



2025 | **16-20**
GIJÓN | **JUNIO**

9º CONGRESO FORESTAL ESPAÑOL

9CFE-1961

Actas del Noveno Congreso Forestal Español
Edita: **Sociedad Española de Ciencias Forestales. 2025.**
ISBN: **978-84-941695-7-1**

Organiza





Integración de datos forestales transnacionales: El grafo de conocimiento Cross-Forest

GIMÉNEZ-GARCÍA, J.M. (1,2), ORDÓÑEZ, C. (2), CRESPO-LERA, N. (1,2), VEGA-GORGOJO, G. (1,2) y BRAVO, F. (2)

(1) Grupo de Sistemas Inteligentes y Cooperativos (GSIC) ETSI de Telecomunicación, Universidad de Valladolid, Valladolid, España.

(2) SMART ecosystems research group, Instituto de Investigación en Gestión Forestal Sostenible (iuFOR), Universidad de Valladolid, Palencia, España.

Resumen

La explotación de datos forestales se ve a menudo obstaculizada por la heterogeneidad de modelos y formatos de publicación, lo que dificulta su integración y análisis. Este artículo aborda el desafío de combinar inventarios forestales y mapas de usos de suelo de España y Portugal utilizando tecnologías de Web Semántica y Datos Abiertos Enlazados, que facilitan la integración y conexión con otras fuentes.

El resultado es el grafo de conocimiento Cross-Forest, que permite realizar consultas complejas sobre los datos agregados de todas las fuentes forestales. Este grafo se encuentra conectado con fuentes externas, como DBpedia, Wikidata o The Plant List, facilitando tareas de visualización y comprensión de los datos, incluso por usuarios no expertos.

Este conjunto de datos integrado, disponible tanto para descarga como para consulta en línea, facilita tareas de investigación y gestión forestal que antaño necesitaban un gran esfuerzo de búsqueda y preprocesamiento de datos y representa un paso importante hacia unos datos más conectados, accesibles y efectivos en ciencia y gestión forestal. Actualmente se utiliza como base del “Explorador Forestal” y “Descarga IFN”, aplicaciones web para visualizar y analizar estos datos y en consultas orientadas a la toma de decisiones en la gestión de recursos forestales.

Palabras clave

Inventarios forestales, mapas de usos de suelo, web semántica, datos abiertos enlazados

1. Introducción

Los conjuntos de datos forestales a gran escala y a largo plazo son esenciales para el avance de la ciencia forestal. La longevidad de los árboles hace que estos conjuntos de datos sean cruciales para investigadores y gestores forestales (PRETZSCH, 2009). Inventarios, mapas y bases de datos forestales a gran escala son herramientas clave para comprender los procesos forestales y evaluar la estructura de los bosques (TOMPPPO et al., 2010). Esta información a largo plazo es



fundamental para una gestión forestal sostenible que garantice los servicios ecosistémicos (RUIZ-PEINADO et al., 2017; BRAVO et al., 2019).

Las entidades públicas son las principales responsables de la adquisición, conservación y difusión de datos forestales a largo plazo. No obstante, estos datos, a menudo complejos y heterogéneos, son difíciles de explotar por usuarios e investigadores no especializados en temas forestales. Su integración y análisis requieren herramientas y conocimientos específicos que no siempre están al alcance de todos los potenciales usuarios. Estos conjuntos de datos también son utilizados por investigadores, educadores, silvicultores y otros actores con impacto en la sociedad.

Los conjuntos de datos forestales a menudo se encuentran aislados y utilizan diferentes vocabularios y formatos, a menudo propietarios. La integración de datos es un desafío clave en la ciencia forestal (ZOU et al., 2019). Esta integración se ve obstaculizada por la diversidad de formatos de datos, las diferencias en las metodologías de recolección de datos, la falta de estándares comunes, e incluso el uso de software propietario. Organizaciones como la Red Europea de Inventarios Forestales Nacionales (ENFIN) trabajan en la armonización de inventarios forestales (VIDAL et al., 2016). Las integraciones de diferentes conjuntos de datos forestales suelen requerir el desarrollo y uso de esquemas comunes, lo que presenta dificultades para generalizar y ampliar su uso a otras fuentes de datos. Sin embargo, aún faltan herramientas para manejar datos forestales heterogéneos.

La Web Semántica ofrece un marco prometedor para la publicación e integración de datos heterogéneos de forma interoperable (HEATH & BIZER, 2011). A través de tecnologías como RDF (SCHREIBER & RAIMOND, 2014), OWL (HITZLER et al., 2012) y SPARQL (HARRIS & SEABORNE, 2013), la Web Semántica no solo facilita la integración de datos, sino que también permite el razonamiento automático y la inferencia de nuevo conocimiento. Las ontologías, que describen conceptos y sus relaciones, permiten la publicación de datos como Datos Enlazados (BIZER et al., 2018). Estos datos pueden consultarse y descargarse online. Los Datos Abiertos Enlazados (LOD) promueven el acceso abierto a los datos y han ganado popularidad en las últimas décadas. Los Datos Abiertos Enlazados son promovidos por organismos nacionales e internacionales y se utilizan en campos como las biociencias y el patrimonio cultural.

Dada la complejidad y heterogeneidad de los datos forestales, y con el objetivo de facilitar su integración y análisis por parte de usuarios de diferentes áreas, este trabajo utiliza tecnologías de la Web Semántica para unificar la información proveniente de los inventarios y mapas forestales de España y Portugal. Este enfoque no solo permite la integración de diferentes fuentes de datos forestales, sino que también facilita la interoperabilidad con otros tipos de datos, como información de administración territorial, socioeconómica o ambiental, y abre las puertas a nuevas aplicaciones y análisis multidisciplinarios.



2. Objetivos

Este estudio se centra en la integración de datos forestales de los inventarios y mapas de uso de España y Portugal. La integración de estos datos es crucial para la investigación, la gestión forestal sostenible y la toma de decisiones informadas. Sin embargo, esta integración se ve obstaculizada por la diversidad de formatos de datos, las diferencias en las metodologías de recolección, la falta de estándares comunes e incluso el uso de software propietario.

Para abordar este desafío, este trabajo utiliza tecnologías de la Web Semántica para crear el Conjunto de Datos Cross-Forest, un recurso unificado y accesible que facilita la integración y el análisis de la información forestal. Para lograr este objetivo, se abordan los siguientes objetivos específicos:

- Desarrollar un conjunto de ontologías (*Cross-Forest Ontology Suite*): Las ontologías son la base para la integración de datos heterogéneos en la Web Semántica. Definición de un conjunto de ontologías que modelen los conceptos y las relaciones presentes en los inventarios forestales y los mapas de uso de España y Portugal. Esto permite armonizar la información proveniente de diferentes fuentes y facilita su explotación conjunta. Además, las ontologías permiten añadir semántica a los datos, lo que facilita el razonamiento automático y la generación de nuevo conocimiento.
- Integrar los mapas de cobertura del suelo y los inventarios forestales en Datos Abiertos Enlazados (LOD) (*Cross-Forest Dataset*): Publicación de los datos integrados en formato de Datos Abiertos Enlazados (LOD), lo que facilita su acceso, integración y reutilización no solo por la comunidad forestal, sino también por otros públicos interesados en la información sobre los bosques, como investigadores de otras áreas, organizaciones no gubernamentales, ciudadanos, etc. La publicación en formato LOD promueve la transparencia y la colaboración entre diferentes actores.

3. Metodología

Para construir el Conjunto de Datos Cross-Forest, se llevan a cabo dos procesos principales: el desarrollo de un conjunto de ontologías para describir los datos y la creación de procesos para transformar las fuentes de datos originales en Datos Abiertos Enlazados. A continuación, se muestran las fuentes originales de datos (Sección 3.1), un breve resumen de las tecnologías de la Web Semántica (Sección 3.2), el desarrollo de las ontologías (Sección 3.3) y la generación de los datos (Sección 3.4).

3.1 Datos de origen

Este trabajo utiliza los inventarios forestales nacionales y mapas de cobertura del suelo de España y Portugal:



- **Inventario Forestal Nacional Español (IFNes):** Este inventario, actualizado aproximadamente cada 10 años, proporciona información estadística y datos de muestreo sobre los recursos forestales españoles. Contiene datos de 1.4 millones de árboles y 99,000 parcelas, incluyendo información sobre especies, posición y medidas dendrométricas. La información se organiza por provincias.
- **Inventario Forestal Nacional Portugués (IFNpt):** Actualizado cada 10 años, este inventario ofrece información estadística agregada por regiones (NUTS 2 y 3). A diferencia del IFNes, no se publican los datos de muestreo.
- **Mapa Forestal Español (MFE):** Este mapa, actualizado cada 10 años, presenta información sobre la cobertura del suelo forestal en España. Contiene 862,900 parcelas con información sobre usos del suelo y especies dominantes, representadas mediante polígonos. Los datos se organizan por provincias.
- **Mapa de Uso del Suelo Portugués (COS):** Este mapa ofrece información sobre la cobertura y uso del suelo en Portugal, incluyendo los usos forestales. Se dispone de cinco versiones del mapa (1995, 2007, 2010, 2015 y 2018). Los datos se representan mediante polígonos con una unidad mínima de 1 hectárea.

A pesar de algunas similitudes, estos conjuntos de datos difieren en las metodologías de recolección, el nivel de detalle de la información, los esquemas de datos, los identificadores y los formatos de publicación.

3.2 Tecnologías de la Web Semántica

La Web Semántica es una extensión de la Web actual que permite dotar de significado a conceptos y entidades representadas mediante IRIs. Esto se logra con el uso de un modelo de datos basado en un grafo dirigido y etiquetado, y en el uso de ontologías que describen los datos y sus relaciones utilizando un lenguaje lógico formal. Esto hace que los datos publicados utilizando estas tecnologías sean autodescriptivos, con un esquema abierto que facilita la integración, el análisis y la reutilización de la información. Las ventajas de la Web Semántica para la integración de datos son numerosas:

- La Web Semántica proporciona un marco común para describir los datos, lo que facilita la integración de información proveniente de diferentes fuentes, independientemente de su formato o estructura original.
- Los datos tienen una semántica explícita, descrita en el mismo formato que los datos. Esto hace que sean comprensibles de manera automatizada por las por agentes artificiales, lo que facilita la interoperabilidad entre diferentes sistemas y aplicaciones.
- La Web Semántica permite realizar inferencias sobre los datos, lo que puede ser útil para descubrir relaciones y patrones que no son explícitos en los datos originales, así como validar la consistencia de los mismos.
- Los datos semánticos pueden ser reutilizados en diferentes proyectos y aplicaciones, lo que facilita el desarrollo de nuevas soluciones.



Debido a estas ventajas, este estudio utiliza tecnologías de la Web Semántica para integrar los diferentes conjuntos de datos forestales. Las tecnologías clave son:

- **RDF (*Resource Description Framework*):** Es un modelo de datos que permite representar información en forma de grafo, donde los datos se expresan como tripletas (sujeto, predicado, objeto). Esto facilita la conexión entre diferentes conjuntos de datos.
- **SPARQL (*SPARQL Protocol and RDF Query Language*):** Es un lenguaje de consulta que permite extraer información de grafos RDF. SPARQL ofrece una forma flexible y potente de acceder a los datos y realizar análisis complejos.
- **RDFS (*RDF Schema*) y OWL (*Web Ontology Language*):** Son extensiones de RDF que permiten añadir semántica a los datos, es decir, definir el significado de los términos y las relaciones entre ellos. Esto facilita el razonamiento automático y la inferencia de nuevo conocimiento.

Además, la Web Semántica facilita la integración de datos forestales con otras fuentes de información, como datos socioeconómicos, ambientales o de infraestructuras, lo que permite un análisis más completo y multidisciplinar.

3.3 Desarrollo de las ontologías

Para integrar la información forestal proveniente de diferentes fuentes y países, se diseña un conjunto de ontologías llamado *Cross-Forest Ontology Suite*. En el desarrollo de estas ontologías se aplican metodologías reconocidas en el campo de la ingeniería ontológica, como las propuestas por Suárez-Figueroa et al. (2012), así como las Mejores Prácticas de Datos en la Web (Farias Lóscio et al., 2016) y las Mejores Prácticas de Datos Espaciales en la Web (Tandy et al., 2017).

En el desarrollo de las ontologías se sigue un enfoque de abajo a arriba (*bottom-up*), consistente en analizar primero las fuentes de datos para identificar los conceptos y relaciones relevantes. Este análisis permite definir el alcance y los requisitos de cada una de las ontologías necesarias para describir cada fuente de datos, capturando sus particularidades para luego. Posteriormente, se crean ontologías de nivel superior que integran los conceptos comunes a todos los conjuntos de datos. Esta estrategia permite el acceso a los datos tanto a nivel local como transnacional.

El diseño de las ontologías es modular, creando diferentes módulos para cada tipo de dato (inventarios forestales, mapas de cobertura del suelo, etc.). Este enfoque facilita la reutilización y el mantenimiento de las ontologías. Para asegurar la coherencia y la flexibilidad en la representación del conocimiento, se utilizó el metamodelado. Esta técnica permite representar conceptos que a veces actúan como clases (categorías) y otras veces como individuos (instancias).

Dado que las fuentes de datos incluyen información geoespacial con diferentes



Sistemas de Referencia de Coordenadas (CRS), se prestó especial atención al modelado de las posiciones espaciales. Para ello, no solo se utiliza GeoSPARQL (Perry and Herring 2012), sino que también se desarrollan ontologías propias para definir posiciones absolutas y relativas, así como la ontología EPSG, basada en el conjunto de datos de mismo nombre, para una representación completa y precisa de los diferentes CRS utilizados en los datos originales.

Un aspecto clave del diseño es la integración con otras fuentes de datos, estableciendo conexiones con conjuntos de datos externos, como Wikidata (<https://www.wikidata.org>), DBpedia (<https://www.dbpedia.org/>) y Cross-Nature (<https://crossnature.eu/>). Esto permite que el Conjunto de Datos Cross-Forest se beneficie de la información disponible en estas fuentes. Además, se incluyen enlaces a Wikipedia (<https://www.wikipedia.org/>) y The Plant List (<http://www.theplantlist.org/>) para enriquecer la información sobre las especies. Esta integración no se limita a conjuntos de datos forestales, sino que también incluye la conexión con datos de ámbito general, lo que amplía el potencial del Conjunto de Datos Cross-Forest para abordar problemáticas multidisciplinares.

3.4 Generación de los datos

Para la transformación de los datos originales a RDF, utilizando las ontologías descritas en la sección 3.3, se desarrolla un proceso automatizado que incluye dos etapas principales:

- **Preprocesamiento:** En esta etapa, se descargan los datos de las diferentes fuentes, se extraen de sus formatos originales (normalmente archivos de Microsoft Access o shapefiles) y se realizan tareas de limpieza y preprocesamiento. Esto incluye la corrección de errores en los datos y el esquema de los mismos, la fusión de información procedente de varios archivos y la simplificación de geometrías. Para ello, se utilizaron herramientas como UnZip, MDB Tools, csvtk y Mapshaper, todas ellas de acceso público con licencias abiertas.
- **Generación de RDF:** Una vez preprocesados los datos, se convierten a RDF utilizando la herramienta SPARQL-Generate (LEFRANÇOIS et al., 2017), también de acceso público y con licencia abierta. Esta herramienta permite convertir diferentes formatos de datos a RDF de forma eficiente y autodescriptiva. Además, se realizan tareas de enriquecimiento de los datos, como la conversión de posiciones relativas a absolutas y la adición de coordenadas WGS84. Para esta tarea se crea epsgrdf, una herramienta de línea de comandos para leer archivos RDF que contienen posiciones relativas y/o posiciones absolutas en un CRS arbitrario, y que genera sus posiciones correspondientes en el mismo CRS que la posición de referencia, así como en WGS84. Esta herramienta está desarrollada en Java y utiliza las bibliotecas Apache Jena, Apache Spatial Information System y JTS Topology Suite.

El proceso de generación de datos se automatiza mediante scripts, lo que garantiza



la reproducibilidad del proceso. Todas las herramientas utilizadas para la manipulación de los datos, incluyendo las desarrolladas en el proyecto, son de acceso público con licencias abiertas, lo que facilita la transparencia y la reutilización del proceso por parte de otros investigadores.

4. Resultados

Este trabajo ha generado dos productos principales: el Conjunto de Ontologías Cross-Forest y el Conjunto de Datos Cross-Forest. Estos productos representan una contribución significativa al campo de la ciencia forestal, ya que proporcionan un marco para la integración y el análisis de datos forestales de España y Portugal utilizando tecnologías de la Web Semántica.

4.

a. *Ontologías*

El Conjunto de Ontologías Cross-Forest está compuesto por cuatro conjuntos de módulos:

- Módulos transversales: Estos módulos describen conceptos generales como medidas, posiciones espaciales y sistemas de referencia de coordenadas.
 - SMO (Ontología de Medidas Simples): Define conceptos como "Entidad Medible", "Medida" y "Unidad". Permite describir las medidas tomadas sobre las entidades, incluyendo su valor y sus unidades.
 - EPSG (Conjunto de Datos de Parámetros Geodésicos EPSG): Contiene una descripción de los Sistemas de Referencia de Coordenadas existentes, incluyendo conceptos como "Sistema de Coordenadas", "Datum" y "Eje".
 - SPO (Ontología de Posiciones Espaciales): Permite representar las posiciones de las entidades, ya sean absolutas o relativas a otra posición. Incluye conceptos como "Entidad Espacial" y "Posición". Utiliza SMO para modelar las medidas de las posiciones y EPSG para referenciar los Sistemas de Referencia de Coordenadas.
- Módulos de Inventario Forestal: Estos módulos describen los inventarios forestales de España y Portugal.
 - IFNes (Inventario Forestal Nacional Español): Describe los datos del Inventario Forestal Nacional Español. Incluye conceptos como "Parcelas", "Árboles", jerarquías de conceptos para clasificarlos y medidas dendrométricas. Utiliza SPO para describir las posiciones de las parcelas y los árboles.
 - IFIS (Estadísticas de los Inventarios Forestales Ibéricos): Describe los datos estadísticos sobre las formaciones dominantes de los inventarios portugués y español. Incluye conceptos como "Unidad NUTS", "Información de Formación Dominante" e "Información de Formación Dominante por Hectárea". Utiliza SMO para describir los datos asociados a las formaciones dominantes.
- Módulos de Mapa de Cobertura del Suelo: Estos módulos describen los mapas de cobertura del suelo de España y Portugal.
 - COS (Mapa de Usos del Suelo Portugués): Describe los datos del Mapa de Usos del Suelo Portugués. Incluye conceptos como "Tesela"



- y "Uso en Tesela". Utiliza SPO para definir las posiciones de las teselas.
- MFE (Mapa Forestal Español): Describe los datos del Mapa Forestal Español. Incluye conceptos como "Tesela", "Uso en Tesela" y clases para definir clasificaciones y medidas dendrométricas. Utiliza SPO para definir las posiciones de las teselas.
 - ILU (Uso del Suelo Ibérico): Contiene conceptos y propiedades comunes a los mapas de cobertura del suelo portugués y español. Incluye conceptos como "Tesela" y "Uso en Tesela".
 - Módulos de Taxonomía: Estos módulos describen las especies arbóreas.
 - Especies: Modela las especies arbóreas y permite la conexión con recursos externos. Incluye conceptos como "Especie", "Género", "Familia", "Orden", "Clase", "Filo" y "Reino".
 - Módulos de Alineamiento: Estos módulos permiten la conexión con ontologías y conjuntos de datos externos, como Wikidata, DBpedia y CrossNature.

Todas las ontologías están disponibles públicamente en <https://github.com/Cross-Forest/Ontologies>.

4.2 Datos Enlazados Abiertos

El Conjunto de Datos Cross-Forest incluye cuatro conjuntos de datos enlazados: IFIS, IFNes, COS y MFE. Estos conjuntos de datos proporcionan una representación en forma de grafo de conocimiento de los inventarios forestales y mapas de cobertura del suelo de España y Portugal. El conjunto de datos IFIS permite comparar las formaciones forestales dominantes en diferentes regiones de España y Portugal. Contiene información estadística agregada sobre los tipos de bosque predominantes en cada región, como coníferas, frondosas o mixtos; las especies arbóreas más abundantes en cada formación; la extensión de cada formación en hectáreas; la organización jerárquica de las regiones administrativas (NUTS 1, NUTS 2, NUTS 3); y datos estadísticos como la superficie total o la superficie por hectárea. Por otro lado, el conjunto de datos IFNes ofrece información detallada sobre los árboles individuales y las parcelas de muestreo en España. Incluye datos sobre la ubicación geográfica y características de las parcelas, la especie, diámetro, altura y otras medidas dendrométricas de cada árbol, las coordenadas geográficas de las parcelas y los árboles (incluyendo posiciones absolutas y relativas), información sobre los diferentes sistemas de referencia utilizados y datos sobre el diseño del muestreo, como el tamaño de las parcelas y la densidad de muestreo.

Los conjuntos de datos COS (Mapa de Uso del Suelo Portugués) y MFE (Mapa Forestal Español) describen la cobertura del suelo en Portugal y España, respectivamente. Ambos conjuntos de datos contienen información detallada sobre la representación geométrica de las áreas (polígonos), las características de cada polígono (como el tipo de cobertura, la superficie y la fecha de la información) y una clasificación jerárquica de los diferentes tipos de cobertura. Sin embargo, existen algunas diferencias clave entre ambos conjuntos de datos. Mientras que COS se centra en la cobertura y uso del suelo en general, MFE se centra



específicamente en la cobertura forestal. Esto se traduce en un nivel de detalle diferente: MFE incluye información más específica sobre la cobertura forestal, como la especie dominante y la densidad de árboles, mientras que COS ofrece una visión más general de los usos del suelo. Además, la escala de los datos también varía: COS tiene una unidad mínima de 1 hectárea, mientras que MFE permite un mayor nivel de detalle.

Los conjuntos de datos están publicados en <https://github.com/Cross-Forest/Data> y se pueden consultar usando SPARQL en <https://crossforest.gsic.uva.es/sparql>. Los datos se encuentran publicados también en el portal de datos europeo (<http://www.europeandataportal.eu/>) y los portales de datos español (<https://datos.iepnb.es/def/sector-publico/medio-ambiente/>) y portugués (<https://snig.dgterritorio.gov.pt/>). Un punto SPARQL para consulta de los datos españoles se encuentra también en el portal del Ministerio para la Transformación Ecológica y el Reto Demográfico (<https://datos.iepnb.es/endpoint-sparql/>)

4.3 Usos del conjunto de datos Cross-Forest

El resultado de este trabajo ha servido como herramienta para facilitar el acceso a los datos de inventarios y mapas forestales en diversas contribuciones científicas. A continuación se presenta un resumen de las más relevantes.

El Conjunto de Datos Cross-Forest se utiliza como fuente en varias aplicaciones. Una de ellas es Forest Explorer (VEGA-GORGOJO, et al. 2022), una aplicación web (<https://forestexplorer.gsic.uva.es/>) que permite la visualización de datos forestales a través de un mapa interactivo. Forest Explorer, al estar basada en tecnologías web, es accesible desde cualquier dispositivo con conexión a Internet, lo que la convierte en una herramienta versátil tanto para un uso profesional, como para la educación y divulgación medioambientales. En el CFE9 presentamos una versión renovada de esta aplicación (VEGA-GORGOJO et al., 2025), que incluye muchas mejoras como la incorporación del IFN2, más opciones de visualización del mapa interactivo y un *dashboard* interactivo de parcelas. En la comunicación (CRESPO-LERA et al., 2025) describimos este *dashboard*, que puede emplearse con cualquier parcela del IFN2 o del IFN3, además de servir para la comparación de inventarios. Incluye un mapa interactivo con los árboles muestreados, figuras y tablas, permitiendo a los usuarios comparar variables como la densidad de arbolado, el área basimétrica y el volumen a lo largo del tiempo.

En (CRESPO-LERA et al., 2025a) presentamos un ejemplo de integración de datos en el Conjunto de Datos Cross-Forest, proponiendo una metodología para estimar las existencias forestales a nivel municipal a partir de los datos del IFN y los mapas de usos del suelo. Como resultado, hemos obtenido información detallada sobre las existencias forestales a escala local, facilitando así la gestión forestal. El municipio tiene una relevancia particular en la gestión de los ecosistemas forestales, ya que permite una planificación y manejo más específicos y adaptados a las necesidades locales (LAZDINIS et al., 2019).



Descarga IFN (VEGA-GORGOJO et al., 2025a) es otra aplicación web (<https://descargaifn.gsic.uva.es/>) que utiliza el Conjunto de Datos Cross-Forest. Está diseñada para que profesionales del sector forestal puedan visualizar y descargar parcelas del IFN2 e IFN3. Para ello, pueden definir sus zonas de interés seleccionando provincias o municipios, o incorporando sus propios polígonos desde un fichero o con una herramienta de dibujo. En la zona resultante pueden aplicarse filtros de especie, incluyendo rangos de valores deseados de existencias. Los formatos de descarga permitidos son GeoJSON, CSV y XSLX, facilitando así su uso profesional y académico.

El caso de Simanfor (BRAVO et al., 2025) ilustra otra aplicación del Conjunto de Datos Cross-Forest. Este simulador utiliza datos de inventarios forestales como estado inicial y combina modelos de crecimiento y rendimiento forestal definidos por las acciones selvícolas del usuario (aclareo y cosecha) para predecir la dinámica forestal a nivel de rodal y de árbol. La preparación de datos es una tarea exigente, que requiere mucho tiempo y es propensa a errores. Afortunadamente, los LOD forestales pueden emplearse para liberar a los usuarios finales de esta responsabilidad. Así, pretendemos utilizar Descarga IFN para obtener los ficheros de entrada que alimenten a Simanfor, facilitando de esta manera su utilización por profesionales del sector forestal.

El Conjunto de Datos Cross-Forest también se utiliza en (CRESPO-LERA et al., 2025b). Aquí presentamos varios casos complejos de gestión forestal que requieren la utilización de información del IFN para la toma de decisiones. Los casos presentados ilustran escenarios de gestión de masas forestales, considerando características estructurales, biodiversidad, perturbaciones o cambios a lo largo del tiempo, abarcando varias ediciones del IFN. Para resolverlos preparamos una serie de consultas en SPARQL al punto Cross-Forest. Es posible adaptar dichas consultas para utilizarlas en nuevos escenarios con sus propias necesidades.

5. Discusión

Este estudio presenta el Conjunto de Datos Cross-Forest, una representación en forma de grafo de conocimiento de los Inventarios Forestales y Mapas de Usos del Suelo de Portugal y España. Este conjunto de datos, en conjunción con el Conjunto de Ontologías Cross-Forest, ofrece una solución robusta para la integración y el análisis de datos forestales, abordando un desafío persistente en la ciencia forestal: la integración de información proveniente de diversas fuentes, con diferentes formatos y niveles de detalle.

La estructura flexible del grafo de conocimiento, inherente a las tecnologías de la Web Semántica, permite la integración de datos heterogéneos de una manera que las bases de datos tradicionales y los Sistemas de Información Geográfica (SIG) no pueden igualar. RDF (*Resource Description Framework*), el modelo de datos subyacente, permite representar la información como un conjunto de tripletes (sujeto, predicado, objeto), lo que facilita la conexión entre diferentes conjuntos de datos y la creación de un grafo de conocimiento interconectado. SPARQL, el



lenguaje de consulta para RDF, proporciona una forma potente y flexible de acceder a los datos y realizar análisis complejos. OWL (*Web Ontology Language*), por otro lado, permite añadir semántica formal a los datos, definiendo el significado de los términos y las relaciones entre ellos.

El Conjunto de Datos Cross-Forest, al proporcionar un único punto final para acceder a los datos forestales, contrasta con la situación actual, donde los conjuntos de datos de origen suelen estar dispersos en diferentes plataformas, formatos y tecnologías. La centralización del acceso a los datos a través de un punto SPARQL simplifica enormemente la tarea de los investigadores y gestores forestales, que pueden ahora formular consultas complejas que involucran diferentes fuentes de datos sin tener que lidiar con la heterogeneidad de formatos y la complejidad de integrar la información manualmente.

La capacidad de integrar datos forestales con otros conjuntos de datos no forestales es otra ventaja clave del Conjunto de Datos Cross-Forest. Al utilizar tecnologías de la Web Semántica, se facilita la conexión con datos de diferentes ámbitos, como información administrativa, socioeconómica o de infraestructuras. Esto permite, por ejemplo, analizar la relación entre la gestión forestal y el desarrollo rural, o estudiar el impacto de las infraestructuras en los ecosistemas forestales. Además, la publicación de los datos en formato RDF y la utilización de ontologías facilita el acceso y la comprensión de la información por parte de usuarios de otras áreas no especializadas en temas forestales, lo que promueve la reutilización de los datos y la colaboración multidisciplinar.

En trabajos recientes, hemos incorporado el Segundo Inventario Forestal Nacional (IFN2) al grafo de conocimiento Cross-Forest, enriqueciendo la información disponible y permitiendo el análisis de la evolución de los bosques españoles a lo largo de un periodo de tiempo mayor. Esta adición, que complementa la información del Tercer Inventario Forestal Nacional (IFN3) ya presente en el conjunto de datos, permite a los investigadores y gestores forestales realizar estudios comparativos entre ambas ediciones del inventario, analizar la dinámica de los bosques y evaluar el impacto de las políticas de gestión forestal a lo largo del tiempo. La integración del IFN2 se ha realizado utilizando las mismas tecnologías y metodología que se emplearon para el IFN3, lo que garantiza la coherencia y la interoperabilidad del conjunto de datos.

Actualmente, se está trabajando en la mejora y ampliación del Conjunto de Datos Cross-Forest. Se están integrando nuevas fuentes de datos, como información climática y datos LiDAR, y se está alineando el conjunto de ontologías con otras ontologías relevantes en el ámbito forestal. Estas mejoras aumentarán aún más el valor del Conjunto de Datos Cross-Forest como herramienta para la investigación y la gestión forestal. Además, se está trabajando en la integración de datos de otros países, con el objetivo de crear una plataforma global para la gestión de la información forestal.



6. Conclusiones

Este estudio ha presentado el Conjunto de Datos Cross-Forest, un recurso innovador que facilita la integración y el análisis de datos forestales de España y Portugal. Mediante el uso de tecnologías de la Web Semántica y los principios de los Datos Abiertos Enlazados, este conjunto de datos ofrece una solución robusta para la gestión de la información forestal.

El Conjunto de Datos Cross-Forest proporciona una serie de beneficios clave:

- Integración de datos heterogéneos de diferentes fuentes.
- Interoperabilidad con otras fuentes de datos y herramientas.
- Capacidad de razonamiento para la generación de nuevo conocimiento.
- Acceso abierto a la información para la comunidad forestal internacional.

Este conjunto de datos tiene el potencial de ser utilizado en una variedad de aplicaciones, como el análisis de la distribución espacial de las especies, la evaluación de la evolución de la cobertura forestal, el estudio del impacto de los incendios forestales y la modelización del crecimiento forestal.

Se espera que el Conjunto de Datos Cross-Forest se convierta en una herramienta importante para la investigación y la gestión forestal, contribuyendo a la toma de decisiones informadas, la gestión sostenible de los bosques y la conservación de la biodiversidad. Al facilitar la integración con otros conjuntos de datos y el acceso a la información por parte de usuarios de diferentes áreas, el Conjunto de Datos Cross-Forest promueve la colaboración multidisciplinar y la generación de nuevo conocimiento.

7. Agradecimientos

Esta investigación ha sido financiada por el proyecto nacional LOD.For.Trees (TED2021-130667B-I00), por el proyecto europeo Small4Good (101135517) y la Consejería de Educación, Junta de Castilla y León (ORDEN EDU/1009/2024).

8. Bibliografía

BIZER, C., VIDAL, M.-E., AND SKAF-MOLLI, H. 2018. Linked open data. Encyclopedia of Database Systems. Springer, 2096–2101, New York

BRAVO, F., FABRIKA, M., AMMER, C., BARREIRO, S., BIELAK, K., COLL, L., ET AL. 2019. Modelling approaches for mixed forests dynamics prognosis. research gaps and opportunities. Forest Systems 28(1) eR002

BRAVO, F., ORDÓÑEZ, C., VÁZQUEZ-VELOSO, A., MICHALAKOPOULOS, S., 2025. SIMANFOR: herramienta renovada para la toma de decisiones forestales en el siglo



XXI. Actas del 9º Congreso Forestal Español

CRESPO-LERA, N., VEGA-GORGOJO, G., GIMÉNEZ-GARCÍA, J., VÁZQUEZ-VELOSO, A., BRAVO, F., RUANO, I., 2025. Dashboard para la Visualización y Consulta de Información Forestal Multi-Inventario. Actas del 9º Congreso Forestal Español

CRESPO-LERA, N., VEGA-GORGOJO, G., GIMÉNEZ-GARCÍA, J., BRAVO, F., RUANO, I., 2025a. Estimación de existencias forestales por municipio a partir de inventarios nacionales y mapas de usos del suelo. Actas del 9º Congreso Forestal Español

CRESPO-LERA, N., VÁZQUEZ-VELOSO, A., VEGA-GORGOJO, G., GIMÉNEZ-GARCÍA, J., BRAVO, F., RUANO, I., 2025b. Unificando el IFN con tecnologías semánticas para su análisis multidimensional. Actas del 9º Congreso Forestal Español

HARRIS, S., AND SEABORNE, A. 2013. SPARQL 1.1 Query Language. Available at: <http://www.w3.org/TR/sparql11-query/> (accessed January 17, 2025).

HEATH, T., AND CHRISTIAN B. 2011. Linked Data: Evolving the Web into a Global Data Space. Springer, 122 p. Cham

HITZLER, P., KRÖTZSCH, M., PARSIA, B., PATEL-SCHNEIDER, P. F., AND RUDOLPH, S. 2012. OWL 2 Web Ontology Language Primer. Available at: <https://www.w3.org/TR/owl2-primer/> (accessed January 17, 2025).

LAZDINIS, M.; ANGELSTAM, P.; PÜLZL, H. 2019. Towards sustainable forest management in the European Union through polycentric forest governance and an integrated landscape approach. *Landscape Ecol* 34, 1737–1749

LEFRANÇOIS, M., ZIMMERMANN, A., AND BAKERALLY, N. 2017. A SPARQL extension for generating RDF from heterogeneous formats. *Proceedings of the 14th Extended Semantic Web Conference*. Springer, 35–50, Portoroz

PRETZSCH, H. 2009. *Forest Dynamics, Growth and Yield: From Measurement to Model*. Springer. 664 p. Heidelberg

RUIZ-PEINADO, R., BRAVO-OVIEDO, A., LÓPEZ-SENEPLEDA, E., BRAVO, F., AND DEL RÍO, M. 2017. Forest management and carbon sequestration in the mediterranean region: a review. *Forest Systems* 26(2) eR04S

SCHREIBER, G., AND RAIMOND, Y. 2014. RDF 1.1 Primer: World Wide Web Consortium. Available at: <https://www.w3.org/TR/rdf11-primer/> (accessed January 17, 2025).



TOMPPO, E., GSCHWANTNER, T., LAWRENCE, M., AND MCROBERTS, R. E. 2010. National Forest Inventories: Pathways for Common Reporting. Springer. 612 p. Dordrecht

VEGA-GORGOJO, G., GIMÉNEZ-GARCÍA, J.M., ORDÓÑEZ, C., BRAVO, F., 2022. Pioneering easy-to-use forestry data with Forest Explorer. Semantic Web, 13(2), 147-162

VEGA-GORGOJO, G., GIMÉNEZ-GARCÍA, J.M., CRESPO-LERA, N., ORDÓÑEZ, C., BRAVO, F., RUANO, I., 2025. Un nuevo Explorador Forestal de los bosques ibéricos. Actas del 9º Congreso Forestal Español

VEGA-GORGOJO, G., EVAGGELATOS, S., PERREAS, N., MICHALAKOPOULOS, S., GIMÉNEZ-GARCÍA, J.M., CRESPO-LERA, N., ORDÓÑEZ, C., BRAVO, F., 2025a. Descarga IFN: aplicación web para la extracción multicriterio de parcelas. Actas del 9º Congreso Forestal Español

VIDAL, C., ICIAR, A., LAURA, H., AND JOHN, R. 2016. National Forest Inventories: Assessment of Wood Availability and Use. Springer. 845 p. Cham

ZOU, W., JING, W., CHEN, G., LU, Y., AND SONG, H. 2019. A survey of big data analytics for smart forestry. IEEE Access 7, 46621–46636.