que van a ser identificados con URIs pueden ser identificados legítimamente por más de un identificador.

La tendencia indica que a lo largo del tiempo se introducen nuevos tipos de identificadores de recursos, los cuales no deben interferir con los ya existentes, permitiendo la existencia de una interpretación semántica común para todos los distintos identificadores.

Para subsistir, la determinación y administración de las URIs son llevadas a cabo descentralizadamente y por parte de variadas organizaciones y ya que es muy difícil mantener un control exhaustivo sobre ellas, las URIs pueden ser asignadas equívocamente a lo largo del tiempo. Se hacen necesarios entonces mecanismos para resolver estos conflictos. Surge entonces el concepto de 'grado de confianza' hacia las organizaciones.



## **5 La recuperación de información con metadatos**

Las técnicas de recuperación de la información clásica asumen que el texto o el objeto de la información en sí es capaz de entregar una descripción de contenido de un recurso con suficiente detalle. Estas técnicas son muy útiles, pero al momento de realizar búsquedas con un mayor grado de sofisticación, si bien los usuarios incurren en estructuración del espacio de búsqueda (de acuerdo a tipo de recurso, localidad e idioma por nombrar algunos), estas tienden a ser infructuosas.

La existencia de estructuras de metadatos permite la generación de herramientas relevantes para la extracción de recursos que incluyan los aspectos semánticos de las consultas. Con los sistemas actuales de recuperación de información de uso masivo, preguntas como "deme todos los recursos que contengan niñas chilotes de 10 años escribiendo un cuento" o "canciones venezolanas que incluyen el sonido de una flauta" no arrojan los resultados apropiados.

La efectividad aumenta en búsquedas en un ambiente en el que existen metadatos, éstos permiten definir múltiples interrelaciones entre los objetos de información, proveyendo de la estructura semántica necesaria para realizar una recuperación útil a los usuarios en términos de calidad y relevancia de los resultados. Además, las estructuras definidas permiten efectuar inferencias sobre éstas y por consiguiente dan cabida a la posibilidad de efectuar consultas más complejas.

Actualmente es muy difícil extraer automáticamente descripciones de metadatos desde el contenido de los recursos léxicos, por lo que no debe descartarse la generación de estos metadatos en forma manual o asistida. Esta generación debe contar con la cooperación por parte de los creadores de los recursos, por lo que la generación de metadatos debe ser simple y directa al grano tomando en cuenta que el propósito de sistemas de metadatos es apoyar las operaciones a efectuar sobre un recurso en particular y no crear descripciones extremadamente completas de éste. Las personas que efectúan marcado de datos van generando con el tiempo habilidades especializadas, generándose cultura de marcado con metadatos.

## **6 Beneficios a la hora de utilizar metadatos**

Los beneficios de utilizar metadatos son diversos y dependen del área en que se utilicen. En términos generales:

Los metadatos adhieren contenido, contexto y estructura a los objetos de información, asistiendo de esta forma al proceso de recuperación de conocimiento desde colecciones de objetos.

Los metadatos permiten generar distintos puntos de vista conceptuales para sus usuarios o sistemas, y liberan a estos últimos de tener conocimientos avanzados sobre la existencia o características del objeto que describen. Estos puntos de vista conceptuales pueden depender del sistema o usuario que los utiliza, así por ejemplo un niño puede visualizar la información de una página en particular de acuerdo a sus necesidades (gracias a los metadatos de la página y perfil del niño)

Los metadatos permiten el intercambio de la información sin la necesidad de involucrar el intercambio de los recursos mismos. Esta particularidad facilita entre otras cosas las búsquedas sobre colecciones distribuidas. Además los metadatos permiten una descripción precisa y discreta de los recursos permitiendo la creación de colecciones virtuales de descripciones donde agrupan los objetos de información para satisfacer requerimientos específicos. Un ejemplo podría ser una institución educacional que recolecta materias de cursos desde distintas instituciones del globo agrupadas por temas, sin importar el formato del material recolectado.

En cada proceso productivo, o en cada etapa del ciclo de vida de un objeto de información, se van generando metadatos para describirlos, y metadatos para describir dichos metadatos (manual o automáticamente). Éstos generan valor agregado al sistema de metadatos y objetos que describen, generando datos desde los cuales es posible extraer conocimiento relevante al sistema mismo y sus procesos.

Los metadatos permiten un acceso a los recursos en forma controlada ya que se conoce con precisión el objeto descrito. Es posible entonces establecer sistemas de filtrado y permiten generar bases para una autenticación y mecanismos para definir grados de confianza sobre las fuentes de información.

Los metadatos permiten preservar los objetos de información permitiendo migrar (gracias a la información estructural) sucesivamente éstos, para su posible uso por parte de las futuras generaciones. La información semántica de los objetos se mantiene, disminuyendo así la pérdida del conocimiento.

Los metadatos son esenciales para sostener un crecimiento de una Web a mayor escala, permitiendo búsquedas e integración del conocimiento desde un mayor número de fuentes heterogéneas.

## **7 Estándares de metadatos**

Un estándar de metadatos es una colección de palabras clave o estructuras que describen distintos conceptos, generalmente cubriendo algún campo del conocimiento que sea estable y no muy grande. Estos conceptos pueden ser usados para describir declaraciones estructuradas dentro de este campo del conocimiento de tal forma que se puedan capturar los aspectos más importantes de su significado.

La generación de estándares de metadatos es una inversión en cuanto a la futura interoperatividad ya que expande las posibilidades de las distintas partes para trabajar efectivamente en el largo plazo, sin importar el cambio de tecnología.

La estandarización de metadatos en la web es en general muy dificultosa, e incluso se genera inconsistencia de metadatos sobre objetos del mismo dominio, el siguiente es un ejemplo breve para ilustrar este punto.

Si catalogamos un pequeño grupo de canciones, e-mails y especímenes biológicos, utilizando una relación persona-objeto, podemos decir que una canción tiene un 'autor' y un 'título' así como una e-mail tiene un 'from' y 'subject'. Siguiendo la misma lógica también podemos decir que un espécimen biológico tiene un 'recolector' y un 'nombre de organismo'. Ahora bien, si una persona busca algo en este grupo de datos (con la relación persona-objeto en mente) es muy posible que encuentre lo que busca, sin embargo esta relación persona-objeto no contempla detalles como el de que un 'recolector' no es quien crea el 'nombre de organismo' al contrario de cómo pasa con un e-mail.

En materia de metadatos las comunidades no acuerdan consenso para establecer criterios y estándares, lo que es lógico ya que existen innumerables formas de organizar objetos. Hasta el día de hoy ningún estándar ha logrado aceptación global, lo que bajo ciertos puntos de vista es una ventaja.

## 8 Clasificación de los estándares de metadatos

Para facilitar la comprensión global de los metadatos existentes es necesario clasificarlos. La clasificación sugerida se realiza mediante grupos o categorías de acuerdo a los propósitos generales de cada marco de metadatos.

Tipo	Definición	Ejemplos
Administrativos	Metadatos utilizados para el manejo y administración de los recursos de información	<ul style="list-style-type: none"><li>• Registros sobre la adquisición</li><li>• Requerimientos de acceso legal</li><li>• Información espacial de los recursos a administrar</li><li>• Control de versión</li><li>• Seguimiento de los usuarios</li><li>• Reutilización de contenidos</li><li>• Condición física de un recurso</li></ul>
Descriptivos y Descubrimiento	Metadatos utilizados para describir, descubrir o identificar recursos de información	<ul style="list-style-type: none"><li>• Metadatos de apoyo para la recuperación de la información</li><li>• Índices especializados</li><li>• Información sobre el esquema de metadatos utilizado</li><li>• Taxonomías</li></ul>
Técnicos, modelos	Metadatos relacionados con cómo un sistema funciona o de cómo se interrelacionan sus componentes	<ul style="list-style-type: none"><li>• Metadatos que describen el funcionamiento de hardware y Software</li><li>• Estructura de formatos digitales</li><li>• Modelos de intercambio de metadatos</li></ul>

### 8.1.1 Administrativos

Se refieren a Información provista para facilitar la administración de los recursos. En este conjunto tienen cabida datos sobre cuando y cómo un objeto fue creado, quien es el responsable de controlar el acceso o registrar su contenido, que actividades de procesamiento fueron efectuados en relación al contenido y que restricciones de acceso o de uso son aplicables. Un ejemplo de éstos son los metadatos utilizados para la preservación que apuntan específicamente a apoyar la retención a largo plazo de los objetos digitales y dependiendo del contexto, a su reconstrucción en caso de pérdida.

### 8.1.2 Descriptivos y de descubrimiento:

Se refieren a la información provista para encontrar, describir y distinguir cada uno de los objetos de información. *Dublin Core* es el ejemplo más claro de este tipo de metadatos.

En esta categoría tienen cabida también los metadatos encargados de describir recursos de dominios específicos del conocimiento. Ejemplos para el campo de las ciencias serían los metadatos de *Darwin Core* que proveen representación para la búsqueda y recuperación de colecciones de historia natural y los pertenecientes al *Data Documentation Initiative* (DDI) el estándar que sirve para describir conjuntos de datos para uso en ciencias sociales

### 8.1.3 Técnicos y modelos

**Técnicos:** Corresponden a los estándares de metadatos relacionados con los elementos que describen como un sistema funciona o debe ser interpretado. Un ejemplo de éstos son los metadatos que describen el formato de alguna imagen digital.

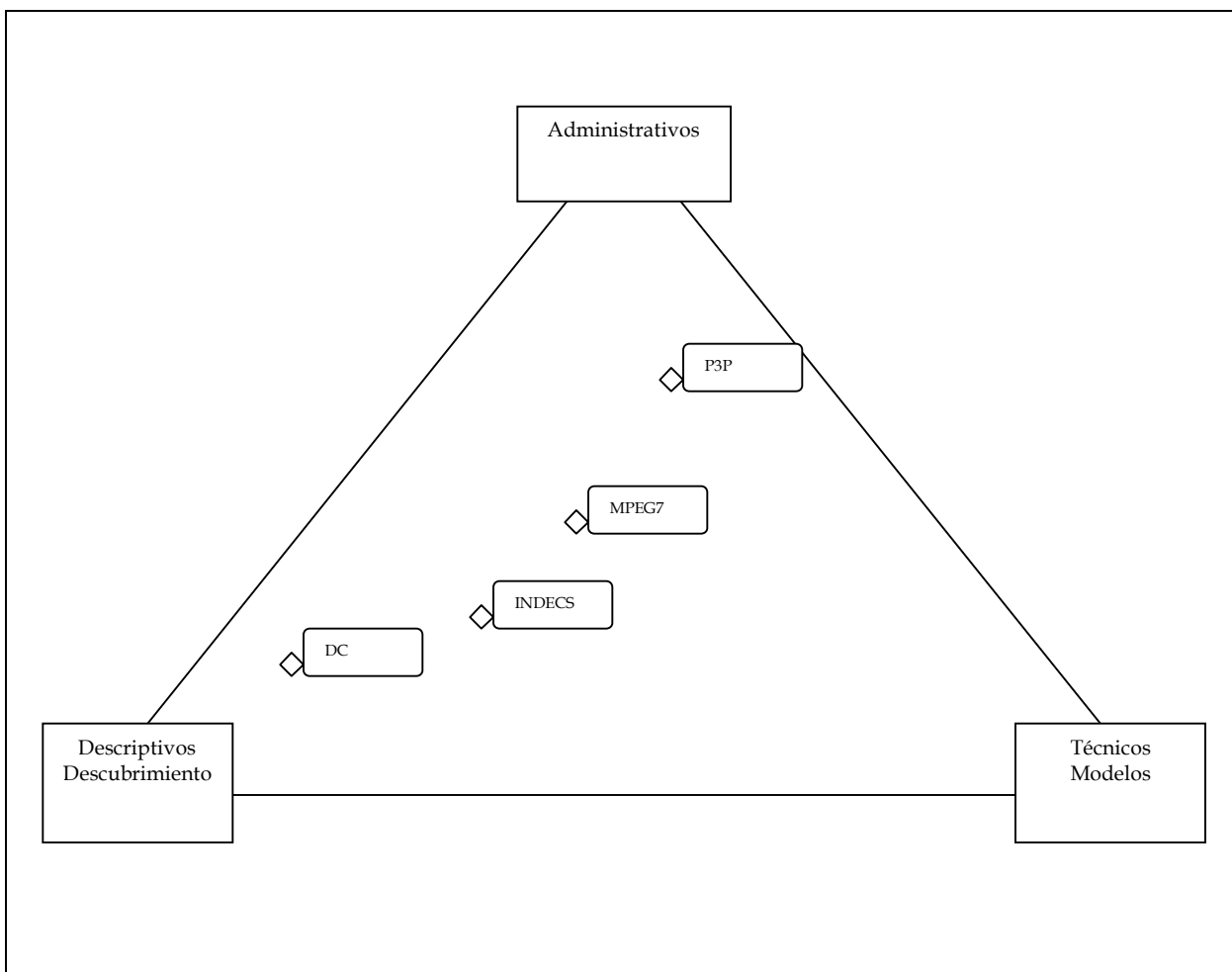
**Modelos:** Tienen relación con las piezas de un objeto de información compuesto, en términos de cómo se interrelacionan cada uno de sus componentes. Por ejemplo un metadato puede describir que, en el contexto de un libro, llegaremos a un tema deseado si seguimos el número de página indicado en el índice y que además las páginas están ordenadas.

Es importante tomar en cuenta que los límites entre estas categorías tienden a ser difusos, por lo que muchos de los metadatos no caben en sólo una de las categorías. Así en un mismo esquema de metadatos se incluyen componentes con distintos propósitos y alcances.

Una clasificación formal en que se agrupan los metadatos en sólo una de estas tres categorías no representa adecuadamente a la realidad, por lo que se puede utilizar un diagrama triangular para visualizar la clasificación.

En el presente diagrama se encuentran clasificados cuatro (de innumerables) estándares, donde un metadato situado cerca una de las categorías indica un mayor número de componentes del que tienen como finalidad cumplir con dicho propósito general.

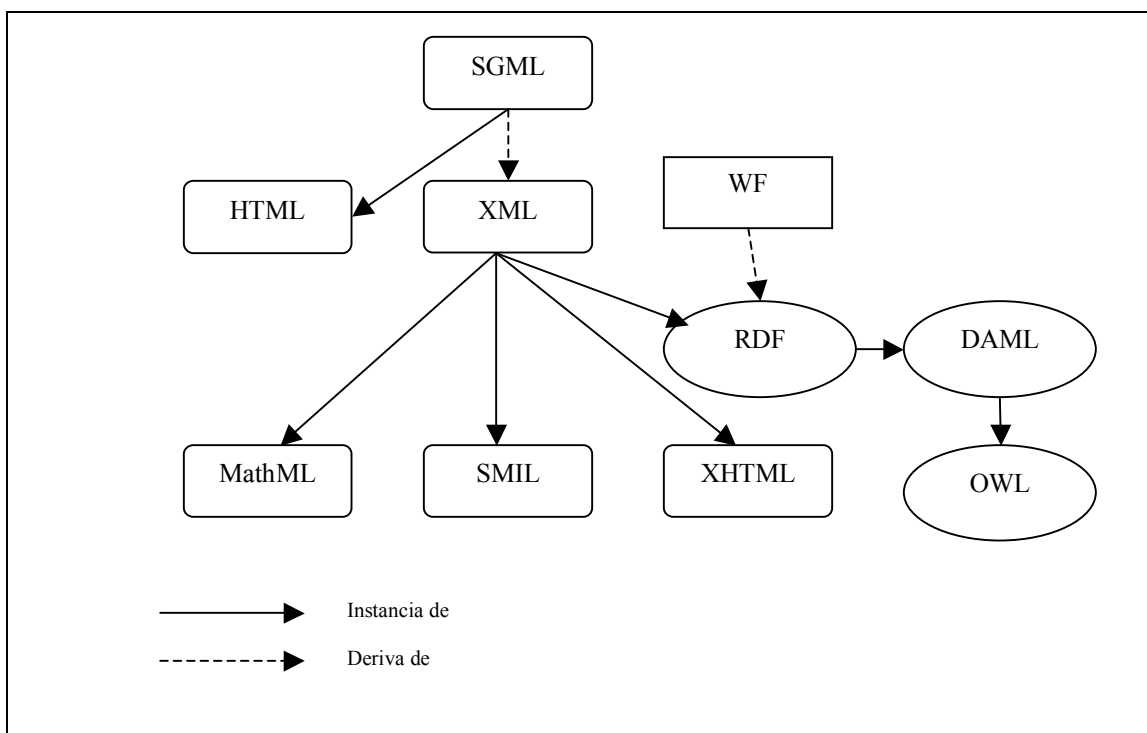
Así es como *Dublin Core* se sitúa (casi) completamente en la categoría de metadatos descriptivos y de descubrimiento, y MPEG7 se sitúa cerca del centro del diagrama por poseer elementos que cumplen con los tres propósitos generales.



## 9 Los lenguajes de metadatos

Para que los metadatos se materialicen es necesaria la existencia de lenguajes que permitan especificar la sintaxis en la que se definen las estructuras, además de proveer medios para las especificaciones semánticas necesarias (que nos digan lo que las expresiones sintácticas significan en términos de un modelo). Estos modelos y sintaxis son las que permiten representar las expresiones, hechos, reglas y consultas sobre las descripciones.

Cada uno de estos distintos lenguajes son derivaciones o instancias de los lenguajes (o esquemas) que los preceden. Como puede observarse en el siguiente diagrama<sup>2</sup>:



<sup>2</sup> Diagrama basado en el esquema del libro Modern Information Retrieval Baeza-Neto, Pág. 174

## **10 Puntos importantes.**

Pueden distinguirse tres aspectos fundamentales en relación a los metadatos en la Web:

### **Es necesaria la identificación:**

Un aspecto muy importante en la descripción de objetos informativos en la Web, es la necesidad imperiosa de identificarlos y de poseer un método para poder acceder a ellos o a descripciones de ellos según sea necesario. A lo largo del estudio no se distinguieron metadatos que no utilizaran URIs (a excepción de metadatos núcleo, que se auto sustentan como DOI), siendo los que permiten el uso de vocabularios controlados y determinan la existencia de los objetos. Para obtener el máximo beneficio en el uso de metadatos es necesario que los identificadores tengan cuatro características: ser único, ser estable y seguro, ser de acceso público y ser persistente. Estas cuatro condiciones son muy difíciles de cumplir, puesto que son necesarias bases sociales para su implantación.

### **Las descripciones son efectuadas mediante sistemas de metadatos:**

La forma en que se generan los sistemas de metadatos en la Web difiere en gran medida a los utilizados en la bibliotecología tradicional. Los metadatos utilizados en la Web no apuntan a realizar descripciones exhaustivas de los recursos (al contrario de los que pasa con los esquemas de marcado como MARC) sino a crear sistemas que utilizan en conjunto los distintos marcos. Los metadatos deben ser en este sentido granulares, siendo una característica necesaria para la subsistencia de estos metadatos. Así es como por ejemplo es común encontrar descripciones realizadas mediante campos Dublin Core e identificadas mediante DOI.

La generación de estos sistemas de metadatos necesita de lenguajes que permitan la agrupación de los distintos componentes. RDF es el candidato principal, por ser un lenguaje de nivel básico que permite la generación de estos sistemas y mediante el uso de DAML (En la actualidad OWL) permitiendo inferir sobre las ontologías.

Dentro de los componentes de una descripción con metadatos, una descripción con Dublin Core es la que ha probado ser la más estable, siendo el candidato a una difusión más extensa, pudiéndose considerar un estándar de propósito general. El grado de estandarización de algunos marcos propuestos para propósitos generales permite un cierto grado de reutilización, sin embargo a nivel de dominios particulares, las comunidades invierten esfuerzos significativos en el desarrollo de estructuras especializadas para satisfacer sus requerimientos, siendo comúnmente disidentes de los sistemas utilizados en otras comunidades del mismo dominio.



La generación de sistemas de metadatos tiene en la actualidad problemas técnicos como una documentación insuficiente de las características de los modelos de los distintos estándares, lo que dificulta la interoperatividad de los sistemas (por ejemplo se dificulta la traducción de los elementos). Además en la actualidad no es común contar con representaciones de los modelos (ontologías) mediante un lenguaje estándar, que permitan entender dichos metadatos.

### **Los metadatos están estrechamente ligados al objeto que describen:**

En los metadatos intrínsecos y extrínsecos, las descripciones de metadatos se encuentran enlazadas estrechamente a las características relevantes de los objetos, generando nuevas dificultades técnicas. Siendo necesario lidiar con los cambios que experimentan los objetos a lo largo del tiempo. En la Web los objetos pueden ser creados dinámicamente, estableciendo dificultades que hasta el día de hoy no han sido resueltas (sin embargo existe tecnología para ello).

### **La barrera:**

La barrera principal en la utilización de metadatos no es tecnológica, dado que se disponen de los elementos para materializar los tres puntos mencionados anteriormente, desde que una URL técnicamente pueda ser utilizada como un identificador persistente (siendo que no lo es) hasta que existan modelos avanzados para la inferencia en disciplinas como de inteligencia artificial. La verdadera barrera para su uso es social, siendo necesario crear una cultura de metadatos. Bajo ciertos puntos de vista esto es posible, apostando a que la evolución de estos metadatos siga un camino similar a la que lo hicieron los catálogos.

### **El contexto:**

En la Web, los metadatos no se gobiernan por autoridades centrales (como una biblioteca nacional), sino que consiste en múltiples grupos, organizaciones y personas trabajando en forma independientemente. Si bien este sistema dificulta el desarrollo de estándares universales, por otro lado entrega la libertad creación y uso suficiente para una explosión de sistemas de metadatos. Se utilizarán entonces lenguajes que permitan esta libertad de creación, como RDF.

Los metadatos pueden ser una evolución en los sistemas de obtención de conocimiento y es muy probable que en el futuro su uso sea masivo, no por el sólo hecho de sus beneficios adicionales, sino porque son esenciales para sostener un crecimiento de una Web a mayor escala. La incorporación de metadatos generará nuevos problemas y desafíos para el futuro.

## 11 Bibliografía

Metadata Principles and Practicalities

Erik Duval et al.

<http://www.dlib.org/dlib/april02/weibel/04weibel.html>

'Setting the Stage', Anne J. Gilliland-Swetland, Introduction to metadata: pathways to digital information, Getty Information Institute, 2000

Metadata: The Future of Information Systems

Keith G jeffery

A Survey of Current Metadata Standards and the Underlying Models'

Ronald Snijder 2001.

<http://www.geocities.com/ronaldsnijder/>

Assessing Metadata Needs for SUL/AIR Digital Collections: A Guide

Nancy Hoebelheinrich

Introduction to Metadata Pathways to digital information y Library Notes Julio/Agosto 1999, número 1286, Harvard University, página 4

Metadata In The World Wide Web

Roberto Galnares

Computer and Information Science Department New Jersey Institute of Technology

<http://www.cs.njit.edu/~galnares/Metadata.html>

About Intellectual Property

World Intellectual Property Organization.

<http://www.wipo.org/about-ip/en>

Extending the Warwick Framework : From Metadata Containers to Active Digital Objects

Daniel, Ron Lagoze, Carl Lagoze (1997)

<http://www.dlib.org/dlib/november97/daniel/11daniel.html>

The organization of information

Arlene G. Taylor

Cataloging Services Department

Stanford University Libraries & Academic Information Resources

<http://www-sul.stanford.edu/depts/catdept/units/metadata/>

Library of Congress Network Development and MARC Standards Office

<http://www.loc.gov/marc/>

Ontologies: Principles, Methods and Applications

M. Uschold, M. Gruninger

<http://www.dcc.uchile.cl/~cguierr/websemantica/uschold96ontology.ps>.

The Semantic Web, A new form of Web content that is meaningful to computers will  
unleash a revolution of new possibilities

Tim Berners-Lee, James Hendler, Ora Lassila

<http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>

## **DC**

Dublin Core Metadata Initiative

<http://www.dublincore.org>

The Dublin Core Metadata Initiative: Mission, Current Activities, and Future Directions  
Weibel, Stuart and Traugott Koch (2000)

<http://www.dlib.org/dlib/december00/weibel/12weibel.html>

The State of the Dublin Core Metadata Initiative April 1999

Weibel, Stuart (1999)

<http://www.dlib.org/dlib/april99/04weibel.html>

Metadata: The Foundations of Resource Description

Weibel, Stuart (1995)

<http://www.dlib.org/dlib/july95/07weibel.html>

A Common Model to Support Interoperable Metadata : Progress report on reconciling  
metadata requirements from the Dublin Core and INDECS/DOI Communities

Bearman, David, Eric Miller

<http://www.dlib.org/dlib/january99/bearman/01bearman.html>

## **URI**

Uniform Resource Identifiers (URI): Generic Syntax

T. Berners-Lee, R. Fielding, L. Masinter

<http://www.ietf.org/rfc/rfc2396.txt>

Guidelines for new URL Schemes, November 1999 RFC2718

<http://www.ietf.org/rfc/rfc2718.txt>

L. Masinter

Registration Procedures for URL Scheme Names

R. Petke

<http://www.ietf.org/rfc/rfc2717.txt>

Hypertext Style: Cool URIs don't change

T. Benders-Lee

<http://www.w3.org/Provider/Style/URI>