

Módulo 2 Comunicando la calidad: Metadatos

Carlos López Vázquez

carlos.lopez@ieee.org

¿Qué es un metadato?

- Describe el contenido, calidad, condición y otras características de los datos
- Usos principales de los metadatos:
 - organizar y preservar las inversiones en datos de una organización
 - proveer información a catálogos de datos y al ClearingHouse
 - proveer información para facilitar la transferencia de datos

¿Qué es un metadato?(2)

- Responde a las preguntas usuales:
 - ¿Qué?
 - ¿Cómo?
 - ¿Quién?
 - ¿Para qué?
 - ¿Dónde?
 - ...
- No es una redacción...

Tendencias en Metadatos

- Vincular los metadatos con archivos de datos, transferencia y sistemas de respaldo
- Deseo de usar tecnologías de SGBD y WEB para publicar metadatos
- Interés creciente en documentar recursos relevantes no digitales (p.ej. proyectos, servicios, expertise)



Estándares...

- También hay iniciativas para generar Metadatos de modelos matemáticos (software). Ver por ejemplo:
www.negia.ucsb.edu/projects/metadata/standard/standard.doc
Content Standard for Computational Models
Version 1.2
Metadata for Models Work Group
Alexandria Digital Earth Prototype Project
- Disponible en el CD

Introducción al estándar de Metadatos FGDC 1998

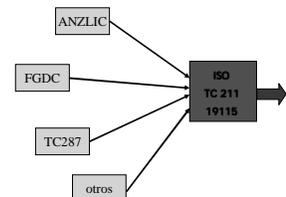
¿Porqué el FGDC 1998?

- Amplia experiencia existente
- Adopción internacional fuera de EEUU
 - Uruguay ☺
 - Australia-Nueva Zelanda
 - México, Canadá, etc.
- Bien documentado, con tutoriales
- Con software específico, gratuito
- La del artillero... (lo conozco en detalle...)

Antecedentes

- Desarrollo del estándar FGDC
 - Junio 8, 1994 Versión 1.0
 - Junio 19, 1998 Versión 2.0
- Desarrollo del estándar ISO
 - Noviembre 2000, Borrador Internacional del Estándar
 - Julio 2003, Versión final del Estándar Internacional
- Otros...

Convergencia



Objetivos (lo que hace)

- Da soporte a usos corrientes de los metadatos
 - inversiones internas - ClearingHouse - transferencia
- Desarrollado con la perspectiva de "¿Qué necesito saber sobre un juego de datos?"
 - disponible - acceso - transferencia - apropiado
- Provee un conjunto común de terminología y definiciones, así como información sobre valores a ser suministrados
- Identifica elementos obligatorios, obligatorios si corresponde y opcionales

Objetivos (lo que no hace)

- El estándar no especifica:
 - la forma de organizar la información dentro de una computadora o sistema
 - los medios para organizar la información para ser transferida
 - los medios a través de los que la información es transmitida, comunicada o presentada al usuario

Decisiones de Implementación

- ¿Qué es un "juego de datos"?
- ¿Cuándo es el mejor momento para generar los metadatos?
- ¿Para quién es todo esto?
 - Gerencia - catálogo - transferencia
 - Detalles, detalles y más detalles
 - El mundo no depende de Ud (¿o sí?)
- Datos preexistentes y el futuro

Aspecto de los metadatos

- Identificación
 - ¿Bibliografía? ¿Arca cubierita? ¿Temas? ¿Actualización? ¿Restricciones?
- Calidad de Datos
 - ¿Nivel de error? ¿Complettitud? ¿Consistencia lógica? ¿Línea?
- Organización espacial de los datos
 - ¿Indirecta? ¿Vector? ¿Raster? ¿Tipo de elementos? ¿Número?
- Referencia espacial
 - ¿Proyección? ¿Sistema de cuadrícula? ¿Datum? ¿Sistema de coordenadas?
- Información de Entidad y Atributos
 - ¿Características? ¿Atributos? ¿Valores de los atributos?
- Distribución
 - ¿Distribuidor? ¿Formato? ¿Medio físico? ¿En línea? ¿Precio?
- Referencia de metadatos
 - ¿Actualización del metadato? ¿Redactor responsable?

Diseción del estándar FGDC 1998

Elementos de la definición

- Secciones
- Elementos compuestos
- Elementos del dato

Secciones

- Constituyen los capítulos principales del estándar
- Están compuestos por:
 - Definición de la sección
 - Lista de elementos, definiciones, tipos y valores
 - Información sobre qué es obligatorio y repetible

Ejemplo de una sección

Información de identificación - Información básica sobre el juego de datos

Tipo: compuesto
Nombre abreviado: idinfo

Información de Identificación =

Cita +
Descripción +
Periodo_Asoado_al_Contenido +
Status + ...

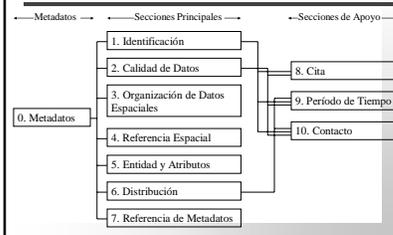
Cita - Información para ser usar para referirse al juego de datos

Tipo: compuesto
Nombre abreviado: citinfo

Descripción - Una caracterización del juego de datos, incluyendo su uso previsto y limitaciones

Tipo: compuesto
Nombre abreviado: descript

Secciones del estándar



Elementos compuestos

- Es un grupo de elementos simples relacionados o de otros elementos compuestos
 - Todos los elementos compuestos están formados en última instancia por elementos simples
- **Formato:**
Nombre del elemento compuesto -- definición
Tipo: compuesto
- **Ejemplo:**
Descripción -- una caracterización del conjunto de datos, incluyendo su uso previsto y limitaciones
Tipo: compuesto

Elemento simple

- Una primitiva lógica de ítems de datos
 - Los elementos simples son los que uno *rellena*
- **Formato:**
Nombre del elemento simple -- definición
Tipo: (elegir entre "entero", "real", "texto", "fecha" u "hora")
Dominio: (lista los valores que pueden ser asignados)
- **Ejemplo:**
Resumen -- una breve descripción del conjunto de datos
Tipo: texto
Dominio: texto libre

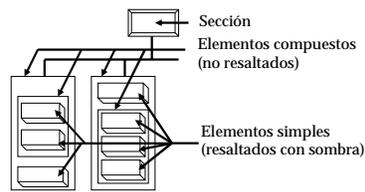
Valores válidos

- El *Dominio* de valores para un elemento simple:
 - puede ser sólo especificado por tipo. En estos casos, se agrega la palabra "libre" (texto libre, entero libre, etc.)
 - puede ser especificado por una lista, referencia a una lista o un rango
 - puede ser parcialmente especificado desde una lista, u opcionalmente ser libre

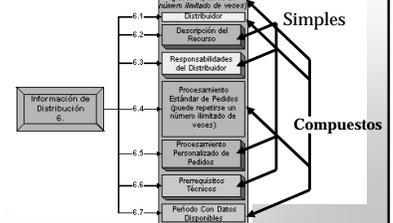
Casos especiales

- El estándar especifica el formato para cuatro casos
 - Fecha
 - Hora del día
 - Latitud y Longitud
 - Direcciones en Internet y nombres de archivos asociados

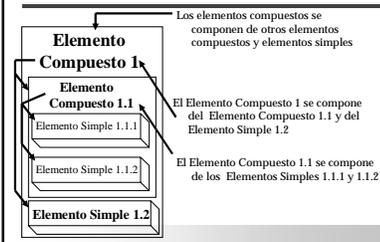
Representación gráfica de los elementos



Ejemplo: Distribución



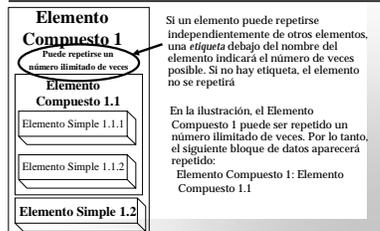
¿Cómo se agrupan?



¿Qué es obligatorio? ¿Qué no?

| Elemento Compuesto | Elemento Simple | Significado |
|--------------------|-----------------|---|
| | | Obligatorio: debe ser completado |
| | | Obligatorio si corresponde: debe ser completado si los datos exhiben la característica definida |
| | | Opcional: suministrado a la discreción del productor de los datos |

¿Qué puede repetirse? ¿Cuántas veces?



Ej.: Gráficos para visualizar

Es opcional, pero si se selecciona los Elementos simples son obligatorios

Otro (buen...) ejemplo: teléfono de contacto

Tomando decisiones al documentar sus datos

Avanzar hacia abajo por la columna

Las decisiones tomadas a niveles altos (cajas exteriores) influyen en las opciones en los niveles más bajos (cajas interiores). Por ejemplo, la Decisión 1.1 influirá en la 1.1.1 y 1.1.2

Las decisiones sobre elementos al mismo nivel son usualmente independientes. La Decisión 1.1 y la 1.2 no se condicionan mutuamente

Usando los colores para tomar decisiones - 1

Elemento Compuesto 1

Elemento Compuesto 1.1

Elemento Simple 1.1.1

Elemento Simple 1.1.2

Elemento Simple 1.2

Todos los elementos están en amarillo, por lo que son obligatorios y deberán ser llenados

Usando los colores para tomar decisiones - 2

Elemento Compuesto 1

Elemento Compuesto 1.1

Elemento Simple 1.1.1

Elemento Simple 1.1.2

Elemento Simple 1.2

El Elemento Compuesto 1 es obligatorio

El Elemento Compuesto 1.1 es opcional. El productor decidirá si desea ingresarlo. Si es así, los Elementos Simples 1.1.1 y 1.1.2 son obligatorios

Si no es así, no se informa sobre el Elemento Compuesto 1 ni los Elementos Simples 1.1.1 y 1.1.2, pasándose directamente a completar el Elemento Simple 1.2

El Elemento Simple 1.2 es obligatorio

Usando los colores para tomar decisiones - 3

Elemento Compuesto 1

Elemento Compuesto 1.1

Elemento Simple 1.1.1

Elemento Simple 1.1.2

Elemento Simple 1.2

El Elemento Compuesto 1 es obligatorio

El Elemento Compuesto 1.1 es obligatorio

El Elemento Simple 1.1.1 es obligatorio

El Elemento Simple 1.1.2 es obligatorio si corresponde. El productor decidirá si los datos tienen la característica descrita por él. Si la tiene, el Elemento debe ser descrito. Si no la tiene, no se completa

El Elemento Simple 1.2 es opcional

Usando los colores para tomar decisiones - 4

Elemento Compuesto 1

Elemento Compuesto 1.1

Elemento Simple 1.1.1

Elemento Simple 1.1.2

Elemento Simple 1.2

El Elemento Compuesto 1 es obligatorio si corresponde. El productor decidirá si los datos tienen la característica descrita por él

El Elemento Compuesto 1.1 es obligatorio

El Elemento Simple 1.1.1 es obligatorio si corresponde. El productor decidirá si los datos tienen la característica descrita por él. Si la tiene, el Elemento debe ser descrito. Si no la tiene, no se completa

El Elemento Simple 1.1.2 es obligatorio

El Elemento Simple 1.2 es opcional

Ej.: Información de Contacto

Cómo leer la representación gráfica

Comience en la Sección con el símbolo 1

Vaya a la columna más a la izquierda (2) y baje por ella (3) hasta completarla

Luego muévase una columna a la derecha (4) y baje por ella (5) hasta completarla

Continúe por la gráfica (6) hasta que todos los elementos hayan sido considerados

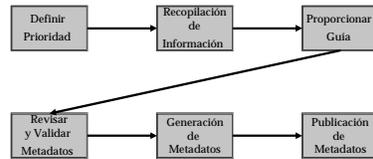
Ejemplo: Sección 1: Identificación

Información de Identificación

Muy lindo... ¿pero cómo lo hago?



¿Qué hacer y en qué orden?



Herramientas para Metadatos

Software para Metadatos

- Para la captura de METADATOS existen varias herramientas como son: MetaLite, CorpsMet95, Tkme, etc.
- MetaLite es un programa que sirve para crear y validar metadatos cubriendo *un conjunto mínimo* de elementos del FGDC.
- CorpsMet95 es otro programa que también permite crear y validar metadatos, cubriendo *todo el conjunto* de campos del FGDC.
- Tkme es un programa que permite crear metadatos, el cual también cubre *todo el conjunto* de campos del FGDC.

Validación de Metadatos: mp

- Es un analizador (Parser) de metadatos, que verifica la sintaxis y crea los archivos de salida.
- Los archivos serán nombrados por un prefijo (ejem.: sidf) con los sufijos .text, .sgml, y .html
 - sidf.text
 - sidf.sgml
 - sidf.html
- MP está disponible para Windows (MS-DOS), Linux y Unix.

Validación de Metadatos

Comando: mp.exe [-e efile] [-t tfile] [-s sfile] [-h hfile] arch_ent

arch_ent: es el nombre del archivo capturado.
efile: archivo donde se graban los errores.
archivos de salida: tfile archivo texto,
sfile archivo sgml,
hfile archivo html.

Ejemplo:

```
mp -e sidf.err sidf.txt
mp -t sidf.text -s sidf.sgml -h sidf.html sidf.txt (MetaLite)
mp -t sidf2.text -s sidf2.sgml -h sidf2.html sidf2.met (CorpsMet95)
```

Presentación de Metadatos

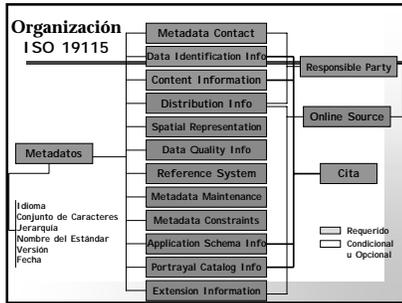
- Las salidas son de varios tipos
 - Versión WEB, para humanos
 - Versión WEB, para humanos expertos
 - Versión SGML, para indizar
 - Versión texto, para desarrollo/validación
- MP permite salidas más o menos bonitas
- Ejemplo:

Aquí va un ejemplo de metadato terminado

ISO 19115 La Próxima Generación

- Se conformó un consenso Internacional sobre metadatos estructurados dentro de un estándar de Geomática más amplio incluido en el Comité Técnico ISO 211 (TC 211)
- ISO 19155 recoge experiencias del FGDC, TC 287, ANZLIC y otros
- También especifica guías para el contenido (vocabulario y estructuras), como el FGDC





Metadatos *Full ISO 19115*

- Incluyen elementos de catálogos y estructuras
- Tienen previsto más detalles que el FGDC en particular para información raster y de imágenes
- En su mayor parte elementos y estructuras condicionales
- Luce familiar para los habituados al FGDC

Núcleo del ISO 19115

- Diseñado para ser usado en catalogar datos para su descubrimiento
- Incluye aproximadamente 50 campos
- Cubre los "quién", "qué", "cuándo", "dónde", "porqué", y "cómo" mejor que el núcleo del FGDC
- Puede no ser adecuado para otros requerimientos más detallados

Comparación FGDC ↔ ISO

- El nivel de detalle de metadatos del FGDC e ISO es comparable
- Los campos obligatorios en el FGDC y el Núcleo ISO son similares, aunque ISO espera más detalle
- FGDC tiene previsto construir herramientas de migración de FGDC V2 a ISO
- *ISO 19115 será conocido como FGDC V3*
Wait and see...

Módulo 2 Comunicando la calidad: Metadatos

Carlos López Vázquez

carlos.lopez@ieee.org