

Big Data y Ordenación de Montes: ¿qué ha cambiado en realidad?

Vega-García C.^{1,2,*}, García R.³

¹Departamento de Ingeniería Agroforestal (EAGROF), Escuela Técnica Superior de Ingeniería Agraria (ETSEA),
Universitat de Lleida, Alcalde Rovira Roure 191, 25198 Lleida.

²Centre Tecnològic Forestal de Catalunya (CTFC), Carretera de Sant Llorenç de Morunys km 2, 25280 Solsona

³Departamento de Informática e Ingeniería Industrial (DIEI),
Universitat de Lleida, Jaume II 69, 25001 Lleida, Spain

*Autor para correspondencia: cvega@eagrof.udl.es/cvega@eagrof.udl.cat

Resumen

Aunque la ordenación de montes ha sido necesariamente multifuncional en nuestras condiciones mediterráneas, nunca ha experimentado más y más diversas demandas que en la actualidad, y es de esperar que éstas aumenten en intensidad en un futuro previsible en el que el desafío radicará en cómo conseguir el aumento de la productividad forestal sin perjuicios para otras funciones.

Los avances experimentados por tecnologías de adquisición y análisis de datos (Big Data) medioambientales y forestales a alta resolución espacial y temporal (teledetección, LIDAR), prácticamente a la carta, suministran un nivel de conocimiento y análisis de los sistemas forestales que permite optimizar la toma de decisiones, la modelización y la simulación de sus dinámicas en el futuro, y a su vez anticipar la provisión de los numerosos servicios ecosistémicos demandados.

Esta ponencia repasa dichos avances tecnológicos, al tiempo que establece su relación con condicionantes estructurales que han lastrado la planificación forestal desde sus inicios en el país: la estructura de la propiedad, y las políticas forestales.

Palabras clave: planificación forestal, *machine learning*, datos georreferenciados.

1. Las demandas actuales sobre la ordenación de montes

Se puede decir que, desde el momento que la humanidad ha sido capaz de considerar su futuro, ha gestionado sus bosques [1]. Todas las civilizaciones alteran su entorno para adaptarlo a sus necesidades y desarrollo técnico, pero las evidencias de iniciativas de compatibilización de conservación y aprovechamiento de los bosques son muy numerosas en la historia de la humanidad, hasta alcanzar el grado de ciencia a partir del siglo XIX. Actualmente la ordenación de montes es una ciencia consolidada por siglos de actividad y desarrollo profesional y científico-técnico.

Qué persigue la ordenación de montes? Según la definición de la *Office national des forêts* francesa (1969): ordenar un monte es fijar primeramente unos objetivos, para a continuación prever las medidas necesarias para alcanzarlos. La ordenación de montes en nuestras limitadas condiciones mediterráneas ha debido tradicionalmente considerar objetivos múltiples. Las demandas sobre la ordenación son incluso más y más diversas en la actualidad, y se espera que aumenten en intensidad en un futuro previsible en el que el desafío radicará en cómo conseguir el aumento de la productividad forestal sin perjuicios para otros objetivos, hoy reformulados de acuerdo a su funcionalidad. La realización de las nuevas demandas multifuncionales sobre bienes y servicios exige formulaciones más precisas, como, por ejemplo, la clasificación de [2], que incluye 23 funciones de regulación, hábitat, productivas y de información.

La necesidad de satisfacer múltiples funciones, y proveer múltiples bienes y servicios demanda la utilización de cada vez más información y de más calidad y mayor resolución. Los inventarios y las cartografías asociadas a los proyectos de ordenación han adquirido una mayor complejidad e importancia.

Afortunadamente, los avances experimentados por tecnologías de adquisición y análisis de datos (Big Data) medioambientales y forestales a alta resolución espacial y temporal (teledetección, LIDAR), prácticamente a la carta, suministran un nivel muy elevado de conocimiento y análisis de los sistemas forestales, en paralelo a estas demandas.

Datos y técnicas permiten teóricamente optimizar la toma de decisiones, la modelización y la simulación de sus dinámicas en el futuro, y poder prescribir actuaciones anticipando la provisión de los numerosos servicios ecosistémicos demandados.

2. Fuentes de datos relevantes:

Entre las fuentes de datos actuales más destacables, y sin ánimo de exhaustividad, deberíamos considerar las posibilidades que ofrecen los datos de los inventarios forestales nacionales (<http://www.mapama.gob.es/es/desarrollo-rural/temas/politica-forestal/inventario-cartografia/inventario-forestal-nacional/default.aspx>), los datos georreferenciados de organismos oficiales de ámbito europeo, como el Joint Research Centre de la Unión Europea (<http://forest.jrc.ec.europa.eu/european-atlas-of-forest-tree-species/>), nacional como el Instituto Geográfico Nacional (<http://pnoa.ign.es>), o autonómico, como el Instituto Cartográfico y Geológico de Cataluña (<http://www>.

icc.cat/vissir3), con productos como LidarCat o PNOA, y sus productos derivados.

De especial relevancia para el sector forestal ha sido la disponibilidad de imágenes digitales, por ejemplo de los satélites Landsat desde 1984, y actualmente del Sentinel-2 (<http://www.nosolosig.com/articulos/618-descarga-los-datos-del-sentinel-2-a-de-forma-facil-y-rapida>); (<https://sentinel.esa.int/web/sentinel/missions/sentinel-2>); (<http://www.icgc.cat/en/Public-Administration-and-Enterprises/Downloads/Aerial-photos-and-orthophotos/Sentinel-2-orthoimages>), con enormes posibilidades de mejora del conocimiento de las masas forestales. Ya hay aplicaciones de visualización desarrolladas de gran interés (<http://84.88.72.6/meta-comparador/?simple=true&id=Ortoimatges-Sentinel-2-Catalunya&num=4>).

“Data literacy”

La abundancia de fuentes de datos espaciotemporales y posibles aplicaciones nos lleva a cuestionar si estamos preparados para procesar y aprovechar adecuadamente estos volúmenes de información, que requieren generalmente conocimientos avanzados, por ejemplo, de minería de datos o geo-estadística.

De acuerdo a expertos en este ámbito, “no todo el mundo es capaz de analizar e interpretar datos”, “hay factores no necesariamente racionales, el convencimiento extendido de que solo el ser humano puede tomar la decisión adecuada”, “la realidad es que hasta ahora no nos habíamos topado con tal cantidad de datos en formatos tan diversos en tiempo real”. A. Pita, nominado mejor científico de datos en los Data Science Awards de 2016, establece que “sin información, solo eres una persona con su opinión. Lo mejor que podemos hacer es preguntar a los datos y valorar si la respuesta es razonable. No nos podemos fiar al 100% de los datos ni de nuestra intuición” (https://retina.elpais.com/retina/2017/05/30/tendencias/1496154380_977310.html).

En consecuencia, la Unión Europea ha definido las técnicas ligadas a Big Data como el nuevo petróleo (N. Kroes, Comisaria Europea Agenda Digital de la UE, 2016).

¿Qué caracteriza a un conjunto de datos como “Big Data”?

Tal y como se muestra en la *Figura 1*, se puede caracterizar lo que se considera como un problema de datos masivos en base a las 4 Vs, ya que las características que los suelen definir empiezan por esta letra:

- **Volumen** (escala de los datos): se refiere al tamaño de los datos, medido normalmente en múltiplos del byte. Por ejemplo un Megabyte corresponde a un millón de bytes y un Gigabyte a mil millones. A su vez, cada byte agrupa 8 bits, que pueden tomar individualmente el valor de 1 o 0 pero que conjuntamente permiten 256 combinaciones distintas. Normalmente, se utilizan para representar una letra o número. Se estima que cada día se generan 2,3 billones de Gigabytes de datos nuevos, aproximadamente el equivalente a 2.300 millones de discos duros de 1 Terabyte (1 billón de bytes).

- **Velocidad (de flujos de datos):** en este caso la dificultad no tiene que ver con el volumen de los datos en su conjunto, el problema es la velocidad a la que “llegan” y se tienen que procesar. Un caso típico es el de los sensores que recogen datos en tiempo real de temperatura, presión, humedad,... Un coche moderno dispone de más de 100 sensores de este tipo. Otro caso son los datos generados en redes sociales, por ejemplo Twitter, o por sistemas como los de la negociación en las bolsas de valores. La *New York Stock Exchange* genera más de un Terabyte de datos durante cada sesión. Y tanto en este caso como en el de los sensores de los coches, el valor está en ser capaz de procesarlos en tiempo real, no en diferido.

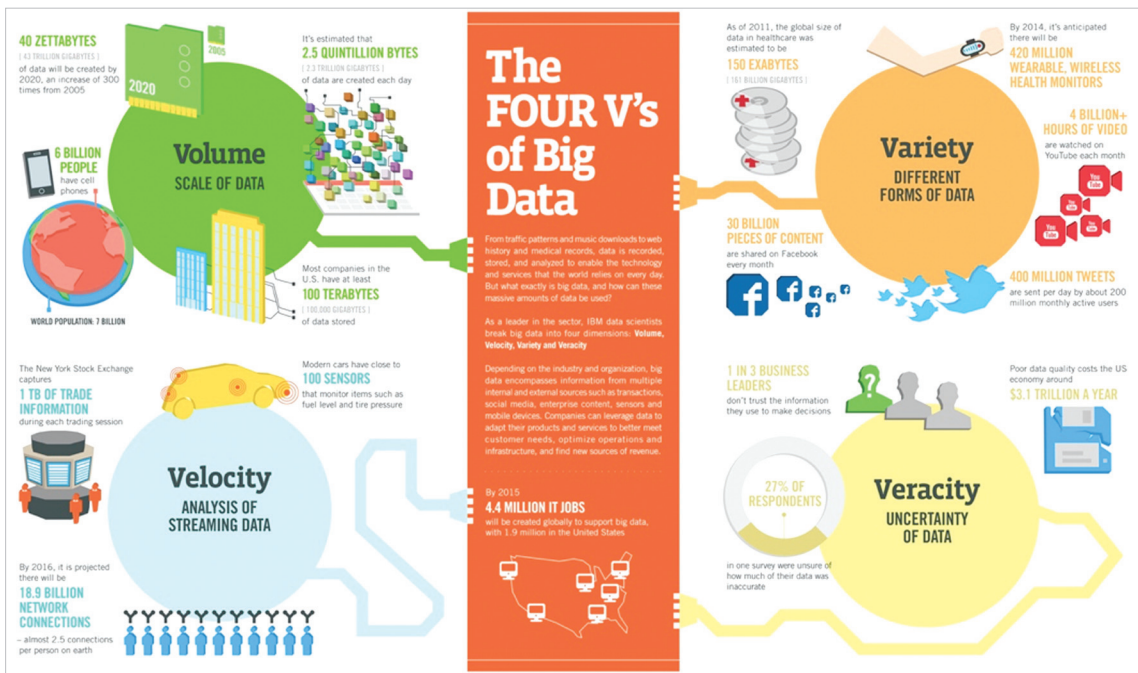


Figura 1. Las 4 Vs del “Big Data”

Fuentes: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC y QAS.

- **Variedad (tipos de datos):** cuando los datos no son uniformes y se combinan tipos diferentes de datos provenientes de múltiples fuentes, el reto es su integración, que potencialmente permitiría descubrir nuevo conocimiento como correlaciones ocultas u otros tipos de patrones no detectables en los datos por separado. En este campo es de gran ayuda el poder automatizar el procesamiento de los datos a nivel semántico, es decir, de su significado. Que el ordenador tenga acceso a información que le permita “saber” que el campo 3 de un conjunto de datos es equivalente al campo 7 de otro conjunto de datos y que por lo tanto se pueden integrar. O que el primero corresponde a una temperatura en grados Celsius y el segundo a una presión en hPa y que por lo tanto tendría sentido buscar correlaciones entre ambos. Este tipo de datos a los que se les incorpora este nivel de “inteligencia” se denomina Smart Data.

- **Veracidad (inexactitudes en los datos):** esta característica está íntimamente relacionada con la calidad de los datos, que en la mayoría de los casos deja bastante que desear. Por ejemplo errores de los sensores, falta de lecturas por pérdidas de conexión o baterías agotadas. O los tan comunes errores humanos si existe algún paso de entrada manual de datos. Se estima que la falta de calidad de los datos cuesta a la economía de EE.UU. 3.100 billones de dólares al año. Por lo tanto, un problema con el que resulta realmente difícil lidiar.

No es necesario que todas las anteriores características se den a la vez para que se trate de un caso de “Big Data”, solo una de ellas ya puede convertir el procesamiento de los datos en un gran reto a nivel computacional. Aunque en muchos casos nos encontraremos con 2 o 3 de ellas. Por lo tanto, se trataría de un problema de “Big Data” si los datos a tratar tienen una o más de estas características. Es decir, gran cantidad de datos (la más típica), un gran flujo de datos, gran heterogeneidad o datos de poca calidad, con muchas inexactitudes que requieren un procesamiento sofisticado para que sean útiles.

Analizando Big Data

Como se ha mencionado anteriormente, el procesamiento de “Big Data” requiere una gran capacidad de cómputo, que se traduce en que normalmente es necesario más de un ordenador para procesar los datos, realmente “nubes” de ellos, y técnicas avanzadas de procesamiento de datos. Estas técnicas van más allá de las puramente estadísticas, aunque el análisis estadístico de “Big Data” es en cualquier caso muy útil y la base de algunas de estas otras técnicas. La *Figura 2*, expone algunas técnicas de análisis de “Big Data”, resaltando su creciente potencial estratégico pero también su grado de madurez. Se destacan una serie de técnicas relacionadas con el Aprendizaje Automático (*Machine Learning*) que permiten la generación de modelos de los datos (con supervisión o de manera totalmente automática) que permiten realizar predicciones sobre los sistemas que generan esos datos sin la necesidad de conocimiento a priori sobre los mismos.

Algunas de las técnicas para el Aprendizaje Automático que están generando más expectativas recientemente son las que imitan algunos de los aspectos de las redes neuronales. Los avances más recientes en esta línea se basan en redes neuronales complejas que permite el denominado Aprendizaje Profundo (*Deep Learning*), muy apropiado para el procesamiento de datos provenientes de sensores, como por ejemplo datos LiDAR.

Tal y como se muestra en la *Figura 3*, y como detallan Guan *et al.*¹, se pueden utilizar redes neuronales y Aprendizaje Profundo para procesar datos de un dispositivo LiDAR móvil y reconocer y clasificar automáticamente árboles. El proceso empieza con un subconjunto de los datos LiDAR en el que se ha anotado manualmente la especie de cada árbol. Se alimenta la red neuronal con estos datos, que aprende a partir de los ejemplos las características propias de cada especie en los datos LiDAR. Después del periodo de entrenamiento, es posible empezar a “preguntar” a la red

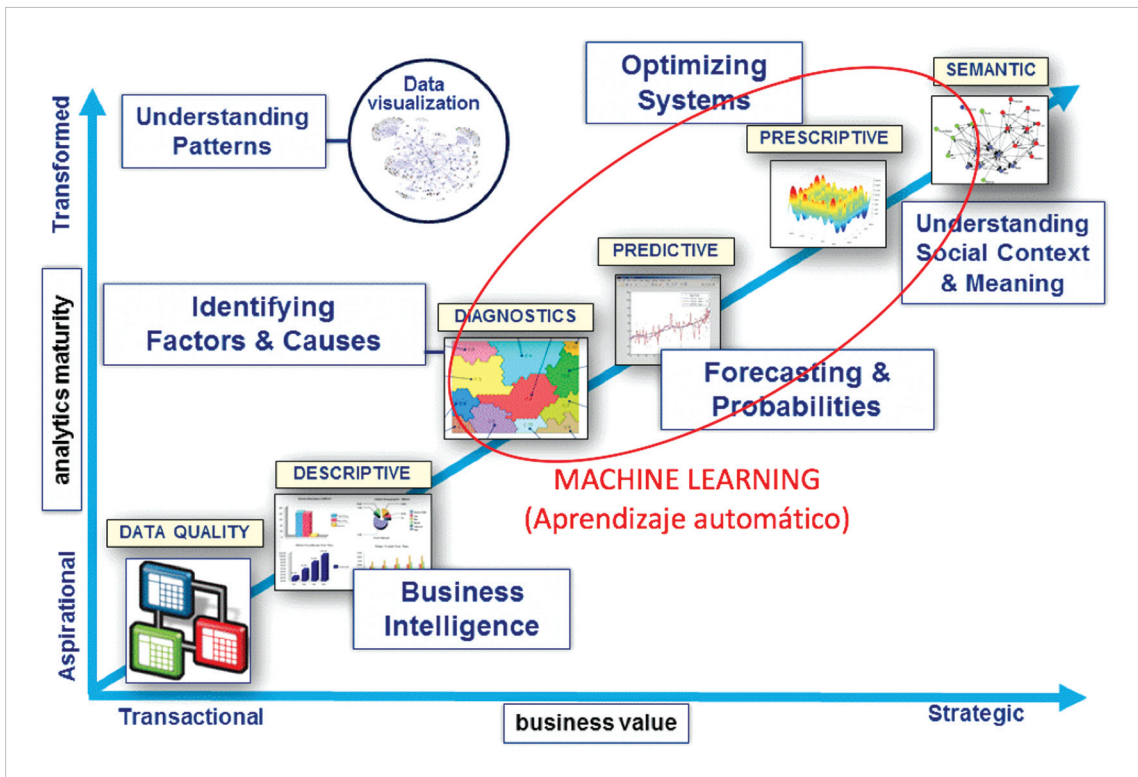


Figura 2. Técnicas para el análisis de “Big Data” según su valor de negocio y nivel de madurez
 Fuente: <http://www.analyticbridge.datasciencecentral.com/profiles/blogs/twelve-emerging-trends-in-data-analytics-part-1-of-4>

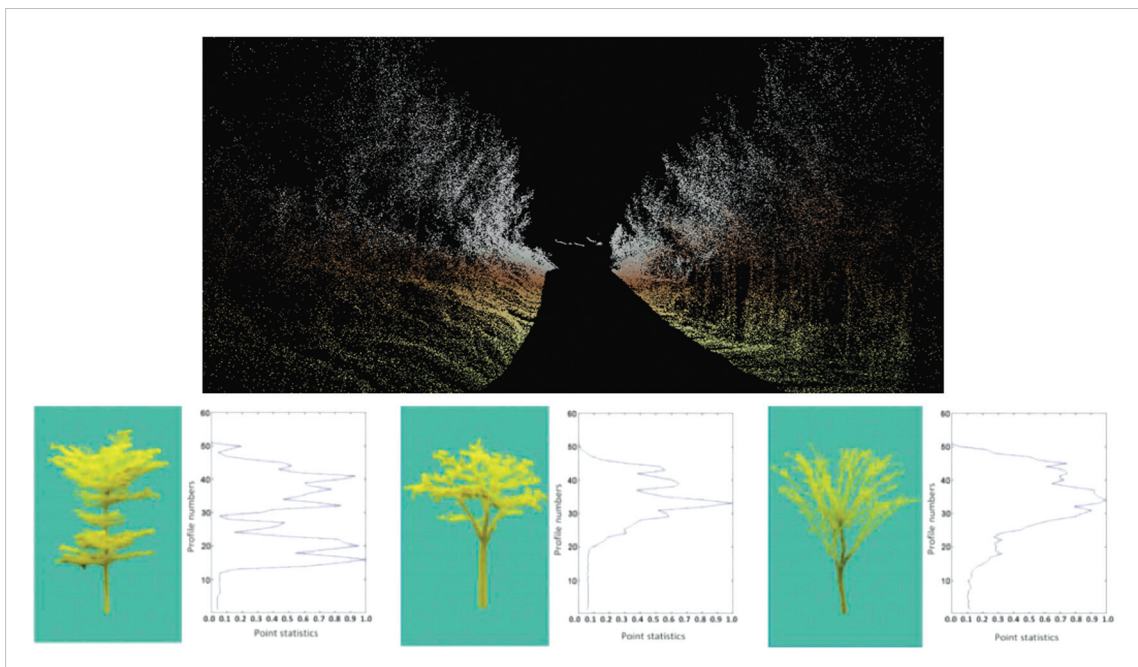


Figura 3. Uso de datos LiDAR para el reconocimiento automático de árboles.
 Fuente: <http://www.analyticbridge.datasciencecentral.com/profiles/blogs/twelve-emerging-trends-in-data-analytics-part-1-of-4>

neuronal a qué especie corresponde un árbol en los datos LiDAR. Normalmente, se utiliza un segundo subconjunto de datos, también anotado manualmente, para verificar las respuestas de la red neuronal y determinar la tasa de acierto conseguida con el entrenamiento. Tal y como informan Guan *et al.* [1] en su artículo, han conseguido una tasa de acierto de casi el 90%.

¿Cómo suele ser “Big Data” en los ámbitos Agrifood y Forestal?

En el contexto agroalimentario y forestal, se pueden identificar ciertos aspectos característicos de los grandes conjuntos de datos relacionados con estos dominios. Por ejemplo, como resultado de un reciente taller sobre “Big Data” para alimentación, agricultura y forestales¹, organizado por *Big Data Europe* y con la participación, entre otros, de la FAO de Naciones Unidas o el Instituto Francés de Investigación en Agricultura (INRA), se identificaron los siguientes patrones como típicos del “Big Data” en estos ámbitos:

- Volumen datos no extremadamente alto.
- Velocidad de datos usualmente baja.
- Veracidad normalmente alta.
- Variedad extremadamente alta, principal problema.
- Combinación de múltiples y heterogéneos tipos de datos y formatos de fuentes variadas.

En base a estas características típicas del “Big Data” agroalimentario y forestal, el principal reto es la integración de datos como respuesta a problemas de la variedad. Las principales propuestas para solventar este reto, y tal y como se muestra en la parte superior derecha de la *Figura 2*, están relacionadas con la semántica (el significado datos) y la representación de conocimiento (utilizando herramientas como los esquemas de datos, los tesauros o las ontologías).

Los principales beneficios de estas propuestas es que facilitan la integración automática o asistida de datos. También que los tratamientos de datos sean más consistente. Por ejemplo, porque los datos van acompañados de metadatos que los describen y dejan claro cómo tienen que interpretarse y sus unidades de medida. Otros ejemplos son metadatos que permiten el seguimiento, su procedencia, o los derechos asociados a su reutilización.

Ejemplo de Uso Semántica y Ontologías

Un ejemplo de aplicación de semántica y ontologías para el tratamiento de “Big Data” en estos ámbitos es el proyecto de la *European Food Safety Authority* (EFSA) monitorización de los medios para la identificación y detección de amenazas para la

¹ “Big data for food, agriculture and forestry: opportunities and challenges”, Septiembre 2015
<https://www.big-data-europe.eu/event/sc2-paris-2015/>

salud de las plantas. En este proyecto, el conjunto de datos de partida no es sencillo, una lista de 140 enfermedades y plagas a monitorizar entre las identificadas en los listados de EPPO *Alerts*² o los de la Unión Europea en la directiva 2000/29 1-A-1 y las EU *Emergency Control Measures*³.

El mayor reto en este caso, tal y como se ha comentado que es típico en el “Big Data” de los entornos agroalimentario y forestal, es la integración de datos relativos a estas plagas. El objetivo es integrar la mayor cantidad de términos relacionados con estas plagas, en multitud de idiomas, para poder detectar menciones en los medios que puedan estar relacionados.

En este caso se integran los datos de 115 de estas amenazas a partir de la *UniProt Taxonomy*, una base de datos de referencia sobre información taxonómica que incluye nombres científicos y también comunes en múltiples idiomas. También se integran los datos extraídos para 45 de ellas de Wikipedia.

Toda esta información se integra gracias a la semántica para conformar una Ontología de Plagas y Enfermedades de las Plantas. Una ontología es una representación formal, que se puede procesar mediante ordenador, de una conceptualización de un dominio. En este caso permite representar los conceptos clave de enfermedad y plaga de las plantas, además de sus interrelaciones con los cultivos afectados, huéspedes, vectores... El foco está en los conceptos, que se asocian entonces a todas las formas de denominarlos con términos científicos y comunes en todos los idiomas recopilados.

Mediante la ontología, es posible entonces organizar de una forma más eficiente toda la información sobre las plagas y enfermedades. Se puede incluso integrar esa información de una manera automática y utilizarla para guiar a los expertos a la hora de capturar información adicional, por ejemplo evitando que la misma información se vuelva a introducir utilizando otro nombre científico equivalente. Información adicional sobre este proyecto esta publicada en [2] y [3].

¿Qué impacto potencial tiene estas técnicas en la Ordenación de montes?

Tras repasar estos avances tecnológicos, debemos sin embargo ponerlos en perspectiva con una realidad ordenadora bastante alejada de la realidad Big Data: A pesar del alto potencial de datos y técnicas para la mejora de la planificación forestal multifuncional, y de la demanda creciente de bienes y servicios forestales, existen condicionantes estructurales muy importantes que limitan su aplicabilidad. Estos condicionantes estructurales han lastrado la planificación forestal desde sus inicios, y están sobre todo sobre todo relacionados con la estructura de la propiedad y las políticas forestales.

Aunque los beneficios para la inventariación forestal, sin embargo, son ya patentes, la toma de decisiones en sí para la prescripción de intervenciones en masas arboladas ordenadas no se ha incorporado al universo de datos en abierto. Las políti-

² https://www.eppo.int/QUARANTINE/Alert_List/alert_list.htm

³ https://ec.europa.eu/food/plant/plant_health_biosecurity/legislation/emergency_measures_en

cas forestales actuales no parecen prever o favorecen todavía la existencia de bases de datos de ordenaciones revisadas y aprobadas o ejecutadas que permitan compartir datos en abierto, formatos, cartografía, prescripciones y sus realizaciones (control). Los proyectos de ordenación están actualmente aislados en sus administraciones e instituciones forestales responsables, a menudo vinculados a aplicativos y software propios de cada comunidad autónoma. Parcialmente, esto es consecuencia de la estructura de la propiedad forestal, mayormente particular, factor que ha acompañado históricamente el desarrollo de la Dasonomía u Ordenación de montes, y que hasta cierto punto impide la divulgación de datos que serían de gran interés. Esta misma estructura de la propiedad genera masas forestales y condicionantes forestales poco favorables a la selvicultura y planificación forestal, que no han sido capaces de resolver las políticas forestales actuales.

Las tendencias a la ordenación en escalas más detalladas de los últimos años mediante ordenación por rodales, sin embargo, abren la puerta a una ordenación más precisa y adaptada a las demandas multifuncionales de bienes y servicios actuales y a la generación de información potencialmente incorporable al universo Big Data.

3. Bibliografía

- [1] Davis LS, Johnson KN. (1987). *Forest management* 3rd ed. McGraw-Hill, New York. ISBN 0070326258.
- [2] De Groot RS, Wilson MA, Boumans RMJ. A typology for the classification, description and valuation of ecosystem functions, goods and services. *Ecological Economics* 41 (2002) 393–408.
- [3] Guan H, Yu Y, Ji Z, Li J, and Zhang Q. (2015). Deep learning-based tree classification using mobile LiDAR data. *Remote Sensing Letters*, 6(11), 864-873.
- [4] Alomar O, Batlle A, Brunetti JM, García R, Gil R, Granollers A, Jiménez S, Laviñ, A, Linge JP, Pautasso M and Reverté C. (2015). Development and testing of the media monitoring tool MedISys for early identification and reporting of existing and emerging plant health threats. *EPPO Bulletin*, 45(2), 288-293.
- [5] Alomar O, Batlle A, Brunetti JM, García R, Gil R, Granollers T, Jiménez S, Laviña A, Reverté C, Riudavets J and Virgili-Gomà J. (2016). Development and testing of the media monitoring tool MedISys for the monitoring, early identification and reporting of existing and emerging plant health threats. *EFSA Supporting Publications*, 13(12).

