

Diapositiva 1

CURSO DE APRENDIZAJE A DISTANCIA
Intercambio de datos interoperables en hidrología

Sección 2 • **Funciones que desempeñan los proveedores de datos en el marco de los procesos de intercambio de datos hidrológicos**

- Lección 2.1** La etapa de recopilación de datos
- Lección 2.2** Las diferentes soluciones o arquitecturas para el almacenamiento de datos
- Lección 2.3** Los metadatos
- Lección 2.4** La publicación y los formatos de los datos
- Lección 2.5** Los servicios web

WEATHER CLIMATE WATER

 WORLD METEOROLOGICAL ORGANIZATION

 WMO HydroHub

 UCAR COMMUNITY PROGRAMS

 COMET

 National Research Council of Italy

Bienvenido a la sección 2 del curso de aprendizaje a distancia Intercambio de Datos Interoperables en materia de Hidrología. Esta sección consta de cinco lecciones.

- En la lección 2.1 se trata la etapa de recopilación de datos.
- En la lección 2.2 se tratan las diferentes soluciones o arquitecturas para el almacenamiento de datos.
- En la lección 2.3 se tratan los metadatos.
- En la lección 2.4 se tratan la publicación y los formatos de los datos.
- En la lección 2.5 se tratan los servicios web.

Diapositiva 2

CURSO DE APRENDIZAJE A DISTANCIA
Intercambio de datos interoperables en hidrología

Lección 2.1 **La etapa de recopilación de datos**

- Describir las propiedades que adquieren los datos a lo largo de su viaje
- Determinar los atributos de los datos que son susceptibles de provocar preguntas de los proveedores de datos
- Distinguir los pasos clave en la etapa de recopilación de datos
- Resumir los objetivos principales de cada uno de los pasos de la recopilación de datos

WEATHER CLIMATE WATER

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

Empecemos por la lección 2.1 — Etapa de recopilación de datos. Los objetivos de aprendizaje de esta lección son los siguientes:

- Describir las propiedades que adquieren los datos a lo largo de su viaje.
- Determinar los atributos de los datos que deberían provocar preguntas de los proveedores de datos.
- Distinguir los pasos clave en la etapa de recopilación de datos.
- Resumir los objetivos principales de cada uno de los pasos de la recopilación de datos.

Diapositiva 3



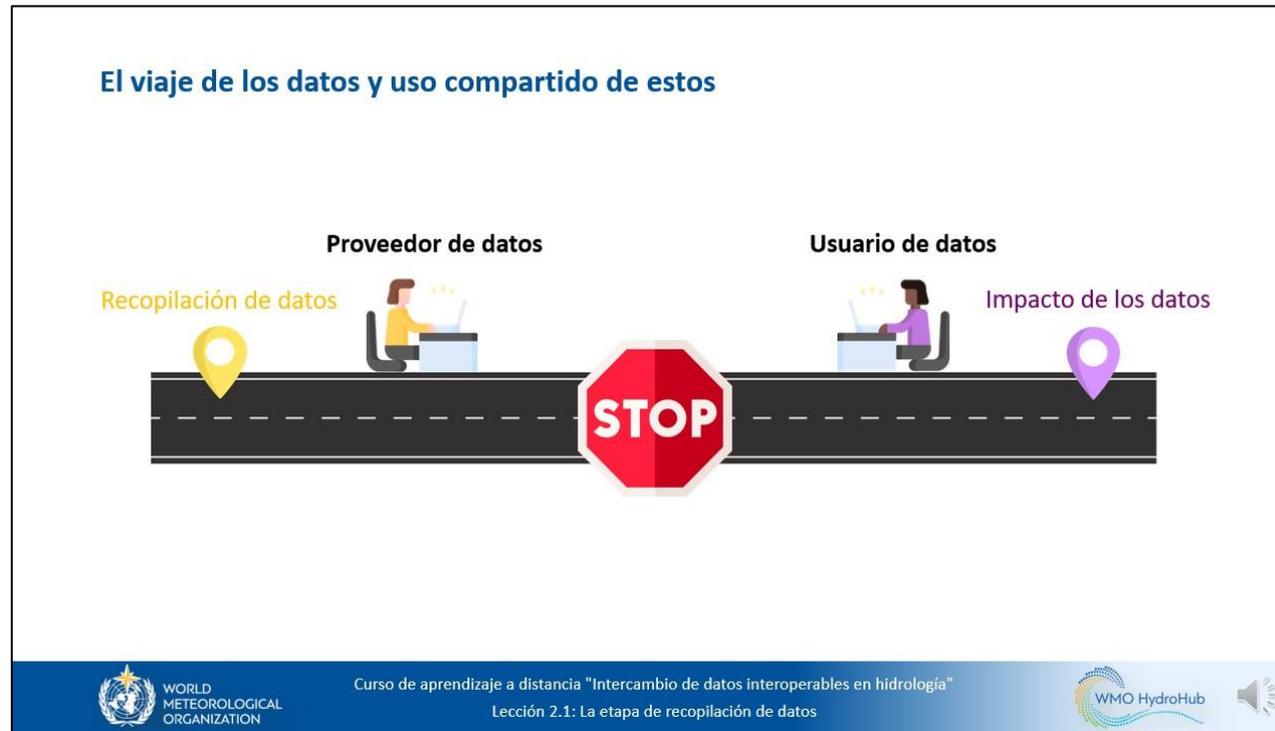
Todos los tipos de datos realizan un largo viaje desde su recopilación hasta su repercusión final en la toma de decisiones y, a medida que avanzan, adquieren más valor.

Diapositiva 4



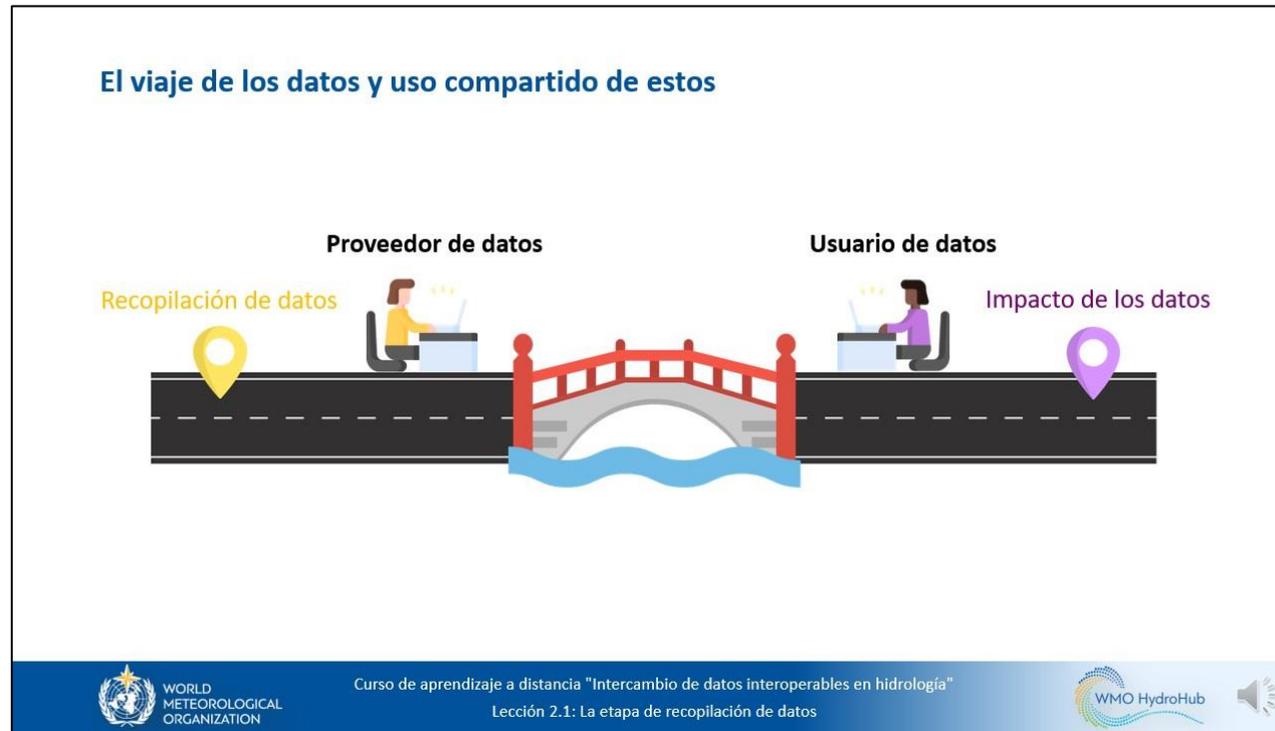
Para iniciar el viaje desde el proveedor de los datos hasta el usuario de los datos, debemos constatar que los datos están presentes (medidos sobre el terreno). A lo largo del viaje, los datos medidos generan información y conocimientos útiles. Esto se debe a que los datos pueden adquirir propiedades útiles: se pueden ver (los datos se registran y almacenan), se pueden leer (los datos y metadatos están disponibles en un formato legible), se pueden entender (los datos y metadatos son comprensibles para diferentes comunidades de usuarios) y los usuarios de los datos pueden confiar en ellos y estos pueden ser útiles para diferentes fines (por ejemplo, elaborar predicciones de crecidas). A medida que los datos adquieren progresivamente estas propiedades, aumenta su aplicabilidad, fiabilidad y valor. Sin embargo, como se ha comentado en la sección 1 de este curso, para llegar al destino final (impacto de los datos), los proveedores deben poner los datos en común con los usuarios.

Diapositiva 5



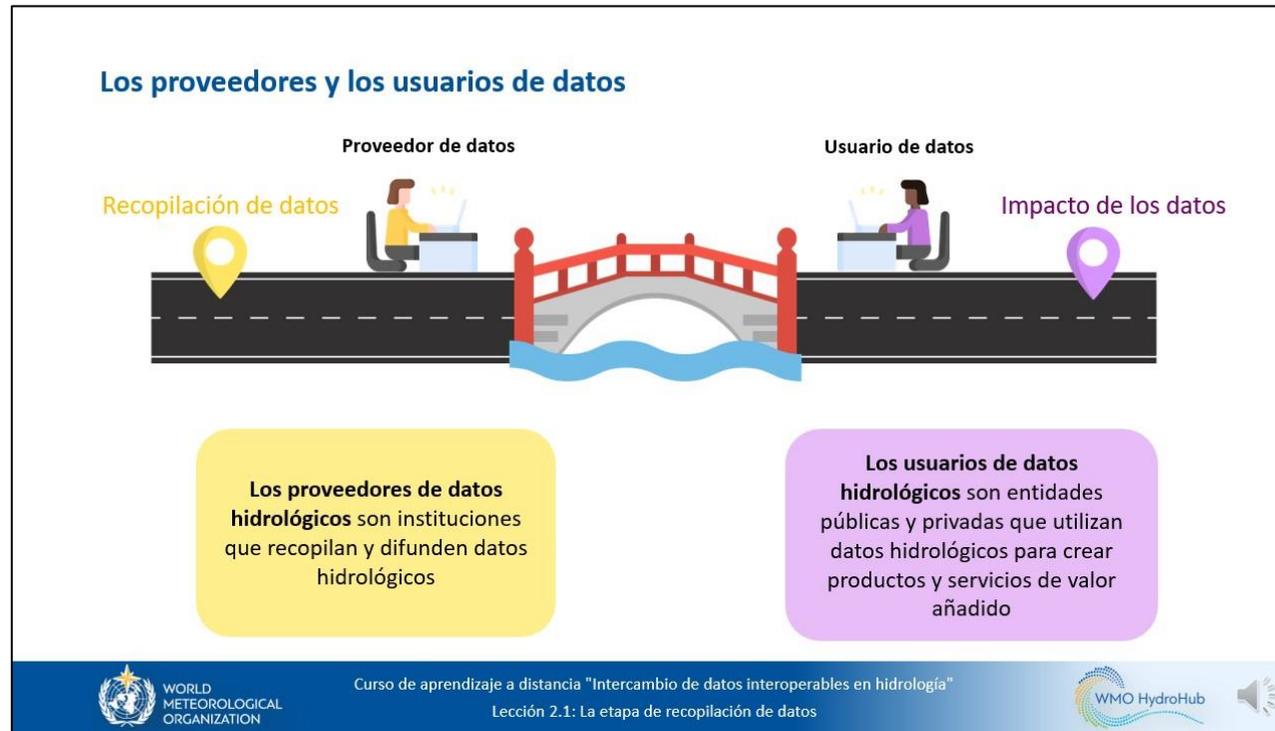
Si los datos no se ponen en común, será imposible que diferentes usuarios reutilicen los datos recopilados, por lo que los datos no adquirirán valor adicional ni aumentará su impacto positivo.

Diapositiva 6



La puesta en común de datos desempeña una función fundamental, puesto que tiende un puente entre proveedores de datos y usuarios de datos, y convierte el viaje de los datos en un proceso más fructífero. Antes de profundizar en los distintos aspectos de la puesta en común de datos, conviene diferenciar los proveedores de datos hidrológicos de los usuarios de datos hidrológicos.

Diapositiva 7



Los proveedores de datos hidrológicos son principalmente los Servicios Hidrológicos Nacionales (SHN), pero también otras instituciones gubernamentales y no gubernamentales, incluidas las entidades del sector privado que recopilan y difunden datos hidrológicos.

Los usuarios de datos hidrológicos son entidades públicas y privadas que utilizan datos hidrológicos para crear productos y servicios de valor añadido que ayuden a adoptar decisiones en todas las cuestiones relativas a la gestión de los recursos hídricos y la reducción del riesgo de desastres, por ejemplo, las predicciones de sequías y crecidas; el diseño de infraestructuras; la gestión de los derechos sobre el agua; la gestión de ecosistemas; la calibración de modelos; y la investigación. Un proveedor de datos también puede ser un usuario que utiliza datos procedentes de otros proveedores de datos.

Diapositiva 8

Preguntas al proveedor de datos

¿Cuáles son las normas sobre metadatos?

¿Cómo deben publicarse y difundirse los datos entre los usuarios?

¿Qué datos se necesitan pero aún no se han facilitado?

¿Se ha alcanzado el nivel necesario de calidad de los datos?

¿Cuáles son los elementos esenciales de los metadatos que hay que documentar?

¿Qué formatos de intercambio de datos deben utilizarse?

¿Qué tipo de datos se deben compartir y con qué fin?

Servicio Hidrológico Nacional X

Múltiples usuarios de datos

Recopilación de datos

Impacto de los datos

WORLD METEOROLOGICAL ORGANIZATION

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

Lección 2.1: La etapa de recopilación de datos

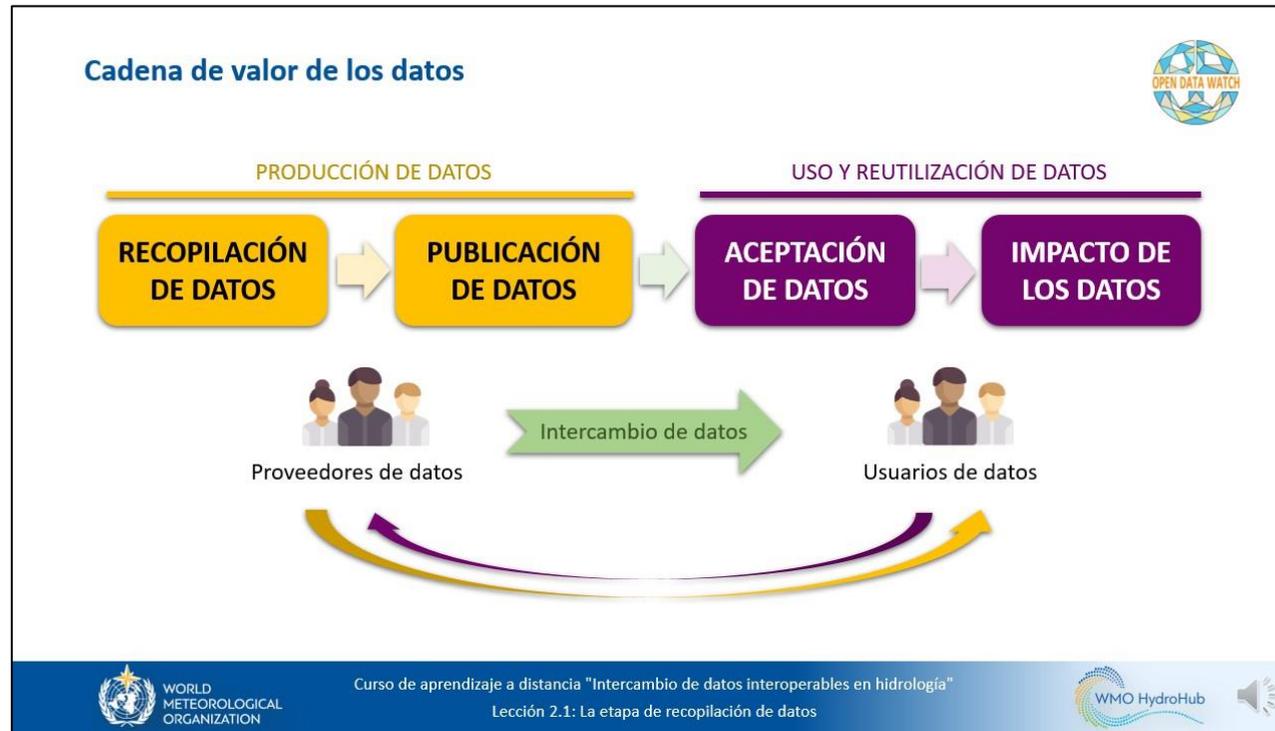
WMO HydroHub

Analicemos ahora un ejemplo en el que el Servicio Hidrológico Nacional del país X (el proveedor de datos) desea poner sus datos en común con múltiples usuarios de forma gratuita y sin restricciones. Teniendo en cuenta que, antes de llegar a su destino final, estos datos tendrán que prepararse y transportarse y que, por último, los usuarios los utilizarán, es posible que surjan múltiples preguntas, tales como:

- ¿Qué tipos de datos deben ponerse en común y con qué fin?
- ¿Los datos han alcanzado el nivel necesario de calidad antes de ponerlos en común?
- ¿Cuáles son los elementos esenciales de los metadatos (datos que describen datos) que hay que documentar antes de ponerlos en común?
- ¿Qué formatos de intercambio de datos deben utilizarse?
- ¿Cuáles son las normas sobre metadatos?
- ¿Cómo deben publicarse y difundirse los datos entre los usuarios?
- ¿Qué son los datos faltantes, es decir, datos que se necesitan pero aún no se han facilitado?

En esta sección del curso, titulada “Función de los proveedores de datos en el marco del intercambio de datos hidrológicos”, le ayudaremos a que halle las respuestas a estas preguntas.

Diapositiva 9



Para responder a estas preguntas, utilizaremos la cadena de valor de los datos presentada anteriormente. A continuación, analizaremos cada paso y descubriremos qué consideraciones importantes deben tener en cuenta los proveedores de datos al poner datos en común.

Nos centraremos en la parte de la cadena de valor dedicada a la producción de datos y destacaremos las funciones y responsabilidades desempeñadas por los proveedores de datos en los pasos previos a la puesta en común. Como se muestra en el gráfico, también cabe señalar que la cadena de valor de los datos puede conllevar que se establezcan comunicaciones bidireccionales. Así pues, los proveedores de datos pueden recibir comentarios de los usuarios sobre las necesidades en materia de datos y la calidad de estos.

Diapositiva 10

La etapa de recopilación de datos



¿Dónde y qué tipo de árboles plantar?



¿Cómo recoger sus frutos?



¿Cómo prepararlos para compartirlos y utilizarlos?



 WORLD METEOROLOGICAL ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"
Lección 2.1: La etapa de recopilación de datos

 WMO HydroHub 

La primera etapa de la cadena de valor de los datos es la recopilación.

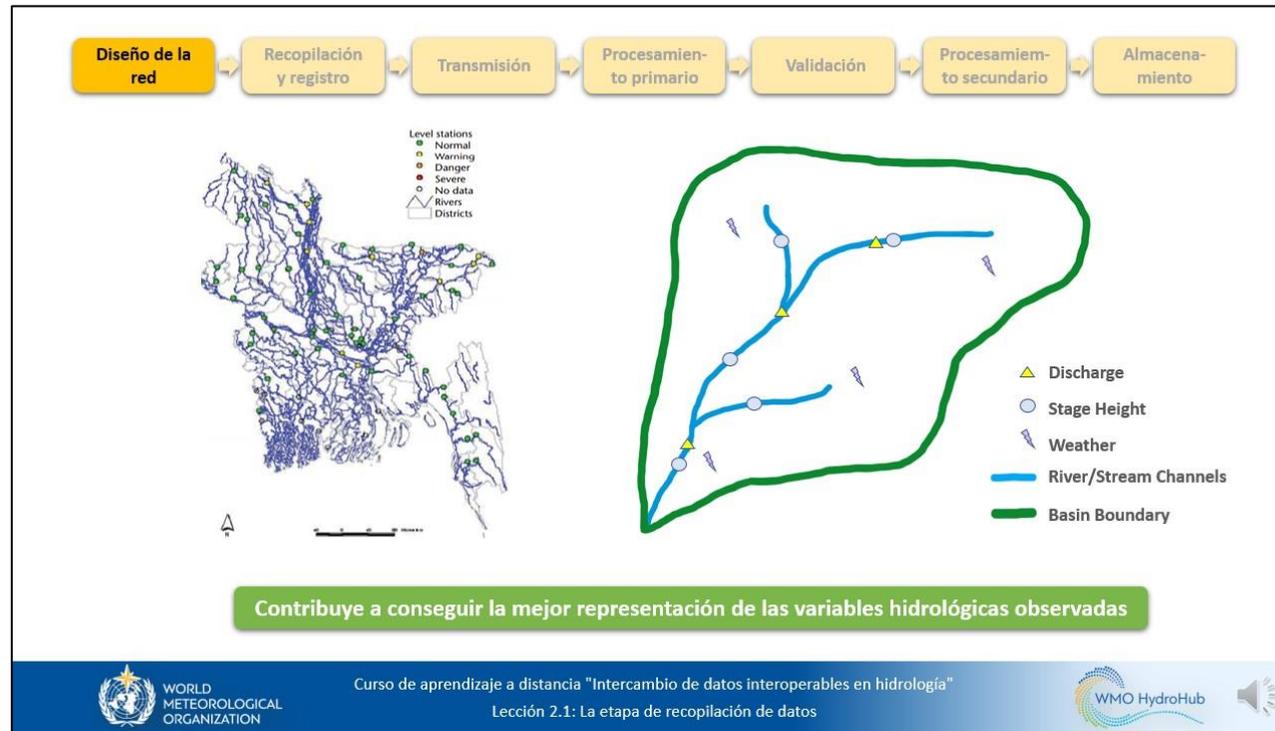
Incluso si los datos cayeran de los árboles, tendríamos que preguntarnos dónde y qué tipo de árboles plantar, cómo recoger sus frutos y cómo limpiarlos y prepararlos para ponerlos en común y utilizarlos.

Diapositiva 11



En hidrología, estas preguntas se responden a través de las siguientes etapas de la recopilación de datos hidrológicos: diseño de redes, recopilación y registro, transmisión, proceso primario, validación, proceso secundario y almacenamiento.

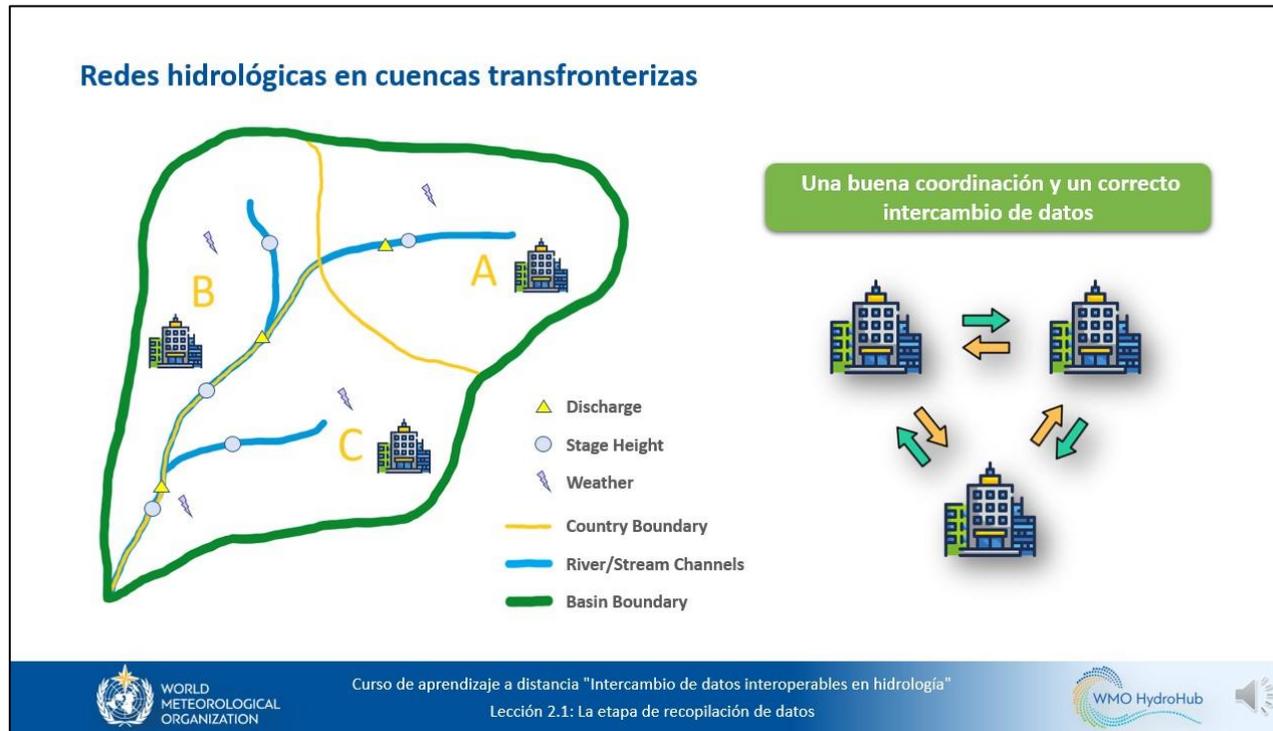
Diapositiva 12



Una red hidrológica bien diseñada contribuye a conseguir la mejor representación de las variables hidrológicas observadas. Una red hidrológica está compuesta por un grupo de estaciones hidrométricas diseñadas y utilizadas para realizar las observaciones necesarias para cumplir objetivos específicos relacionados con el agua, como la evaluación de los recursos hídricos, la predicción de crecidas o la gestión de sequías. El diseño de la red se basa en las necesidades de los usuarios y en las cuestiones prácticas determinadas por el emplazamiento, como el acceso y la disponibilidad de energía eléctrica.

Una estación hidrométrica concreta puede formar parte de más de una red si sus datos contribuyen a las necesidades de varias redes; este caso suele ser el más frecuente en la mayor parte del mundo. Alternativamente, una única red puede constar de varios tipos de estaciones si todas ellas aportan información al objetivo de la red. Así, por ejemplo, una red de alerta de crecidas podría contener tanto limnímetros como estaciones meteorológicas. Tenga en cuenta que en esta lección solo mostramos un ejemplo sencillo de diseño de red, pero existen otros más complejos.

Diapositiva 13



Con frecuencia, la gestión de las redes hidrológicas compete a varias instituciones independientes y, en el contexto transfronterizo, esas instituciones pueden operar en distintos países que compartan una cuenca común. Para lograr el objetivo de conseguir una gestión integrada y sostenible de los recursos, es fundamental que los países ribereños establezcan una buena coordinación y un correcto intercambio de datos, especialmente en el contexto transfronterizo.

Diapositiva 14

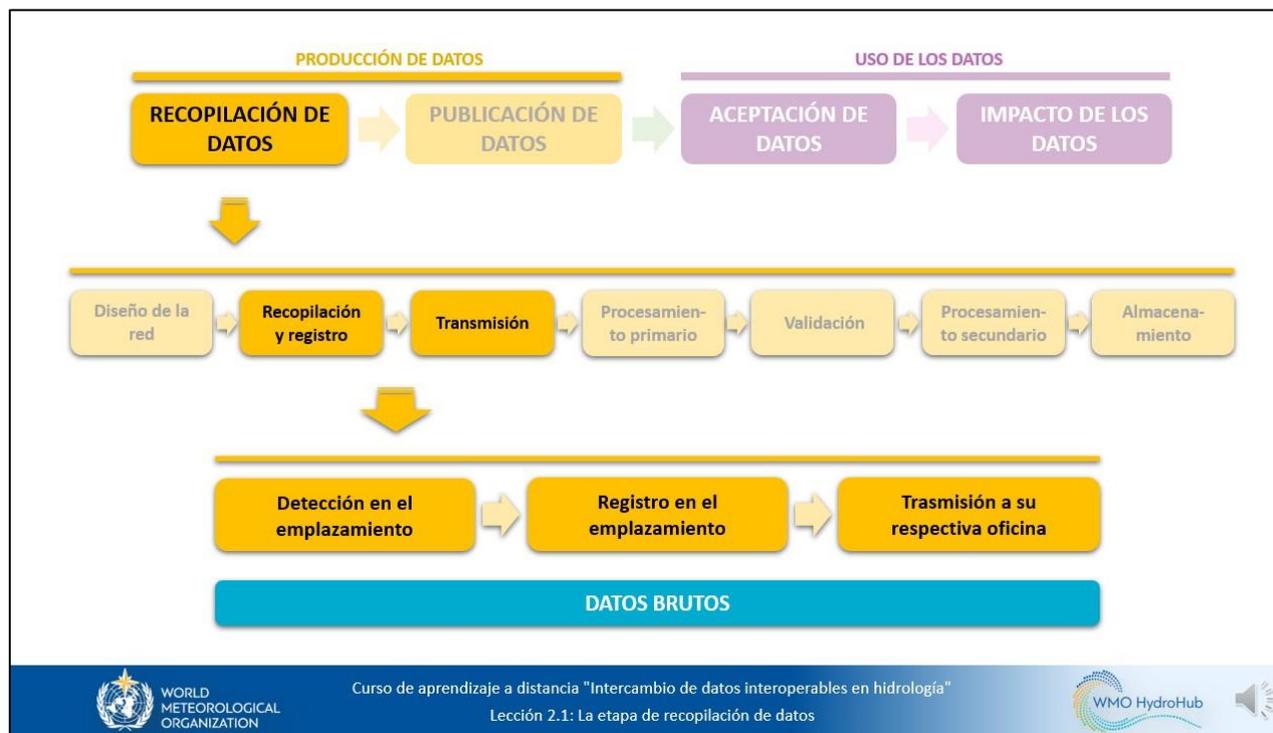


Se deben recopilar dos tipos de datos en cada punto de recolección de datos de una red hidrológica: 1) metadatos de observación; y 2) observaciones hidrológicas (los valores medidos). En cuanto a los metadatos de observación, el elemento más importante es la localización geoespacial, que describe la posición de la estación de observación en el momento de la observación. Los metadatos de observación también deben incluir información sobre el emplazamiento, su cubierta superficial y cualquier otro elemento que pueda influir en las observaciones. Todos estos elementos son fundamentales para que los usuarios puedan interpretar y utilizar correctamente las observaciones.

El segundo tipo de datos recopilados en el emplazamiento son las observaciones hidrológicas. En el caso del emplazamiento de recolección de datos que se muestra en la imagen, la variable observada es el nivel de agua medido por un observador hidrológico mediante una escala limnométrica en un momento determinado.

Es importante señalar que la verificación y la calibración periódicas de los instrumentos sobre el terreno, así como la verificación de las curvas de gasto, son fundamentales para conseguir que los datos tengan una calidad estable antes de ponerlos en común con los usuarios.

Diapositiva 15



En todos los emplazamientos de recopilación de datos se comenzará por detectar las observaciones hidrológicas para, seguidamente, cifrarlas o registrarlas y, por último, transmitir las a la oficina correspondiente. Existen diferentes tipos de detección, registro y transmisión que se describen detalladamente en la *Guía de prácticas hidrológicas* (OMM-Nº 168), volumen I: Hidrología – De la medición a la información hidrológica, disponible en la sección de referencias de este curso.

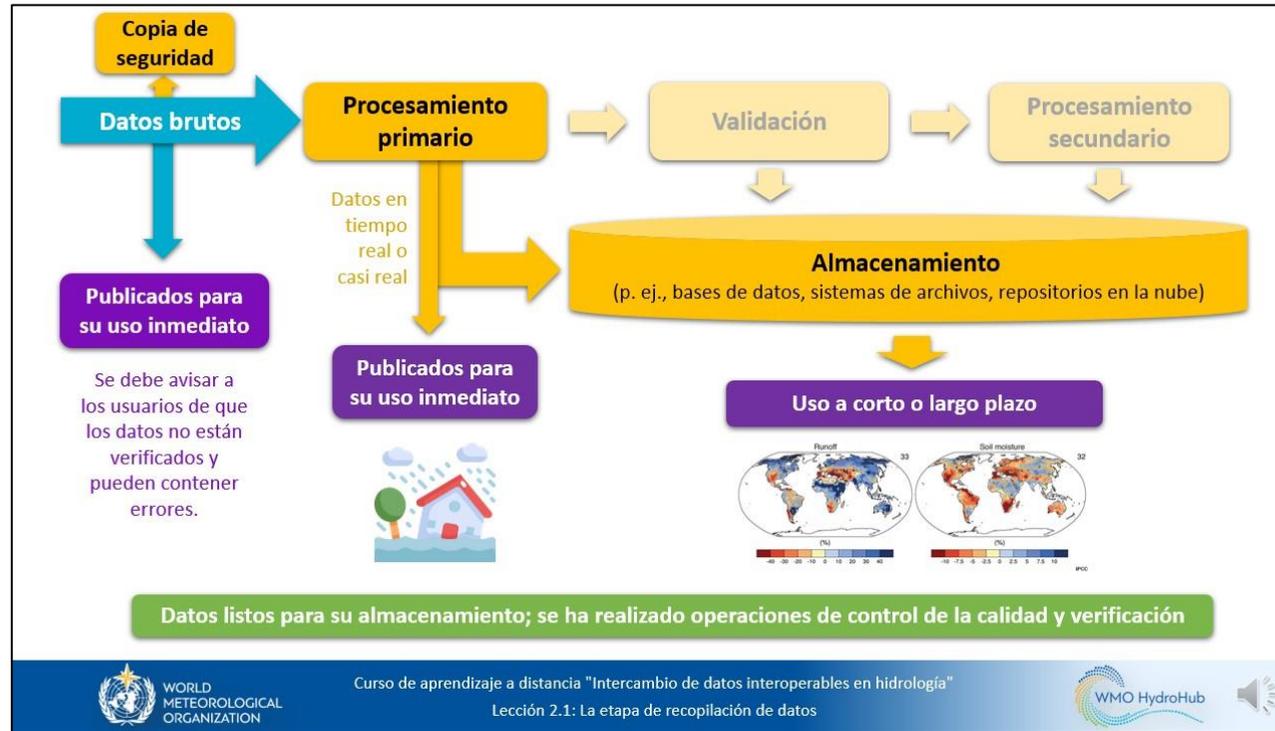
Los datos detectados, registrados y transmitidos son los datos brutos que aún no han sido sometidos a proceso, control de calidad y validación, y deben almacenarse en una base de datos protegida independiente.

Diapositiva 16



Los tres pasos siguientes de la recopilación de datos son el proceso primario, la validación y el proceso secundario.

Diapositiva 17



Como ya se ha mencionado, se debe realizar una copia de seguridad de los datos brutos transmitidos tal y como se reciben del emplazamiento de recopilación de datos antes de someterlos al proceso primario, el cual comprende operaciones de control de calidad y verificación.

El proceso primario, junto con los dos pasos siguientes —validación y proceso secundario—, constituye la interfaz fundamental entre la recopilación de los datos y su uso, ya que permite lograr un almacenamiento eficaz de los datos y vela por su adecuada calidad y por su disponibilidad para fines de puesta en común.

Si concurren circunstancias especiales, como la predicción de una crecida a corto plazo, se pueden publicar y utilizar inmediatamente los datos brutos, antes de someterlos al proceso primario, pero se debe avisar a los usuarios de que los datos no se han verificado y pueden contener errores.

El proceso primario de los datos es necesario para prepararlos para su almacenamiento, por ejemplo, en una base de datos, en un sistema de archivos o en un repositorio en la nube, de modo que puedan utilizarse a corto o largo plazo. Durante ese proceso se efectuarán normalmente operaciones de control de calidad y verificación. En función del tipo de datos, la fase del proceso primario podrá conllevar una serie de operaciones, como la comprobación preliminar de los datos, la identificación y conservación de los registros originales, el ajuste de datos por errores conocidos, la agregación e interpolación de datos, el cálculo de variables derivadas, etc.

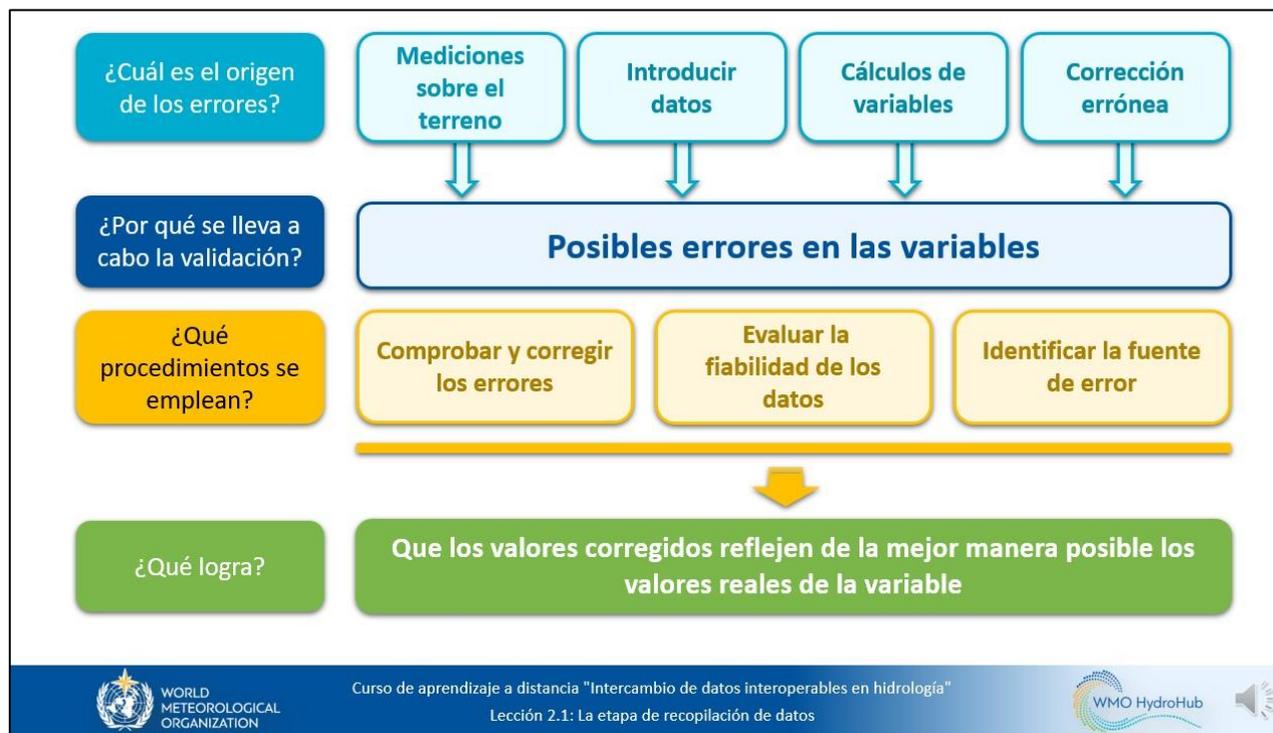
En el caso de los datos en tiempo real o casi real, estos suelen someterse únicamente al proceso primario, probablemente de forma automática, y a un control de calidad relativamente sencillo antes de publicarse para su uso inmediato, por ejemplo, para la predicción de crecidas.

Diapositiva 18



A continuación pasaremos al importante paso de la validación de los datos.

Diapositiva 19



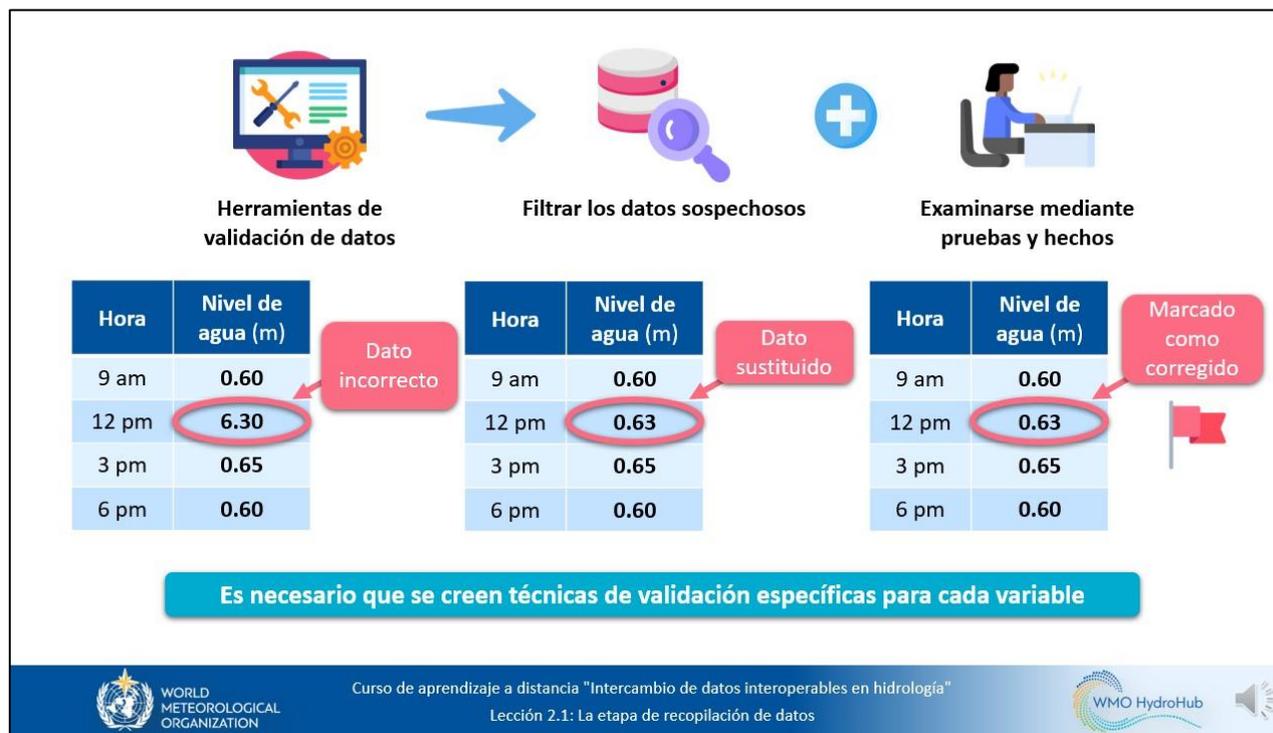
Después del proceso primario se emplean procedimientos de validación de datos. La necesidad de validar los datos surge porque las mediciones sobre el terreno están sujetas a errores, derivas en la calibración de instrumentos y fallos de funcionamiento. También pueden producirse errores al introducir los datos, durante los cálculos y, esperamos que solo en contadas ocasiones, al corregir de forma errónea datos "válidos".

La validación de los datos se lleva a cabo para: 1) corregir, de ser posible, errores en los valores observados; 2) evaluar la fiabilidad de los datos, aunque no sea posible corregir los errores; y 3) identificar la fuente de error para conseguir que este no se repita.

La validación de los datos es la forma de conseguir que los valores corregidos reflejen de la mejor manera posible los valores reales de la variable.

Es importante destacar que, si los datos no se validan, los usuarios corren el riesgo de tomar decisiones basadas en datos erróneos que no representan fielmente la situación en cuestión. Por ejemplo, introducir datos incorrectos en un modelo puede dar lugar a orientaciones erróneas que, en el caso de predicciones de crecidas graves, acarrearían consecuencias devastadoras.

Diapositiva 20



Los procedimientos de validación de datos son, en esencia, herramientas que permiten determinar los datos sospechosos que han de examinarse más a fondo mediante otras pruebas y hechos corroborantes. Cuando se detecte que un valor concreto es incorrecto, se sustituirá por el valor alternativo que pueda acercarse más al valor verdadero de la variable, y este se deberá marcar como corregido utilizando el banderín apropiado. Dado que cada variable hidrológica tiene características diferentes, es necesario que se creen técnicas de validación específicas para cada una de ellas. Cabe destacar que la validación nunca debe considerarse un mero ejercicio estadístico, pues hay que tener siempre en cuenta las propiedades y el comportamiento de la variable en cuestión. Sin embargo, los datos originales y la corrección aplicada se deberán registrar debidamente.

Tenga en cuenta que esta información se facilita únicamente para que conozca mejor el proceso. Para más detalles sobre los procedimientos de validación de datos, consúltese la *Guía de prácticas hidrológicas* (OMM-Nº 168), volumen I: Hidrología – De la medición a la información hidrológica, y el *Manual on Stream Gauging* (WMO-No. 1044) (Manual sobre aforo de caudales), volumen 1 — *Fieldwork* (Trabajo de campo).

Diapositiva 21



Tras la validación de datos, el proceso secundario se considera un paso necesario para producir datos en un formato convertido, resumido o reducido, por ejemplo, el nivel del agua en metros puede ser el valor medido. Durante el proceso secundario, el nivel del agua puede convertirse en caudal fluvial, en metros cúbicos por segundo, mediante una curva de gasto.

Se incluyen también en ese concepto la edición secundaria tras una validación más compleja y la inserción de datos sintéticos (ni medidos ni observados) para completar datos faltantes en el registro. Si no es posible insertar datos sintéticos, los datos faltantes se marcarán con el banderín “*missing*” (faltante), y de ese modo los datos “medidos” se pondrán más claramente de manifiesto.

Además, es posible reagrupar datos e introducir niveles adicionales de cifrado, así como convertir las unidades de medida en las unidades normalizadas, por ejemplo, las adoptadas en la base de datos. La conversión de series temporales irregulares en regulares es también, en muchos casos, una de las operaciones necesarias.

Diapositiva 22

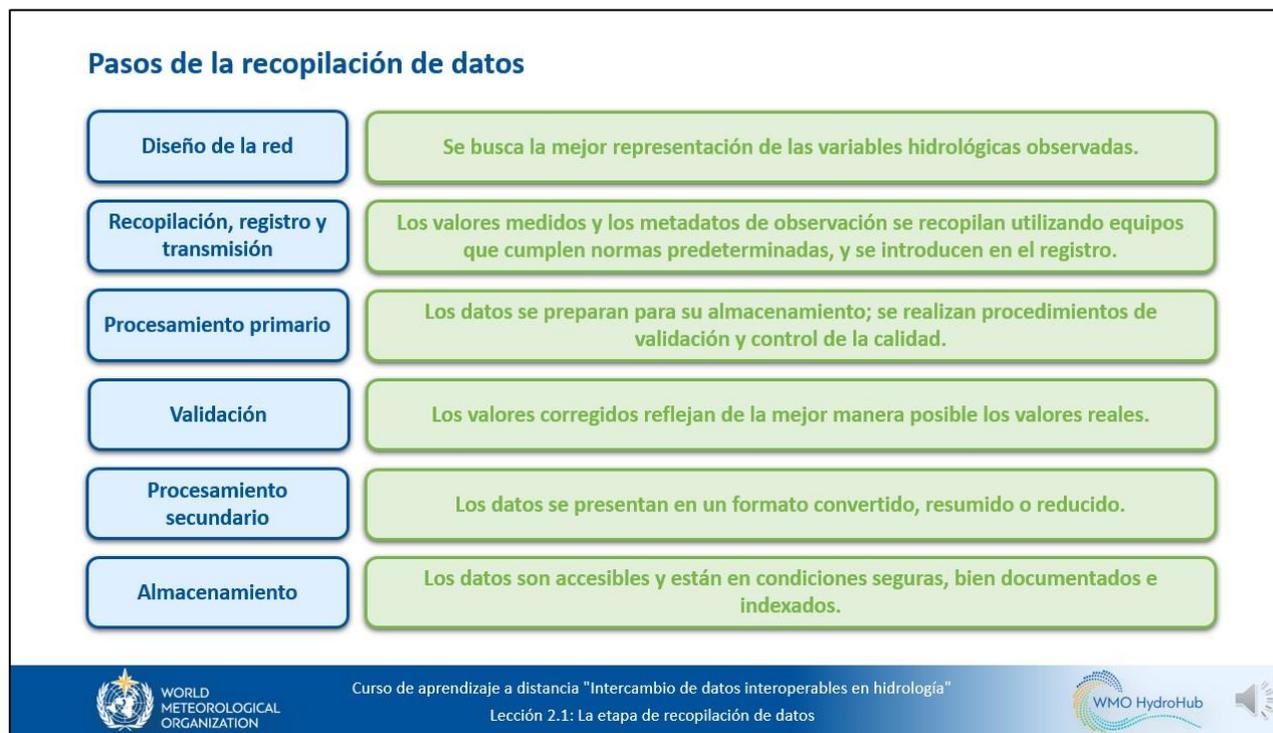


Una vez finalizados el proceso, la validación y los controles de calidad de los datos, estos se almacenarán. Es importante poner de relieve que los numerosos conjuntos de datos producidos en las distintas etapas deberán almacenarse por separado, incluidos los datos brutos tal como se recibieron del terreno.

Los datos deberán almacenarse de modo que sean accesibles, aunque deberán estar protegidos, y también deberán documentarse e indexarse. Los metadatos también deberán archivar, pues permiten un acceso rápido y sencillo a los conjuntos de datos.

Una de las consideraciones más importantes cuando se almacenan datos digitalmente es la solución o arquitectura de almacenamiento de datos utilizada.

Diapositiva 23



En resumen, los objetivos principales de los diversos pasos de la recopilación de datos son los siguientes:

Paso 1 — Diseño de la red: se busca la mejor representación de las variables hidrológicas observadas.

Pasos 2, 3 y 4 — Recopilación, registro y transmisión: los valores medidos y los metadatos de observación se recopilan utilizando equipos que cumplen normas predeterminadas, y se introducen en el registro.

Paso 5 — Proceso primario: los datos se preparan para su almacenamiento; se aplican procedimientos de control de calidad y verificación.

Paso 6 — Validación: los valores corregidos reflejan de la mejor manera posible los valores reales.

Paso 7 — Proceso secundario: los datos se presentan en un formato convertido, resumido o reducido.

Paso 8 — Almacenamiento: los datos son accesibles y están protegidos, bien documentados e indexados.

A continuación estudiaremos la lección 2.2, que trata sobre diferentes soluciones o arquitecturas para el almacenamiento de datos.