



Curso 5

Versión no interactiva

Gestión de datos en un inventario forestal nacional

La versión interactiva de este curso está disponible gratuitamente en la siguiente dirección_
<https://elearning.fao.org/>



Algunos derechos reservados. Esta lección está bajo una licencia CC BY-NC-SA 3.0 IGO
(https://creativecommons.org/licenses/by-nc-sa/3.0/igo/deed.es_ES).

En este curso

Lección 1: Conceptos básicos de la gestión de datos	5
Introducción de la lección	5
Terminología básica de la gestión de bases de datos	5
Requisitos de los sistemas de información forestal	14
Resumen.....	16
Lección 2: Recopilación y gestión de los datos de campo.....	17
Introducción de la lección	17
Herramientas de datos de los IFN: Una visión general	17
Recopilación de datos para los IFN.....	19
Gestión de datos para los IFN.....	21
Validación de datos	24
Resumen.....	29
Lección 3: Elaboración de informes en línea y archivo de datos.....	30
Introducción de la lección	30
Cuadros de mando y portales de datos.....	30
Procesamiento analítico en línea (OLAP)	31
Archivo de datos.....	33
Resumen.....	36

Este curso ofrece una visión general de la recogida de información y la gestión de datos para los inventarios forestales nacionales (IFN).

¿A quién va dirigido este curso?

Este curso está dirigido principalmente a personas implicadas en los IFN, especialmente con interés en los métodos y herramientas de gestión, procesamiento e intercambio de datos de los IFN.

Específicamente, este curso está dirigido a:

1. Responsables de la toma de decisiones que planifican inversiones para ejecutar el IFN.
2. Encargados de la gestión de datos y jefes de equipo de los IFN.
3. Técnicos responsables del análisis de datos.
4. Estudiantes, como material curricular en las escuelas forestales.
5. Jóvenes y nuevas generaciones de forestales.

Estructura del curso

Lección 1: Conceptos básicos de la gestión de datos

Esta lección describe el papel de las Tecnologías de la Información y la Comunicación (TIC) en la gestión de datos forestales. Como esta lección también pretende definir algunos términos forestales relacionados con las TIC, le servirá de introducción a las lecciones más complejas de este curso..

Lección 2: Recopilación y gestión de los datos de campo

Esta lección se centra en la gestión de datos de campo y en los procesos necesarios para que los datos estén listos para su almacenamiento y análisis a largo plazo.


Lección 3: Elaboración de informes en línea y archivo de datos

Esta lección explora la tecnología OLAP (Procesamiento Analítico en Línea) y describe la importancia de los metadatos y microdatos en la recopilación y almacenamiento de información forestal.

Acerca de la serie

Este curso concluye la serie de ocho cursos a su propio ritmo que cubren diversos aspectos de un IFN.

Aquí puede ver la serie completa

Curso	Aprenderá sobre el curso
Curso 1: ¿Por qué un inventario forestal nacional (IFN)?	Objetivos y propósito de un IFN, y cómo los IFN contribuyen a la formulación de las políticas y a la toma de decisiones en el sector forestal.
Curso 2: Preparación de un inventario forestal nacional (IFN)	La planificación y el trabajo necesarios para establecer un IFN eficiente o un Sistema nacional de monitoreo forestal (SNMF).
Curso 3: Introducción al muestreo	Aspectos generales del muestreo en los inventarios forestales.
Curso 4: Introducción al trabajo de campo	Consideraciones para el trabajo de campo, variables a nivel de parcela y mediciones a nivel de árbol.
 Curso 5: Gestión de datos en un inventario forestal nacional	(Este es el curso que está estudiando actualmente).
Curso 6: Garantía de calidad y control de calidad en un inventario forestal nacional	Procedimientos de GC y CC en la recopilación y gestión de datos de inventarios forestales.
Curso 7: Elementos del análisis de datos	Enfoques/cálculos típicos en los análisis de datos y temas relacionados.
Curso 8: Resultados de los inventarios forestales nacionales Presentación de informes y difusión	Presentación de informes de los IFN y la importancia de la presentación de informes en el contexto de las acciones de REDD+.

Lección 1: Conceptos básicos de la gestión de datos

Introducción de la lección

En esta lección comprenderemos el papel de las Tecnologías de la Información y la Comunicación (TIC) en la gestión de datos forestales.

Para ello, analizaremos los términos, técnicas y métodos relacionados con las TIC que se utilizan para gestionar los datos de los IFN. También examinaremos los requisitos de un sistema de información forestal.

Objetivos

Al final de esta lección, usted podrá:

1. Describir la terminología y los conceptos básicos de la gestión de datos en la tecnología de bases de datos.
2. Explicar los requisitos clave de un sistema de información en materia de monitoreo forestal nacional.

Terminología básica de la gestión de bases de datos

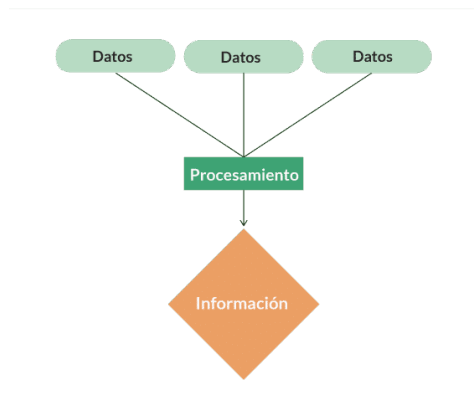
Comprender la terminología clave asociada a la gestión de bases de datos le ayudará a dominar conceptos más complejos a medida que avancemos en el curso. Comencemos este curso examinando la terminología de las TIC que es relevante para la recopilación y organización de datos de los IFN.

Tecnologías de la información y la comunicación (TIC)

Tecnologías de la información y la comunicación (TIC) es un **término genérico** que incluye el diseño, desarrollo, aplicación, soporte y gestión de sistemas de información computarizados. En esencia, las TIC se ocupan del uso de computadoras y programas informáticos para convertir, almacenar, proteger, procesar, transmitir y recuperar información.

Datos

Colección de hechos, como números, palabras, mediciones, imágenes, observaciones o simplemente descripciones de cosas. La **información** se refiere al resultado significativo obtenido tras procesar los datos. Nuestras cifras no cambiarán nada si no las traducimos en historias significativas.



Integridad de los datos

El término integridad de los datos hace referencia a su exactitud y coherencia. Al crear bases de datos, hay que prestar atención a la integridad de los datos y a cómo mantenerla. Una buena base de datos garantizará la integridad de los datos siempre que sea posible.

Por ejemplo, un usuario podría intentar introducir accidentalmente un código de tipo de vegetación en un campo de fecha. Si el sistema aplica la integridad de los datos, evitará que el usuario cometa estos errores, mediante una advertencia o un mensaje de error cuando se introduzcan datos erróneos en un campo específico. Ésta es una de las ventajas de utilizar formularios de campo en línea en la recopilación de datos de campo. Por supuesto, no todas las entradas erróneas pueden ser identificadas por esta comprobación formal.

Gestión de datos

Proceso administrativo que incluye la adquisición, la validación, el almacenamiento, la protección y el procesamiento de datos para garantizar su accesibilidad, fiabilidad y vigencia para los usuarios. Un principio fundamental de la gestión de datos forestales es almacenar todos los datos tal y como se recogieron en su forma original. Esto permite flexibilidad en la forma de procesar los datos y garantiza que todos los cálculos se reproduzcan a partir de los datos originales. Los datos originales también sirven como "datos de base" y pueden ser importantes para contrarrestar los errores detectados durante la validación y/o el análisis de los datos.

Base de Datos y Sistema de Gestión de Bases de Datos (SGBD)

Una base de datos es una **colección de datos almacenados en un sistema informático de forma**

organizada para facilitar el acceso y la gestión. Las bases de datos son básicamente contenedores de datos. En los IFN, las bases de datos se utilizan para gestionar y archivar datos del inventario de campo, fotografías de campo, mapas y datos de teledetección, así como documentos relacionados (como manuales de campo, directrices, informes del inventario).

Los Sistemas de Gestión de Bases de Datos, comúnmente denominados SGBD, son **programas informáticos que permiten realizar diversas operaciones con las bases de datos**. Los SGBD permiten a los usuarios acceder a las bases de datos, así como manipular, elaborar informes y representar datos. También ayudan a controlar el acceso a la base de datos. Algunos ejemplos de software de bases de datos o SGBD populares son MySQL, Microsoft Access, Microsoft SQL Server, PostgreSQL, FileMaker Pro, Oracle Database e InterBase.

Lenguaje de acceso a las bases de datos

Un lenguaje de acceso a bases de datos se utiliza para **acceder, ingresar, actualizar o recuperar datos** del SGBD. Es una herramienta informática clave para organizar datos, crear bases de datos y controlar datos mediante lenguajes de consulta. Existen varios lenguajes de este tipo que pueden utilizarse con este fin, entre ellos SQL (Lenguaje de Consulta Estructurada).



Videos

[What is Database & SQL? *https://www.guru99.com/introduction-to-database-sql.html*](https://www.guru99.com/introduction-to-database-sql.html)

[Learn Basic SQL in 10 Minutes \(*https://www.youtube.com/watch?v=bEtnYWuo2Bw&t=1s*\)](https://www.youtube.com/watch?v=bEtnYWuo2Bw&t=1s)

Esquema

La **estructura de la base de datos** se denomina esquema y define qué tipo de datos se almacenan. Puede pensar en ella como la estructura de un esqueleto que representa la vista lógica de toda la base de datos. Define cómo se organizan los datos y cómo se relacionan sus componentes. Formula todas las restricciones que deben aplicarse a los datos.

Los factores clave que hay que tener en cuenta sobre los datos a la hora de diseñar una base de datos son:

1. **Tipos de datos**, que se utilizan en cada campo y clasifican valores de datos que tienen

propiedades comunes.

2. **Validación**, que son las reglas para aceptar datos.
3. **Campo clave, o la clave primaria** (aprenderás más sobre claves más adelante en esta lección).

Los desarrolladores planifican un esquema de base de datos con antelación, para saber qué componentes son necesarios y cómo se conectarán entre sí. Esto también es importante en el contexto de los IFN: diseñar un esquema de base de datos inteligente facilitará los análisis.

Entidad y atributo

Una **entidad** de datos es un **objeto de un repositorio de datos**. Las entidades de datos son los objetos del modelo de datos, como "árbol" o "suelo". Las entidades no representan ningún dato por sí mismas, sino que son contenedores de atributos y relaciones entre objetos.

Un **atributo** de datos es una **unidad de información dentro de una entidad de datos**, es decir, es un descriptor de un solo valor para un punto de datos u objeto de datos. Son como propiedades de las entidades de datos: por ejemplo, una entidad "árbol" puede contener los siguientes atributos: especie, altura y edad.

Modelos de datos

Los modelos de datos definen la **estructura lógica de una base de datos**. Definen cómo se conectan los datos entre sí y cómo se procesan y almacenan dentro del SGBD.

Un modelo de datos ayuda a percibir, organizar y describir los datos en un esquema conceptual que incluye tanto los datos como las operaciones para manipular el conjunto de datos. Es conveniente considerar ahora los dos principales modelos de datos: el **Modelo Entidad-Relación**; y el Modelo Relacional.

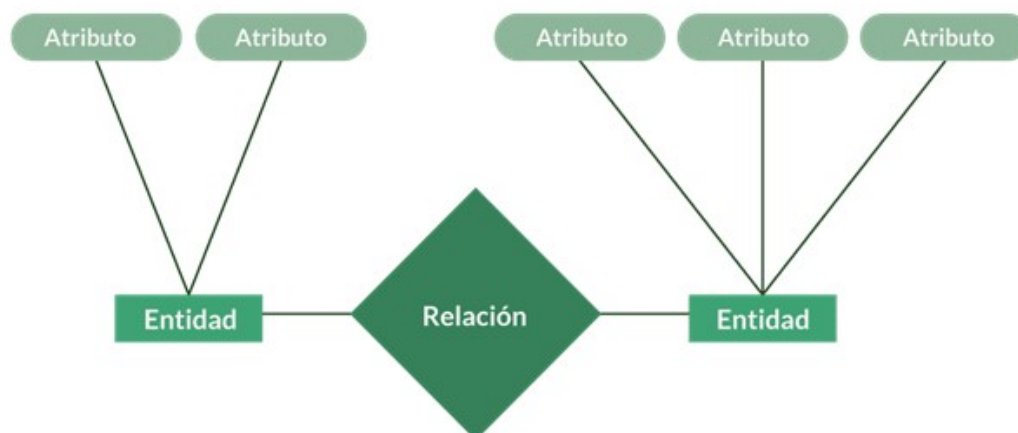
Modelo entidad-relación

Como su nombre indica, el Modelo Entidad-Relación (ER) se **basa en la noción de entidades del mundo real y las relaciones entre ellas**. Al formular escenarios del mundo real en el modelo de base de datos, el Modelo ER crea un conjunto de entidades, un conjunto de relaciones, atributos generales y restricciones.

El Modelo ER es el más adecuado para el diseño conceptual de una base de datos. El diseño conceptual es la primera etapa del proceso de diseño de una base de datos. El objetivo en esta etapa es diseñar una

base de datos que sea independiente del software de base de datos y de los detalles físicos. El resultado de este proceso es un modelo conceptual de datos que describe las principales entidades, atributos, relaciones y restricciones de datos de un determinado dominio de problemas.

Veamos una representación visual del modelo entidad-relación.



Modelo relacional

Las bases de datos relacionales, que son el **modelo de datos más utilizado en los SGBD**, se han convertido en la tecnología líder para gestionar bases de datos, aplicables desde smartphones y computadoras personales hasta mainframes. El modelo de datos relacional **representa los datos en forma de tablas**. Una relación es una tabla de datos bidimensional con nombre. Cada relación (o tabla) consta de un conjunto de columnas con nombre y un número arbitrario de filas. Un atributo es una columna identificada de una relación. Cada fila de una relación corresponde a un registro que contiene valores de datos (atributos) de una entidad concreta. Veamos una representación visual de una base de datos relacional.

id	species_name	dbh	tree_height_calc	tree_basal_area	tree_volume_stem	tree_volume_bole	tree_biomass_ag	tree_biomass_bg
1001	Trichila prieuriana	28	12	0.061575216	0.038792386	0.024630086	0.041559922	0.011221179
1002	Acacia nilotica	40	18	0.125663706	0.118752202	0.100530965	0.123853538	0.033440455
1003	Acacia nigrescens	22	14	0.038013271	0.027939754	0.025145779	0.030169727	0.008145826
1004	Trichila emetica	24	14	0.045238934	0.033250617	0.029925555	0.035754825	0.009653803
1005	Toona ciliata	28	10	0.061575216	0.032326988	0.024630086	0.034785146	0.009391989
1006	Trichila prieuriana	32	10	0.080424772	0.042223005	0.032169909	0.045143384	0.012188714
1007	Trichila dregeana	32	14	0.080424772	0.059112207	0.048254863	0.062692425	0.016926955
1008	Acacia nilotica	28	14	0.061575216	0.045257784	0.024630086	0.048307525	0.013043032
1009	Trichila emetica	68	18	0.363168111	0.343193865	0.217900866	0.348935144	0.094212489
1010	Trichila emetica	20	16	0.031415927	0.026389378	0.025132741	0.02853468	0.007704364

■ Atributo ■ Columna ■ Registro (fila)

Vista de una relación (tabla) en una base de datos

Un registro es una fila completa de elementos de datos en una tabla. Veamos un ejemplo. El árbol número 1005 de la tabla tiene un registro (resaltado en naranja en la tabla anterior) en el que todos los datos de entrada se registran en una tabla. Esto difiere de una tupla (que suele surgir a través de una consulta), que es un conjunto de registros que incorpora datos de múltiples tablas. Los datos de una tupla tienen en común un determinado valor de referencia, por ejemplo, todos los registros variables (volumen, biomasa, valor del tronco, etc.), a través de distintas tablas, pertenecen al árbol número 1005. Por lo tanto, los datos, procedentes de varias tablas, que forman una tupla, se asocian todos a un mismo árbol.



¿Sabía qué?

¿Cuáles son las principales características de los modelos de bases de datos relacionales?

Las principales características de este modelo son:

- Los datos se almacenan en tablas (denominadas relaciones).
- Las relaciones se pueden "normalizar".
- Cada relación de una base de datos debe tener un nombre distinto o único.
- Una relación no puede tener dos atributos con el mismo nombre. Cada atributo debe tener un nombre diferente.
- Dentro de una relación, cada fila debe ser única.
- Cada columna de una relación contiene valores del mismo dominio. Un dominio es un conjunto único de valores permitidos para un atributo en una tabla. Por ejemplo, un dominio de mes del año puede aceptar enero, febrero y hasta diciembre como valores posibles, y un dominio de números enteros puede aceptar números enteros negativos, positivos y cero.



Videos

Vea un vídeo (en inglés) que explica la normalización de bases de datos para principiantes.

[Basic Concept of Database Normalization - Simple Explanation for Beginners](#)

[Database Schema: Entity Relationship Diagram](#)

[Converting ER Diagrams to Schemas | SQL | Tutorial 23](#)

Tipos de datos

La tipificación de datos es una forma de **clasificar los valores de datos que tienen propiedades comunes**.

Cuando se crea una base de datos, es necesario establecer tipos de datos para cada campo. Por ejemplo, en una base de datos de inventario forestal puede necesitar campos de texto para "Nombre del Registrador", pero números para "Edad del Árbol". Los campos suelen estar restringidos a un tipo de datos específicos.

Los distintos tipos de valores de datos también necesitan distintas cantidades de memoria para almacenarlos y tienen distintas operaciones que pueden realizarse con ellos. Por ejemplo, si el tipo de datos es una cadena de caracteres, puede introducir números, pero no podrá realizar cálculos, ya que el software entiende las cifras como caracteres. Los tipos de datos más compatibles son:

- **Enteros** o números enteros, como 5, 27, 65575.
- **Números de coma flotante** (con decimales, a veces llamados números reales o flotantes), por ejemplo 5,2; 37,4; 196,247.
- **Caracteres** como c, F, 3 \$, #.
- **Cadenas de caracteres** como abc, deu496, 3erf08!@.
- **Valores booleanos**, por ejemplo, Verdadero/Falso, o Sí/No.

Una base de datos de inventario forestal suele contener todos estos tipos de datos. Además, también puede contener archivos (fotos, videos, archivos de audio, documentos digitales), fechas, horas y tipos complejos (es decir, atributos combinados). Los ejemplos de tipos de datos complejos en una base de datos de inventario forestal se dividen en dos categorías principales: **datos de coordenadas** y **datos taxonómicos**. Leamos un poco más sobre estos dos.

los **datos de coordenadas** (puntos) pueden consistir en un sistema de referencia de coordenadas horizontal (SRC), una coordenada X y una coordenada Y. El SRC vertical contiene la tercera dimensión, como Z o altura o altitud. En los SRC compuestos pueden unirse datos de sistemas horizontales y verticales. Nota: Sistema de referencia de coordenadas (SRC) y sistema de referencia espacial (SRE) son sinónimos y suelen intercambiarse. Es muy importante no sólo especificar las coordenadas, sino también indicar qué sistema de referencia se ha utilizado. Sin esta información, los valores de las coordenadas carecen prácticamente de sentido.

Los datos taxonómicos pueden contener datos sobre familia, género, especie, subespecie, variedad. Los datos taxonómicos también pueden contener la cita del autor que hace referencia a la lista de la persona (o equipo) que pone a disposición por primera vez el nombre científico de un taxón.



Consejos prácticos

Además, una base de datos del IFN se puede ampliar a una **base de datos espacial** que almacene datos sobre objetos cartografiados (en forma de puntos, líneas y polígonos). Estas bases de datos permiten utilizar consultas espaciales y análisis espaciales con los datos de los IFN.

Identificador y clave primaria

Cada entidad requiere una etiqueta que la identifique de forma única. En una base de datos relacional, esto se denomina **clave primaria**. Por lo tanto, una clave primaria es **una columna (o combinación de columnas) especial de una tabla de base de datos relacional designada para identificar de forma exclusiva cada registro de la tabla**. Sin los conceptos de clave primaria y clave externa, estrechamente relacionados, las bases de datos relacionales no funcionarían. Por eso es necesario definir una clave primaria para cada entidad cuando se construye un esquema de base de datos. Recuerde que puede haber más de una clave primaria para una entidad.

Base de datos de archivos planos y almacén de datos

Una base de datos de archivos **planos almacena los datos en un archivo de texto sin formato**. Cada línea del archivo de texto contiene un registro, con campos separados por delimitadores, como comas o tabulaciones.

Aunque utiliza una estructura simple, una base de datos de archivos planos no puede contener múltiples tablas como una base de datos relacional. Las bases de datos de archivos planos fueron desarrolladas e implementadas a principios de los años 70 por IBM. Los proyectos de almacenamiento de datos utilizan archivos planos para importar datos.

Los almacenes de datos **recopilan contenidos de varias bases de datos operativas**, como las personales, las de grupos de trabajo, las de departamentos y las de planificación de recursos empresariales (ERP). Son sistemas utilizados para la elaboración de informes y el análisis de datos y se consideran un componente esencial de la inteligencia empresarial.

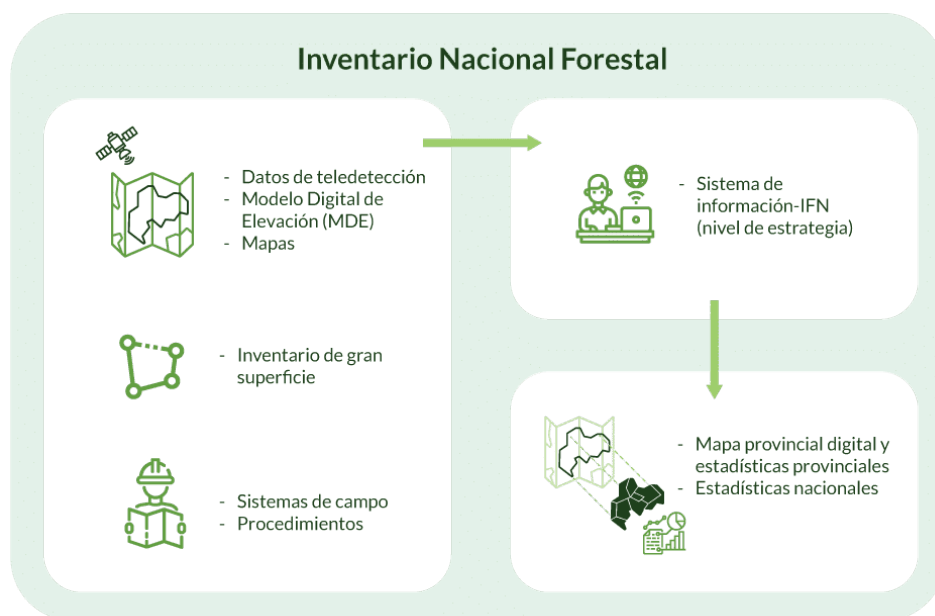
Ejemplo de país: Instituto de Investigación Forestal de Corea

Este ejemplo muestra cómo diseñar un almacén de datos para crear una infraestructura de intercambio

de datos espaciales y de estudios de campo en un sistema de información forestal. Como se ve en el ejemplo, la arquitectura de un almacén de datos puede ser compleja. En la actualidad, los sistemas de gestión de datos de los IFN más grandes de los países con un SNMF institucionalizado (por ejemplo, NAFIDAS para el IFN suizo y TaxWebb en Suecia), pueden compararse con una arquitectura de almacén de datos. Ninguno de estos sistemas es mejor o peor que el otro, y la decisión de qué sistema es adecuado para cada situación local se basa totalmente en los recursos disponibles y las necesidades.

Requisitos de los sistemas de información forestal

Los sistemas de información forestal proporcionan una plataforma para ofrecer a los responsables de la toma de decisiones información exhaustiva y de alta calidad que ha sido sometida a procesos de control de calidad. Los IFN y otros datos de los inventarios forestales pueden considerarse elementos esenciales dentro de este concepto.



Un sistema moderno de gestión de la información sobre recursos forestales requiere lo siguiente:

1. **Adaptabilidad a diferentes entornos de hardware y software**, diferentes condiciones y diferentes escalas geográficas de inventario.
2. **Capacidad de utilizar los datos existentes** en todas las fases del inventario.
3. **Escalabilidad** de ser transferible a un sistema operativo más grande donde pueda aprovechar al máximo dicho sistema en términos de rendimiento.

4. **Flexibilidad para dar cabida** a diversos modelos de datos.
5. **Adaptabilidad para alojar datos de IFN repetidos** con listas de variables posiblemente cambiantes y otras adaptaciones.

Además, el sistema debe ser seguro, fácil de usar, accesible en línea y ofrecer herramientas flexibles para la elaboración de informes. La documentación de los sistemas y sus procedimientos debe ser transparente y permitir acomodar futuras mejoras.

Competencias necesarias para la gestión de datos

La gestión y el procesamiento de datos forestales requieren personal con sólidos conocimientos sobre inventarios forestales, información relativa a las características de los bosques (y la naturaleza) y conocimientos sobre las herramientas y los métodos necesarios para realizar todas las tareas relacionadas con estas áreas. Básicamente, se requieren conocimientos específicos en los tres campos que se indican a continuación:

C conocimientos en estadística	Los conocimientos estadísticos, en particular, el muestreo estadístico y la modelización estadística, son necesarios para definir los procedimientos y las fórmulas matemáticas en las fases de planificación del análisis de datos. El experto debe comprender los objetivos generales del IFN, participar en las fases iniciales de planificación y tener un conocimiento claro de las variables utilizadas.
Diseño de bases de datos	Para diseñar, crear y mantener la base de datos, el diseñador de la base de datos no sólo debe programar e implementar el sistema al principio, sino que también debe mantener y mejorar el sistema de la base de datos, lo que requiere calificaciones informáticas específicas para el cargo.
Administración de la base de datos	La administración de la base de datos requiere recibir datos desde el terreno, supervisar su introducción en el sistema, llevar a cabo la comprobación y corrección de errores y gestionar las consultas ad hoc de los usuarios. El administrador de la base de datos debe

tener conocimientos informáticos y de bases de datos, y también es importante que haya participado en el trabajo de campo y esté familiarizado con los procedimientos utilizados, así como con los objetivos generales del inventario.

Resumen

Antes de finalizar, aquí están los puntos clave de aprendizaje de esta lección:

- Un Sistema de Gestión de Bases de Datos (SGBD) es un conjunto de programas que permite a sus usuarios acceder a bases de datos, manipularlos, elaborar informes y representarlos. Los SGBD son la tecnología predominante para toda la gestión de datos.
- El Modelo Entidad-Relación es el más adecuado para el diseño conceptual de una base de datos. Las bases de datos relacionales se han convertido en la tecnología dominante para la gestión de bases de datos.
- Al planificar una base de datos, necesitamos diseñar un esquema que contenga entidades, atributos, relaciones, campos clave y reglas de validación. La gestión y el procesamiento de datos forestales requieren personas con buenos conocimientos sobre inventarios forestales, características de la información forestal y conocimientos sobre herramientas y métodos. Las funciones asociadas a las competencias necesarias son las de biométrico o estadístico forestal, diseñador y mantenedor de bases de datos, y analista de datos/administrador de bases de datos.

Lección 2: Recopilación y gestión de los datos de campo

Introducción de la lección

Esta lección se centra en la gestión de los datos de campo y describe los procesos necesarios a fin de que estos datos estén listos para su almacenamiento y análisis a largo plazo.

También explica cómo construir una base de datos operativa para alojar los datos del inventario forestal. Dado que la calidad de los datos es siempre una prioridad, esta lección también revisa los procesos de validación de datos.

Objetivos

Al final de esta lección, usted podrá:

1. Describir cómo se recopilan, almacenan y gestionan los datos de campo.
2. Identificar los requisitos específicos de los sistemas de las TIC para el inventario y el monitoreo forestal.
3. Explicar cómo construir una base de datos operativa para alojar los datos del inventario forestal.
4. Identificar métodos para mejorar la calidad general de una base de datos del IFN.

Herramientas de datos de los IFN: Una visión general

Elección de herramientas de TIC para la gestión de datos de los inventarios forestales

Desde un punto de vista tecnológico, ahora es fácil adquirir los equipos de TI de oficina necesarios, ya que las computadoras de escritorio y portátiles de oficina son capaces de gestionar los datos de los IFN.

A nivel corporativo, se necesitan servidores locales o servicios en la nube (alojados por un proveedor de servicios en la nube). En la mayoría de los casos, la principal razón para adquirir un servidor es el aumento del número de puestos de trabajo del personal que utiliza la red de forma intensiva.

Supongamos que prevé aumentar a más de siete el número de trabajadores de una oficina, cada uno de los cuales utiliza una computadora. En ese caso, es posible que desee considerar un servidor para gestionar mejor su fuerza de trabajo y los datos que producen.

Los **equipos de campo de los inventarios forestales realizan la mayor parte de sus tareas al aire libre**, exponiendo sus equipos a condiciones meteorológicas variables y a una manipulación poco adecuada.

Por eso, **trabajar con computadoras resistentes es una ventaja**. Sin embargo, los dispositivos resistentes suelen ser más caros que los dispositivos estándar.

Uso de tecnologías de pantalla táctil

Existen dos variedades de tecnologías de pantalla táctil que se ofrecen actualmente: la **táctil resistiva**, que utiliza la **presión del cuerpo humano como input**, y la **táctil capacitiva**, que utiliza las **propiedades eléctricas del cuerpo humano como input**. A la hora de elegir entre estas dos tecnologías, hay que tener en cuenta el entorno en el que se va a utilizar el dispositivo.

Si su dispositivo se va a utilizar en condiciones más adversas y lluviosas, que pueden requerir limpiar la pantalla reiteradamente, **un panel táctil resistivo puede funcionar mejor**, ya que puede utilizarlo tanto con las manos protegidas por guantes como con un lápiz táctil.

Si, por el contrario, **va a utilizar el dispositivo para aplicaciones más sofisticadas**, opte por un panel táctil capacitivo. Con los paneles táctiles capacitivos, la multitáctil no es un problema, puede desplazarse con facilidad y tiene una sensibilidad excelente.

Aplicaciones móviles de recopilación de datos

Existen varias aplicaciones para la recopilación móvil de datos, entre ellas:

[Open Data Kit \(ODK\)](#) (en inglés)

[Survey Solutions](#) (en inglés)

Open Foris (OF) [Collect Mobile](#) (en inglés)

[ArcGIS Survey123](#) (en inglés)

Algunas aplicaciones son más adecuadas para entrevistas y encuestas socioeconómicas (como ODK y Survey Solutions), mientras que otras funcionan mejor para los estudios sobre la tierra (por ejemplo [ArcGIS AppStudio](#) en inglés and [Q Field](#))— que es una herramienta gratuita diseñada específicamente para manejar datos de inventarios forestales.

Presentación: Open Foris Collect y Collect Mobile

Open Foris (OF) Collect es una aplicación móvil para Android que permite recopilar datos para encuestas sobre el terreno. Funciona como una solución de gestión de datos de campo y permite la personalización completa de la estructura del inventario de datos, las variables y las reglas de validación

de datos. La Interfaz de Usuario (UI) de entrada de datos de OF se genera automáticamente.

El sistema puede utilizarse en un entorno autónomo -sin necesidad de conexión a Internet- o puede instalarse en un servidor en el que varios usuarios puedan trabajar simultáneamente con acceso a los mismos datos de las encuestas. En un proyecto de mayor envergadura, como ocurre con los IFN, es necesaria la instalación de un servidor, que proporciona el espacio de almacenamiento para alojar los datos.

OF también permite la validación sobre la marcha para mejorar la calidad de los datos y se integra con OF Collect para la gestión y el análisis de datos.

Recopilación de datos para los IFN

Los datos de campo se pueden recopilar y registrar utilizando dispositivos electrónicos o formularios de campo impresos. Los registradores de datos móviles se han utilizado en los IFN desde finales de la década de 1980 y, en la actualidad, las tablets resistentes o los smartphones se utilizan habitualmente para recopilar datos de campo. La introducción de datos en una computadora tras la evaluación de campo se puede hacer manualmente (con un teclado), a través de un cable o de forma inalámbrica desde un dispositivo electrónico.

Si se dispone de comunicación móvil sobre el terreno, los datos pueden transferirse directamente a la base de datos central o a la nube. En cualquier caso, **es ventajoso realizar la introducción de datos lo antes posible (en términos de tiempo y espacio) al lugar donde se generan los datos**. Los métodos para comprobar los datos directamente sobre el terreno permiten a los equipos de campo corregir errores e incoherencias inmediatamente, lo que elimina fuentes de error y mejora la calidad de los datos.

La comunicación por teléfono móvil aumenta la seguridad de los equipos de campo, ya que pueden comunicarse más fácilmente en situaciones de emergencia, y también permite la introducción de datos en línea en una base de datos central. Sin embargo, la cobertura de las redes de telefonía móvil es escasa en muchas regiones, sobre todo en las zonas rurales poco pobladas donde se pueden encontrar las parcelas de campo, lo que hace técnicamente imposible la comunicación móvil directa desde el terreno. En tales situaciones, es crucial que el jefe del equipo compruebe los datos recogidos y haga copias de seguridad utilizando dispositivos de almacenamiento portátiles, y más tarde, cuando el equipo llegue a las zonas con conexión móvil, los datos se puedan cargar o enviar a la base de datos central.



Consejos prácticos

Los inventarios forestales suelen requerir la recopilación de datos de atributos que contienen información taxonómica y de localización espacial (puntos). Los inventarios forestales suelen requerir la recopilación de datos de atributos que contienen información taxonómica y de localización espacial (puntos). SQLite es una base de datos multiplataforma muy popular en dispositivos móviles.

Veamos ahora el proceso de flujo de datos desde el terreno hasta una base de datos depurada y lista para el análisis.

Datos del formulario de campo

Mientras se procesan los datos de los formularios de campo, se recomienda mantener fotocopias de estos en una oficina local, si es posible. Los formularios de campo se transfieren a la oficina principal y los datos se introducen manualmente en la base de datos.

Hay dos opciones para transportar los datos a la base de datos principal.

Opción 1: Después de un día de trabajo, los datos se exportan de la tablet y se copian en una computadora. Se almacena una "copia de seguridad" de los datos. A continuación, el archivo de datos se envía a la oficina principal por correo electrónico, almacenamiento en la nube o se transporta en una memoria USB.

Datos principales

Los datos principales se componen de:

Datos de entrada: Se validan, se elimina la "lista de errores" y se comprueban todas las entradas erróneas.

Datos de validación: Se corrigen todos los errores y se depuran los datos. A continuación, se puede avanzar a la siguiente fase.

Análisis de datos: Los datos depurados están listos para el análisis.

Transferencia e introducción de datos

La transferencia de datos desde tablets o smartphones se puede organizar de muchas maneras.

1. En las aplicaciones en la nube, los datos se almacenan automáticamente en el servidor en la nube.
2. Supongamos que primero se recogen los datos offline y el dispositivo dispone de conexión a Internet. En ese caso, los datos se pueden almacenar en el servidor en la nube (Google Drive, OneDrive, DropBox, etc.) o transferirse por correo electrónico a la oficina.
3. En algunos sistemas, los datos recogidos también se pueden exportar primero al almacenamiento de la tablet y, a continuación, copiarse o enviarse de forma inalámbrica (por ejemplo, mediante Bluetooth) a una computadora portátil y transferirse posteriormente a la base de datos central.
4. En algunos sistemas, los datos también se pueden enviar directamente a un servidor de red a través de una red de telefonía móvil.
5. El uso convencional de formularios de campo en papel requiere que el funcionario de campo entregue los formularios en la oficina, donde se introducen manualmente en la base de datos principal.

Gestión de datos para los IFN

Requisitos específicos de los sistemas de TIC para el inventario y el monitoreo forestal

A continuación, se exponen algunas consideraciones típicas para un sistema de información viable para los IFN:

1. Los requisitos para los datos, el análisis y la elaboración de informes vienen determinados por las distintas necesidades de información que existen para las diferentes situaciones. Los bosques son complejos, por lo que el SGBD puede contener datos sobre árboles, otra vegetación, fauna, características del suelo, flujos de agua, etc. También es necesario recopilar información espacial (como las coordenadas de los atributos).
2. Los resultados de un análisis de requisitos de datos suelen incluir temas como objetos, datos, relaciones, procesos, vías de acceso, integridad de los datos, diseño de la información, intercambio de datos y seguridad de los datos.

3. A la hora de crear estos sistemas, hay que tener en cuenta las normativas gubernamentales, los estándares y los compromisos de elaboración de informes.
4. Estos sistemas suelen contener datos sobre cambios en el entorno y conjuntos de datos multitemporales.

Bases de datos en los IFN

Las bases de datos son herramientas eficaces para analizar, almacenar y gestionar grandes conjuntos de datos del IFN, mientras que una hoja de cálculo es más útil para realizar análisis *ad hoc* que para alojar y analizar datos del IFN. Sin embargo, crear una base de datos para un estudio requiere más tiempo que utilizar hojas de cálculo.

Dado que existen muchas formas de almacenar los datos recopilados de los IFN, no hay una conclusión sencilla sobre cuál es la óptima. Los requisitos de eficiencia y calidad exigen un sistema flexible, sostenible y de bajo mantenimiento. Por lo general, los datos de los IFN se almacenan en una base de datos centralizada que funciona como punto de entrada principal para todos los datos recopilados.

Una base de datos del IFN se aloja en un SGBD para crear y gestionar varias bases de datos. Un SGBD permite crear, leer, actualizar y eliminar datos de una base de datos. El SGBD sirve de interfaz segura entre las bases de datos y los usuarios finales o los programas de aplicación, garantizando que los datos estén organizados de forma coherente y sigan siendo fácilmente accesibles. Existen numerosos programas de gestión de bases de datos en el mercado, tanto comerciales como de código abierto.

Datos categóricos

Los SGBD pueden alojar muchos tipos de datos diferentes, y uno de los más populares son los datos categóricos.

Una variable categórica es una variable que contiene un número fijo de valores posibles divididos en grupos o categorías. Una base de datos del IFN suele contener varias variables categóricas (como la clase de uso de la tierra, el tipo de vegetación o el color del suelo) u otras variables que se han convertido a esa forma, por ejemplo, como datos agrupados (por ejemplo, la agrupación de especies y la clase de cubierta de copas).

En la gestión de datos, los datos categóricos se pueden almacenar como una lista o, más comúnmente,

como una tabla de consulta. Además, los datos categóricos se pueden organizar en forma de tabla(s) **plana(s)** o **jerárquica(s)**. Veamos ejemplos de ambos tipos.

Crear un esquema para una base de datos

Cuando se empieza a construir un esquema para una base de datos de inventarios forestales, es necesario definir cada objeto (entidad) y describir las variables (atributos) que se deben medir. Antes de empezar a trabajar con el esquema del IFN, es necesario tener una idea clara de la estructura lógica del estudio, una lista detallada de las variables que deben medirse durante el trabajo de campo y decidir la forma óptima de recopilar datos sobre cada variable.

Al elaborar un esquema, hay que examinar el diseño (documento) del muestreo y la parcela y el manual de campo. Los documentos correspondientes pueden indicarle qué entidades se observan, cómo están definidas y cómo se relacionan entre sí y qué propiedades (es decir, atributos) contienen.

Generalmente, los diseñadores de la base de datos se sientan junto con el diseñador del muestreo y el diseño de la parcela y repasan la lista de variables que la base de datos tiene que acomodar

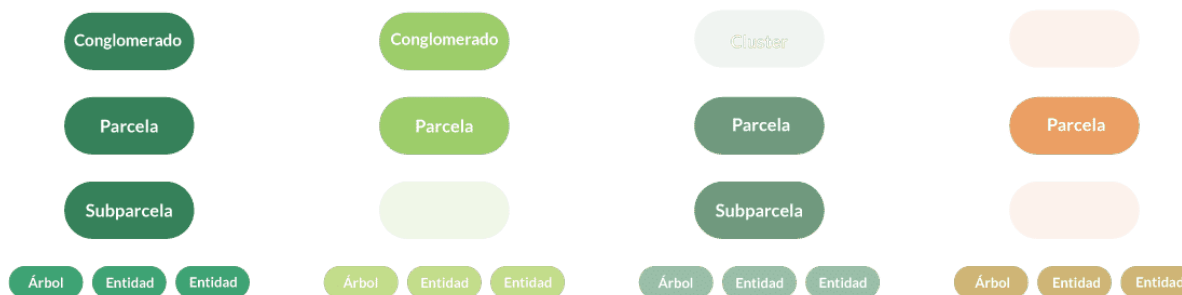


Consejos prácticos

Se recomienda realizar este diseño del modelo conceptual de datos en paralelo a la creación de los formularios de campo para evitar dificultades en la fase de procesamiento de los datos. Con demasiada frecuencia, el diseño de la base de datos se realiza después de crear los formularios de campo, y la estructura de datos que funciona bien para el campo (formularios) puede no funcionar eficazmente en un sistema de gestión de datos.

Se recomienda estudiar cómo se construyen las bases de datos en otros inventarios forestales o utilizar plantillas ya preparadas para esquemas de BD. Con el fin de asegurarse de que el esquema de la base de datos permita obtener las estimaciones o resultados deseados, se recomienda probar todo el flujo de trabajo con algunos datos de prueba (simulados) antes de que el esquema final se implemente y sea utilizado por los equipos de campo. Los datos de las evaluaciones forestales suelen ser jerárquicos. Así, el modelo de base de datos jerárquica se parece a un organigrama o a un árbol genealógico: tiene un único segmento raíz (Nivel 1) conectado a segmentos de nivel inferior. A veces, la jerarquía consta de

varios niveles.



Sin embargo, en una base de datos de inventario forestal real, la estructura del esquema suele ser mucho más compleja que los diagramas simples que acabamos de ver. Esto se debe a que grandes cantidades de datos diversos requieren más entidades. Por ejemplo, si se recogen datos sobre perturbaciones forestales en parcelas de muestreo con múltiples variables, se requiere una nueva entidad.

Además, la agrupación de atributos puede crear una nueva entidad en la base de datos. La agrupación es necesaria para que las Interfaces de Usuario (UI) sean más fáciles de usar cuando se utilizan dispositivos móviles con pantallas más pequeñas. De este modo, el recopilador de datos sólo puede ver un número menor de atributos en la pantalla a la vez.

Validación de datos

Por validación de datos se entiende la comprobación de la exactitud y calidad de los datos introducidos antes de utilizarlos, importarlos o procesarlos de alguna otra forma. Se pueden realizar diversos tipos de validación. La validación de datos es una forma de depuración de datos. El objetivo de la validación de datos es garantizar que los datos registrados se ajusten a las definiciones y los requisitos de exactitud fijados en el protocolo del inventario. El objetivo es crear una base de datos coherente, exacta y completa.

Podemos añadir varios tipos de reglas de validación en un SGBD para ayudar a mejorar la calidad de los datos. En algunos sistemas existen dos clases de gravedad y mensajes: **advertencias** y **errores**. Las advertencias son para casos en los que el usuario del sistema sólo tiene que comprobar (por ejemplo, una relación dap-altura anormal, o un árbol muy alto), y los errores son entradas que deben corregirse.

Los tipos de validación típicos son los siguientes:

1. Sin valores nulos, es decir, no se permiten datos omitidos.
2. Tipo de datos (por ejemplo, entero, cadena).
3. Singularidad, es decir, no se permiten duplicados.
4. Validación por comparación, en la que un atributo de datos se compara con otro atributo de datos o con un literal mediante operadores de comparación (es decir, <, <=, ==, !=, >=, >) y valores de atributos.
5. Expresiones coherentes, por ejemplo, el correo electrónico de una persona se debe escribir correctamente.

Validación de las listas de especies vegetales

Cada inventario forestal necesita una lista de especies arbóreas y más listas para los bambúes, otras plantas o incluso para otros taxones como los animales (por ejemplo, en el IFN de Papúa Nueva Guinea). Cuando se elabora una lista de especies para un IFN o se compara o fusiona cualquier lista de especies arbóreas (o vegetales), los nombres taxonómicos deben comprobarse para evitar errores tipográficos o un aumento artificial debido a los sinónimos. La estandarización de los nombres de las plantas es un paso crítico en un IFN y en toda investigación de estudio ecológico.

Existen **múltiples bases de datos mundiales en Internet que abarcan millones de especies**. Se puede acceder libremente a la mayoría de estas bases de datos en línea o descargándolas en el equipo local. Sin embargo, estas bases de datos se han compilado a menudo a partir de fuentes de datos heterogéneas que varían en cuanto al momento de publicación y el lugar de origen. Hasta la fecha, la lista de referencia de nombres de plantas vasculares más utilizada es [WFO Plant List](#) (en inglés), alojada en el Real Jardín Botánico de Kew. Sin embargo, la TPL no se actualiza desde hace una década y se originó en una época en la que no existía nueva información filogenética sobre muchos géneros.

En el caso de listas largas de especies, un método eficaz es utilizar R para recuperar y procesar datos taxonómicos con el fin de validar los nombres de las plantas. R es la herramienta líder en estadística, análisis de datos y aprendizaje automático. Es más que un paquete estadístico, es un lenguaje de programación. Una ventaja de utilizar R en comparación con las hojas de cálculo es que los scripts de R pueden funcionar como un documento de la cadena de procesamiento de datos.

Existen varios paquetes de R (por ejemplo, `taxonst` y `taxize`, `RBIEN`, `rentrez`) o herramientas en línea (por ejemplo, [Global Names Resolver](#) (en inglés) o el [Taxonomic Name Resolution Service](#) (en inglés)) que ayudan a los investigadores a comprobar su información taxonómica. Sin embargo, algunas de estas herramientas se basan en la TPL como lista de referencia, por lo que habría que comprobar si la base de datos de respaldo está actualizada.

La validación de los nombres de las especies puede basarse en la coincidencia directa y difusa. La validación de los nombres de las especies puede basarse en coincidencias directas y difusas. Algunos métodos de R pueden presentar tanto los resultados como las tasas de éxito para seleccionar las mejores coincidencias individuales esperadas. Con la ayuda de R, los algoritmos de búsqueda pueden encadenarse para que una lista de plantas se valide con la ayuda de múltiples repositorios. Cabe señalar que algunos de estos repositorios requieren una clave de Interfaz de Programación de Aplicaciones ([API](#)) que identifica al usuario. Del mismo modo, el proveedor de servicios puede establecer límites a los usuarios de la API para que no superen un determinado número de solicitudes por segundo (por ejemplo, véanse las [condiciones de uso de Kew](#) (en inglés)).

Codificación de especie

Se recomienda encarecidamente la codificación de especies para almacenar y procesar datos de taxones en un SGBD. La razón de utilizar códigos es crear un identificador único para cada entrada en una tabla de especies y simplificar el procesamiento y la agregación de datos en la fase de elaboración de informes. Además, si hay un error tipográfico en el nombre de la especie (en la tabla de taxones), ese error se puede corregir fácilmente en cualquier momento. Así pues, todas las especies deben tener un código único, y los géneros pueden codificarse si es necesario.

Si existen diferentes tipos de plantas (o animales) en el inventario, cada uno de ellos puede tener su propia tabla de taxones, como se muestra en la pantalla a continuación. Veamos ahora los diferentes métodos de codificación de especies.

Código	Código	Nombre científico	Sinónimos
ACACI/AURIC	Especie	Acacia auriculiformis	
ACACI/CRASS	Especie	Acacia farnesiana	
ACACI/FARNE	Especie	Acacia farnesiana	
ACACI/HOLOS	Especie	Acacia holosericea	
ACACI/LEPTO	Especie	Acacia leptocarpa	
ACACI/MANGI	Especie	Acacia mangium	
ACACI/MEARN	Especie	Acacia mearnsii	
ACACI/SIMSI	Especie	Acacia simsii	
ACACI	Especie	Acacia sp.	

A continuación, veremos las especies codificadas como números

Código	Código	Nombre científico	Sinónimos
33296	Especie	Abutilon pictum	
25213	Especie	Abutilon Sinense	
25214	Especie	Abutilon theophrasti	
38000	Especie	Abutilon andamanica	
10001	Especie	Abutilon auriculiformis	
23003	Especie	Acacia catechu	
10002	Especie	Acacia comosa	
23004	Especie	Acacia concinna	



Consejos prácticos

Supongamos que el país no dispone de ningún sistema de codificación de especies. En ese caso, un método aplicable es utilizar un sistema en el que el código se forme combinando dos cadenas de texto: la primera identifica el género y la segunda, la especie. En caso necesario, pueden añadirse códigos más largos para formas, subespecies, variantes, etc..

Ventajas de los códigos de texto

Utilizar códigos de texto en lugar de números tiene varias ventajas:

1. La codificación de las especies puede seguir el mismo orden alfabético que los nombres de las especies.
2. Las nuevas especies se pueden añadir o eliminar fácilmente de la lista.
3. Si se añaden nuevas especies, es más fácil relacionarlas, integrarlas en los códigos de especies afines del mismo género. En un sistema de numeración, esta conexión se perdería.
4. La codificación eficiente de las especies permite realizar fácilmente operaciones combinadas (como con tablas de consulta o ecuaciones externas).
5. La agregación por género es rápida en las fases de análisis y elaboración de informes (por ejemplo, al calcular índices de biodiversidad y números de géneros arbóreos por parcela).
6. Los códigos de texto son más legibles que los numéricos.

Una vez validada y aprobada la lista de especies y comenzada la recogida de datos, sólo se pueden editar los nombres o añadir más sinónimos. Si faltan especies, éstas también se pueden añadir (con nuevos códigos y nombres) a la lista durante el ciclo del inventario forestal.

Para más información sobre la codificación de especies con la ayuda de scripts R, consulte los siguientes materiales [Open Foris' materials](#). (en inglés).

Resumen

Antes de finalizar, aquí están los puntos clave de aprendizaje de esta lección:

- Los equipos de los inventarios forestales realizan la mayor parte de sus tareas al aire libre, exponiendo sus herramientas a condiciones meteorológicas variables y a una manipulación poco adecuada. Por eso, se recomienda el uso de dispositivos resistentes.
- Las computadoras de escritorio y portátiles estándar suelen ser suficientes para el trabajo diario de oficina con datos forestales. A nivel corporativo se necesitan servidores locales o servicios en la nube alojados por un proveedor de servicios en la nube.
- La recopilación de datos móvil mejora la calidad de los datos porque podemos utilizar métodos para verificar los datos directamente sobre el terreno.
- Antes de empezar a trabajar con el esquema del IFN, es necesario tener una idea clara de la estructura lógica del estudio y una lista detallada de las variables que se van a medir. Consultar el protocolo del inventario es primordial.
- La validación de los datos significa comprobar la exactitud y la calidad de los datos de origen antes de utilizarlos, importarlos o procesarlos de alguna otra forma.
- Una lista de especies validada y una codificación adecuada de las especies harán que el procesamiento y la elaboración de informes sobre los resultados sean más eficaces. También mejorará la calidad de la información sobre los resultados.

Lección 3: Elaboración de informes en línea y archivo de datos

Introducción de la lección

Esta lección muestra algunos de los métodos, técnicas y herramientas actuales para la elaboración de informes sobre los datos de los IFN. También describe el concepto de metadatos y ofrece lineamientos sobre la importancia de archivar y almacenar los datos para su uso futuro.

Objetivos

Al final de esta lección, usted podrá:

1. Describir la función de la tecnología de cuadros de mando de datos en la elaboración de informes sobre inventarios forestales
2. Explicar la tecnología de "Procesamiento analítico en línea" (OLAP) y su funcionamiento.
3. Definir el concepto y los requisitos básicos de los metadatos y el archivo de datos.

Cuadros de mando y portales de datos

¿Qué es un cuadro de mando de datos?

Los cuadros de mando y los portales tienen la misma funcionalidad, pero usos diferentes. Los portales proporcionan un repositorio centralizado de información clave para una organización o el público, y suelen contener texto enriquecido, accesos directos, imágenes interactivas y mapas. Algunos portales pueden proporcionar análisis en tiempo real de los datos subyacentes.

Los cuadros de mando, por su parte, proporcionan una visibilidad rápida para facilitar la comprensión, con fácil acceso a los cuadros, gráficos e informes que se necesitan con más frecuencia. Además, por regla general, los portales producen representaciones estáticas de los resultados en tablas y mapas predefinidos, mientras que todos los cuadros de mando proporcionan un contenido más dinámico mediante el uso de modelos de datos y análisis de datos en tiempo real. Aquí tiene algunos ejemplos de cuadros de mando y portales de IFN para que los explore (en versión original):

1. [Evaluación de los recursos forestales mundiales 2020](#)
2. [Satellite Land Monitoring Systems \(SLMS\), empowered by UN- R EDD/FAO](#)
3. [Bangladesh Forest Inventory Results](#)

4. [Germany data dashboard \(various forest inventories\)](#)
5. [L'outil dataIFN France](#)
6. [Canada NFI portal](#)
7. [Papua New Guinea Climate Change and Forest Monitoring Web portal](#)
8. [National Forest Monitoring System Portal \(Democratic Republic of Congo \)](#)
9. [National Land Monitoring System of Suriname](#)

Procesamiento analítico en línea (OLAP)

¿Qué es OLAP?

El procesamiento analítico en línea (OLAP) es una tecnología de procesamiento informático que permite la ejecución rápida de consultas analíticas complejas. Permite al usuario cortar y filtrar datos, y producir resultados específicos sin tener que implementar cálculos separados. Por estas razones, las bases de datos OLAP son populares para la elaboración de informes.

OLAP forma parte de la categoría más amplia de inteligencia empresarial, que engloba las bases de datos relacionales, la redacción de informes y la minería de datos. Las aplicaciones OLAP se implementaron por primera vez en la elaboración de informes empresariales, pero hoy en día esta técnica se aplica en múltiples sectores, entre ellos la silvicultura y la agricultura. Los informes de inventarios forestales y los resultados de los IFN pueden procesarse y mostrarse con OLAP.

Sin embargo, existen **retos asociados a OLAP: utiliza una terminología diferente de la terminología estadística y de bases de datos estándar**. No obstante, no profundizaremos en estas diferencias, sino que nos centraremos en las principales ventajas y aplicaciones de las técnicas OLAP. A continuación, veremos tres ejemplos de uso de OLAP en el procesamiento de datos de los IFN.

Sistema de elaboración de informes y análisis del INF sueco

Una base de datos relacional permite al IFN sueca utilizar la técnica OLAP. Esta base de datos permite elaborar informes y análisis rápidos de datos para productos estándar, utilizados para la elaboración de informes nacionales e internacionales. El IFN sueco almacena los datos utilizados con más frecuencia en Microsoft (MS) Analysis Services, que es una herramienta OLAP y de minería de datos. Las herramientas utilizadas para la elaboración de informes y el análisis de datos son MS Excel, MS

PowerPivot y diferentes herramientas de MS Reporting Services.

Para la elaboración de informes ad hoc y tareas específicas, el IFN sueco utiliza herramientas de SAS Institute y Microsoft Power Pivot para extraer datos directamente de la base de datos y llevar a cabo análisis. Sin embargo, la mayoría de los demás paquetes estadísticos pueden extraer y manejar datos desde la base de datos MS SQL Server.

TaxWebb (en inglés) es la herramienta de análisis interactivo del IFN sueco cuyo objetivo es proporcionar un acceso rápido y sencillo a las estadísticas. Permite a los usuarios elaborar sus propios informes.

NFI Suizo

El Sistema Nacional de Análisis de Datos para el Inventario Forestal (NAFIDAS) fue desarrollado para el IFN suizo. NAFIDAS utiliza un diseño similar al aplicado en el IFN sueco: consta de un almacén de datos operativos con interfaces a sistemas de datos de origen (fuentes de datos internas y externas), un **área de almacenamiento de datos y metadatos** (en inglés) y herramientas de presentación para el usuario final.

NAFIDAS elabora tablas y mapas a partir de los datos de los IFN suizos o de inventarios regionales, y consta de tres componentes principales:

1. una aplicación web para la gestión, la documentación y la elaboración de informes;
2. bases de datos para el almacenamiento; y
3. la aplicación de análisis de datos para el análisis..

Open FORIS calc con SAIKU analytics

El **vídeo de demostración de Open Foris** (en inglés) muestra un ejemplo de OLAP en la elaboración de informes de resultados de inventarios forestales en Open Foris Calc con la ayuda del software Saiku Analytics.

Saiku Server es un software de código abierto alojado en la web que facilita la visualización y consulta de datos. Aunque existe una versión del software disponible gratuitamente en el sitio web de Saiku, se ha personalizado una versión especial para una mayor compatibilidad con Collect Earth. Saiku Server se incluye en el instalador de Collect Earth. La historia de Saiku comienza en el punto temporal 3:44.

Archivo de datos

Toda organización es responsable de la preservación de los datos. En general, los formularios en papel bien organizados y almacenados son bastante duraderos, pero son difíciles de reutilizar, o incluso de hacer copias de seguridad.

Cuando se utilizan soportes digitales, los discos duros se estropean, los archivos se extravían, los formatos de los datos pueden quedar desactualizados, las contraseñas se olvidan, los detalles específicos necesarios para utilizar los datos se olvidan, los soportes en los que se almacenan quedan obsoletos (por ejemplo, los disquetes) e incluso las personas responsables se pueden trasladar, jubilar o fallecer. Afortunadamente, los avances tecnológicos y la aparición de archivos de acceso público han facilitado y hecho más fiable la preservación de datos a largo plazo.

Un **archivo de datos** es una **colección de conjuntos de datos con metadatos adjuntos** almacenados de forma que diversos usuarios puedan localizar, adquirir, comprender y utilizar los datos. Los datos archivados están a salvo de catástrofes naturales o provocadas por el hombre y se preservan en una forma que seguirá siendo accesible a medida que cambie la tecnología.

Ventajas del archivo de datos

Hay muchas razones por las que el archivo de datos es valioso, siendo algunas de las más contundentes las siguientes:

1. reducir la pérdida de datos;
2. utilizarlos para nuevos análisis y estudios. Los datos antiguos también pueden servir como valiosos datos de referencia para analizar tendencias a largo plazo;
3. utilizarlos como material de formación para estudiantes, ya que son un medio eficaz en términos de costos para aumentar la productividad científica de un país. Los datos archivados pueden ayudar a desarrollar un cuadro de científicos postdoctorales altamente formados y productivos;
4. disipar las inquietudes sobre la exportación de propiedad intelectual y la falta de participación de los científicos locales en la recopilación de datos que suelen afectar a los científicos extranjeros que trabajan en países tropicales;

5. verificar resultados y corregir errores; y
6. cumplir los mandatos y requisitos de donantes y gobiernos.

Metadatos

Los metadatos son **información estructurada que describe, explica, localiza o facilita la recuperación, el uso o la gestión de un recurso de información**. Los metadatos suelen denominarse datos sobre datos o información sobre la información. Los metadatos proporcionan información que permite a los usuarios comprender plenamente los **datos** (por ejemplo, documentos, imágenes, conjuntos de datos), los **conceptos** (por ejemplo, esquemas de clasificación) y las **entidades del mundo real** (por ejemplo, organizaciones, lugares, protocolos de muestreo). Un ejemplo típico de inventario forestal sería un protocolo de campo detallado que describiera cómo se realizan las mediciones y qué categorías (códigos) se utilizan.

En el contexto de un IFN, los metadatos se pueden utilizar para responder a preguntas como qué datos se recopilaron, cómo se recopilaron, por qué se recopilaron, hasta qué punto son fiables y qué cuestiones se deben tener en cuenta al trabajar con ellos. Los metadatos también pueden describir cómo obtener los datos, qué herramientas se necesitan para trabajar con ellos y otros temas relacionados.

El objetivo de documentar los datos es proporcionar suficiente información sobre el conjunto de datos para que alguien pueda trabajar fácilmente con ellos dentro de 20 años y hacer que los resultados sean reproducibles. Cuantas más propiedades de metadatos se adjunten a los datos, más valiosos e inteligentes serán éstos.

Los metadatos son importantes para cualquier repositorio de información porque permiten:

- determinar la disponibilidad, ubicación, antigüedad de creación, propietario de los datos y accesibilidad de estos;
- comprender y utilizar correctamente los datos;
- gestionar los datos de forma más eficiente; y
- garantizar la interoperabilidad de los datos.

IFN y metadatos

El sistema de información de los IFN para la gestión de datos debe estar bien documentado con metadatos. Todas las descripciones deben ser compatibles con los manuales de campo que describen los datos.

Para una base de datos del IFN, los metadatos pueden contener documentación detallada de los campos que componen los conjuntos de datos, incluida la definición, el tipo de medida, las unidades cuando proceda y cualquier vocabulario controlado o lista de códigos presentes en los datos. El estándar de los metadatos es un conjunto común de términos y definiciones que describen los datos, describiendo las propiedades características que se deben registrar y los valores que deben tener dichas propiedades. Los estándares de los metadatos están especialmente desarrollados para los datos geoespaciales.



Consejos prácticos

A la hora de crear metadatos para la información forestal, podemos citar ejemplos aplicados en el sector forestal y en otros sectores. En el caso del IFN, podemos fijarnos en las [documentaciones de las bases de datos](#) (en inglés) de datos del Programa Nacional de Inventario Forestal y Análisis (FIA) de Estados Unidos. El [Programa de Inventario Forestal y Análisis de Vermont](#) (en inglés) y el [Inventario Forestal Nacional sobre Tierras Forestales en Inglaterra en 2018](#) (en inglés) proporcionan más ejemplos de metadatos.

Microdatos estadísticos

Microdatos es un concepto que significa pequeños conjuntos de datos o datos agregados en niveles específicos (espacial o contextualmente). Se aplica especialmente en algunas organizaciones internacionales, como el [Banco Mundial](#) (en inglés), [Eurostat](#) y algunas organizaciones de las Naciones Unidas, incluida la [FAO](#).

Los microdatos pueden incluir características de unidades de una población, como individuos, hogares o establecimientos, recogidas por un censo, una encuesta o un experimento. Como tal, el concepto se aplica más en encuestas socioeconómicas que en inventarios forestales (biofísicos), pero se trata de un concepto nuevo y puede ser beneficioso entenderlo en el contexto de la gestión de datos.

Los catálogos de microdatos son conjuntos de datos para almacenar y compartir conjuntos de datos recopilados. Los datos del conjunto de datos pueden provenir de la recopilación de datos primarios (véase el [ejemplo de los microdatos](#) de las parcelas de muestreo del IFN sueco) o de datos secundarios mediante agregación o síntesis. Los resultados se suelen publicar como agregados, tanto por razones de privacidad como por la gran cantidad de datos que contienen. Por lo general, se eliminan todos los identificadores personales para garantizar la privacidad y la confidencialidad. Los catálogos de microdatos pueden dar cabida a la información forestal como formas de estadísticas procesadas.

Los microdatos agregados también pueden ser muy útiles para el análisis de políticas, la investigación y las estadísticas muy desagregadas (por ejemplo, por localidades, poblaciones o grupos etarios).

Permiten obtener una imagen clara de los problemas estudiando las relaciones e interacciones entre fenómenos. Así pues, los microdatos son fundamentales para diseñar proyectos y formular políticas, orientar las intervenciones y el monitoreo, y medir el impacto y los resultados de las intervenciones.

[El Catálogo de microdatos sobre la Alimentación y la Agricultura \(FAM\) de la FAO](#) proporciona un inventario de los conjuntos de datos recopilados a través de encuestas agrícolas y de hogares que contienen información relacionada con la agricultura, la silvicultura, la seguridad alimentaria y la nutrición.

Resumen

Antes de finalizar, aquí están los puntos clave de aprendizaje de esta lección:

- Los portales suelen proporcionar información clave al equipo de una organización o al público.
- Los cuadros de mando proporcionan una rápida visibilidad del estado de la empresa, facilitando así el acceso para ver las principales métricas, cuadros, gráficos e informes.
- El procesamiento analítico en línea (OLAP) es una tecnología de procesamiento informático que permite la ejecución rápida de consultas analíticas complejas.
- Preservar los datos es responsabilidad de toda organización. Los archivos de datos han hecho más fácil y fiable la preservación de datos a largo plazo.
- Los metadatos son información estructurada que describe, explica, localiza o facilita la recuperación, el uso o la gestión de un recurso de información.
- Los catálogos de microdatos son conjuntos de datos para almacenar y compartir pequeños conjuntos de datos recopilados.