

Diapositiva 1



CURSO DE APRENDIZAJE A DISTANCIA
Intercambio de datos interoperables en hidrología

Lección 2.4 **Publicación y formatos de datos**

- 1. Enumerar y describir los tres pasos de la etapa de publicación de datos
- 2. Conocer los formatos estándar de datos e interpretar un ejemplo
- 3. Describir el formato de datos WaterML 2.0 y las razones para utilizarlo
- 4. Interpretar algunos elementos esenciales de WaterML 2.0

WEATHER CLIMATE WATER

 WORLD METEOROLOGICAL ORGANIZATION

 WMO HydroHub

 UCAR COMMUNITY PROGRAMS

 COMET

 National Research Council of Italy

Bienvenido a la lección 2.4 — Publicación de datos y formatos. Los objetivos de aprendizaje de esta lección son los siguientes:

- Enumerar y describir los tres pasos de la etapa de publicación de datos.
- Conocer los formatos normalizados de datos e interpretar un ejemplo.
- Describir el formato de datos WaterML 2.0 y las razones para utilizarlo.
- Interpretar algunos elementos esenciales de WaterML 2.0.

Diapositiva 2






La publicación de datos es la segunda etapa de la cadena de valor de los datos. Una vez concluida la etapa de recopilación de datos, los proveedores de datos deben publicar los datos y los metadatos que los acompañan de forma que los usuarios puedan acceder a ellos. La etapa de publicación comprende tres pasos: el análisis y la recuperación de datos de la base de datos para extraer información útil; la publicación de datos a través de servicios web; y la difusión de los datos a posibles usuarios.

Para optimizar el valor de los datos, es importante pensar estratégica y cuidadosamente cómo se analizarán, divulgarán y difundirán para fomentar su uso. ¿Están llegando los datos a los usuarios previstos a través de los canales de difusión adecuados?, ¿están en formatos abiertos y accesibles?, ¿se facilitan metadatos y se ajustan a las normas internacionales?



Diapositiva 3

El paso del análisis y la recuperación de datos

Análisis	determinar cuál es el subconjunto de información útil que conviene difundir	
Recuperación	recuperar metadatos y datos identificados en un formato específico desde el almacenamiento	

 WORLD METEOROLOGICAL ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"
Lección 2.4: Publicación y formatos de datos

 WMO HydroHub 

Los proveedores de datos, antes de publicar los datos, deben determinar cuáles desean o están obligados a difundir. Para ello, se pasa la fase de análisis y recuperación, que consta de dos partes:


- 1) Análisis: para determinar cuál es el subconjunto de información útil que conviene difundir y cómo presentarla al público.
- 2) Recuperación: para recuperar metadatos y datos identificados en un formato específico desde el almacenamiento (aplicando posiblemente la conversión de formato de datos) a fin de satisfacer las necesidades de los usuarios.

Diapositiva 4

El formato de los datos

“is a pre-established layout for data. A computer program accepts data as input in a certain format, processes it, and provides it as output in the same or another format. All data are stored in some format with the expectation that it will be processed by a program that knows how to handle that format.”

TechTarget.com©



Normas para los formatos de datos → **Grado de utilidad que tendrán los datos para los distintos usuarios**

WORLD METEOROLOGICAL ORGANIZATION

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

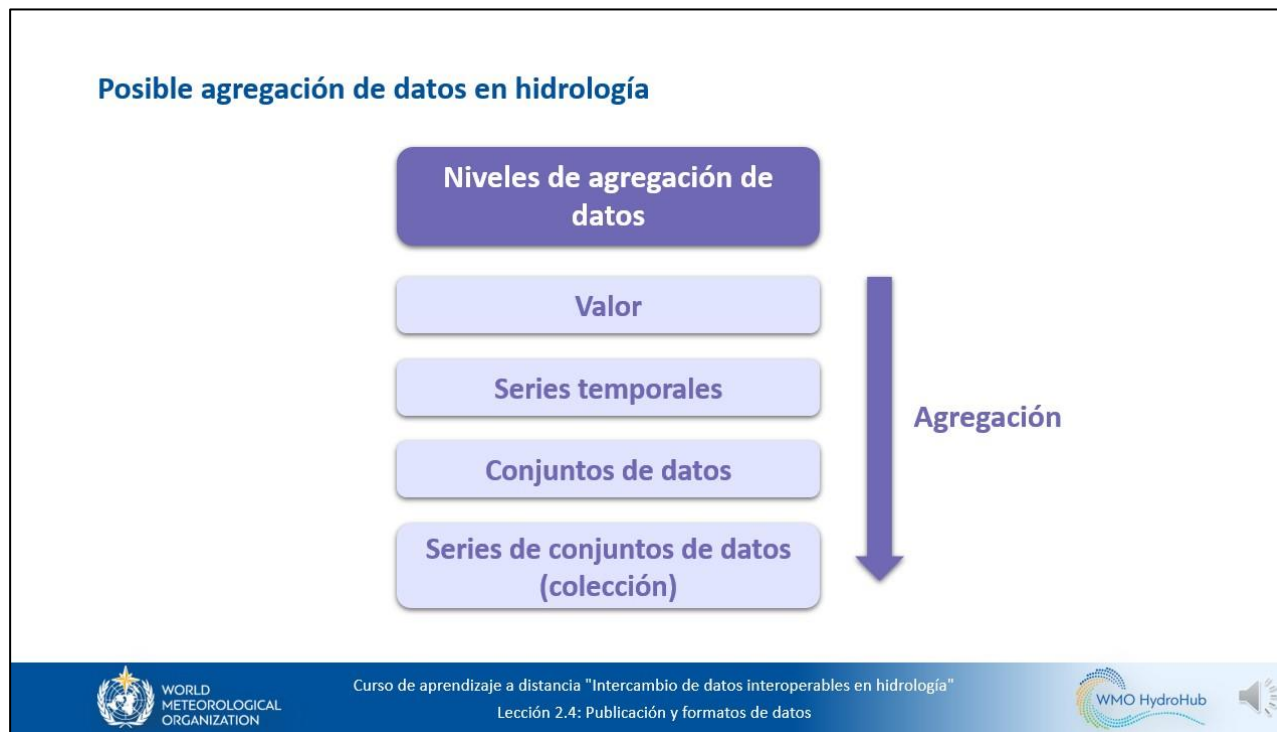
Lección 2.4: Publicación y formatos de datos

WMO HydroHub

De acuerdo con la empresa TechTarget.com[©], el formato de los datos es una configuración preestablecida para los datos. Un programa informático permite la entrada de datos en un formato determinado, los procesa y los devuelve en el mismo formato o en otro. Todos los datos se almacenan en algún formato a la espera de que los procese un programa capaz de gestionar ese formato.

Como ocurre con los metadatos, existen normas para los formatos de datos que pueden establecerse a nivel organizativo, comunitario e internacional. La norma que adopte un proveedor de datos determinará el grado de utilidad que tendrán sus datos para los distintos usuarios.

Diapositiva 5

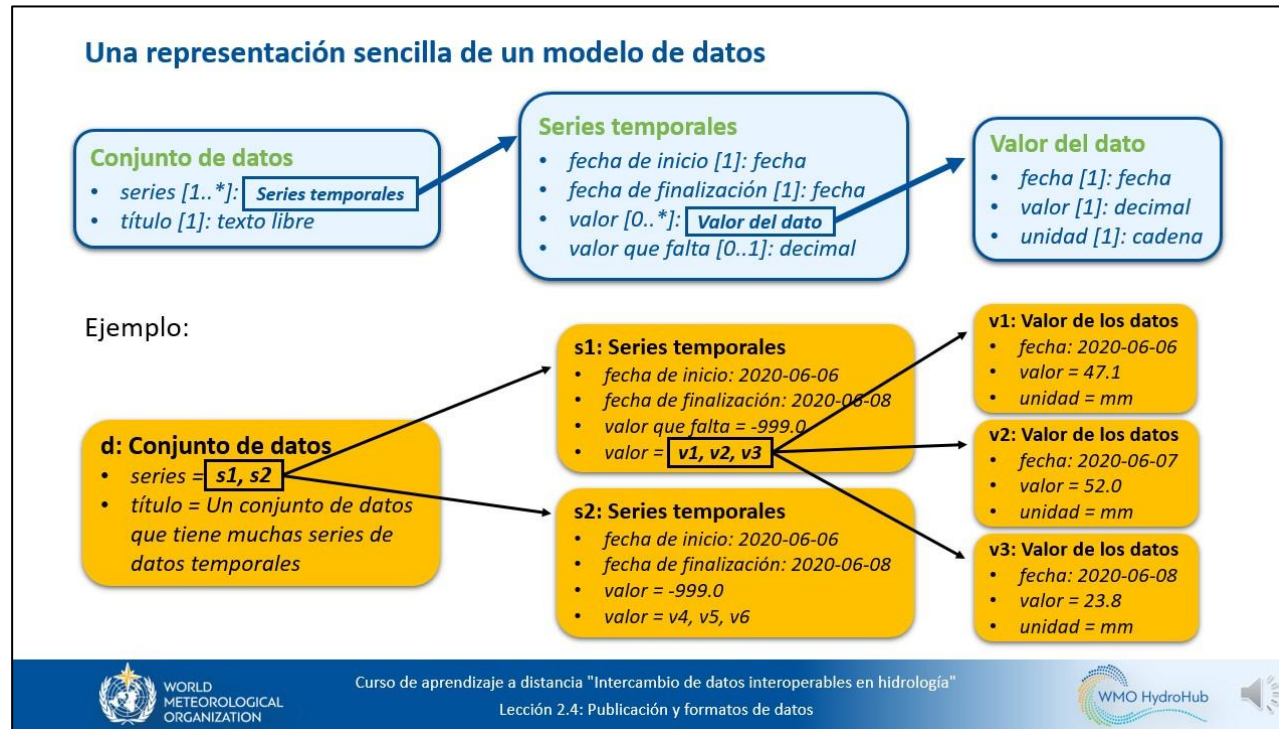


Antes de presentar las diferentes normas y especificaciones técnicas relativas a los datos, es importante recordar los distintos niveles de agregación de datos posibles:

- de valores simples a series temporales;
- a conjuntos de datos;
- a series de conjuntos de datos.

Las especificaciones técnicas de modelización y codificación de datos presentan diferencias según los niveles de agregación de datos.

Diapositiva 6



A continuación se muestra, a modo de ejemplo, una representación sencilla de un modelo de datos para obtener información de los datos. Un modelo de este tipo se compone de tipos de datos, en color verde, y los elementos correspondientes que se enumeran debajo del encabezamiento del tipo.

Los recuadros naranjas muestran un ejemplo de registro de datos, en el que los elementos se documentan con valores ficticios.

Diapositiva 7

Normalización para los formatos de datos

Normalización para el modelo general de datos

➤ describe el modelo de datos y especifica los tipos de datos y los elementos que deben documentarse


Data value

- *fecha [1]: fecha*
- *valor [1]: decimal*
- *unidad [1]: cuerda*

Normalización para la codificación de datos

➤ aborda cómo codificar el modelo general de datos en un formato legible por máquina

```
<dataValue>
<time>1990-01-01</time>
<value>8.6</value>
<units>mm</units>
```



STANDARDS

WORLD METEOROLOGICAL ORGANIZATION Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología" WMO HydroHub

Lección 2.4: Publicación y formatos de datos

Al igual que con los metadatos, también existen dos tipos de normalización para los formatos de datos:

- 1) La normalización para el modelo general de datos describe el modelo de datos y especifica los tipos de datos y los elementos que deben documentarse, así como el significado de los elementos. Por ejemplo, el tipo “valor de datos” contiene elementos como “tiempo”, “valor” y “unidad”.
- 2) La normalización para la codificación de datos aborda cómo codificar el modelo general de datos en un formato de lectura mecánica. Por ejemplo, en formato XML.

Normalmente, existen dos documentos distintos para cada uno de los dos formatos de datos, denominados especificaciones técnicas. Sin embargo, a veces solo se dispone de una especificación técnica que cubre el modelo de datos y la codificación de datos.

Diapositiva 8

FORMATOS DE DATOS	Ejemplos
CSV (archivos de valores separados por comas)	➤ USGS RDB : formato delimitado por tabuladores específico para hidrología
JSON (JavaScript Object Notation)	➤ GeoJSON : diseñado para representar accidentes geográficos simples y sus atributos no espaciales
XML (Lenguaje de marcado extensible)	➤ WaterML : principalmente para observaciones hidrológicas ➤ Observation & measurements : para observaciones generales
Archivos binarios	➤ NetCDF-CF : principalmente para observaciones climáticas ➤ BUFR : principalmente para observaciones meteorológicas ➤ GeoTIFF

WORLD METEOROLOGICAL ORGANIZATION

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

Lección 2.4: Publicación y formatos de datos

WMO HydroHub

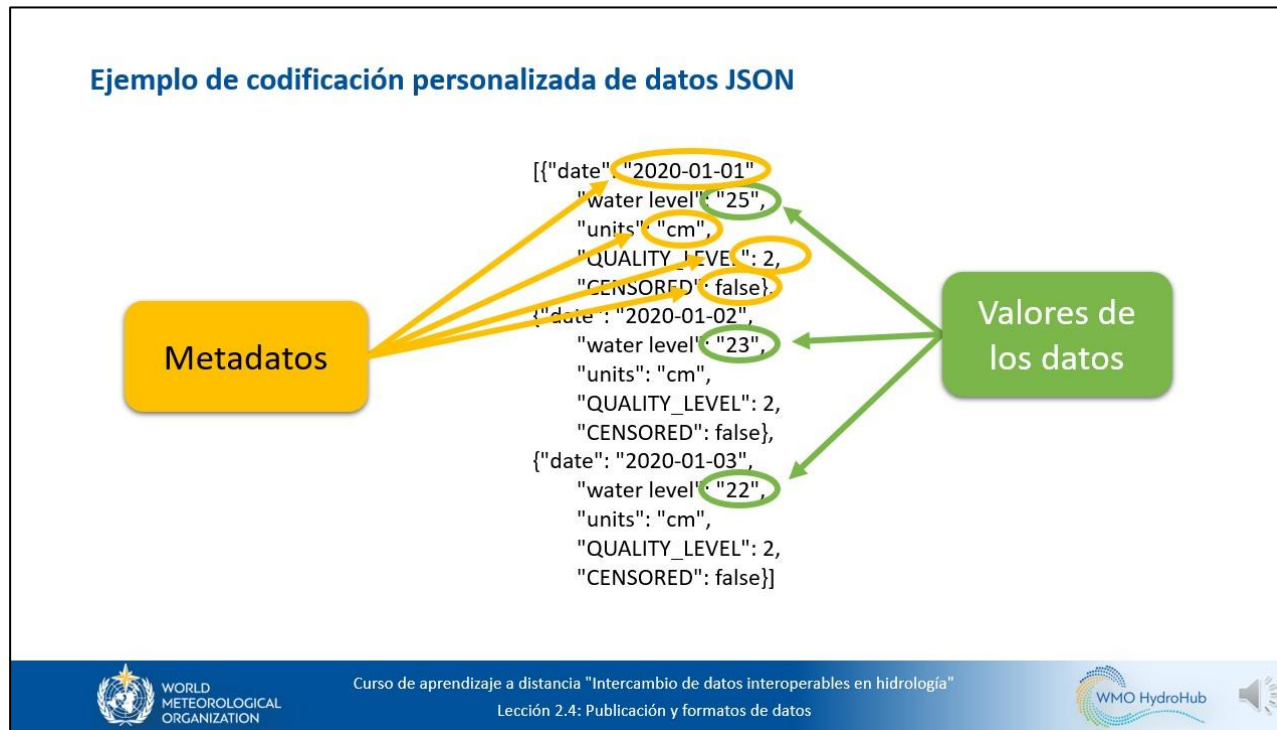
Los formatos de datos pueden dividirse según su codificación en formato de lectura mecánica, a saber, archivos de valores separados por comas (CSV), notación de objetos en JavaScript (JSON), lenguaje de marcado extensible (XML) y archivos binarios.

- Los archivos CSV tienen un formato de datos delimitado en el que los campos o las columnas están separados por comas y los registros o filas por saltos de línea.
- JSON es un formato de archivo normalizado abierto. Es un buen formato para el intercambio de datos debido a su independencia respecto al lenguaje. Se trata de un formato legible por humanos y también de lectura mecánica.
- XML es un formato normalizado para archivos de datos a un nivel más general. Establece una serie de reglas para la codificación de documentos en formato legible por humanos y de lectura mecánica. Fue desarrollado por el World Wide Web Consortium (W3C).

Estos son los ejemplos de formatos de datos normalizados para cada tipo específico de codificación de datos. Las especificaciones técnicas de estos formatos de datos normalizados ya se han desarrollado y están a disposición del público.

Sin embargo, los proveedores de datos pueden, en ocasiones, adoptar un modelo de datos y una codificación personalizados para cada tipo de codificación de datos. Esta solución no favorece el intercambio de datos, ya que no abarca a una comunidad de usuarios extensa.

Diapositiva 9



Aunque los formatos de datos se limiten principalmente a codificar los valores de los datos, a menudo también incorporan metadatos esenciales, como se muestra en este ejemplo.

Diapositiva 10

Formatos de intercambio de datos en hidrología

OGC®
Making location count.

WaterML 2.0



WORLD METEOROLOGICAL ORGANIZATION

Aprobadas

- Parte 1, sobre las series temporales
- Parte 2, sobre el gasto, el aforo y las secciones
- Parte 4 sobre las aguas subterráneas (GroundWaterML2)

En fase de desarrollo

- Parte 3, sobre las características de la hidrología de superficie
- WaterML 2.0 para datos relativos a la calidad del agua

WaterML 2.0 permite el intercambio de datos entre diversos sistemas de información

WORLD METEOROLOGICAL ORGANIZATION

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

Lección 2.4: Publicación y formatos de datos

WMO HydroHub

Analicemos ahora la realidad de los formatos para el intercambio internacional de datos hidrológicos.


En 2008, en la *Guía de prácticas hidrológicas* (OMM-Nº 168), volumen I: Hidrología – De la medición a la información hidrológica, se estableció que "no hay actualmente normas aplicables a los formatos de intercambio de datos hidrológicos".

Para resolver este problema, la OMM y el Open Geospatial Consortium (OGC) redactaron la norma WaterML 2.0. Las siguientes partes de WaterML 2.0 ya han sido aprobadas por el OGC y a través de la Resolución 11 (EC-69) de la OMM: la parte 1, sobre las series temporales, y la parte 2, sobre el gasto, el aforo y las secciones. El OGC también ha aprobado la parte 4, sobre las aguas subterráneas (GroundWaterML2). La parte 3, sobre las características de la hidrología de superficie, y el perfil WaterML 2.0 para datos relativos a la calidad del agua, se hallan en fase de desarrollo. WaterML 2.0 es un modelo de información normalizado para la representación de datos hidrológicos creado para permitir el intercambio de datos entre diversos sistemas de información. WaterML 2.0, codificado en XML y basado en las normas del OGC, ofrece un formato de intercambio hidrológico interoperable que puede utilizarse para satisfacer una gran variedad de necesidades de los usuarios.


Diapositiva 11


WaterML 2.0

- Se implementa como un esquema de aplicación del Lenguaje de Etiquetado Geográfico (GML) versión 3.2.1;
- hace uso de la norma Observaciones y Medidas del OGC (Observations & Measurements);
- permite que los datos procedentes de fuentes dispares se entiendan y reutilicen mutuamente;
- diseñado para brindar soporte a una gran variedad de escenarios de intercambio.




© CartoCite 2004-2007 (Creative Commons)






WORLD METEOROLOGICAL ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"
Lección 2.4: Publicación y formatos de datos



WMO HydroHub



WaterML 2.0 se implementa como un esquema de aplicación del lenguaje de marcado geográfico (GML), versión 3.2.1, haciendo uso de la norma sobre observaciones y mediciones del OGC. La adecuación a dicha norma, que establece un esquema conceptual y un vocabulario transversal para la descripción de datos y metadatos de observación, permite la comprensión mutua y la reutilización de datos procedentes de fuentes dispares.

WaterML 2.0 está diseñado para soportar una gran variedad de escenarios de intercambio, a saber, el funcionamiento de infraestructuras, el intercambio transfronterizo de datos de observación, la difusión de datos para divulgación o la mejora de la gestión de desastres a través del intercambio de datos, entre otros.

Diapositiva 12



Los elementos de la parte 1 de WaterML 2.0 permiten definir de forma normalizada las propiedades fundamentales relativas a las series temporales hidrológicas, a saber:

- los fenómenos observados;
- la extensión espacial;
- los límites temporales;
- los procedimientos utilizados para generar las series temporales (por ejemplo, los datos brutos de un sensor);
- los metadatos específicos, como los calificadores de series temporales, los tipos de interpolación, los comentarios, las claves de calidad, etc.;
- los puntos de monitoreo;
- las colecciones de objetos relacionados;
- los vocabularios de términos específicos de distintos ámbitos.

Diapositiva 13

Algunos elementos esenciales de WaterML 2.0			Ejemplo
observedProperty	Describe el fenómeno que se está observando	Se recomienda documentar la propiedad observada usando términos adecuados de vocabularios u ontologías controlados y consensuados	<pre><om:observedProperty xlink:href="http://codes.wmo.int/wmdir/0/observedVariableTerrestrial/171" xlink:title="River discharge">Volume of water flowing through a river per unit of time</om:observedProperty></pre>
UnitOfMeasure	Describe la unidad de medida	Se recomienda documentar con conceptos adecuados de vocabularios/ontologías controlados	<pre><wml2:uom xlink:href="http://codes.wmo.int/common/unit/m3_s-1" xlink:title="cubic metres per second" code="m3/s"/></pre>
interpolationType	Describe el procedimiento para realizar la estimación que representa el valor	Se recomienda documentar con conceptos adecuados del vocabulario controlado del OGC	<pre><wml2:interpolationType xlink:href="http://www.opengis.net/def/waterml/2.0/interpolationType/Continuous" xlink:title="Continuous"/></pre>

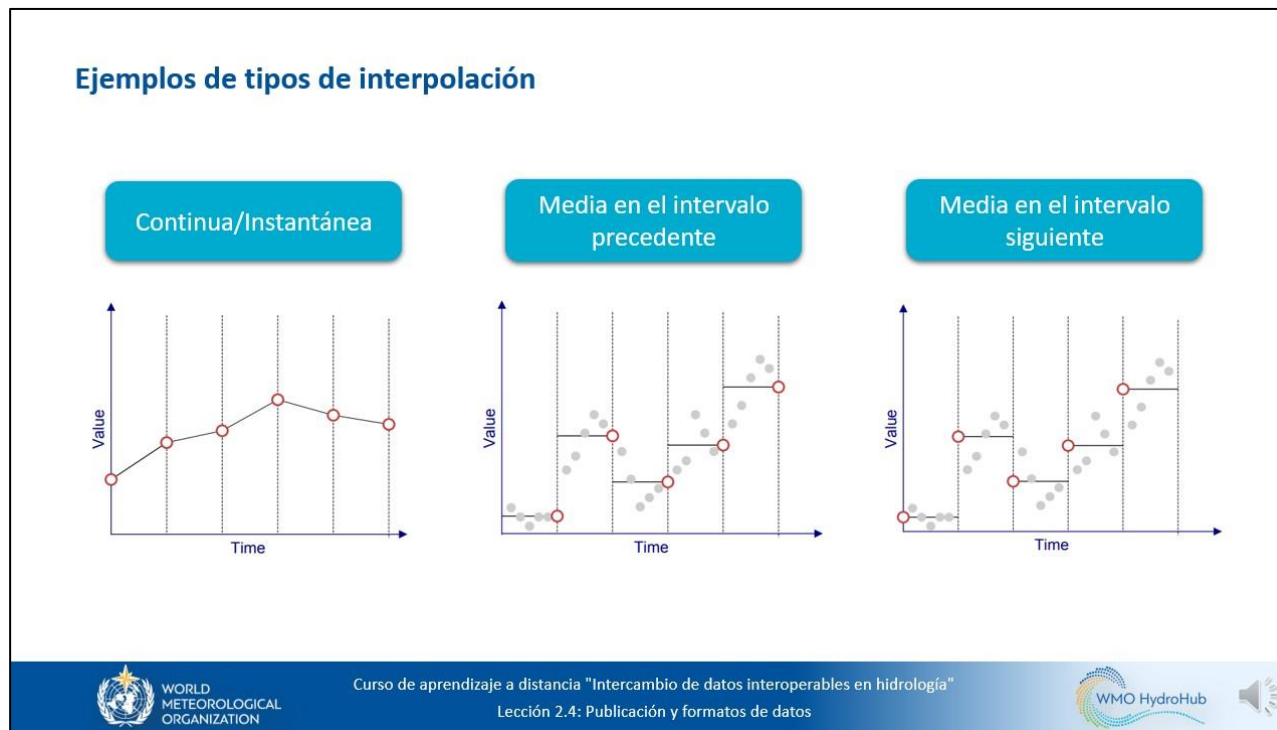
WORLD METEOROLOGICAL ORGANIZATION Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología" WMO HydroHub

Lección 2.4: Publicación y formatos de datos

Algunos de los elementos esenciales de WaterML 2.0 que los proveedores de datos deben documentar son los siguientes:

- "Observed property" (Propiedad observada). Esta propiedad se utiliza para describir el fenómeno que se está observando, por ejemplo, la temperatura, el nivel o el caudal. Se recomienda documentar la propiedad observada usando términos adecuados de vocabularios controlados u ontologías consensuados por una comunidad o a nivel internacional, como el registro de claves de la OMM y la ontología hidrológica descritos en la lección anterior. Por lo tanto, como se muestra en el ejemplo, la propiedad observada debe hacer referencia a: 1) un enlace (resaltado en amarillo) al concepto de un vocabulario controlado u ontología específicos; 2) su etiqueta (escrita en color verde en el ejemplo); y 3) su descripción (escrita en azul). Se recomienda separar el fenómeno observado de las unidades utilizadas para la medición.
- Las unidades utilizadas para la medición se documentan con el elemento de WaterML 2.0 denominado "Unit of Measure" (Unidad de medida). La unidad de medida, al igual que sucede con la propiedad observada, debe hacer referencia a un enlace al concepto adecuado de un vocabulario controlado u ontología específicos. En nuestro ejemplo, en color amarillo se hace referencia a un concepto del registro de claves de la OMM. Además, la unidad de medida debe contener una etiqueta de unidad de preferencia, escrita en color verde, y el código de unidad apropiado del sistema de códigos del Código Unificado para Unidades de Medida (UCUM), escrito en azul.
- Otro elemento esencial de WaterML 2.0 es "Interpolation type" (Tipo de interpolación). Una de las características fundamentales de las series temporales de mediciones es la naturaleza de la relación entre el instante temporal y el valor registrado. Esta relación viene determinada por el procedimiento utilizado para realizar la estimación que representa el valor. Representarlo es esencial para poder interpretar correctamente los valores de las series temporales. Por ejemplo, el valor podría representar un promedio a lo largo del período de tiempo. El ejemplo de la diapositiva muestra cómo se codifica el tipo de interpolación en WaterML 2.0, haciendo referencia a un enlace (resaltado en amarillo) a un término adecuado del vocabulario controlado del OGC, así como a una etiqueta de preferencia del tipo de interpolación (escrita en verde). Los tipos de series temporales se definen en WaterML 2.0 y se describen en la documentación de WaterML 2.0 disponible en la sección de referencias del curso. En la siguiente diapositiva presentaremos tres de estos tipos de series temporales: continua/instantánea; media en el intervalo precedente; y media en el intervalo siguiente.

Diapositiva 14

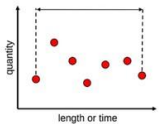
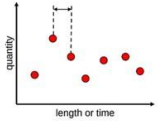
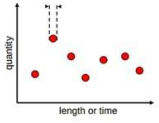


- Una serie temporal continua emplea todos los valores observados a lo largo de un período de tiempo. Los puntos de referencia observados están conectados y puede haber interpolación entre ellos.
- En el tipo “media en el intervalo precedente”, el valor representa el valor medio en el intervalo precedente, por ejemplo, el caudal medio diario.
- En el tipo “media en el intervalo siguiente”, el valor representa el valor medio en el intervalo siguiente.

Para conocer otros tipos de series temporales definidas por WaterML 2.0, consulte la documentación de WaterML 2.0.

Diapositiva 15

Algunos elementos esenciales de WaterML 2.0

phenomenonTime	Describe la extensión temporal de la serie temporal que es el resultado de la observación		Debe documentarse de acuerdo con la norma ISO 8601	<pre><om:phenomenonTime> <gml:TimePeriod gml:id="series_3_time_acquisition"> <gml:beginPosition>1980-01-01T00:00:00Z</gml:beginPosition> <gml:endPosition>2021-01-01T00:00:00Z</gml:endPosition> </gml:TimePeriod> </om:phenomenonTime></pre>
Intended Observation Spacing	Define el intervalo de tiempo previsto entre observaciones individuales		Debe documentarse de acuerdo con la norma ISO 8601	<pre><wml2:intendedObservationSpacing>PT15M</wml2: intendedObservationSpacing></pre>
Aggregation Duration	Define la duración temporal a lo largo de la cual se han agregado los datos		Debe documentarse de acuerdo con la norma ISO 8601	<pre><wml2:aggregationDuration>P1D</wml2:aggregation Duration></pre>

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"
Lección 2.4: Publicación y formatos de datos

WMO HydroHub

Otros elementos esenciales de WaterML 2.0 que los proveedores de datos deben documentar son:

- **Phenomenon Time (Intervalo de tiempo del fenómeno):** esta propiedad describe la extensión temporal de la serie temporal que es el resultado de la observación. WaterML2.0 limita este elemento a un período de tiempo, en el que se definen el momento de inicio y de fin de la serie temporal. En la diapositiva se muestra un ejemplo de cómo se codifica el intervalo de tiempo del fenómeno en WaterML 2.0, en el que el momento de inicio y de fin de la serie temporal se resaltan en amarillo. Tenga en cuenta que el tiempo se debe documentar siguiendo la notación de la norma ISO 8601.
- **IntendedObservationSpacing (Espaciado previsto entre observaciones):** se trata de otro elemento esencial de WaterML 2.0. Este define el intervalo de tiempo previsto entre observaciones individuales. Es habitual que las observaciones se realicen con frecuencia, como las llevadas a cabo por sensores automatizados o visitas periódicas; esta característica permite determinar el tiempo previsto entre mediciones. Así lo reflejan los puntos individuales que componen la serie temporal resultante. En la diapositiva se muestra un ejemplo de cómo se codifica el elemento IntendedObservationSpacing en WaterML 2.0. Tenga en cuenta que este elemento se debe documentar de acuerdo con la norma ISO 8601. En nuestro ejemplo, se documenta como "PT15M", que significa "tiempo del período 15 minutos".
- El siguiente elemento fundamental de WaterML 2.0 es "Aggregation Duration" (Duración de la agregación). Si el proceso conlleva la agregación temporal de un conjunto de resultados, en este punto deberá expresarse la duración temporal a lo largo de la cual se han agregado los datos. Por ejemplo, agregación por hora o por día. Tenga en cuenta que la duración de la agregación también debe documentarse de acuerdo con la norma ISO 8601. En nuestro ejemplo, se documenta como "P1D", que significa "período de un día".

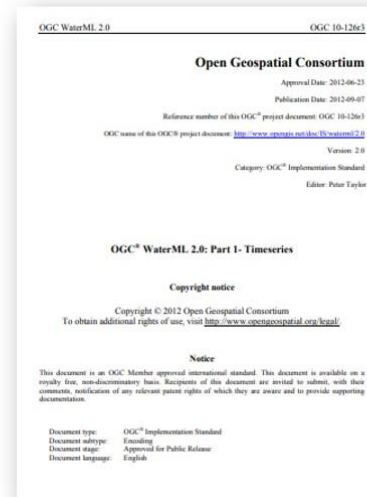
WaterML 2.0 contiene otros elementos fundamentales que no se tratarán en esta lección. Para saber más sobre ellos, consulte la documentación de WaterML 2.0.

Diapositiva 16

WaterML 2.0

- Reduce la heterogeneidad de los formatos de datos;
- permite gestionar, compartir y analizar mejor grandes cantidades de datos hidrológicos

Se alienta a los proveedores de datos hidrológicos a que apliquen esta norma y establezcan un sistema de publicación de datos de observación hidrológica en formato WaterML 2.0 a través de servicios web



WORLD
METEOROLOGICAL
ORGANIZATION

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

Lección 2.4: Publicación y formatos de datos



La adopción de la norma WaterML 2.0 reduce la heterogeneidad de los formatos de datos. También permite gestionar, poner en común y analizar mejor grandes cantidades de datos hidrológicos.

Se alienta a los proveedores de datos hidrológicos a que utilicen esta norma y establezcan un sistema de publicación de datos de observación hidrológica en formato WaterML 2.0 a través de servicios web.

La lección 2.5. tratará sobre los servicios web.