

Diapositiva 1

CURSO DE APRENDIZAJE A DISTANCIA
Intercambio de datos interoperables en hidrología

Lección 2.3 **Los metadatos**

- Definir los metadatos y describir los tipos utilizados en hidrología
- Proporcionar un ejemplo de por qué se necesitan metadatos para compartir datos
- Explicar los elementos y los tipos de metadatos
- Explicar la necesidad de normas sobre metadatos y proporcionar ejemplos
- Definir la semántica; describir ontologías y vocabularios controlados

WEATHER CLIMATE WATER

WORLD METEOROLOGICAL ORGANIZATION | WMO HydroHub | UCAR COMMUNITY PROGRAMS COMET | National Research Council of Italy

Bienvenido a la lección 2.3 — Metadatos. Los objetivos de aprendizaje de esta lección son los siguientes:

- Definir los metadatos y describir los tipos utilizados en hidrología.
- Proporcionar un ejemplo de por qué se necesitan metadatos para poner datos en común.
- Describir los tipos de metadatos y sus elementos.
- Explicar por qué se necesitan normas sobre metadatos y proporcionar ejemplos.
- Definir la semántica y describir ontologías y vocabularios controlados.

Diapositiva 2

La importancia de los metadatos

¡El nivel de agua es de **891 unidades de altura!**



¿Teme que se produzca una crecida?



Los metadatos proporcionan información necesaria para utilizar los datos de forma correcta y responsable

Necesita más información para comprender el significado de los datos



Los metadatos esenciales pueden ser útiles:



- 891 (cm) = 8,91 metros
- Una presa libera agua a diario exponiendo el instrumento a una subida temporal del nivel del agua.
- El limnómetro se encuentra 30 km río abajo y 50 metros por debajo en elevación de la ubicación del propietario.

WORLD METEOROLOGICAL ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"

Lección 2.3: Los metadatos

WMO HydroHub



Empecemos esta lección con un ejemplo sencillo: un propietario local consulta los datos de una estación hidrológica y observa que el nivel del agua en un río es de 891 unidades de altura.

Es una cifra altísima. Este propietario teme que se produzca una crecida. ¿Se trata de un temor razonable?

Una vez que haya reflexionado sobre si los datos reflejan la posibilidad de que se produzca una crecida, ¿a qué conclusión llega?

Puede que haya llegado a la conclusión de que necesita más información, y está en lo cierto. Necesita metadatos para entender qué significa que el nivel del agua esté a 891 unidades de altura. Por ejemplo:

- ¿En qué unidades se puede medir la altura del nivel de agua? Digamos que en centímetros; entonces, el nivel de agua en metros es de 8,91.
- ¿Cuál es la exposición del instrumento de la estación a las influencias externas? Tal vez en la descripción del emplazamiento proporcionada por los metadatos haya observado que una presa aguas arriba libera agua a diario exponiendo el instrumento a una subida temporal del nivel del agua. El momento de esa subida temporal coincide con el de la lectura de 8,91 metros.
- ¿Dónde se encuentra el limnómetro? Los metadatos pueden informarle de que se encuentra 30 km río abajo y a una elevación 50 metros inferior a la de la ubicación del propietario.

Es probable que exista toda esa información, además de otros datos sobre la observación, para describir los datos. Esa información son los metadatos esenciales que permiten a los usuarios utilizar los datos de forma correcta y responsable.

Diapositiva 3

¿Qué son los metadatos y por qué se utilizan?

Los metadatos son los datos que describen a los datos

- Las observaciones que carecen de metadatos presentan un uso muy limitado.
- Los metadatos son necesarios para que los usuarios tengan garantías de que la utilización de esos datos es adecuada para su aplicación.
- Los metadatos pueden ser datos en sí mismos en función de la naturaleza de las necesidades y los objetivos de los usuarios.



los metadatos deberán documentarse y recibir tanta atención como los datos mismos



WORLD
METEOROLOGICAL
ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"

Lección 2.3: Los metadatos



Los metadatos son los datos que describen los datos. En hidrología, se suelen utilizar para describir la variable observada, las condiciones en las que se realizó la observación, el modo en que se efectuó o se clasificó la medición, y el método de proceso de los datos.

La utilidad de las observaciones que carecen de metadatos es muy limitada. Únicamente se puede aprovechar todo el potencial de las observaciones si estas se acompañan de suficientes metadatos.

Los metadatos son necesarios para que los usuarios tengan garantías de que la utilización de esos datos es adecuada para su aplicación.

La línea divisoria entre metadatos y datos no siempre está clara. Los metadatos pueden ser datos en sí mismos en función de la naturaleza de las necesidades y los objetivos de los usuarios. Por ejemplo, un estadístico puede utilizar metadatos sobre la ubicación para analizar la distribución de las estaciones. En este caso, y para este usuario, los metadatos sobre la ubicación son datos, y pueden comprender la latitud, la longitud y la elevación.

Por todo ello, los metadatos deberán documentarse y recibir tanta atención como los propios datos.

Diapositiva 4

Tipos de metadatos

 **Los metadatos de localización** permiten a los usuarios localizar datos y responder a:

- ¿Qué contiene un conjunto de datos?
- ¿Dónde se recopilaron los datos?
- ¿Cuándo se realizaron las observaciones?
- ¿Quién es el proveedor de datos?

 **Los metadatos de observación** permiten interpretar los datos en contexto

- Unidades de medida
- Estado operativo de la estación
- Método de medición y de observación
- Hora base diurna

 **Los metadatos de uso** permiten a los usuarios finales acceder a los datos y utilizarlos

- Políticas de datos
- Endpoint de acceso
- Protocolo para acceder a servicios

 WORLD METEOROLOGICAL ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"

Lección 2.3: Los metadatos

 WMO HydroHub



En hidrología, se pueden encontrar tres tipos principales de metadatos: 1) metadatos de localización; 2) metadatos de observación; y 3) metadatos de uso.

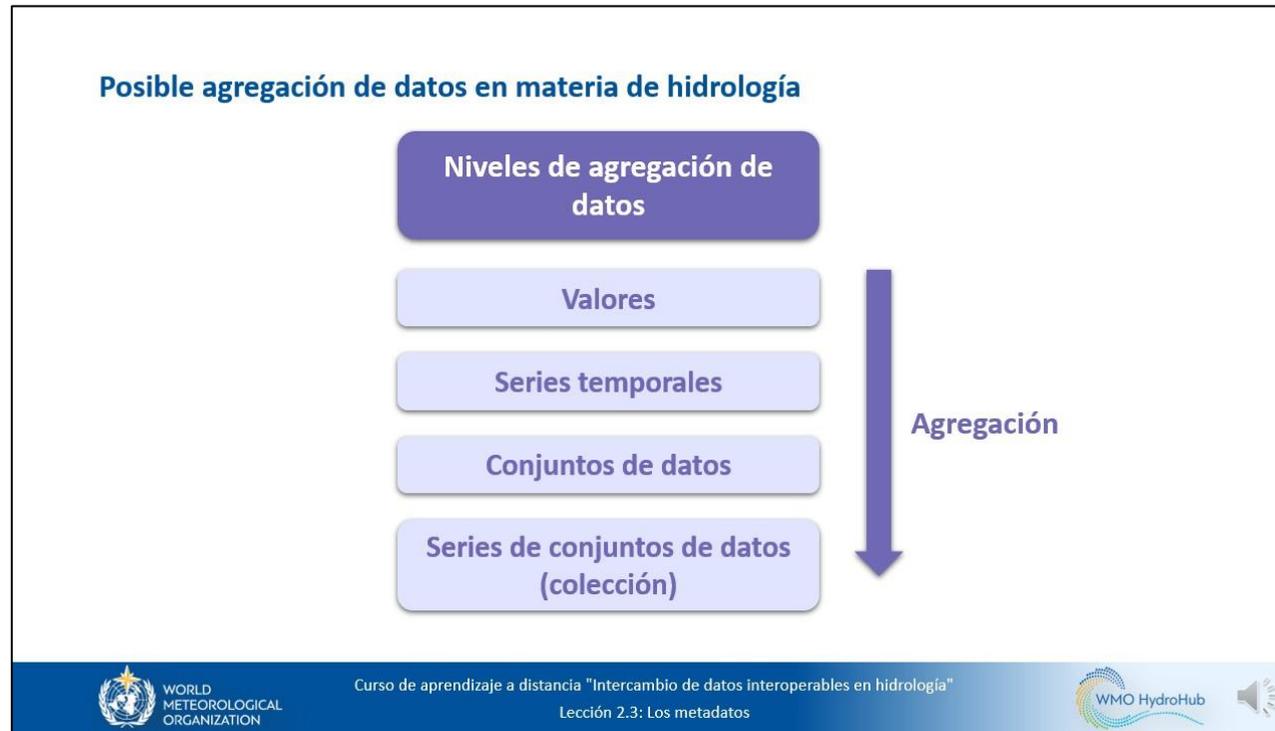
Los metadatos de localización permiten a los usuarios localizar un conjunto de datos y responder a preguntas como las siguientes:

- ¿Qué contiene un conjunto de datos?
- ¿Dónde se recopilaron los datos?
- ¿Cuándo se realizaron las observaciones?
- ¿Quién es el proveedor de los datos?

Los metadatos de observación, también llamados metadatos de interpretación o descripción, permiten interpretar los valores de los datos en contexto. Los metadatos de observación deben describir la variable observada, las condiciones en las que se observó, el modo en que se midió y clasificó, y el método de proceso de los datos. Esto contribuye a que los usuarios tengan garantías de que la utilización de esos datos es adecuada para su aplicación. Estos metadatos comprenden información como las unidades de medida, el método de medición y de observación, el estado operativo de la estación y la hora base diurna.

Los metadatos de uso proporcionan información que permite a los usuarios finales acceder a los datos y utilizarlos. Ejemplos de metadatos de uso son las políticas de datos, el punto extremo de servicios de acceso y el protocolo de servicio.

Diapositiva 5

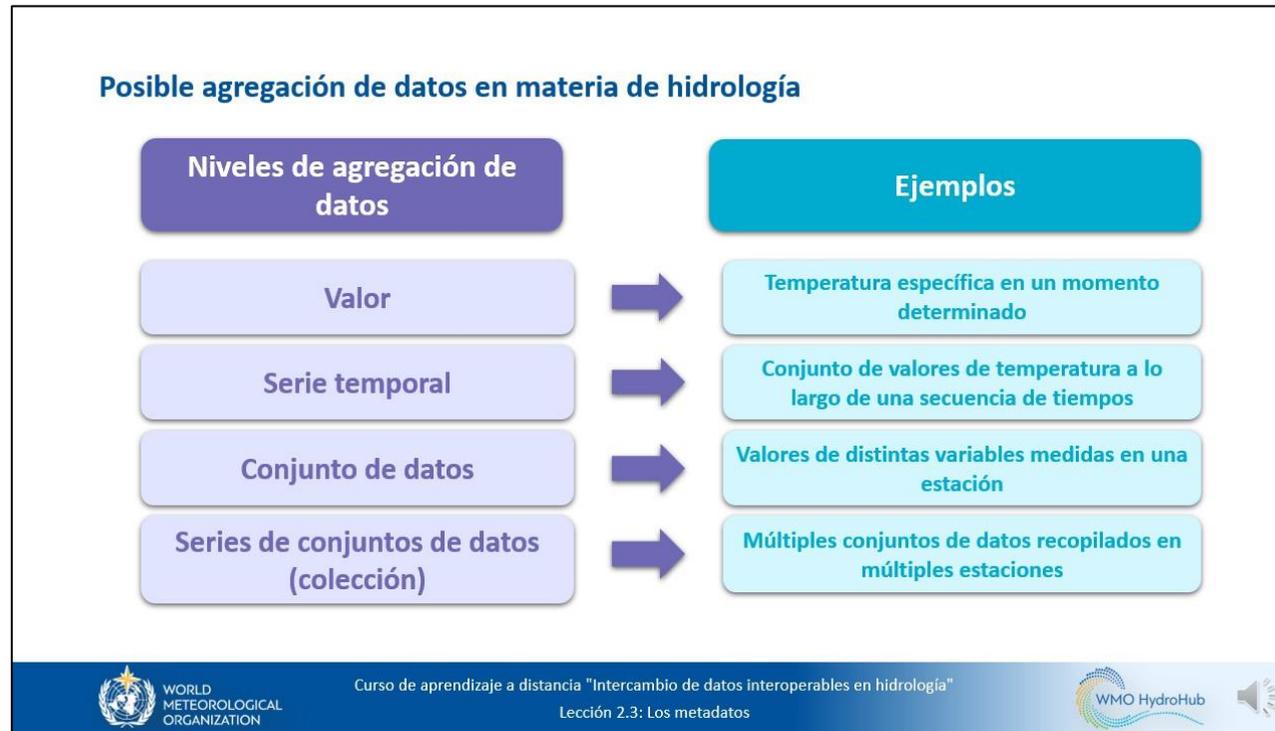


Antes de presentar los diferentes tipos de metadatos y sus elementos, es importante señalar los diferentes niveles de agregación de datos posibles en el ámbito de la hidrología.

Los datos pueden agregarse a distintos niveles:

- de valores simples a series temporales, por ejemplo, series de valores a lo largo del tiempo para los parámetros dados;
- a conjuntos de datos, por ejemplo, agregación lógica de múltiples series temporales de una única estación que mide varios parámetros;
- a series de conjuntos de datos, por ejemplo, todos los conjuntos de datos de varias estaciones de una zona geográfica o cuenca determinada.

Diapositiva 6



He aquí un ejemplo de agregación de datos.

El valor es una temperatura específica en un momento determinado.

La serie temporal es un conjunto de valores de temperatura a lo largo de una secuencia de momentos.

El conjunto de datos es un conjunto de valores de distintas variables medidas en una estación concreta, por ejemplo, temperatura, nivel del agua, precipitación, etc.

Las series de conjuntos de datos están formadas por múltiples conjuntos de datos recopilados en múltiples estaciones.

Para cada nivel de agregación de datos existen tipos de metadatos y, esos tipos de metadatos contienen elementos. Estos tipos y elementos pueden no ser los mismos para cada nivel de agregación, pero deben ser coherentes entre ellos.

Diapositiva 7

Ejemplos de elementos de metadatos de un conjunto de datos

| | Nombre | Definición | Multiplicidad | Dominio | Ejemplo |
|----------------------------|---------|--|----------------------|----------------|---|
| Elementos de los metadatos | Título | Nombre por el que se conoce el conjunto de datos | Es necesario 1 o más | Texto libre | Título = Series temporales de precipitaciones de estación Blue Lake |
| | Latitud | Latitud de la estación en grados decimales | Es necesario 1 | Número decimal | Latitud = 34,63 |
| | ... | ... | ... | ... | |

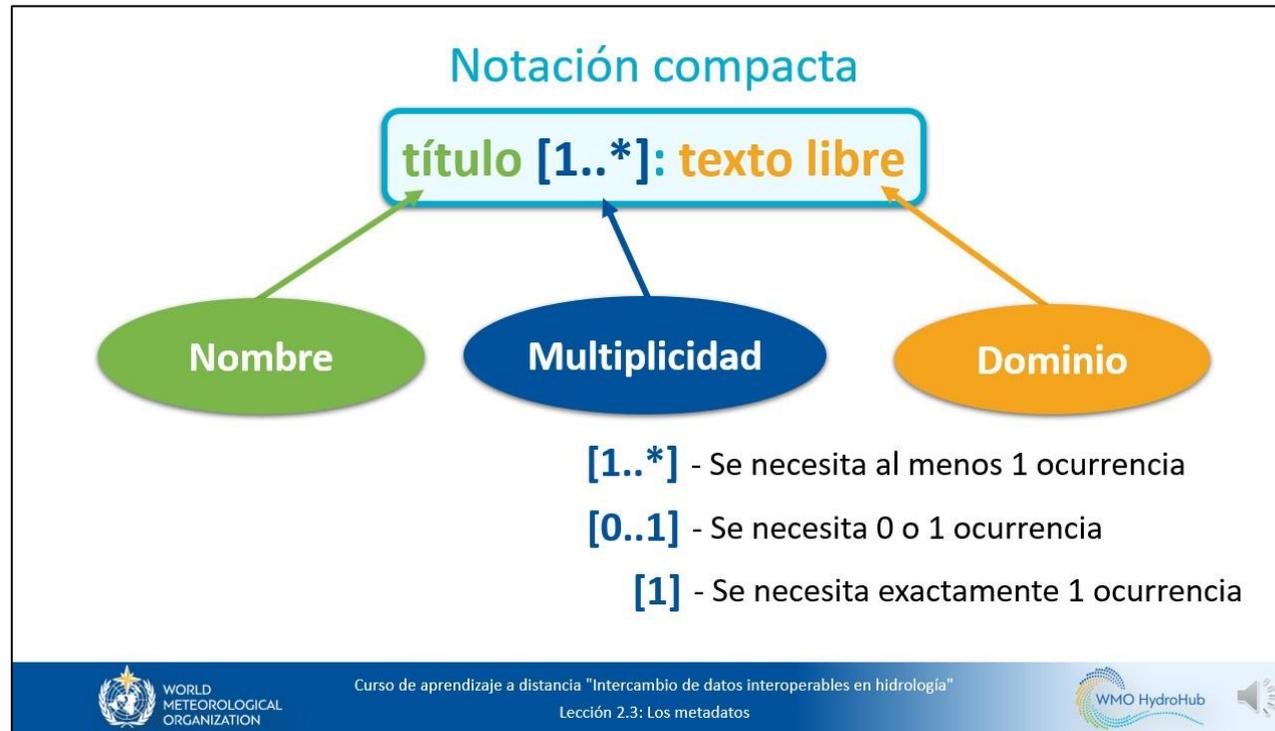
 WORLD METEOROLOGICAL ORGANIZATION Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología" Lección 2.3: Los metadatos  

Un conjunto de datos puede contener elementos de metadatos, por ejemplo, título, latitud, longitud, etc. Cada uno de estos elementos de metadatos está definido por:

- un nombre;
- una definición, esto es, una explicación detallada de lo que documenta el elemento;
- la multiplicidad que define el número permitido de ocurrencias del elemento descrito, por ejemplo, ¿cuántos títulos puede tener el conjunto de datos?;
- el dominio que define un conjunto de valores permitidos, que podría ser la URL, texto libre, un valor numérico, etc.

En las casillas de color verde claro se muestran ejemplos de los valores de elementos como el título y la latitud.

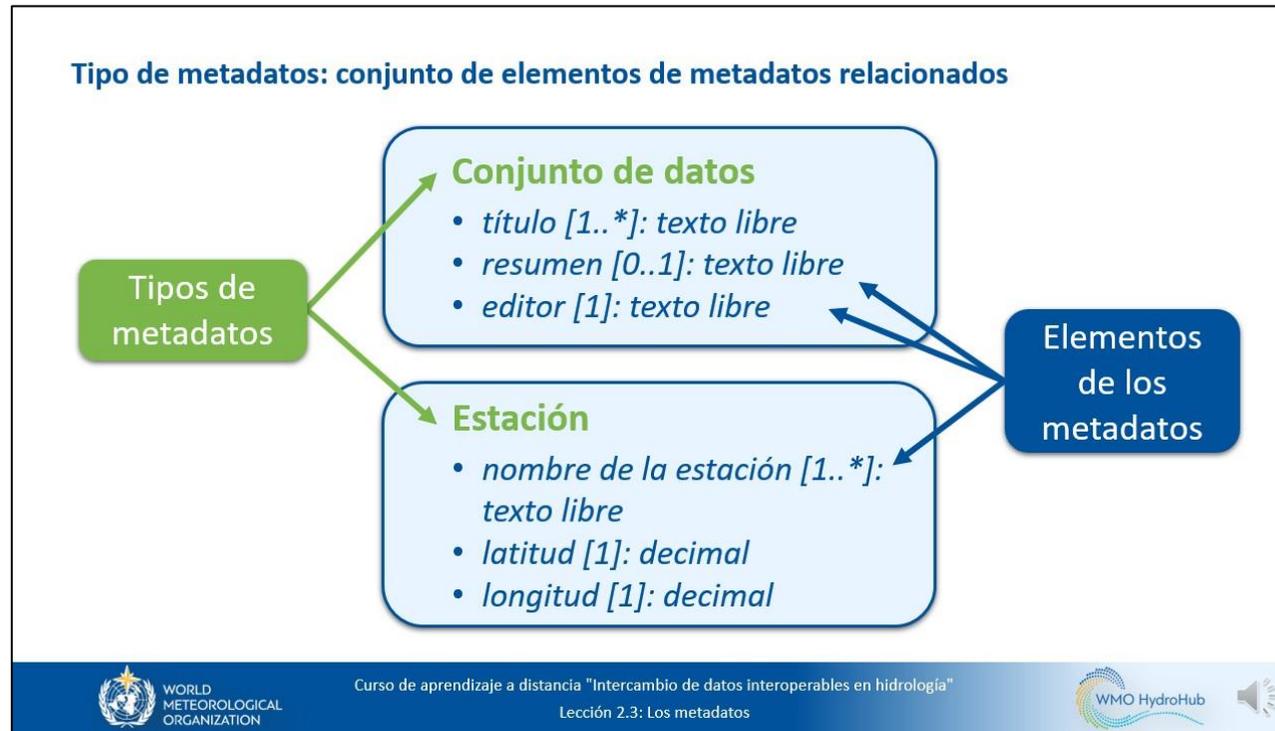
Diapositiva 8



A menudo se utiliza una notación compacta para definir los elementos de metadatos de forma concisa. Como puede observar, la notación compacta incluye el nombre del elemento, el dominio y la indicación de la multiplicidad permitida.

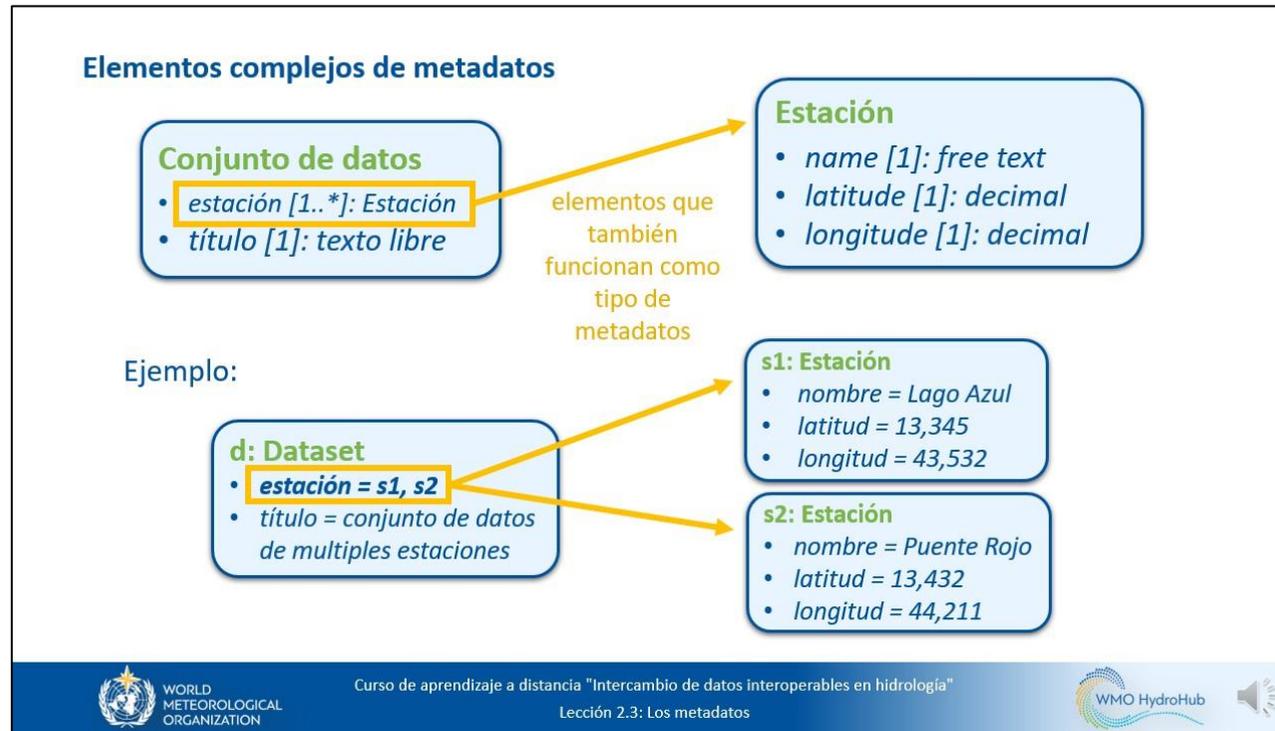
Veamos cómo se interpreta la notación de multiplicidad. [1 punto punto *] nos dice que se necesita al menos una ocurrencia, pero puede haber múltiples ocurrencias, por lo que el asterisco puede ser reemplazado por dos, tres, cuatro, etc. [0 punto punto 1] nos dice que no puede haber más de una, pero es posible que haya 0 ocurrencias. ¿Qué nos dice [1]? Que debe haber exactamente una ocurrencia.

Diapositiva 9



Un conjunto de elementos de metadatos relacionados forma un tipo de metadatos. Por ejemplo, todos los elementos relativos a una estación podrían recopilarse formando un tipo específico. En este ejemplo, los dos cuadros muestran los nombres de los tipos de metadatos "conjunto de datos" y "estación" de color verde, y los correspondientes elementos de metadatos de color azul.

Diapositiva 10



Los tipos de metadatos permiten crear relaciones estructuradas de metadatos mediante el uso de elementos complejos. Los elementos complejos son elementos de metadatos que también pueden servir como tipos de metadatos. Por ejemplo, los elementos de metadatos de un conjunto de datos podrían contener información sobre más de una estación.

Diapositiva 11

Normas sobre metadatos

- Suelen comenzar como esquemas desarrollados por una comunidad de usuarios concreta;
- La comunidad suele almacenar y mantener de forma controlada las especificaciones de los metadatos;
- las especificaciones publicadas suelen conservarse en una ubicación central;
- contienen definiciones semánticas de los elementos y formas normalizadas de codificarlos.



Necesitamos normas sobre metadatos para definir de forma normalizada los distintos tipos de metadatos y sus elementos, así como su codificación en un formato de lectura mecánica.

- Las normas sobre metadatos suelen comenzar como esquemas elaborados por una comunidad de usuarios concreta para conseguir la mejor descripción posible de los tipos de metadatos y sus elementos a fin de satisfacer sus necesidades.
- La comunidad de usuarios suele almacenar y mantener de forma controlada las especificaciones de los metadatos.
- Las especificaciones publicadas suelen conservarse en una ubicación central, por ejemplo, en un documento de referencia en un sitio web o en un registro de metadatos.
- Por lo general, contienen definiciones semánticas de los elementos y formas normalizadas de codificarlos en formatos digitales como bases de datos y lenguaje de marcado extensible (XML).

Diapositiva 12

Normas sobre metadatos

Normalización del modelo de metadatos

- para definir los tipos y elementos de metadatos permitidos, sus significados, dominios, multiplicidad, etc.

Normalización para la codificación de metadatos

- proporciona información sobre cómo traducir el modelo de metadatos a un formato legible por máquinas.



STANDARDS

WORLD METEOROLOGICAL ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"

Lección 2.3: Los metadatos

WMO HydroHub

Existen dos estrategias distintas para normalizar metadatos:

- 1) Normalización del modelo de metadatos para definir los tipos y elementos de metadatos permitidos, sus significados, dominios, multiplicidad, etc. Esta normalización suele ser específica de un tema o una comunidad, por ejemplo, la hidrología.
- 2) Normalización para la codificación de metadatos en un formato de lectura mecánica que permita la localización de metadatos y el acceso a los mismos por parte de las aplicaciones de usuario. Este tipo de normalización proporciona información sobre cómo traducir el modelo de metadatos a un formato de lectura mecánica.

Ambas estrategias pueden aplicarse por separado o siguiendo las mismas normas sobre metadatos.

Diapositiva 13

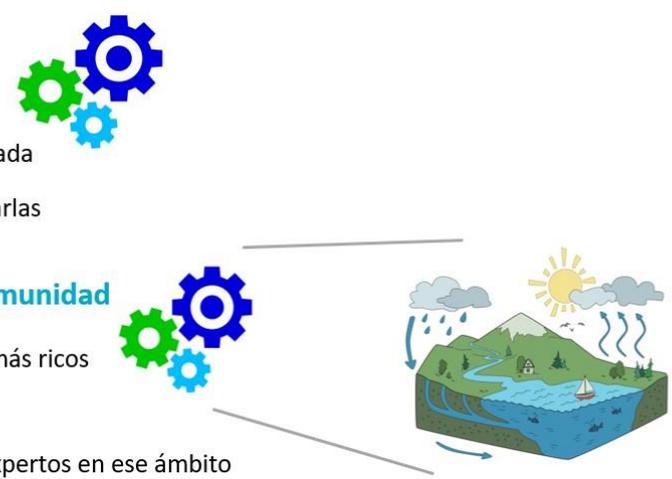
Normas sobre metadatos

❖ **Genéricas**

- ✓ Fáciles de usar
- ✓ Adoptadas de forma generalizada
- ✓ A menudo es necesario ampliarlas

❖ **Específicas de un tema o comunidad**

- ✓ Vocabulario y una estructura más ricos
- ✓ Sumamente especializadas
- ✓ Solo comprensibles para los expertos en ese ámbito



WORLD METEOROLOGICAL ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"

Lección 2.3: Los metadatos

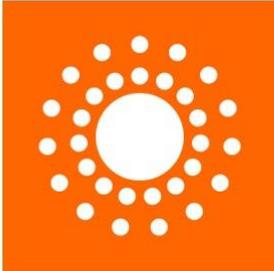
WMO HydroHub

Existe un gran número de normas sobre metadatos. Algunas son genéricas, mientras que otras son específicas de un tema o comunidad. Las genéricas suelen ser fáciles de usar y se han adoptado de forma generalizada pero, a menudo, es necesario ampliarlas para que abarquen información más específica. Las normas específicas de un tema o comunidad poseen un vocabulario y una estructura mucho más ricos, pero tienden a ser sumamente especializadas y solo los expertos de ese ámbito suelen comprenderlas. A continuación presentamos algunos ejemplos de normas sobre metadatos genéricas y específicas de un tema o comunidad.

Diapositiva 14

Dublin Core Metadata Element Set (ISO Standard 15836)

- Puede comprenderse y aplicarse fácilmente;
- consta de 15 elementos necesarios para identificar de forma única un recurso digital;
- puede utilizarse para facilitar la localización de datos y productos;
- no permite la implementación de metadatos de observación específicos de la hidrología.



WORLD METEOROLOGICAL ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"

Lección 2.3: Los metadatos

WMO HydroHub



La norma ISO 15836 define el conjunto de elementos de metadatos Dublin Core. Se trata de una norma básica que puede comprenderse y aplicarse fácilmente y, como tal, es una de las normas sobre metadatos más conocidas. Dublin Core consta de 15 elementos que abordan los elementos descriptivos, administrativos y técnicos más básicos necesarios para identificar de forma única un recurso digital. Actualmente se hace hincapié en la localización de recursos en distintos ámbitos.

En el ámbito hidrológico, Dublin Core puede utilizarse para codificar metadatos de localización con el fin de facilitar la localización de datos y productos. Esta norma no permite la implementación de metadatos de observación específicos de la hidrología.

Diapositiva 15



ISO 19115 — Información geográfica

- Una de las normas de metadatos más conocidas;
- define cómo describir la información geográfica y los servicios asociados;
- proporciona un procedimiento claro para describir los conjuntos de datos geográficos digitales;
- promueve el uso apropiado y la recuperación eficaz de los datos geográficos.





WORLD METEOROLOGICAL ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"

Lección 2.3: Los metadatos



WMO HydroHub



La norma ISO 19115 es una de las normas sobre metadatos más conocidas. Define el modo de describir la información geográfica y los servicios asociados, incluidos los contenidos, las referencias espaciales y temporales, la calidad de los datos, el acceso y los derechos de uso. El objetivo de esta norma internacional es proporcionar un procedimiento claro para describir los conjuntos de datos geográficos digitales, de modo que los usuarios puedan determinar si los datos les serán de utilidad y cómo acceder a ellos. Además, promueve el uso apropiado y la recuperación eficaz de los datos geográficos mediante el establecimiento de un conjunto común de terminología, definiciones y procedimientos de extensión de los metadatos.

Diapositiva 16

**ISO 19115 — Información geográfica**

- Secciones de metadatos obligatorias y condicionales, entidades de metadatos y elementos de metadatos.
- Conjunto mínimo de metadatos necesarios para servir al rango completo de aplicaciones de metadatos.
- Elementos de metadatos opcionales, para permitir una descripción estándar más amplia de datos geográficos, si fuera necesario.
- Método para ampliar los metadatos a fin de satisfacer necesidades especializadas.

WORLD
METEOROLOGICAL
ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"
Lección 2.3: Los metadatos



La norma ISO 19115 contiene cuatro disposiciones fundamentales:

- 1) secciones de metadatos obligatorias y condicionales, entidades de metadatos y elementos de metadatos;
- 2) el conjunto mínimo de metadatos requeridos para prestar servicio a la mayoría de las aplicaciones de metadatos (localización de los datos, determinación de la idoneidad de los datos para el uso, acceso a los datos, transferencia de los datos y uso de datos y servicios digitales);
- 3) elementos de metadatos opcionales para permitir una descripción normalizada más amplia de recursos, si es necesario;
- 4) un método para ampliar los metadatos a fin de satisfacer necesidades especializadas.

Aunque la norma ISO 19115 se aplica a los datos y servicios digitales, sus principios pueden extenderse a muchos otros tipos de recursos, como mapas, gráficos y documentos de texto, así como a datos no geográficos.

Diapositiva 17

Norma sobre metadatos del Sistema Mundial Integrado de Sistemas de Observación de la OMM (WIGOS)

- Elaborado por la OMM;
- describe los **metadatos de observación** necesarios para utilizar de forma eficaz las observaciones de todos los componentes del WIGOS;
- permite determinar las condiciones en las que se realizó la observación o medición, y todos los aspectos que puedan afectar a su empleo o interpretación.



Consulte la página de referencia en el sitio web del curso



Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"

Lección 2.3: Los metadatos



En lo que respecta a las normas específicas de un tema o comunidad, la OMM ha elaborado la *Norma sobre metadatos del Sistema Mundial Integrado de Sistemas de Observación de la OMM* (OMM-Nº 1192), en la que se describen los metadatos de observación necesarios para utilizar de forma eficaz las observaciones de todos los componentes del Sistema Mundial Integrado de Sistemas de Observación de la OMM (WIGOS), incluidas las observaciones hidrológicas. Cualquier usuario debe ser capaz de utilizar los metadatos del WIGOS para determinar las condiciones en las que se realizó la observación o medición, y todos los aspectos que puedan afectar a su empleo o interpretación, es decir, averiguar si las observaciones son adecuadas para su propósito.

La publicación está disponible en línea y se puede acceder a ella a través de la sección de referencias del curso.

Diapositiva 18

| Categorías de metadatos del WIGOS | |
|--|---|
| 1. Variable observada | Especifica las características básicas de la variable observada y de los conjuntos de datos resultantes |
| 2. Objetivo de la observación | Especifica las principales esferas de aplicación de la observación y el programa/s y red/es de observación a los que está afiliada la observación |
| 3. Estación/plataforma | Especifica la instalación de observación, en particular, la estación fija, el equipo móvil o la plataforma de teledetección en la que se realiza la observación |
| 4. Entorno | Describe el entorno geográfico en el que se lleva a cabo la observación |
| 5. Instrumentos y métodos de observación | Especifica el método de observación y describe las características del instrumento o instrumentos utilizados para llevar a cabo la observación |
| 6. Muestreo | Especifica cómo se emplea el muestreo o el análisis para obtener la observación notificada o cómo se recoge una muestra |
| 7. Procesamiento e informe de los datos | Especifica cómo se transfieren los datos brutos a las variables observadas y cómo se informa de ellos a los usuarios |
| 8. Calidad de los datos | Especifica la calidad de los datos y la trazabilidad de la observación |
| 9. Propiedad y política de datos | Especifica quién es el responsable y el propietario de la observación |
| 10. Contacto | Especifica dónde se puede obtener información sobre la observación o el conjunto de datos |

WORLD METEOROLOGICAL ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"

Lección 2.3: Los metadatos

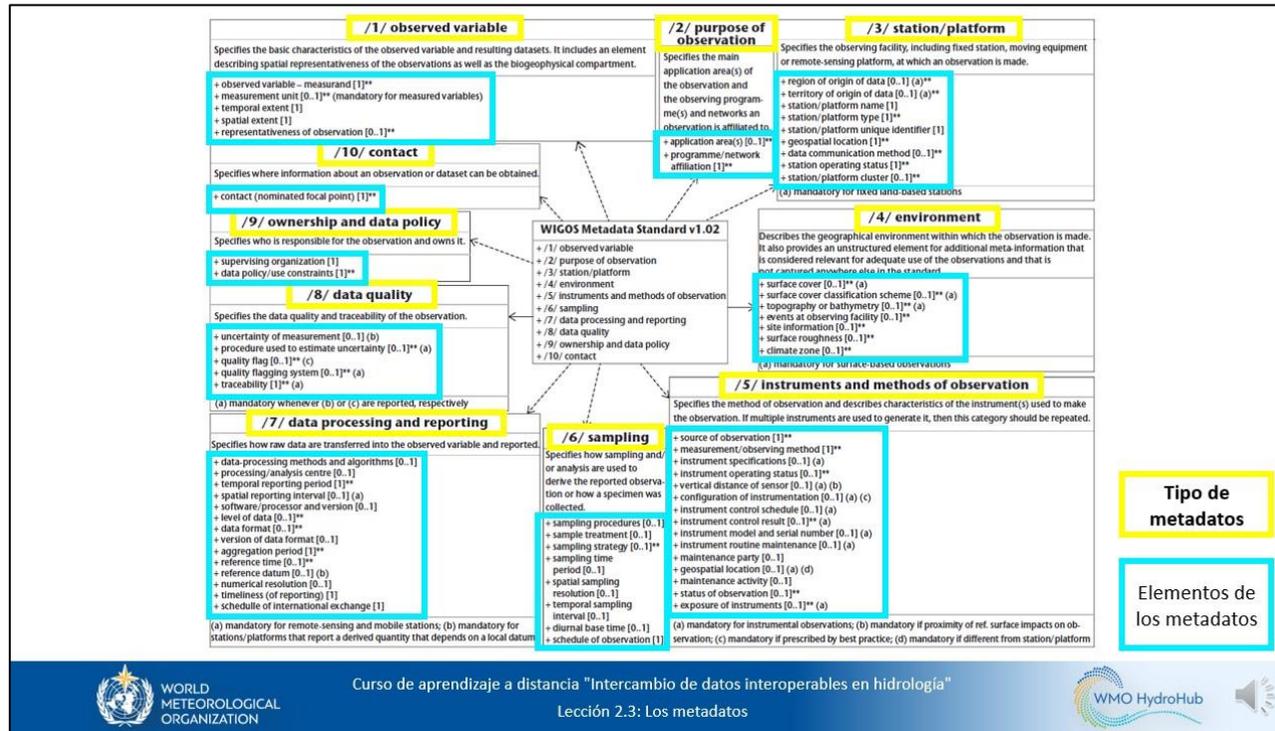
WMO HydroHub

Se han establecido diez categorías de metadatos del WIGOS que definen la norma sobre metadatos del WIGOS. Cada categoría consta de un elemento de metadatos, como mínimo. Se considera que todas las categorías enumeradas son importantes para documentar e interpretar las observaciones realizadas, e incluso para su utilización en un futuro lejano. Las categorías no siguen ningún orden en concreto, pero reflejan la necesidad de:

- indicar la variable observada;
- señalar por qué, dónde y cómo se realizó la observación;
- describir el modo de proceso de los datos brutos;
- evaluar el nivel de calidad de la observación;
- especificar la propiedad;
- definir dónde se puede obtener información sobre la observación o el conjunto de datos.

Estas diez categorías se muestran en la siguiente diapositiva mediante un detallado diagrama en lenguaje de modelización unificado (UML) de la Norma sobre Metadatos del WIGOS.

Diapositiva 19



Este es el diagrama en lenguaje UML de la Norma sobre Metadatos del WIGOS. En cada tipo de metadatos se enumeran sus correspondientes elementos y la multiplicidad definida para cada elemento. Una multiplicidad de [0..1] representa un elemento opcional o condicional. Una multiplicidad de [1] denota un elemento obligatorio.

Diapositiva 20



Los elementos obligatorios, condicionales y opcionales de los metadatos se clasifican de acuerdo con la norma ISO.

Los elementos obligatorios de los metadatos se facilitarán siempre. Los campos correspondientes no podrán dejarse en blanco; deberán contener el valor del metadato o, en casos concretos, la razón por la que no figura ningún valor. La mayoría de los elementos de la Norma sobre Metadatos del WIGOS se consideran obligatorios, habida cuenta de que permitirán utilizar adecuadamente las observaciones en todas las esferas de aplicación de la OMM. En algunos casos, si los Miembros no pueden suministrar un elemento obligatorio, se indicará que el elemento en cuestión no se aplica, es desconocido o no se encuentra disponible como razón por la que no figura ningún valor. La justificación de esta instrucción reside en que se brinda más información al señalar la razón por la que un elemento obligatorio de los metadatos no está disponible que si no se notifica nada al respecto. Los elementos condicionales de los metadatos se notificarán cuando se cumplan las condiciones especificadas, en cuyo caso el contenido de los campos correspondientes nunca se dejará en blanco: se hará constar el valor del metadato o, en el caso de los elementos marcados con el símbolo de almohadilla, la razón por la que no figura ningún valor. Por ejemplo, el elemento "Spatial reporting interval" (Intervalo espacial de notificación de datos) se clasifica como condicional porque solo se aplica a las observaciones de teledetección y las plataformas móviles. Por lo tanto, los elementos de esta categoría deben considerarse obligatorios para los sistemas de observación por teledetección y móviles, pero no para las estaciones terrestres en superficie.

Los elementos opcionales de los metadatos también deberían facilitarse siempre. Proporcionan información útil que puede contribuir a entender mejor una observación. En la Norma sobre Metadatos del WIGOS muy pocos elementos se consideran opcionales. Los elementos opcionales suelen ser importantes para una comunidad en particular.

Se recomienda encarecidamente que consulte las especificaciones detalladas de los elementos obligatorios, condicionales y opcionales de los metadatos del WIGOS que figuran en la publicación *Norma sobre metadatos del Sistema Mundial Integrado de Sistemas de Observación de la OMM* (OMM-Nº 1192) a la que puede acceder desde la sección de referencias del curso.

Diapositiva 21

Principales beneficios de adoptar la Norma sobre metadatos del WIGOS

- Aumentar la longevidad de los datos;
- facilitar el intercambio y la reutilización de datos por parte de distintos usuarios;
- contribuir al uso eficaz de los datos;
- conseguir que los usuarios finales confíen en los datos.



Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"

Lección 2.3: Los metadatos



Los principales beneficios de que los Servicios Hidrológicos Nacionales adopten la Norma sobre Metadatos del WIGOS son las siguientes:

- aumentar la longevidad de los datos;
- facilitar el intercambio y la reutilización de datos por parte de distintos usuarios;
- contribuir al uso eficaz de los datos;
- conseguir que los usuarios finales confíen en los datos.

El panorama mundial del modelado de metadatos

- Enfoques no diseñados con la intención de compartir datos;
- priorizar las necesidades internas por encima de las necesidades de las comunidades de usuarios más extensas;
- no existe una única forma correcta de representar la información;
- el mismo contenido suele representarse de diversas formas;
- las mismas normas sobre metadatos se pueden aplicar de forma diferente.



En la actualidad, distintas organizaciones, e incluso distintos departamentos dentro de las organizaciones, suelen gestionar el modelado de metadatos según cada caso en particular, adoptando enfoques que no están diseñados con la intención de poner datos en común. Estos modelos suelen priorizar las necesidades internas por encima de las necesidades de las comunidades de usuarios más extensas. El hecho de que no exista una única forma correcta de representar la información supone un reto. Este reto se ve agravado por la tendencia de las personas a modelar los datos de forma aislada, pensando en una aplicación concreta. Como resultado, el mismo contenido informativo suele representarse de diversas formas que, por lo general, son incompatibles entre distintos sistemas y organizaciones. Además, las mismas normas sobre metadatos se pueden aplicar de forma diferente en distintas organizaciones debido a la ambigüedad de las normas.

Asimismo, algunos de los retos relacionados con las normas sobre metadatos están asociados a su vez a retos relacionados con la semántica.

Diapositiva 23

La semántica
hace referencia al significado y a la interpretación de las palabras, los símbolos y la estructura de las oraciones

Los distintos proveedores de datos pueden utilizar:

- términos diferentes para los mismos parámetros;
- idiomas diferentes para los mismos términos;
- el mismo idioma y el mismo término, pero el significado puede seguir siendo diferente.

Permite que los metadatos y los datos sean más localizables, accesibles, interoperables y reutilizables

WORLD METEOROLOGICAL ORGANIZATION
Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"
Lección 2.3: Los metadatos
WMO HydroHub

Definamos el concepto de semántica. La semántica hace referencia al significado y a la interpretación de las palabras, los símbolos y la estructura de las frases.

En materia de hidrología, los distintos proveedores de datos pueden utilizar:

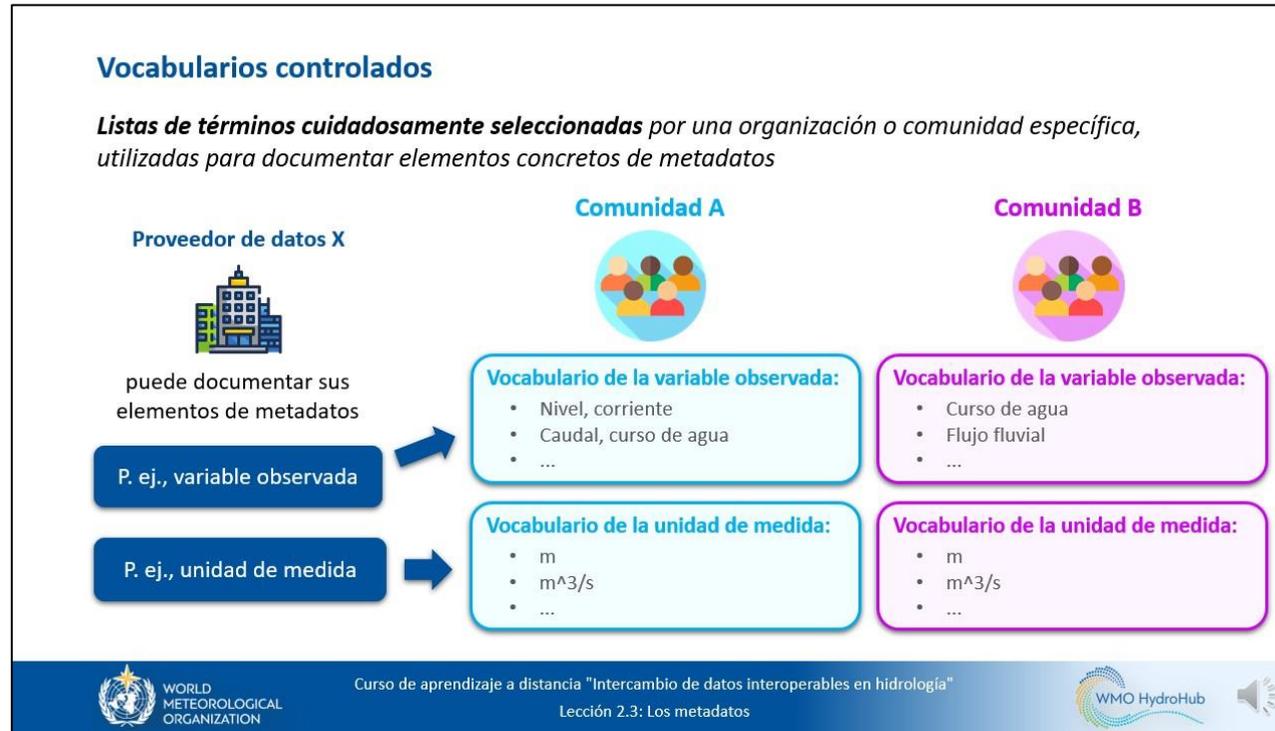
- términos diferentes para los mismos parámetros;
- idiomas diferentes para los mismos términos;
- el mismo idioma y el mismo término, pero el significado puede seguir siendo diferente.

El objetivo de mejorar la semántica es identificar de forma unívoca aquellos conceptos que todo proveedor de datos describe mediante los metadatos. La mejora de la semántica permite que los metadatos y los datos sean más localizables, accesibles, interoperables y reutilizables. A continuación se indican dos maneras importantes de mejorar la semántica:

- 1) utilizar vocabularios controlados;
- 2) implementar ontologías.

En las siguientes diapositivas explicaremos el significado de los vocabularios controlados y las ontologías y proporcionaremos ejemplos.

Diapositiva 24

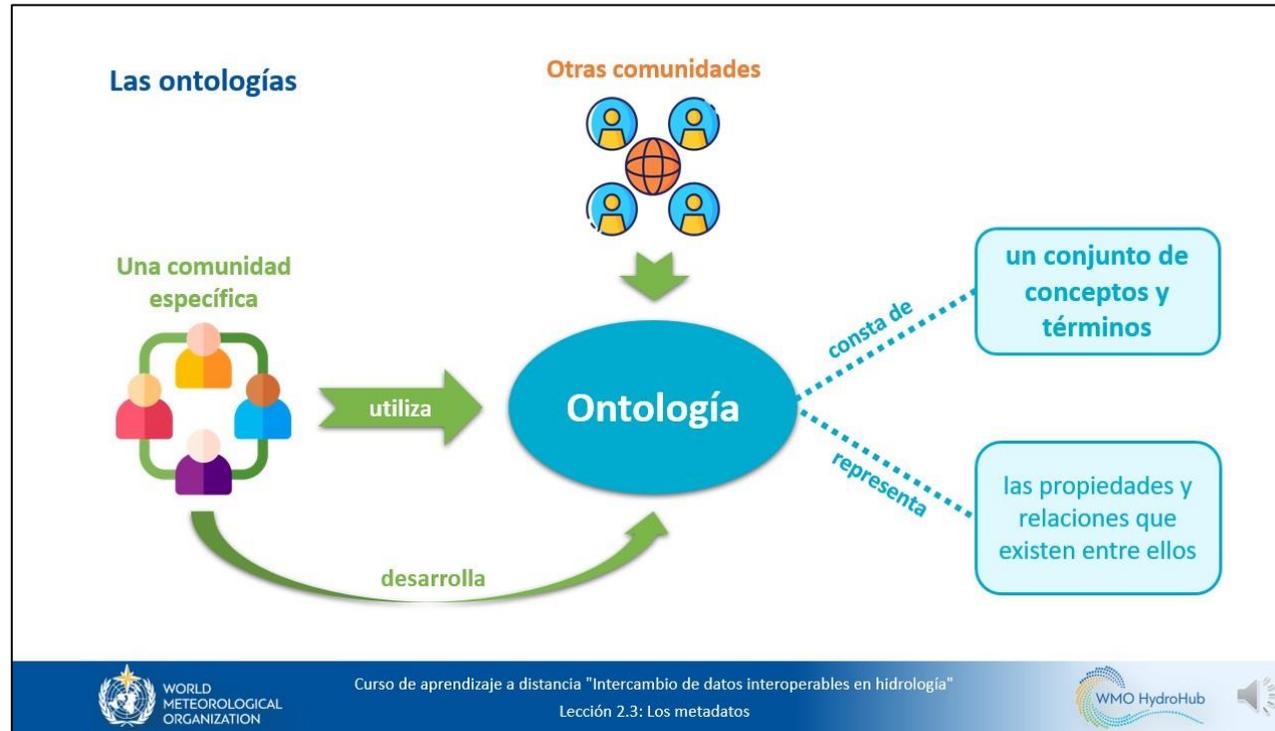


Los vocabularios controlados son listas de términos —palabras y oraciones— cuidadosamente seleccionadas por una organización o comunidad específica, utilizadas para documentar elementos concretos de metadatos.

Por ejemplo, el proveedor de datos X puede documentar sus elementos de metadatos, como una variable observada, usando términos de un vocabulario establecido por la comunidad A. Estos términos serían: “nivel, corriente”; “caudal, corriente”, etc.

Puede que la comunidad B utilice un vocabulario diferente para designar los mismos términos empleados por la comunidad A. La comunidad B podría referirse a “nivel, corriente” como “nivel del agua” y a “caudal, corriente” como “flujo fluvial”. Aunque el vocabulario de parámetros sea diferente en las comunidades A y B, las unidades de este ejemplo son las mismas.

Diapositiva 25



La ontología es un conjunto de conceptos y términos creados y utilizados por una comunidad específica que, además, muestra las propiedades y relaciones que existen entre ellos. Otras comunidades y sus usuarios también pueden beneficiarse del uso de ontologías existentes. Por ejemplo, para documentar elementos de metadatos, los proveedores de datos pueden utilizar conceptos definidos en una ontología específica. Es importante señalar que la norma sobre metadatos que utilice el proveedor de datos debe permitir el uso de ontologías para describir sus elementos de metadatos.

Diapositiva 26

La ontología hidrológica del WHOS

➤ **Conceptos** (de la Ontología CUAHSI)

- **etiquetas de preferencia** (de la Ontología CUAHSI)
- **etiquetas alternativas** (sinónimos, incluso en diferentes idiomas)

➤ **Relaciones entre conceptos** (de la Ontología CUAHSI)

La hidro-ontología está disponible a través de un endpoint SPARQL (lenguaje de consulta semántica) para permitir la interacción entre máquinas



CUAHSI
universities allied for water research

WORLD METEOROLOGICAL ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"

Lección 2.3: Los metadatos

WMO HydroHub

A continuación presentaremos una ontología hidrológica del WHOS —una hidro-ontología— que contiene conceptos pertenecientes al campo de la hidrología, sus propiedades y las relaciones existentes. Esta hidro-ontología puede aplicarse para mejorar el intercambio de datos hidrológicos y la interoperabilidad a diferentes escalas, desde la local hasta la mundial.

La hidro-ontología consta de:

- 1) conceptos de la ontología del Consorcio de Universidades para el Progreso de la Ciencia Hidrológica (CUAHSI) con sus etiquetas de preferencia, que se han complementado con etiquetas alternativas (sinónimos), también en diferentes idiomas;
- 2) relaciones entre conceptos.

En la sección de referencias del curso encontrará enlaces a la ontología del CUAHSI, así como a material adicional relacionado con las ontologías. La hidro-ontología está disponible a través de un punto extremo en lenguaje de consulta semántica (SPARQL) para permitir la interacción entre máquinas. De este modo, los clientes compatibles con el protocolo SPARQL pueden hacer consultas a la hidro-ontología mediante la programación. Actualmente existen dos clientes de este tipo:

- LodLive;
- y el intermediario del WHOS, del que se hablará en la sección 4 del curso.

Diapositiva 27

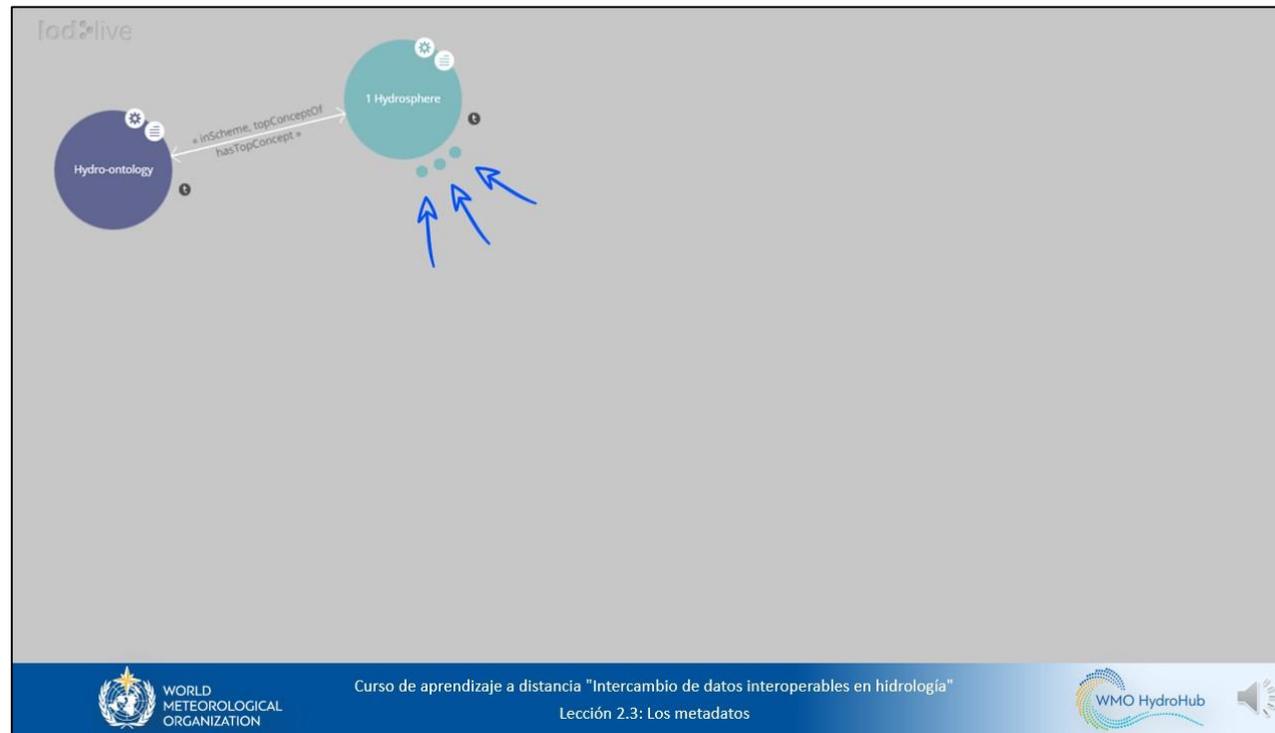
The screenshot displays the LodLive web interface. At the top, a yellow banner contains the URL <http://hydro.geodab.eu/ontology-browser/hydro-ontology.html>. The interface is divided into several sections:

- simple search:** Includes a "CHOOSE A DATASET" dropdown menu with "choose..." selected, an "INSERT A KEYWORD" input field, and a "start >>" button.
- insert URI:** Includes a "PASTE A RESOURCE ADDRESS" input field and a "start >>" button.
- browse your data:** Includes a "PASTE SOME RDF DATA" input field, a "PASTE AN ..." input field, and a "start >>" button. A purple banner with the text "coming soon" is overlaid on this section.
- Live on LodLive:** Includes a "CHOOSE AN ENDPOINT" dropdown menu with "choose..." selected.
- Hydro-ontology:** A central panel showing "Hydro-ontology" and a "RESOURCE" dropdown menu. The selected option is "example - Hydro-ontology concept scheme", which is highlighted with a yellow oval and a yellow arrow pointing to it. Below the dropdown is the URL "http://hydro.geodab.eu/hydro-ontology/concept/" and a "start >>" button with a hand icon.

At the bottom of the interface, there is a blue footer containing the WMO logo, the text "WORLD METEOROLOGICAL ORGANIZATION", the course title "Curso de aprendizaje a distancia 'Intercambio de datos interoperables en hidrología'", the lesson title "Lección 2.3: Los metadatos", and the WMO HydroHub logo.

LodLive es una herramienta de visualización gráfica del marco de descripción de recursos (RDF). Al conectarse a un punto extremo SPARQL de una ontología, LodLive hace que sus datos sean fácilmente accesibles y legibles en forma de representaciones gráficas. Para visualizar la hidro-ontología, se debe acceder a la herramienta web LodLive a través del siguiente enlace, luego seleccionar "Example – Hydro-ontology concept scheme" (Ejemplo — Esquema conceptual de hidro-ontología) en el cuadro y hacer clic en "Start" (Iniciar).

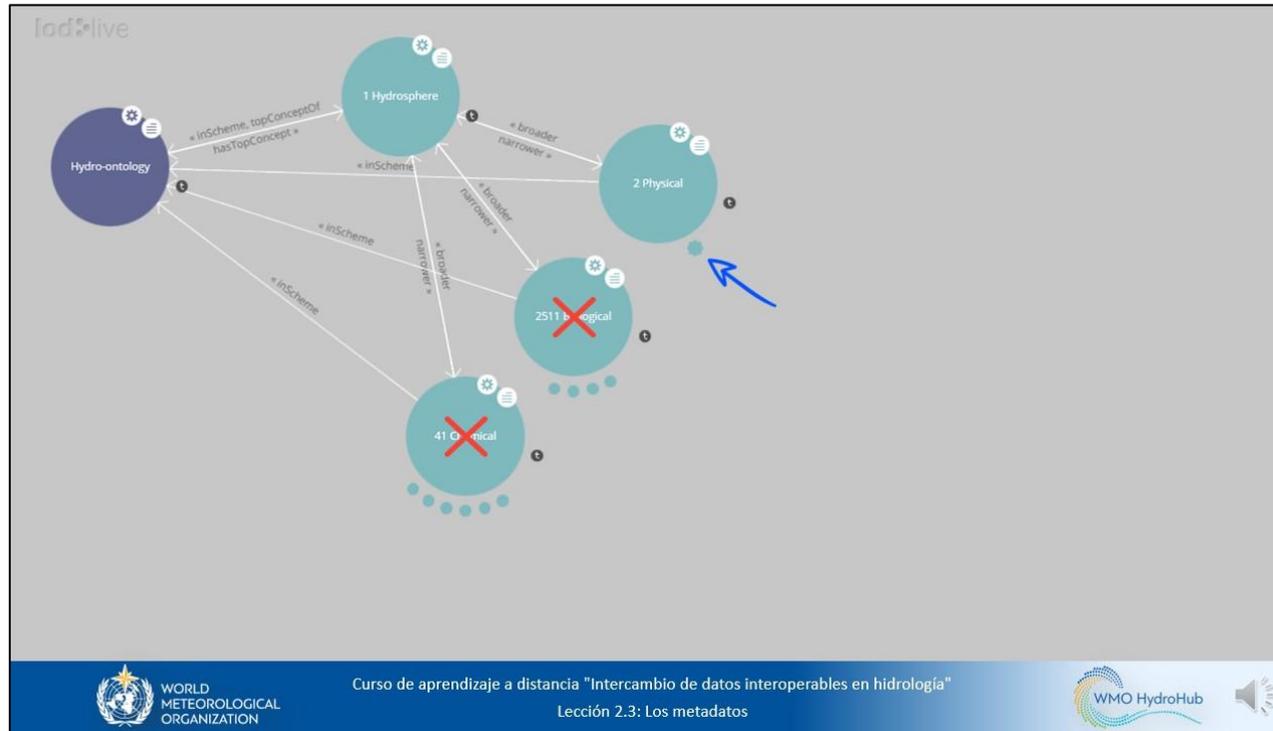
Diapositiva 28



En LodLive, los conceptos se representan como círculos y las relaciones entre conceptos como flechas.

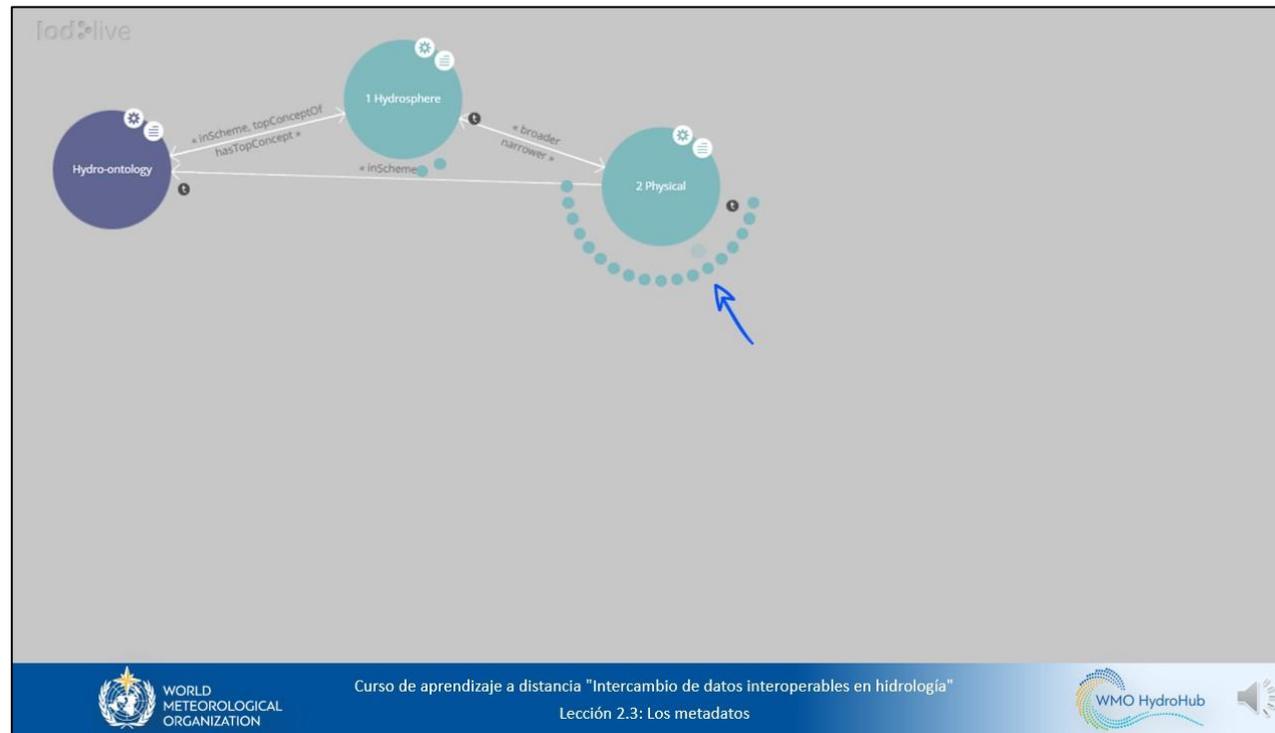
Al comenzar a explorar los conceptos, encontramos la "hidrosfera" como el concepto principal de la hidro-ontología. Partiendo de este punto, podemos concretar ese concepto al adentrarnos en sus tres conceptos relacionados: físico, químico y biológico.

Diapositiva 29



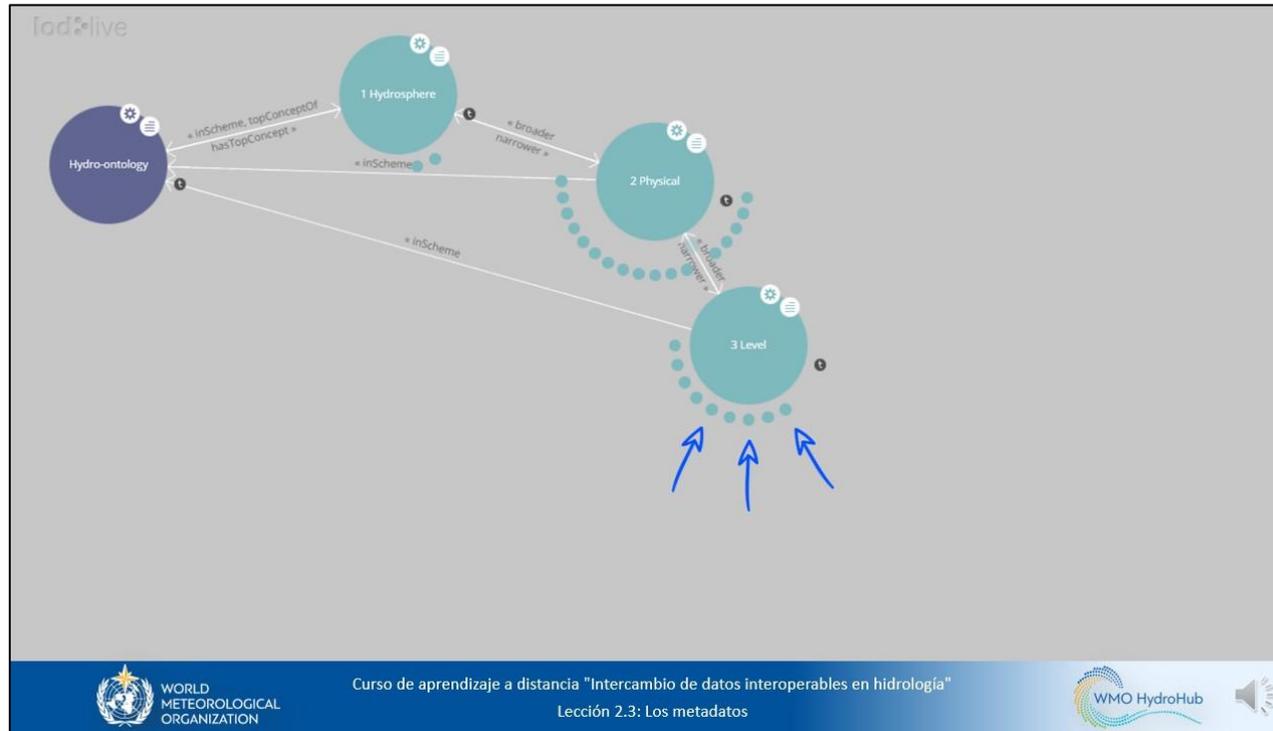
En nuestro ejemplo vamos a buscar las propiedades físicas asociadas al nivel del agua, por lo que cerraremos los conceptos químico y biológico, y ampliaremos el concepto físico.

Diapositiva 30



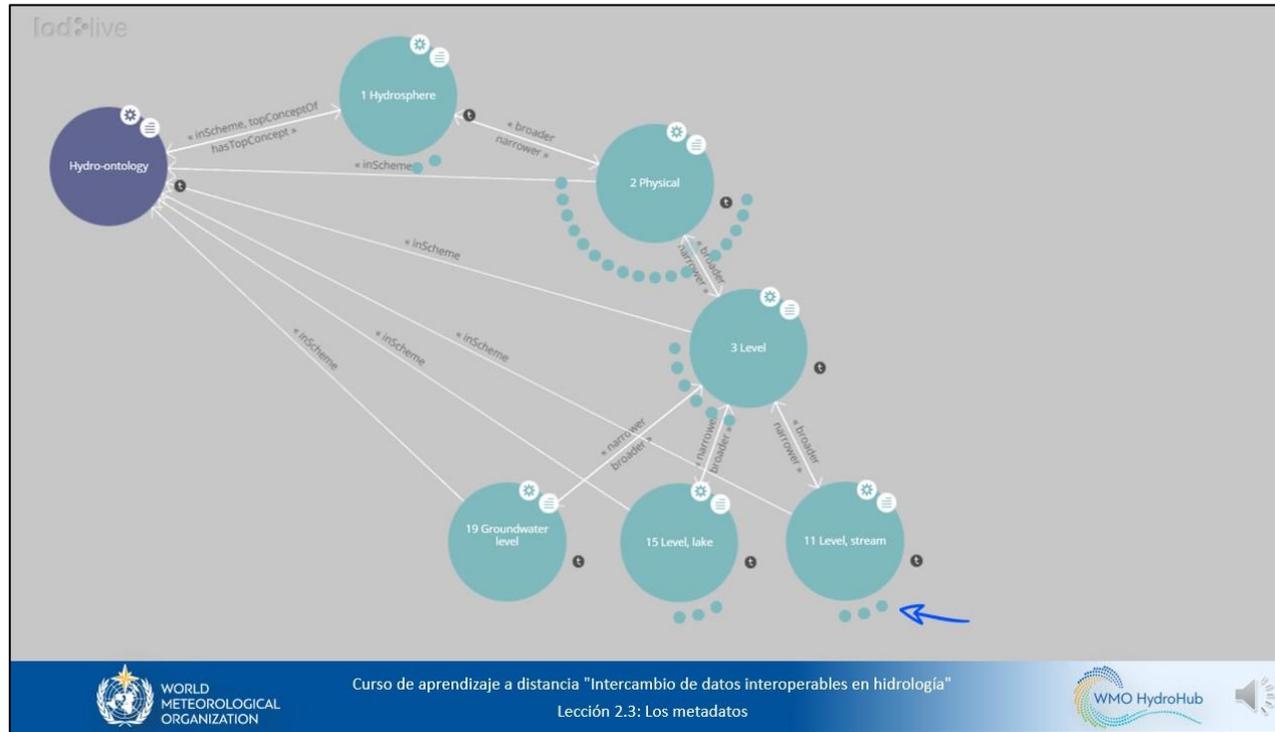
Entre las múltiples opciones de conceptos de la hidrosfera física más concretos, esta corresponde a "nivel".

Diapositiva 31



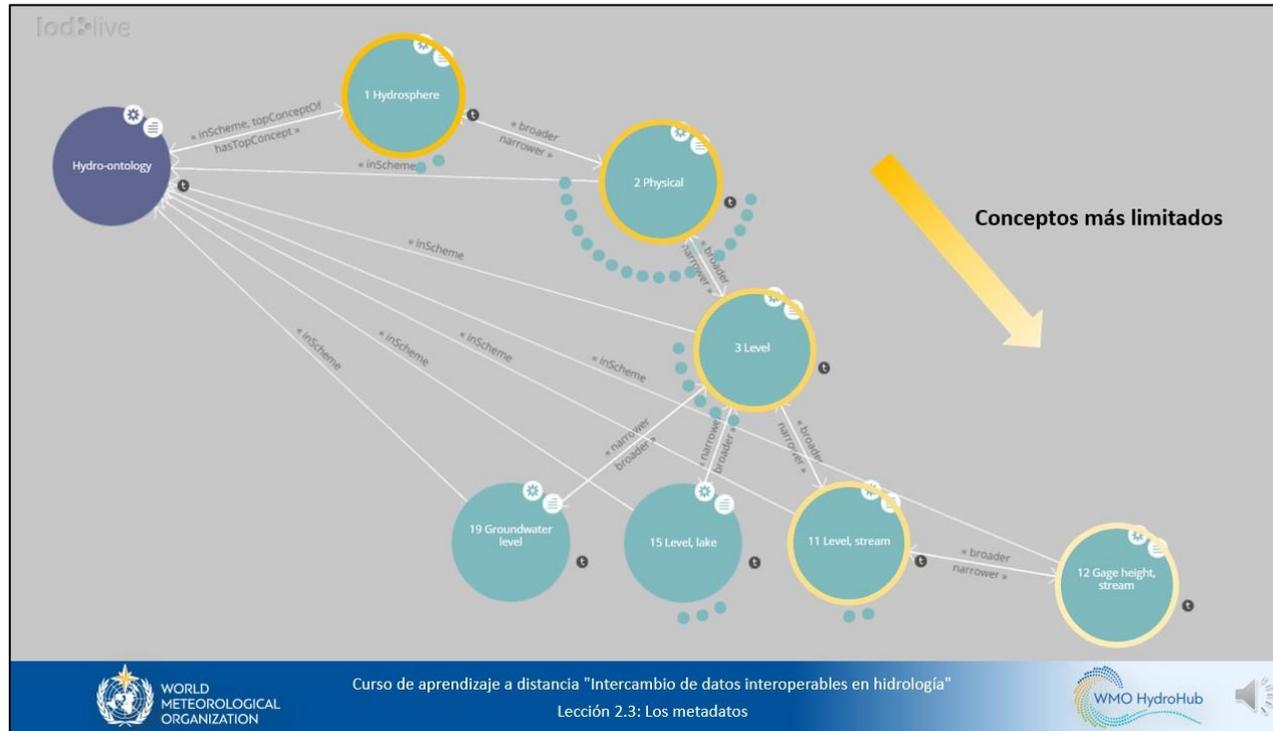
A partir de "nivel", podemos buscar conceptos todavía más concretos, como el nivel de una corriente, un lago o de aguas subterráneas.

Diapositiva 32



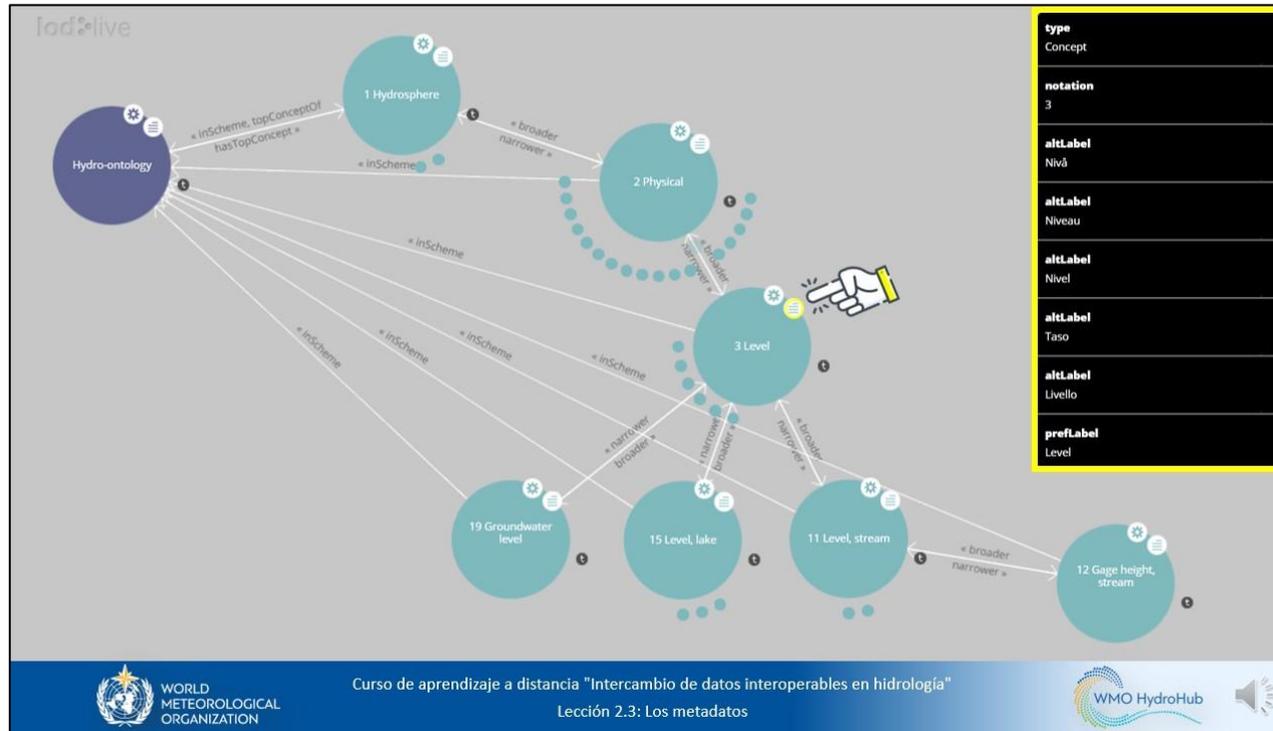
Si elegimos "nivel, corriente", podemos acotar aún más nuestra búsqueda y descubrir el concepto "altura de la escala, corriente".

Diapositiva 33



Disponer de relaciones lógicas entre los conceptos ayuda a los usuarios a poder relacionarlos con mayor facilidad, incluso si las distintas fuentes de datos los describen de forma diferente. Las relaciones lógicas también ayudan a armonizar diferentes instancias que tienen los mismos nombres o descripciones. En nuestro caso, utilizando el concepto 12 "altura de la escala, corriente" de la hidro-ontología que está relacionado con los conceptos más amplios "hidrosfera", "físico", "nivel" y "nivel, corriente", los proveedores de datos pueden describir sus datos con mayor precisión.

Diapositiva 34



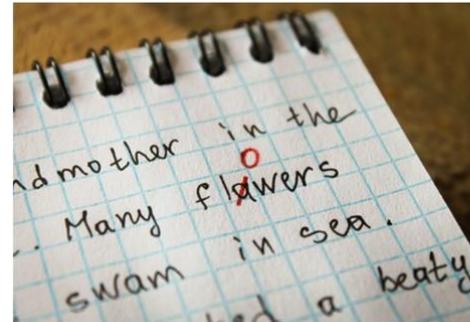
Podemos hacer clic en cada concepto para obtener información adicional, como las etiquetas alternativas, y también verlo en diferentes idiomas. Ahora que hemos definido la ontología y hemos explicado cómo está estructurada la hidro-ontología, vamos a presentar los vocabularios controlados.

Principales beneficios de utilizar vocabularios controlados y ontologías

Mejorar la interoperabilidad semántica



Garantizar la coherencia ortográfica



El uso de vocabularios controlados y ontologías para documentar elementos de metadatos permite:

- mejorar la interoperabilidad semántica de los datos a nivel organizativo, comunitario o internacional;
- garantizar la coherencia ortográfica.

Diapositiva 36



En realidad, los vocabularios controlados y las ontologías existen a distintos niveles. Estos distintos niveles representan los vocabularios y las ontologías consensuados y utilizados por las respectivas organizaciones, comunidades o a escala internacional. En la diapositiva se muestran algunos ejemplos.

Diapositiva 37



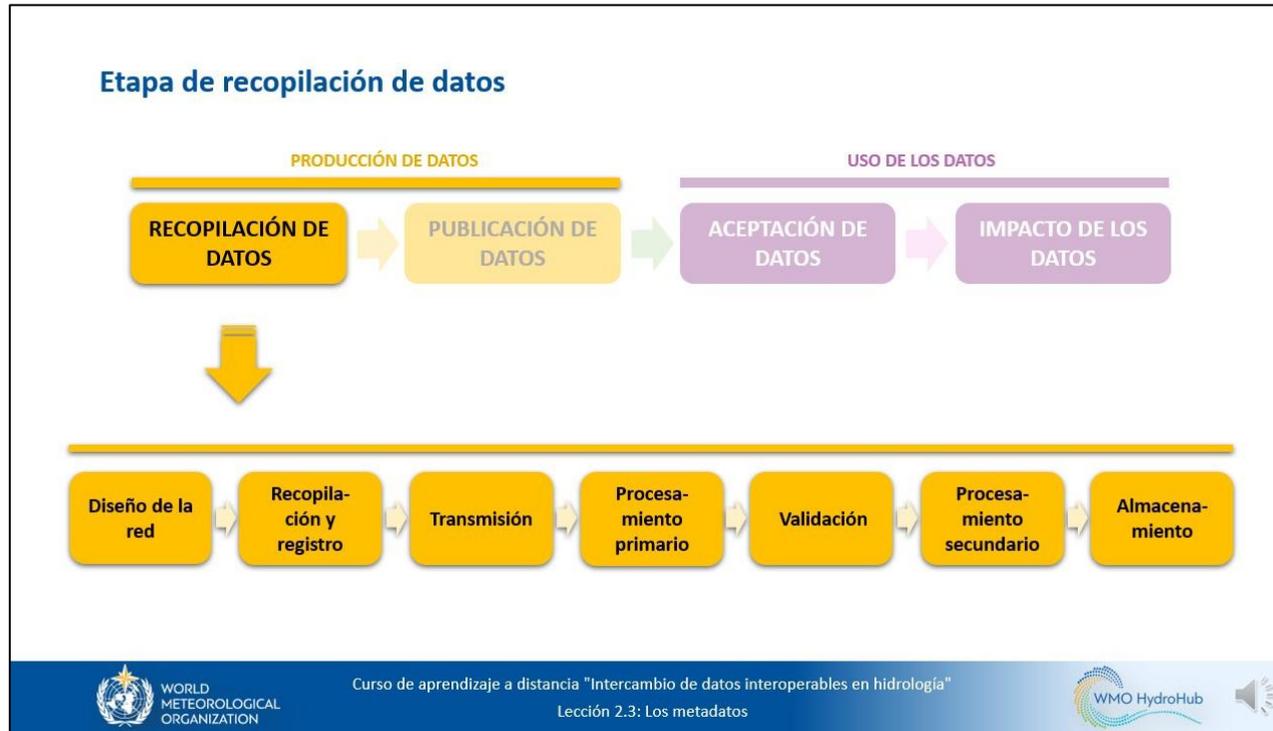
Veamos ahora cómo pueden los proveedores de datos aplicar los vocabularios controlados y las ontologías con objeto de mejorar la interoperabilidad semántica y, por ende, la reutilización de sus datos y metadatos.

- Un proveedor de datos puede codificar los datos y metadatos en un formato hidrológico de lectura mecánica, como WaterML (WML), utilizando vocabularios y lenguaje locales. En este caso, si la mayoría de los usuarios de datos no locales no aplica vocabularios controlados y ontologías, tendrán una capacidad limitada para interpretar cabalmente el significado de la información.
- Para aumentar la capacidad de interpretación de datos y metadatos por parte de los usuarios no locales, y mejorar la interoperabilidad semántica, los proveedores de datos podrían codificar los datos y metadatos en un formato hidrológico de lectura mecánica mediante el empleo de vocabularios controlados y ontologías creados y consensuados por la comunidad hidrológica o, aún mejor, por la comunidad internacional. De este modo, la mayoría de los usuarios de datos de todo el mundo podrá interpretar cabalmente el significado de los metadatos.
- Para mejorar aún más la interoperabilidad semántica de los datos y metadatos, los proveedores de datos podrían asociar los valores de los metadatos con conceptos y relaciones de ontologías muy conocidas en el formato específico del marco de descripción de recursos (RDF). Así, los usuarios de los datos podrán disponer de toda la información semántica sobre los metadatos procedente de diferentes ontologías.

Existen otras formas de mejorar la semántica, pero no se explicarán en este curso.

Es importante señalar que los aspectos relativos a la normalización deberían orientar la elección de vocabularios y ontologías.

Diapositiva 38



Ya hemos estudiado los distintos pasos y consideraciones importantes de la etapa de recopilación de datos. Tras la cadena de valor de los datos, la siguiente parte del curso aborda la fase de publicación de datos.