

Regina Motz

InCo- Fing

Universidad de la República



El objetivo de esta Unidad es comprender la utilidad de los metadatos para la web y para los sistemas de información en general.



MOTIVACIÓN PARA ESTUDIAR LOS METADATOS

"LOS METAADATOS SON PELIGROSOS Y UTILES A LA VEZ"

[CASO OBAMA-2013] En el año 2013, durante la presidencia de Barack Obama en Estados Unidos salió a la prensa un episodio muy comentado respecto a Metadatos. Te invitamos a leer lo que publicaba al respecto el periódico The Guardian en su artículo del 21 de junio 2013.

NSA surveillance: don't underestimate the extraordinary power of metadata

Puedes encontrar el artículo en el siguiente enlace:

<https://www.theguardian.com/technology/2013/jun/21/nsa-surveillance-metadata-content-obama>

Puedes también ver una opinión sobre el uso de metadatos en ese contexto en el video que encuentras en el siguiente enlace: https://youtu.be/_o2djiZOxyA

[CASO COVID-19] Actualmente también se habla de metadatos en la lucha contra el coronavirus. Te invitamos a ver el artículo en el siguiente enlace:

<https://techcrunch.com/2020/03/27/telco-metadata-grab-is-for-modelling-covid-19-spread-not-tracking-citizens-says-ec/>

Sin duda que la información geocalizada es muy útil, pero para saber que los metadatos pueden decir mucho más que información geolocalizada te invitamos a navegar por esta Unidad.

¡Esperamos que la disfrutes! Recuerda que cualquier duda o comentario puedes enviarlo al foro.

METADATOS

Comencemos analizando la epistemología de la palabra "Metadatos"

La etimología de este término consta de dos palabras, una griega y otra latina. Por un lado la palabra griega "meta", que significa después de o más allá de, y por otro lado el vocablo latino "datum", que significa dato. Así mismo, se forma la expresión metadatos como más allá de los datos. En base a esto, metadatos son un conjunto de datos que describen el contenido informativo de un recurso, de archivos o de información de los mismos. Es decir, es información que describe otros datos. No obstante, la palabra metadatos no tiene una definición exclusiva, ya que son varias las expresiones con las que se conoce, tales como datos sobre datos, informaciones sobre datos, informaciones sobre informaciones o datos sobre informaciones.

Vale la pena detenernos en repasar la distinción entre *datos*, *información* y *conocimiento* con el siguiente video: [DATOS et al.](#)

Observar que tanto la información como el conocimiento en algún momento pueden ser tratados como datos.

¿Cuál es el desafío de trabajar con datos de la web?

Recuperar datos para generar información.

Información es el conjunto de datos disponibles, en cualquier representación (numérica, texto, audio, visual, gráfica, etc.) que cuando combinada por un conjunto de procesos manuales o automáticos, asiste a la toma de decisiones.

Ejemplos de metadatos:

Base de Datos Relacional

Esquema Relacional (METADATOS) Información

PERSONAL (cedula, nombre, celular)

4.235 / 596-2	HU Moreno	0985443215
4.297 / 654-3	Pedro Rodríguez	099045652
3.876 / 543-2	Mauro Sosa	098212345
4.567 / 890-1	Rosario Martínez	099023456
3.765 / 765-7	Silvia Filian	094785972

TRABAJAN (sucursal, cedula, entrada, salida)

Sucursal	Cedula	Entrada	Salida
Montevideo	3.875.543-2	15:00	23:00
Montevideo	4.987.654-3	7:00	15:00
Punta del Este	4.567.890-1	12:00	22:00
Punta del Este	3.765.765-7	22:00	6:00

El procesamiento de datos para generar información dió un salto cualitativo en el momento de implantarse los sistemas de bases de datos relacionales.

En una base de datos relacional se distinguen claramente los datos y el esquema de la base de datos.

El esquema de la base de datos es quien describe la relación existente entre los datos.

El esquema de una base de datos relacional corresponde entonces a **metadatos**.

Ejemplos de metadatos:

Página web

Los metadatos también son las anotaciones que se le colocan a los textos para facilitar su procesamiento. Si observan cualquier página web el marcado HTML que posee consiste en:

- (i) Información visual (tipo y tamaño de letras, colores)
- (ii) Enlaces a documentos con contenido relacionado

Veamos por ejemplo la página web de presentación de la Facultad de Ingeniería <https://www.fing.edu.uy/institucional/presentaci%C3%B3n> (Figura 1).

Inicio / Institucional

Presentación

La Facultad de Ingeniería de la Universidad de la República es la única institución pública de educación superior en ingeniería del Uruguay. Su objetivo es la formación de profesionales capacitados para afrontar las necesidades y demandas del sector productivo, así como contribuir al desarrollo de la sociedad.

Esta Casa de Estudios fue creada el 14 de Julio de 1885 bajo el nombre de Facultad de Matemática y Ramas Anexas, tomando en 1975, por resolución del Ministro de Educación y Cultura, el nombre por el que actualmente la conocemos.

Como parte de la Universidad de la República, a la Facultad de Ingeniería le incumbe, a través de todos sus órganos, en sus respectivas competencias, acrecentar, difundir y defender la cultura; impulsar y proteger la investigación científica y las actividades artísticas y contribuir al estudio de los problemas de interés general y propender a su comprensión pública; defender los valores morales y los principios de justicia, libertad, bienestar social, los derechos de la persona humana y la forma democrático-republicana de gobierno.

Podrá encontrar más información en la [síntesis histórica](#).

Ubicación

La sede central de la Facultad está ubicada en Montevideo en la calle Julio Herrera y Reissig, 565.
Coordenadas: -34.9181706;-56.1665725

Fachada de la Facultad de Ingeniería



Figura 1: Página web de presentación de la Facultad de Ingeniería.

Observando el código de la página (Figura 2) podemos encontrar las etiquetas HTML de formato como:

<head> or <title> y las navegacionales como href. (Además de otras muchas etiquetas en las que nos enfocaremos más adelante)

```
<!DOCTYPE html>
<html lang="es" dir="ltr" prefix="content: http://purl.org/rss/1.0/modules/content/ dc: http://purl.org/dc/terms/ foaf: http://xmlns.com/foaf/0.1/ og:
<head>
  <meta charset="utf-8" />
  <script>var _paq = _paq || [];(function(){var u=("https:" == document.location.protocol) ? "https://www.fing.edu.uy/matomo/" : "https://www.fing.edu.uy
  <meta name="Generator" content="Drupal 8 (https://www.drupal.org)" />
  <meta name="MobileOptimized" content="width" />
  <meta name="HandheldFriendly" content="true" />
  <meta name="viewport" content="width=device-width, initial-scale=1.0" />
  <link rel="shortcut icon" href="/sites/default/files/Logo-1.png" type="image/png" />
  <link rel="alternate" hreflang="es" href="https://www.fing.edu.uy/es/institucional/presentaci%C3%B3n" />
  <link rel="canonical" href="https://www.fing.edu.uy/es/institucional/presentaci%C3%B3n" />
  <link rel="shortlink" href="https://www.fing.edu.uy/es/node/6724" />
  <link rel="revision" href="https://www.fing.edu.uy/es/institucional/presentaci%C3%B3n" />
  <link rel="prev" href="/es/node/39402" />
  <link rel="up" href="/es/node/39402" />
  <link rel="next" href="/es/institucional/autoridades" />

  <title>Presentación | Facultad de Ingeniería</title>
  <link rel="stylesheet" media="all" href="/sites/default/files/css/css_seSmukjQ-00rjguuetaas2_7dx1crzI_gORCy48YcIQ.css" />
  <link rel="stylesheet" media="all" href="/sites/default/files/css/css_Dx5BBD6C2i0qjRtunyJzmVbVrLe9qR6V8l0fXhph4NI.css" />

<!--[if lte IE 8]>
<script src="/sites/default/files/js/js_VtafjXmRvoUGazqzYTA3WjKx9WcWhjP0G4ZnnqRamA.js"></script>
<![endif]-->
<script src="https://use.fontawesome.com/releases/v5.13.1/js/all.js" defer crossorigin="anonymous"></script>
<script src="https://use.fontawesome.com/releases/v5.13.1/js/v4-shims.js" defer crossorigin="anonymous"></script>

</head>
<body class="page-node-6724 page-node-type-book">
  <a href="#main-content" class="visually-hidden focusable">
    Pasar al contenido principal
  </a>

  <div class="dialog-off-canvas-main-canvas" data-off-canvas-main-canvas>
    <div class="page">

      <div class="container">
        <div class="row">
          <div class="col-sm-6">
```

Figura 2. Código correspondiente a la página web de la Figura 1.

Al leer nosotros la página podemos fácilmente centrarnos en el contenido (Figura 3) y podemos encontrar que la fecha de creación de la Facultad de Ingeniería fue el 14 de julio de 1885, del contenido de la página en el texto: "Esta Casa de Estudios fue creada el 14 de Julio de 1885"

Presentación

La Facultad de Ingeniería de la Universidad de la República es la única institución pública de educación superior en ingeniería del Uruguay. Su objetivo es la formación de profesionales capacitados para afrontar las necesidades y demandas del sector productivo, así como contribuir al desarrollo de la sociedad.

Esta Casa de Estudios fue creada el 14 de Julio de 1885 bajo el nombre de Facultad de Matemática y Ramas Anexas, tomando en 1975, por resolución del Ministro de Educación y Cultura, el nombre por el que actualmente la conocemos.

Como parte de la Universidad de la República, a la Facultad de Ingeniería le incumbe, a través de todos sus órganos, en sus respectivas competencias, acrecentar, difundir y defender la cultura; impulsar y proteger la investigación científica y las actividades artísticas y contribuir al estudio de los problemas de interés general y propender a su comprensión pública; defender los valores morales y los principios de justicia, libertad, bienestar social, los derechos de la persona humana y la forma democrático-republicana de gobierno.

Figura 3. Contenido que nosotros podemos leer de la página web de presentación de la Facultad de Ingeniería.

Si es un agente de software es el encargado de procesar los datos del contenido de la página, lo que obtiene en principio es un conjunto de caracteres sin ningún significado. Deberá realizarse un procesamiento de lenguaje natural para identificar oraciones, palabras, sustantivos, verbos, conectores, utilizar diccionarios y corpus de textos hasta llegar a identificar que la posible fecha de creación de la facultad, en el contenido, es el 14 de julio de 1885.

Pero si pudiéramos que a los caracteres "14 de julio de 1885" de la oración "Esta Casa de Estudios fue creada el 14 de Julio de 1885" se los anotara con metadatos que describieran que esos caracteres corresponden a "fecha de creación" de esta forma:

" Esta Casa de Estudios fue creada el <fecha de creación> 14 de julio de 1885 <\ fecha de creación >."

¿Qué ventajas/desventajas crees que obtendríamos?

Realiza la **actividad individual 2.2** para expresar tu punto de vista.

¿CÓMO DEFINIMOS METADATOS EN SISTEMAS DE INFORMACIÓN?

En el contexto de los sistemas de información, la definición más útil es la de:

Metadata es

**Información estructurada sobre un recurso de información
que puede ser usada para dar soporte a distintas operaciones.**

Aclaraciones sobre algunos términos usados en la definición de Metadatos

La información no estructurada es generalmente descripciones en lenguaje natural sin formato definido. La información semi-estructurada es aquella donde la definición de la estructura de la información sirve también para describir a la propia información, es el caso típico de los archivos html. Se habla de **información estructurada** cuando es posible distinguir claramente entre la estructura de la información y la información en sí como es el caso de las tablas relacionales del modelo relacional de bases de datos.

Especificar que los Metadatos deben ser estructurados apoya su facilidad de procesamiento automático.

Los **recursos de información** comprenden, en sentido amplio: **datos, operaciones y relaciones**. El concepto genérico de dato se aplica tanto a datos simples como el nombre de calles hasta datos complejos como una factura, un expediente, una historia clínica, un web-service o un software. El concepto de operaciones se aplica a las acciones que se pueden realizar en el sistema de información, por ejemplo la administración de workflows o la integración semántica de datos. Por último, el concepto genérico de relaciones se aplica a las relaciones entre diferentes datos como por ejemplo una relación llamada “vivienda” entre una persona y una dirección. También se aplica a relaciones semánticas entre diferentes conceptos, en donde se puede llegar a determinar que “cliente” para un organismo es lo mismo que “paciente” o “contribuyente” para otro.

Una característica extra de los recursos de información que se encuentran en la web es que son generados por distintos autores/instituciones/organismos en forma autónoma. Esta característica de autonomía en la creación de los recursos de información hace que aparezcan heterogeneidades en todo su proceso de publicación en la web. Estas heterogeneidades hacen imprescindible que para poder descubrir e integrar estos recursos de información se utilicen metadatos estandarizados de acuerdo a su dominio de aplicación.

Al especificarse en la definición de Metadatos que éstos pueden ser usados para dar soporte a distintas operaciones, se hace referencia al objetivo final de los metadatos. El volumen de información estructurada que se puede asociar a cualquier recurso de información no está limitado, pero se considerará solo aquellos que pueden ser usados para dar soporte al trabajo con los recursos de información (donde el trabajo puede ser: identificar, recuperar, extraer, integrar, evaluar, entre otras muchas actividades).

Cualquier recurso de información tiene como mínimo tres características: el contexto de su existencia, información de contenido y la forma en que ese contenido se relaciona entre sí o con otros recursos de información. Esquemáticamente se puede ver en la Tabla 1 algunas de las

preguntas que buscan satisfacer los metadatos en estos tres aspectos.

Contexto	Para qué se creó Cómo se creó Quién lo creó Cuándo se creó
Contenido	Sobre qué es el objeto o qué contiene
Estructura	Conjunto de relacionamientos dentro del objeto o con otros.

Contexto

Para qué se creó
Cómo se creó
Quién lo creó
Cuándo se creó

Contenido

Sobre qué es el objeto o qué contiene

Estructura

Conjunto de relacionamientos dentro del objeto o con otros.

Clasificación de metadatos:

Según el objetivo operacional de los metadatos se acostumbra clasificarlos en: descriptivos, estructurales o administrativos.

- **Metadatos descriptivos** son aquellos que describen qué es el objeto y cuál es el contenido del recurso de información. Generalmente se asocian a las operaciones de descubrimiento, recuperación e integración de información. Ejemplos de metadatos descriptivos de un recurso de información son: autor, fecha de creación, clasificación, asunto.
- **Metadatos estructurales** son los que se centran en describir el conjunto de relacionamientos dentro del recurso de información o con otros recursos. Por ejemplo, en el caso de documentos complejos formados por diversos archivos, los metadatos estructurales describen las relaciones existentes entre esos archivos como pueden ser expedientes electrónicos con actuaciones o correos electrónicos con archivos adjuntos.
- **Metadatos administrativos** son los que describen el contexto del recurso de información y son los que auxilian en la gestión, la preservación y los derechos de los recursos de información. Ejemplos de metadatos administrativos son: indicador de versión del recurso, fecha de generación, fecha de validez, derechos de acceso (confidencialidad, propiedad intelectual).

Generación de Metadatos

Independiente del tipo de metadatos (descriptivos, estructurales o administrativos) para que sean un apoyo al intercambio y la integración de datos es necesario que la definición de los metadatos sea una tarea de consenso en los dominios de aplicación. Surgen así los Comités de Definición de Metadatos por dominio específico.

Mira ahora con un poco más de detalle el ejemplo de la base de datos relacional EJEMPLO 2-.BASE DE DATOS

EJEMPLO 2: BASE DE DATOS RELACIONAL

Observar que el esquema de la base de datos relacional dice a través de la definición de las tablas relacionales descripciones de las "entidades" o "recursos de información" como en este caso lo son "PERSONAL" y "TRABAJAN".

Los atributos de las tablas son identificados por su nombre. Por ejemplo "cédula", "celular".



En el contexto de Uruguay entendemos la semántica de "cédula" como la de identificador nacional de identidad de los ciudadanos y usamos comúnmente el término "celular" para referirnos al número personal de teléfono móvil.

Un tema un poco más complejo es la semántica implícita en el dominio del negocio ¿Cómo pudieron ustedes comprender la semántica de los atributos "entrada" y "salida" de la tabla TRABAJAN?

¿Cómo se entenderían "cedula", "celular" y todos los atributos si la empresa necesita en algún momento compartir estas tablas con una filial de Colombia o de España?

¿Cómo se simplificaría el proceso de intercambio de información si se usaran nombres con semántica consensuada en el dominio del negocio! ¡Metadatos estándares!

Mira lo que hay de Metadatos en AGESIC (Agencia de Gobierno Electrónico y Sociedad de la Información y del Conocimiento) en este enlace: AGESIC-METADATOS (<https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/comunicacion/publicaciones/modelo-referencia-metadatos-personas>)

----- FIN EJEMPLO 2 BASES DE DATOS RELACIONALES -----

CALIDAD DE LOS METADATOS

Al definir los metadatos deben tenerse en cuenta ciertos criterios de calidad para su diseño. Son buenos criterios de calidad de metadatos: Completitud, Minimalidad, Consistencia y Accesibilidad.

- **Completitud** se refiere a que los metadatos deben dar soporte a todas las operaciones que se deseen realizar con los objetos de información.
- **Minimalidad** se refiere a que en la definición de los metadatos no deben existir redundancias y contener el número mínimo de elementos para apoyar a las operaciones para las que está definido. (Observar que Dublin Core sólo tiene 15 elementos básicos).
- **Consistencia** se refiere a que la definición de los metadatos no puede permitir la descripción de objetos de información que no sean válidos. (Un claro ejemplo es que no se puede pedir para un hombre el número de partos que tuvo).
- **Accesibilidad** se refiere a que se debe garantizar el acceso a los metadatos a lo largo del tiempo. Es preferible que el acceso sea vía una publicación on line abierta de los mismos.
- **Exactitud** se refiere que hasta donde sea posible se verifique de forma automática la exactitud de los valores de los elementos de los metadatos. Por ejemplo si hay un elemento en el esquema de los metadatos que es "e-mail" los valores para este elemento deben ser secuencias de caracteres que contienen el carácter "@".

¿Qué sucede si por el criterio de minimalidad el standard de metadatos que selecciono para usar no cubre todas mis necesidades?

Las iniciativas de metadatos pueden ser extendidas para ser usadas en aplicaciones específicas de dos formas : por extensión o por creación de un *profile*.

Extensión: Es la adición de elementos a un esquema existente para soportar la descripción de recursos particulares a un grupo o a un tema.

Profile: Son subconjuntos de un esquema que son implementados por un grupo de interés particular. Los perfiles pueden restringir el número de elementos a ser usados, refinar definiciones de elementos y especificar valores que pueden tomar los elementos.

En ambos casos, se deben seguir los lineamientos de la iniciativa que se está usando, tener cuidado con no incluir redundancias ni inconsistencias.

¿Cuánto se puede hacer para seguir el criterio de exactitud?

Usar aplicaciones de referentes, diccionarios y tesauros, pueden ser de gran ayuda. Por ejemplo WordNet (<https://wordnet.princeton.edu/>).

Realiza la **actividad individual 2.3**, que aborda el criterio de exactitud.

INICIATIVAS DE METADATOS

A nivel mundial la definición de los metadatos usados para catalogar documentos está fuertemente soportada por la iniciativa DUBLIN CORE (<https://dublincore.org/>)

Del mismo modo que como presenta AGESIC los metadatos para personas o para empresas,

DUBLIN CORE presenta los metadatos para describir documentos definiendo la Estructura y la Semántica a usar:

Elementos de Dublin Core

<i>Subject</i>	<i>tema relacionado al objeto descripto</i>
<i>Title</i>	<i>nombre del objeto</i>
<i>Author o Creator</i>	<i>responsable por la creación del objeto</i>
<i>Publisher</i>	<i>agente responsable por hacer accesible el objeto</i>
<i>OtherAgent</i>	<i>personas que contribuyeron al contenido del objeto</i>
<i>Date</i>	<i>fecha de publicación</i>
<i>ObjectType</i>	<i>género del objeto</i>
<i>Form</i>	<i>formato de los datos del objeto</i>
<i>Identifier</i>	<i>identifica el objeto de forma única</i>
<i>Relation</i>	<i>indica tipo de relacionamiento con otros objetos</i>
<i>Source</i>	<i>objetos de los cuales el objeto es derivado</i>
<i>Language</i>	<i>idioma relativo al contenido intelectual del objeto</i>
<i>Coverage</i>	<i>ubicación espacial y duración temporal del objeto</i>
<i>Rights</i>	<i>derechos de propiedad</i>
<i>Description</i>	<i>descripción textual del objeto</i>

Ejemplo en Dublin Core

Title="Metadata Demystified"
Creator="Brand, Amy"
Creator="Meyers, Barbara"
Subject="metadata"
Description="Presents an overview of metadata conventions in publishing"
Publisher="NISO Press"
Publisher="The Sheridan Press"
Date="2003-07"
Type="Text"
Format="application/pdf"
Identifier=
 "http://www.niso.org/standards/resources/Metadata_Demystified.pdf"
Language="english"

Varios esquemas de metadatos estandares se han desarrollado para atender a las diferentes disciplinas, tales como:

CDWA : Categories for the Desccription of Works of Arts

FGDC : Geospatial Metadata – Federal Geographic Data Commiter

MPG : Metadata para archivos de video

TEI Text Encoding Initiative- Metadatos para marcar electrónicaente textos (novelas, poesías, obras teatrales) para soportar investigación en las áreas humanísticas.

Existen también varias propuestas de metadatos para el mismo dominio específico. Ejemplo de sólo algunas para salud:

CPT(Current Procedural Terminology),

MESH es un tesauro (vocabulario controlado y jerárquicamente organizado) producido por la Biblioteca Nacional de Medicina. Se utiliza para indexar, catalogar y buscar información biomédica y relacionada con la salud.

UMLS (Unified Medical Language System)

CPT(Current Procedural Terminology),

CSP(CRISP thesaurus),

HCPT (HCPDS Version of Current Procedural Terminology),

LNC (Logical Observation Identifiers Names and Codes terminology),

MEDCIN(medical terminology of symptoms, history, physical examination, tests, diagnoses, and therapies),

PDQ (Physician data query).

SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms) proporciona la terminología general básica para registros de salud electrónicos.

La existencia de varias iniciativas de Metadatos para los mismos dominios lleva a la necesidad de hacer transformaciones (o “Crosswalks”) entre las iniciativas de Metadatos.

Referente a los estandards para publicar datos en la web, es el World Wide Web Consortium (W3C <https://www.w3.org/>) la comunidad internacional encargada de desarrollar estándares abiertos para garantizar el crecimiento a largo plazo de la Web. En su sección sobre Metadatos publican el standard RDF (Resource Description Framework) que lo estudiaremos en la Unidad 4 de este curso.

Algunos repositorios que registran Metadatos :

<http://www.dcc.ac.uk/resources/metadata-standards>

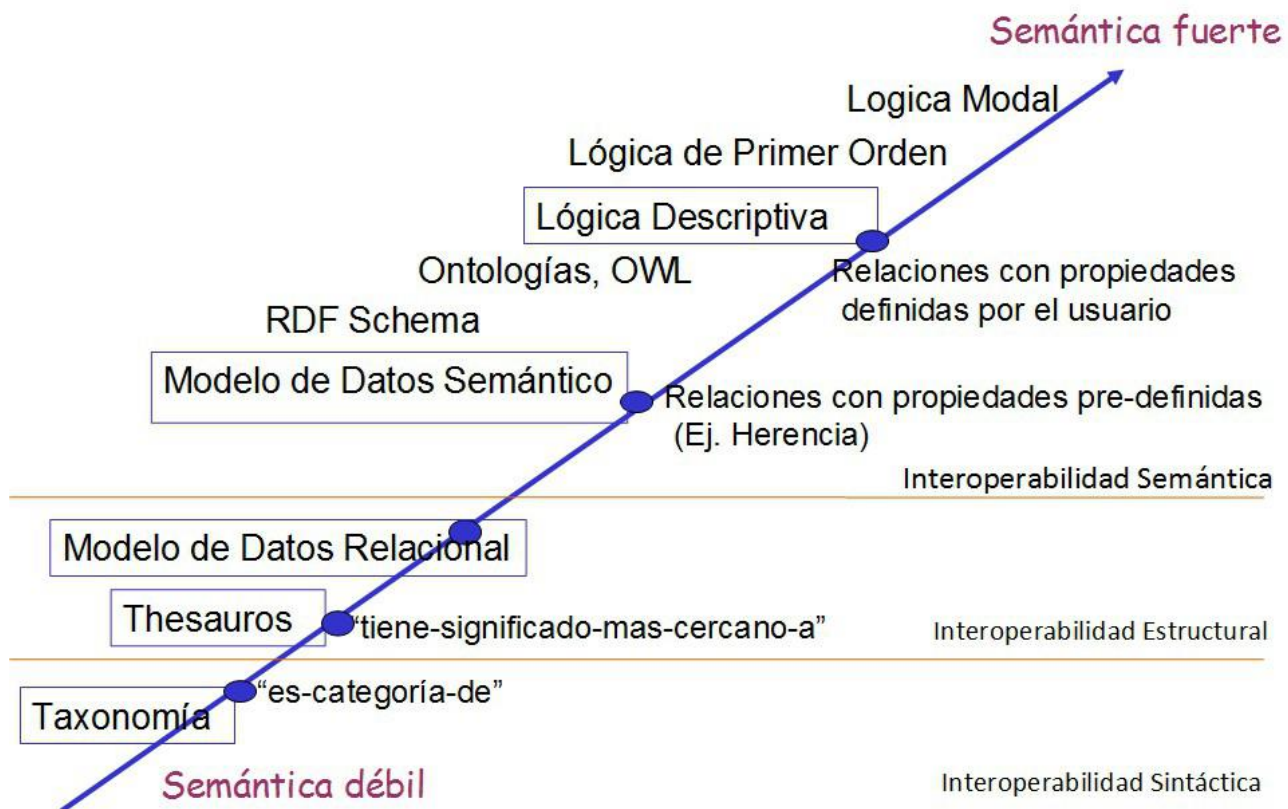
<http://metadataregistry.org/>

Resumen METADATOS

Desde su definición los metadatos persiguen ser apoyo a ciertas operaciones a realizar con los datos. Si nos fijamos en la operación de integrar datos y de interoperar aplicaciones podemos ver que las diferentes representaciones que se le han dado a los metadatos apuntan a resolver distintos niveles semánticos de la interoperabilidad.

Desde el nivel más básico de taxonomías, con semántica muy débil, pasando por modelos de datos conceptuales como pueden ser el modelo de bases de datos relacional o el orientado a objetos, a modelos semánticos como RDF y ontologías basadas en lógica de descripciones y aún siguiendo a modelos basados en lógicas temporal, modal.

Espectro de Metadata



En las siguientes Unidades de este curso caminaremos avanzando por esta gráfica hasta llegar a las ONTOLOGÍAS.

En la próxima Unidad veremos XML desde el punto de vista de Modelo de Datos y lo ubicaremos en la estructura de la WEB SEMANTICA.
