



Seguimiento de la biodiversidad en la era del Big Data

Laura Hernández^{1,2,*} , Jose M. Álvarez-Martínez³ , Cristina Gómez Almaraz^{4,5} , Rut Sánchez de Dios² , Borja Jiménez Alfaro⁶ , Flor Álvarez-Taboada⁷ 

- (1) ICIFOR, Institute of Forest Research (INIA, CSIC). Ctra. La Coruña, Km. 7.5, ES-28040, Madrid, España.
(2) Departamento de Biodiversidad, Ecología y Evolución. Universidad Complutense de Madrid. C/ José Antonio Novais, 12. ES-28040 Madrid, España.
(3) Instituto de Hidráulica Ambiental de la Universidad de Cantabria "IH Cantabria". C/ Isabel Torres, 15, ES-39011 Santander, Cantabria, España.
(4) IuFOR-EiFAB, Universidad de Valladolid (UVA) Av. de Madrid, 503, ES-34004 Palencia, España.
(5) Department of Geography and Environment, School of Geoscience, University of Aberdeen, King's College, Aberdeen, AB24 3FX, Scotland, UK.
(6) Vegetation & Biodiversity Lab. Biodiversity Research Institute (IMIB). Universidad de Oviedo. C/ Gonzalo Gutiérrez Quirós s/n E-33600 Mieres, Spain.
(7) Escuela de Ingeniería Agraria y Forestal. Universidad de León. Av. Astorga, 15, 24401 Ponferrada, León, España.

* Autor de correspondencia: L.Hernández [hernandez.laura@inia.csic.es]

> Recibido el 12 de septiembre de 2022 - Aceptado el 16 de septiembre de 2022

Como citar: Hernandez, L., Álvarez-Martínez, J. M., Gómez, C., Sánchez de Dios, R., Jiménez Alfaro, B., Álvarez-Taboada, F. 2022. Seguimiento de la biodiversidad en la era del Big Data. *Ecosistemas* 31(3): 2450. <https://doi.org/10.7818/ECOS.2450>

"Digital data are accumulating at unprecedented rates. These contain a lot of information about the natural world, some of which can be used to answer key ecological questions."

(Jarić et al. 2020)

Introducción

La diversidad biológica o biodiversidad es de vital importancia para la persistencia de los ecosistemas terrestres, ya que constituye el pilar que da estabilidad funcional a los sistemas naturales y proporciona una gran variedad de servicios ecosistémicos esenciales para el bienestar humano. Pero, ¿cuál es el estado actual de la biodiversidad? ¿Cómo estamos progresando con respecto a los objetivos de conservación establecidos para limitar o reducir la extinción de especies? ¿Cuáles son las principales presiones y amenazas para la biodiversidad derivadas del Cambio Global? Estas preguntas sólo pueden responderse si existe un conocimiento sólido sobre el estado y tendencias de aspectos esenciales que gobiernan los patrones y procesos de la biodiversidad a escalas espacio-temporales complementarias (Kühl et al. 2020). Este conocimiento es crucial para desarrollar políticas de conservación eficaces y una gestión medioambiental que reverta las tendencias de muchas poblaciones y comunidades actualmente en declive (IPBES 2019). Por este motivo, disponer de documentación exhaustiva derivada de programas de seguimiento de la distribución, estructura y funcionamiento de la biodiversidad, así como los efectos del Cambio Global sobre su conservación, son esenciales para alcanzar los objetivos de muchos reglamentos internacionales. Entre estas disposiciones destacan los objetivos estratégicos de las metas de Aichi, definidas para evaluar el progreso hacia los objetivos del Convenio sobre la Diversidad Biológica (CBD 2014). En concreto, el objetivo B: "Reducir las presiones directas sobre la biodiversidad y promover su uso sostenible" y C: "Mejorar la situación de la biodiversidad salvaguardando los ecosistemas, las especies y la diversidad genética", requieren un seguimiento exhaustivo del estado y tendencias de la biodiversidad. Igualmente, la Agenda 2030 para los Objetivos de Desarrollo Sostenible, concretamente el ODS 14 "La vida bajo el agua" y el ODS 15 "La vida en la tierra", exigen un amplio esfuerzo para medir el

progreso hacia el uso sostenible de los recursos naturales y la conservación de la biodiversidad en el agua y en la tierra (UN 2015).

Dada la importancia de los datos para dar respuesta a estos desafíos, la Ecología, como otras ramas de la Biología y Ciencias de la Tierra, se ha hecho eco del fenómeno denominado "Big Data" o "Data Science". El volumen mundial de datos se ha multiplicado nueve veces en los últimos años y continúa creciendo de forma exponencial (Farley et al. 2018), lo que confirma el establecimiento y continuidad futuros de este paradigma que brinda nuevas oportunidades, pero también retos, ya que requiere de nuevas herramientas, técnicas, formas de trabajo y marcos teóricos (Kitchin 2014).

Una mayoría de autores acepta el término Big Data, que empezó a usarse décadas recientes en el ámbito del comercio electrónico cuando las empresas privadas buscaban nuevas formas de desarrollar y controlar grandes cantidades de datos, principalmente para mejorar su rendimiento (McAfee y Brynjolfsson 2012; Kitchin 2014). El término se ha extendido después a muy diversos ámbitos, desde la política a la industria, y a diferentes ramas científicas, generando la necesidad de una definición universal del mismo. Aunque esta definición es compleja y varía entre disciplinas, sí existe un consenso general sobre los elementos clave que caracterizan al Big Data. En términos generales, se define como la capacidad de gestionar y analizar datos, que a su vez deben caracterizarse a través de las denominadas 4Vs: volumen, velocidad, variedad y veracidad (Farley et al. 2018; Musvuugwa et al. 2021). El volumen se refiere a la gran cantidad de datos recopilados; la velocidad a la tasa con que se recopilan; la variedad alude a su estructura o heterogeneidad y la veracidad indica su fiabilidad. En el contexto de este estudio, relacionado con el seguimiento de la biodiversidad, el Big Data se ha definido como una "herramienta tecno-política para gestionar la distribución de las especies y comunidades biológicas", y como "la acumulación intensiva de información digitalizada de la distribución espacial y temporal de especies y comunidades" (Devictor y Bensaude-Vincent 2016). En esta monografía nos referimos a Big Data como un concepto amplio que considera la capacidad de gestionar y analizar una gran cantidad de datos espacio-temporales y heterogéneos sobre seguimiento de la biodiversidad, flexibilizando así las características de volumen y velocidad.

Pronosticado por Jay Grey como el cuarto paradigma de la ciencia (Hey et al. 2009), la aparición del Big Data, Data Science, o, e-Science, como nueva disciplina, más allá de la ciencia teórica, la experimental y la de simulación computacional, ha dado lugar también a un interesante debate científico. A raíz de la aparición y consolidación de la disciplina, Anderson (2008) la presentó como el fin de la ciencia teórica y promovió una intensa discusión comparativa sobre la investigación de datos versus la investigación a través de hipótesis. En el campo del estudio de la biodiversidad se ha discutido ampliamente sobre el impacto epistemológico de la cuantificación sistemática de la naturaleza con la acumulación de grandes conjuntos de datos. Devictor y Bensaude-Vincent (2016) lo definen como la “datificación de la biodiversidad”. Señala su raíz en la búsqueda de biodiversidad global promovida para fundamentar la política medioambiental con una sólida base empírica con plataformas como GBIF (Global Biodiversity Information Facility). Sin embargo, en concordancia con otros autores como Bowker (2000), se advierte que en ocasiones los registros son transformados en datos digitales en un proceso que cambia su significado. Según estos autores, la creación de bases de datos globales empieza a ser un fin en sí mismo, y en esa “datificación”, promoviendo la uniformidad y compatibilidad, las especies y comunidades pierden su contexto y se crean nichos artificiales desvinculados de las realidades biofísicas de los organismos. Sin embargo, otros autores como Succi y Coveney (2019) sugieren que la capacidad de reconocimiento de patrones de análisis del Big Data proporciona una oportunidad para solventar obstáculos como, por ejemplo, el análisis de la no linealidad o el caos, frecuentes en la naturaleza, y que limitan, en ocasiones, el avance del conocimiento de muchos de sus complejos sistemas.

Lo cierto es que en este marco del Big Data y con la actual capacidad de generación, almacenamiento y procesamiento de datos, la ecología está evolucionando rápidamente hacia una ciencia de “datos intensivos”, de acceso abierto, más colaborativa e interdisciplinar. La necesidad de integrar y analizar cantidades masivas de datos ha dado pie incluso a una nueva rama científica, la eoinformática o bioinformática (Michener y Jones 2012). El número de publicaciones científicas donde conceptos como biodiversidad y Big

Data coinciden se ha sextuplicado en los últimos 10 años (ver Fig.1). Y en la actualidad, la publicación de artículos de datos o “data papers” con el objetivo de fomentar la preservación, accesibilidad y reutilización de los datos usados en investigación (Pérez-Luque y Ros-Candeira 2019), es cada vez más frecuente en todas las áreas de conocimiento y grupos editoriales (ver caso de Ecosistemas, Puerta-Piñeiro et al. 2020). Un ejemplo en este número es el artículo de datos sobre mamíferos marinos de la región subantártica de Chile de Sánchez-Jardón et al. (2022a). Sin embargo, muy relacionado con las características intrínsecas del Big Data, su aplicación en ecología presenta aún grandes retos y hay actualmente una interesante discusión sobre su aplicabilidad a escalas relacionadas con la gestión de espacios y especies (Farley et al. 2018; Bayraktarov et al. 2019; Kühl et al. 2020). Estos desafíos están relacionados con la capacidad de gestión de grandes volúmenes de datos, problemas de comparabilidad e integridad debidos a la alta heterogeneidad y carácter multi-escala de la información, diferentes metodologías de recopilación y procesamiento de datos, la incompleta cobertura taxonómica y espacial, y, el bajo nivel de conocimiento técnico en Big Data de muchos científicos y gestores, entre otros aspectos. Algunas de las soluciones incluyen el desarrollo de plataformas de intercambio de datos y código abierto a diferentes escalas, computación de alta velocidad en la nube para el análisis de gran volumen de datos, procesos de armonización en la toma o de manejo de datos y la aplicación de modelos estadísticos flexibles que sean capaces de manejar datos heterogéneos y fuentes de incertidumbre variables, todo ello integrado en campañas de formación y sensibilización que permitan acercar a la comunidad científica y gestora al mundo del Big Data aplicado al seguimiento de la biodiversidad (Pereira y Cooper 2006; Callaghan et al. 2019).

En este contexto, esta revisión trata de dar una visión general de las fuentes de información disponibles en la actualidad para el estudio de la biodiversidad usando Big Data, así como de la intensidad, extensión y temporalidad que abarcan, y los ecosistemas y formas de vida, mejor o peor representados. El objetivo último ha consistido en hacer una valoración del estado actual de conocimiento y uso del Big Data para el seguimiento de la biodiversidad de ecosistemas terrestres, así como de sus retos y oportunidades.

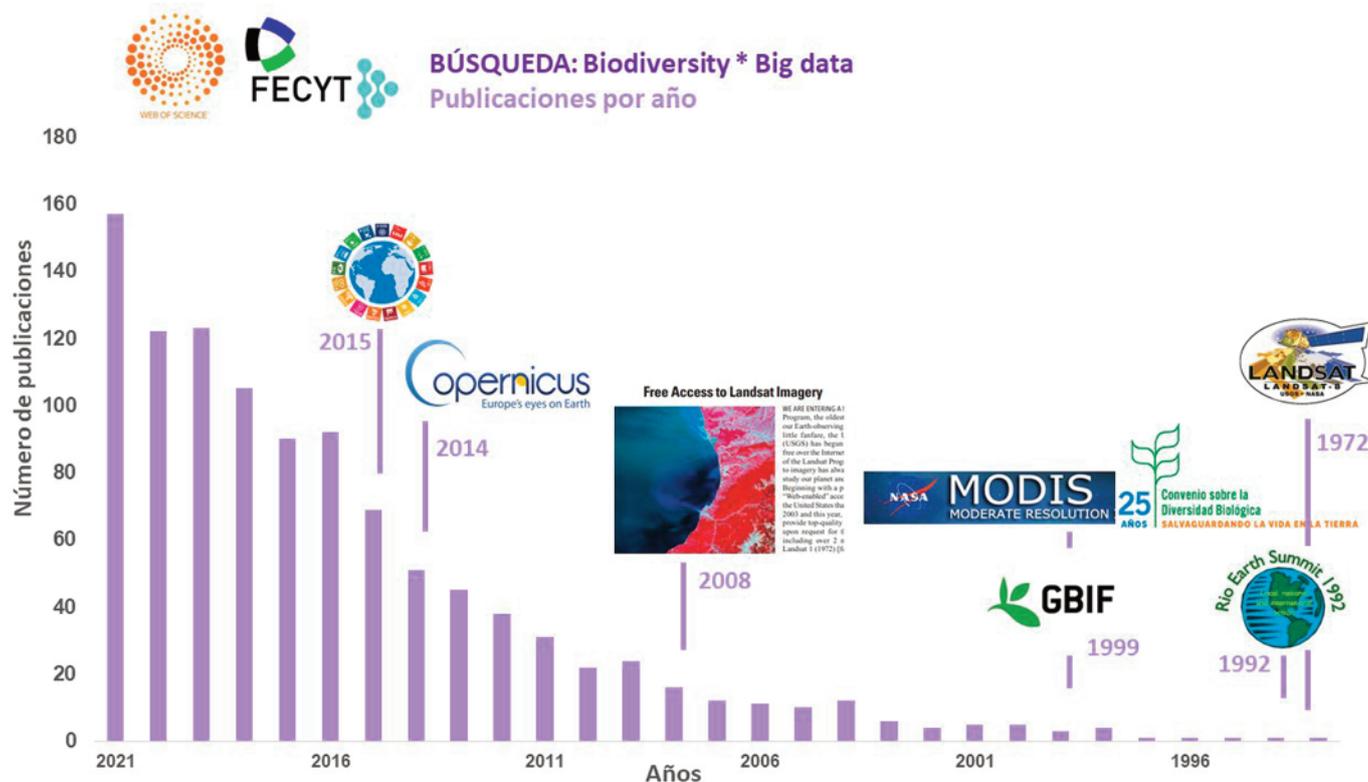


Figura 1. Evolución del número de publicaciones por año donde convergen conceptos como biodiversidad y Big Data e hitos relacionados (Fuente: Web of Science, junio 2022).

Figure 1. Evolution of the number of publications per year where concepts such as biodiversity and Big Data converge. Historical milestones are indicated (Source: Web of Science, June 2022).

Datos sobre biodiversidad y nuevos sistemas emergentes de Big data

En un contexto histórico, la “datificación” de la biodiversidad, como lo definen [Devictor y Bensaude-Vincent \(2016\)](#), surge por la necesidad de dar respuesta a problemas globales, en concurrencia con una fuerte corriente de concienciación medioambiental, que fueron el fruto de grandes hitos como la Cumbre de Río de 1992 y el Convenio sobre Diversidad Biológica de 1993. Estos instrumentos internacionales, a través de los objetivos de la Agenda 21, actualmente muchos de ellos comprendidos en los ODS ([UN 2015](#)), pretenden avanzar hacia un desarrollo sostenible, frenando la pérdida de biodiversidad mediante la evidencia científica. Así, en 1999 se crea la GBIF, una organización intergubernamental financiada por la OCDE (Organización para la Cooperación y el Desarrollo Económico), cuyo objetivo es proporcionar acceso abierto a los datos sobre la biodiversidad mundial. A partir de esta iniciativa, otros hitos importantes como el acceso abierto y gratuito a gran cantidad de datos de sensores remotos ([Landsat](#) en 2008, [Copernicus](#) en 2014) o la corriente de Open Data Science, que también fue promovido por la OCDE ([OCDE 2007](#)) y que se empieza a promover en ecología a partir de esos mismos años ([Reichman et al. 2011](#)), han contribuido al incremento exponencial del uso del Big Data en el seguimiento de la biodiversidad en las últimas décadas ([Fig.1](#)).

Como muchas otras ramas de la ciencia, el seguimiento y estudio de la biodiversidad necesita datos, que tras una serie de procesos se conviertan en información y conocimiento. Este proceso es designado por varios autores como el ciclo de vida de los datos ([Michener y Jones 2012](#)). Según identifican estos mismos autores, en ciencia experimental este ciclo comienza con la planificación de

la toma de datos, seguido del resto de etapas clave (recopilación, preservación, control de calidad, creación y gestión de metadatos, integración, análisis y difusión). En cambio, en el marco del Big data o Data Science, el ciclo comienza normalmente con la gestión e integración de diferentes bases de datos.

La información necesaria para hacer seguimiento de la biodiversidad es muy heterogénea. Esta información puede organizarse en sistemas de datos o mediciones recopiladas por comunidades de investigadores o técnicos de diferentes disciplinas. Cada comunidad puede presentar diferentes sistemas de información, formatos, así como diferentes niveles de experiencia e inversión histórica en series de datos. El primer reto para el seguimiento de la biodiversidad es comprender y abarcar toda esta heterogeneidad ([Farley et al. 2018](#)). A esta variedad del flujo de información hay que añadir las características que presentan algunos sistemas emergentes en esta disciplina como son el volumen, la veracidad y la velocidad ([Fig 2](#)).

Así, en Ecología podemos identificar en la actualidad, entre otros, los siguientes sistemas de información: a) *series temporales de datos de programas de observación remota*; b) *la agregación de datos de campo o experimentales locales o regionales* (dendroecología, fisiología, fenología, censos); c) *las redes a largo plazo de datos de campo nacionales o internacionales*; d) *redes de sensores automatizados* (fenología, foto-trampeo, meteorológicas), y, e) *iniciativas de Ciencia Ciudadana*. Conjuntamente, toda esta información está amplificando la escala espacio-temporal de las cuestiones relacionadas con el seguimiento de la biodiversidad ([Peireira y Cooper 2006](#)) y su gestión ([Callaghan et al. 2021](#)) ([Fig. 2](#)). A continuación presentamos brevemente estos sistemas de información identificados.



Figura 2. Ejemplo de algunos sistemas de recolección de datos sobre biodiversidad utilizados en ecología y que presentan características de Big Data. Sensores remotos (gran Volumen de datos); Ciencia Ciudadana (Veracidad), donde se recogen datos por parte de personas con diferentes grados de experiencia; Sensores a tiempo real (Velocidad) que pueden suministrar datos continuos a baja latencia y alta velocidad; Observaciones de campo y experimentos (Variedad) por parte de los científicos, en una amplia variedad de mediciones, sistemas, formatos y escalas. (Modificado de [Farley et al. 2018](#)) (Ilustraciones diseñadas por Freepik).

Figure 2. Example of data collection systems on biodiversity used in ecology that present characteristics of Big data. Remote Sensing (Volume); Citizen Science (Veracity), where data is collected by people with different degrees of experience; Real-time sensors (Speed) that can deliver continuous data at low latency and high speed; Field observations and experiments (Variety) by scientists, in a wide variety of measurements, systems, formats and scales. (Modified from [Farley et al. 2018](#)) (Illustrations designed by Freepik).

Series temporales de datos de observación remota

Los sensores remotos sobre plataformas satelitales, aviones, o drones, adquieren gran cantidad de datos de forma continua y con variedad de resoluciones espaciales, espectrales y temporales. Los datos espectrales procesados proporcionan información de aspectos relacionados con la biodiversidad como son la cobertura arbórea, el verdor, el contenido de clorofila del océano, la productividad primaria neta o la extensión del hábitat (Turner et al. 2003). Este monográfico cuenta con ejemplos del uso de datos espectrales de diversos tipos de sensores remotos para la detección y seguimiento de flora invasora (Vázquez de la Cueva et al. 2022) o para la caracterización y seguimiento de diferentes tipos de vegetación (López Trullén et al. 2022). Actualmente las tecnologías de observación remota ofrecen una gran oportunidad para el seguimiento de la biodiversidad, dado que los programas internacionales de Observación de la Tierra, como el europeo Copernicus con los satélites Sentinel, o los estadounidenses Landsat y MODIS, proporcionan enormes cantidades de datos (del orden de petabytes) accesibles de forma gratuita para todos los usuarios. Además, cada vez más estos programas proporcionan estándares de procesamiento para facilitar su uso (Zhu 2019). Los datos adquiridos de forma regular y frecuente, y procesados de forma correcta y comparable, permiten explorar características de la cubierta terrestre con frecuencia temporal a la carta, desde unos pocos días a varios años, facilitando el conocimiento de los procesos de cambio a distintas escalas temporales (Gómez et al. 2016; Xu et al. 2022). Mientras que los datos ópticos y LiDAR se usan de forma generalizada en el seguimiento de diferentes aspectos de la biodiversidad (Álvarez-Martínez et al. 2018), los datos radar o hiperespectrales ofrecen nuevos retos y oportunidades todavía por explorar (Gómez et al. 2019). Uno de los actuales retos para un seguimiento más completo de la biodiversidad es la integración de estos datos espacialmente continuos con otros datos ecológicos más precisos a nivel taxonómico (Schimel et al. 2013; Pereira y Cooper 2006).

Datos de campo experimentales a escala local y regional

Tradicionalmente, los datos recogidos sobre biodiversidad han tenido como objeto medir los patrones y procesos que gobiernan la distribución y estructura de especies y comunidades en el espacio y el tiempo (Farley et al. 2018). Se han recogido de forma local o regional sin estandarización y en muchas ocasiones se trata de observaciones limitadas sin continuidad temporal. A menudo, los datos se han ido recogiendo en hojas de cálculo, pequeñas bases de datos locales y desagregadas con poca o ninguna interoperabilidad, o se presentaban en volúmenes y listas de presencia y abundancia de individuos (censos de flora o fauna) (Musvuugwa et al. 2021). Este es el caso también de gran cantidad de información sobre biodiversidad de las colecciones históricas de museos y herbarios. Sin embargo, en las últimas décadas se está produciendo un cambio progresivo hacia la recopilación digital de los datos de estos métodos tradicionales. Actualmente, los datos se generan y recogen a través de instrumentos automatizados como “tabletas” u ordenadores de campo que hacen más fácil la digitalización e interoperabilidad con el fin de agregar información a gran escala. Además, muchos datos estáticos de museos se están digitalizando (ver ejemplo de iDigBio), facilitando así el acceso a datos contemporáneos e históricos, lo que resulta clave para el estudio de la dinámica y conservación de la biodiversidad a largo plazo (Page et al. 2015). Esta digitalización de colecciones históricas e iniciativas de Ciencia abierta (Riechman et al. 2011) están generando millones de datos de presencia y otras variables de especímenes con diferentes niveles taxonómicos. En este sentido, plataformas y repositorios de datos globales (GBIF, Data One, Neotoma (paleoecology), GenBank) o específicas (TRY (plant trait database), DEN (Dendro-Ecological Network), NPN (phenology), ESDB (soils), Global Plants (digitized plant specimens), DoPi (pollinators), WORMS (marine species), GLORIA (alpine plant species), the Global Vegetation Project) están jugando un papel importante para la agregación y acceso a esta información. Los mayores desafíos que presentan

estos sistemas de información, además del volumen, son la veracidad y la variedad de los datos ya que sus registros dependen de identificaciones taxonómicas y espacio-temporales recogidas por redes dispersas en todo el mundo, en muchas ocasiones desconectadas a nivel nomenclatural (Farley et al. 2018) o con unidades de medida de referencia diferentes.

Redes a largo plazo de datos nacionales e internacionales

Las redes de seguimiento a largo plazo proporcionan observaciones y datos experimentales, en algunos casos, en ecosistemas o lugares específicos. Algunas de las más conocidas son, en Europa, la Red de Investigación Ecológica a Largo Plazo (LTER), y en USA la Red Nacional de Observatorios Ecológicos (NEON) y los Observatorios de Zonas Críticas (CZO), entre otras. Estas redes cubren el seguimiento de diferentes formas de vida y variables ambientales en ecosistemas como bosques, matorrales, pastizales, desiertos, tundra, sistemas acuáticos y entornos urbanos. Comprenden muchos tipos de sensores y mediciones, por lo que la variedad de datos es una de sus principales características y retos para su uso. Por ejemplo, en LTER hasta hace poco cada nodo o zona de monitoreo tenía sus propios protocolos de toma de datos, lo que ha dificultado los análisis ecológicos a gran escala. En NEON y CZO, de más reciente creación, este problema no existe.

En el caso específico de los bosques existen también dos fuentes relevantes de información a largo plazo: la red europea de seguimiento de daños en bosques (ICP Forests) y los inventarios forestales nacionales (IFNs). ICP Forest, activa desde hace más de tres décadas en Europa, es la única red paneuropea existente que hace un seguimiento del estado de los bosques utilizando indicadores comunes y métodos armonizados para la recolección de datos (ver ejemplo en este número del uso de esta red para el seguimiento de la defoliación a escala nacional en Adame et al. (2022)). El objetivo de la red es recabar información sobre la salud de los bosques europeos y los factores abióticos y bióticos involucrados en la dinámica de los mismos. Por otro lado, los IFNs son las principales fuentes de información sobre los bosques a gran escala, y generan millones de datos sobre la composición, estructura, dinámica y funcionamiento forestal (Alberdi Asensio 2015). Sus datos pueden ser usados para hacer un seguimiento del estado de conservación de diferentes ecosistemas forestales de forma estandarizada a escala nacional como proponen Pescador et al. (2022) en este monográfico. Sin embargo, a escala internacional son proyectos, en origen independiente, con diversas metodologías de toma de datos, con diferentes objetivos y que reflejan muy diversos ecosistemas forestales (Hernández et al. 2016). Debido a estas peculiaridades es difícil estandarizar los datos para hacer cálculos comparables. Sin embargo, al ser la fuente de la que se nutren muchas de las estadísticas forestales a escala global a través de diferentes iniciativas como las de la red europea ENFIN o la FAO, se ha avanzado en la armonización y comparabilidad de la estimación de indicadores (McRoberts et al. 2012). En este número existe una reseña sobre un libro de acceso libre que describe la reciente creación de una red de IFNs en la región de Latinoamérica y el Caribe (Hernández 2022). El acceso a parte de los datos globales de estas grandes fuentes de información sobre los bosques es posible mediante plataformas como la del JRC, el GBIF, el LFC, o nuevas infraestructuras de software abierto y “data sharing” como Github (Matragkas et al. 2014).

Además de la heterogeneidad en la toma de datos y variedad de los registros y las dificultades que estas características suponen para la integración y uso de la información de la mayoría de estas redes de seguimiento a gran escala, otros dos retos importantes en estos sistemas de información son la velocidad y la accesibilidad. Como discuten Farley et al. (2018), dado el impacto económico que puede tener los efectos del Cambio Global en la biodiversidad, hay una fuerte presión para que la velocidad en el seguimiento y el análisis sea mayor. Para que el conocimiento ecológico sea útil, debe obtenerse a tiempo. Respecto a la accesibilidad, aunque la tendencia actual es hacer el acceso en línea gratuito

y libre a toda la información de estos sistemas de datos, el acceso a algunos de sus registros es aún restringido (por ejemplo, la localización exacta de las parcelas) o lento (derivado de unos tiempos de procesamiento y control de calidad largos y con diferentes políticas gubernamentales al respecto (Vidal et al. 2016).

Redes de sensores automatizados sobre biodiversidad

Estas redes de vigilancia automatizada *in situ* de datos biológicos y ecológicos están aumentando rápidamente su volumen a medida que la tecnología de sensores mejora y disminuyen los costes. Por ejemplo, la red Fluxnet de torres de covarianza de flujos de Foucault tiene más de 500 sitios de análisis y 200 variables registradas durante más de 20 años con resolución subhoraria, con un volumen de datos brutos del orden de petabytes. Otros sensores cada vez más frecuentes son los de fototrampeo y seguimiento vía radio o satélite para registrar comportamiento y movimiento animal, sensores de temperatura y composición de agua, micrófonos de sonido, dendrómetros automáticos, conjuntamente con mediciones más clásicas como las relacionadas con estaciones meteorológicas o hidrológicas. Estas redes automatizadas suelen registrar datos a alta velocidad y de forma continua, a escala diaria u horaria, por lo que uno de los retos de los sistemas de información que recogen los datos es la actualización y mantenimiento de los repositorios y las vías de acceso a usuarios finales e intermedios. Por otro lado, estas redes instrumentadas están sujetas a la deriva del sensor, a errores de calibración y a problemas con el procesamiento automatizado, lo que genera desafíos de veracidad (Farley et al. 2018). Estos datos suelen almacenarse en repositorios nacionales (Agencia Estatal de Meteorología (AEMET) e internacionales (CRU, WorldClim) como ocurre, por ejemplo, con los datos meteorológicos. El acceso a los datos brutos suele ser libre vía web o bajo petición para aquellos datos más actualizados y para algunos indicadores y estimaciones específicos sujetos a licencias de uso o propiedad.

Ciencia Ciudadana e iEcology

Los avances en el desarrollo de la tecnología móvil inteligente y el acceso general y global a internet han hecho posible que las iniciativas de ciencia ciudadana sean cada vez más utilizadas en todas las ramas y escalas de la Ecología (Louv y Fitzpatrick 2012). Esta nueva corriente consiste en la acumulación, gestión y posterior análisis de datos registrados por el público general, lo que permite la realización de muestreos intensivos basados en datos de alta resolución a diferentes escalas, con un coste inferior al de los métodos tradicionales con expertos (Farley et al. 2018). El desarrollo de algunas plataformas de ciencia ciudadana en línea (por ejemplo, eBird, plantNet, iNaturalist, Flickr) promueve la participación de numerosos sectores sociales así como la interacción entre aficionados y expertos para una mejor identificación y calidad de los datos georegistrados. De la misma forma, estas plataformas cuentan con software específico que permite la calibración y perfeccionamiento de sistemas de aprendizaje automático para la identificación de especies y comunidades, patrones y procesos ecológicos de interés. Todas estas características convierten estas nuevas plataformas de ciencia ciudadana en herramientas versátiles y útiles para el seguimiento de la biodiversidad, y su uso por aficionados y expertos se incrementa día a día (ver ejemplos sobre seguimiento de biodiversidad en la Patagonia chilena y la flora amenazada del Pirineo en Sánchez-Jardón et al. (2022b) y Pizarro et al. (2022) en este número.

Por último, es importante nombrar a la iEcología (definida como la ecología de Internet), un enfoque de investigación emergente que trata sobre el estudio de los patrones y procesos ecológicos utilizando datos en línea generados para otros fines y almacenados digitalmente (Jarić et al. 2020). Esta nueva corriente de investigación aprovecha el flujo de datos generado por las redes sociales (*social media stream*) para abordar cuestiones ecológicas y analizar los procesos a distintas escalas espacio-temporales y en diversos contextos, tratando de complementar métodos tradicionales de obtención de datos.

El uso de estas nuevas fuentes de información presenta aún grandes desafíos relacionados con el manejo y almacenaje del gran volumen de datos generados, la velocidad de actualización y análisis, y sobre todo, aspectos relacionados con la veracidad (ya que los datos son recopilados por personas con diferentes grados de experiencia) y posibles sesgos de la población de datos de la que proviene la información de las diferentes redes sociales analizadas.

Representatividad de programas de observación de biodiversidad en tipos de ecosistemas y formas de vida

En el seguimiento de la biodiversidad es muy frecuente disponer de bases de datos afectadas por diferentes tipos de sesgo, tanto taxonómico como a nivel de ecosistema, que pueden alterar significativamente las conclusiones de nuestros análisis espacio-temporales. Esto se debe a la variabilidad de conocimiento que existe entre diferentes organismos y ecosistemas debido principalmente al interés que suscita su estudio, o dicho de otro modo, al grado en el que son objetivo de la investigación científica (Troudet et al. 2017). Se ha demostrado sin embargo que centrarse únicamente en el estudio de ecosistemas u organismos "carismáticos" no es suficiente para identificar problemas globales y desarrollar planes de conservación eficaces y sostenibles, ya que todos los organismos y ecosistemas son esenciales para el funcionamiento global del sistema Tierra. Así pues, examinar el sesgo taxonómico en el seguimiento de la biodiversidad y comprender sus causas subyacentes es fundamental para intentar identificar esos vacíos de conocimiento y encontrar soluciones que permitan hacer un seguimiento más completo y global a largo plazo (Domínguez-Lozano 2019).

Trabajos previos han definido una sobrerrepresentación de algunos organismos como plantas o vertebrados en bases de datos globales como GBIF (Troudet et al. 2017). Nosotros obtenemos un resultado similar al hacer una búsqueda de artículos indexados en la Web of Science desde los años 90 sobre diferentes ecosistemas y organismos en el marco del Big Data y el seguimiento de la biodiversidad (Fig 3). Así, encontramos una sobrerrepresentación de los ecosistemas terrestres frente a los acuáticos, a pesar de que los segundos son mucho más extensos a escala global. Dentro de los terrestres, los bosques son objeto de muchos más estudios (42%) que los pastizales (15%) o los matorrales (4%) en el marco de la Data Science. Si hacemos una comparación con los tipos de hábitats que cubre la Directiva hábitat mediante la Red Natura 2000, cuyo objetivo es un marco para protección de los hábitats más vulnerables, pero también representativos de Europa (EU 2008), observamos que los tipos de ecosistemas que estarían más infra-estudiados son los de matorral (representa 11% de los hábitats en la RN200) frente a los bosques (35%) o pastizales (13%). Entre los ecosistemas acuáticos, los ecosistemas de agua dulce parecen estar más estudiados y entre estos, sobre todo los ligados a sistemas riparios. En el caso de las diferentes formas de vida u organismos, nuestra búsqueda corrobora los resultados de Troudet et al. (2017) o Gaiji et al. (2013), siendo mucho mayor el número de publicaciones centradas en animales, sobre todo aves, y también plantas, frente a hongos y otros microorganismos que son mucho más numerosos a escala global y que pueden ejercer un papel ecológico clave en muchos sistemas de la Tierra.

Gestión, integración y análisis de programas de adquisición de datos

Tras la recolección de los datos e información asociada para el seguimiento de la biodiversidad, uno de los grandes retos de la Data Science o ciencia del Big Data es la gestión e integración de grandes bases de datos, que, por último, serán analizadas con el objetivo final de generar y difundir conocimiento.

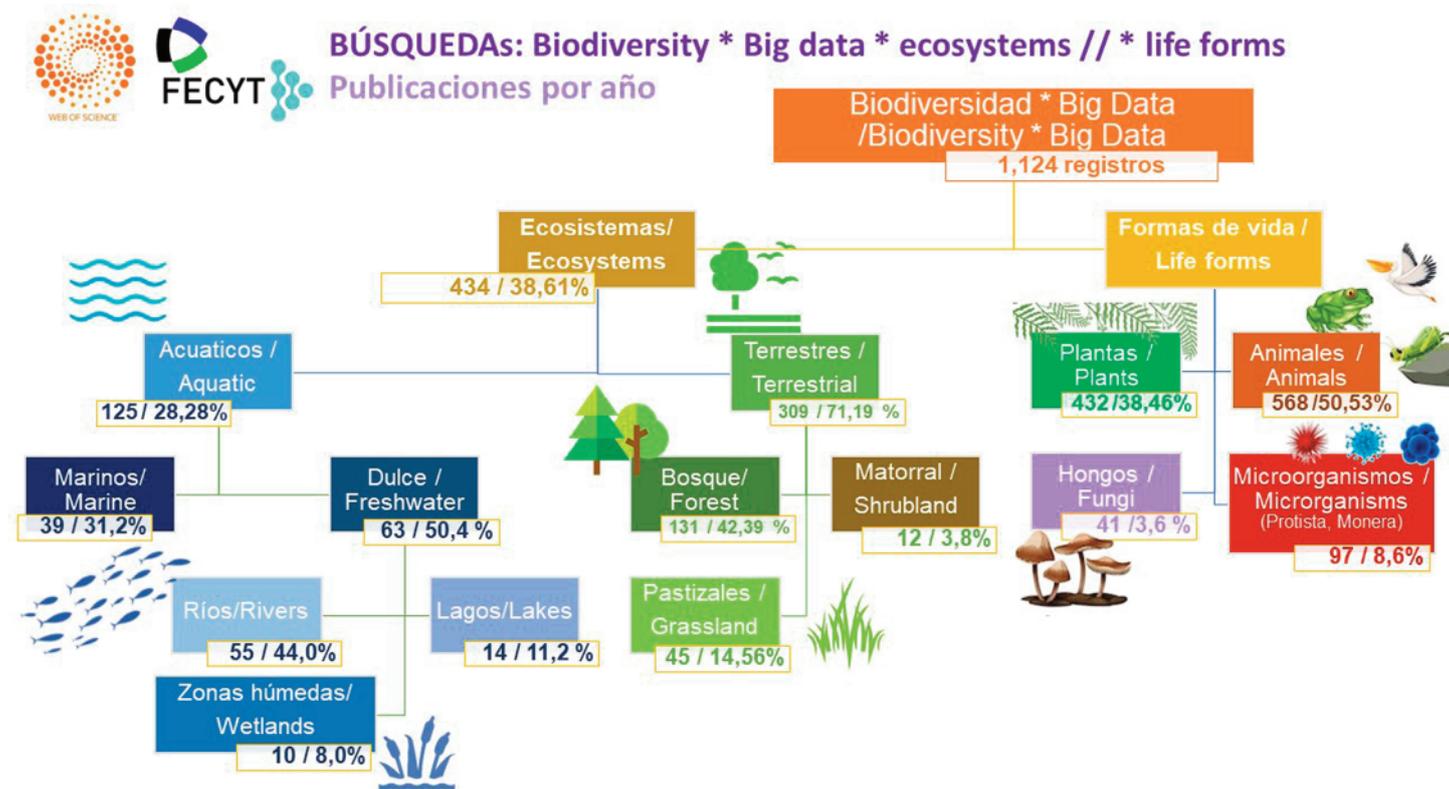


Figura 3. Esquema de la representatividad de la información sobre biodiversidad en un contexto de Big Data en diferentes tipos de ecosistemas y formas de vida. Esta representatividad se estima a través de las publicaciones científicas indexadas por los temas señalados en la figura (Biodiversidad * Big Data * ecosystems // * life forms) y recogidas en la *Web of Science* (último acceso en junio 2022). Pueden existir solapamientos o ausencias entre tipos. (Ilustraciones diseñadas por Freepik).

Figure 3. Diagram of the representativeness of information on biodiversity in a context of Big Data in different types of ecosystems and life forms. This representativeness is estimated through the scientific publications indexed by the topics indicated in the figure (Biodiversidad * Big Data * ecosystems // * life forms) and collected in the *Web of Science* (last accessed June 2022). (Illustrations designed by Freepik).

Esta digitalización de datos sobre biodiversidad está generando grandes volúmenes de información a dos niveles. Por un lado, se dispone de registros históricos que provienen de la digitalización de herbarios y museos, y por otro, las series temporales de datos procedentes de nuevas tecnologías de adquisición de información espacial (por ejemplo, teledetección) y de la nueva corriente de datos abiertos en muchas ramas científicas (Matragkas et al. 2014). El acceso a la información sobre biodiversidad es cada vez mayor, pero por otro lado se están creando nuevas necesidades para su almacenamiento, procesamiento, gestión y uso. Uno de los mayores desafíos en este sentido es la integración y gestión, en una misma plataforma, de información muy heterogénea y proveniente de diferentes fuentes, donde ha sido clave el desarrollo durante la última década de las denominadas ciberinfraestructuras (CI). Las CI están diseñadas para la conservación, el acceso, la conexión y la reutilización de datos científico-técnicos, usualmente de acceso libre (Farley et al. 2018). Sin embargo, para que estas infraestructuras sean útiles es muy importante mejorar los protocolos, códigos y normas sobre estandarización de datos para que su almacenamiento garantice precisión, coherencia y fiabilidad (Musvuugwa et al. 2021).

De forma paralela, la información sobre seguimiento de biodiversidad está muchas veces dispersa y tiene diversos orígenes. Por esta razón, para maximizar la usabilidad de los datos ecológicos, la documentación rigurosa de esta información disponible es clave. Ejemplos de buenas prácticas en este sentido, con protocolos estandarizados para la recogida y el almacenamiento de datos en diferentes ecosistemas, regiones y sistemas de referencia, son la red europea ICP Forest y la americana NEON. En esta línea, muchas organizaciones están invirtiendo esfuerzo y recursos en los últimos años para construir una arquitectura abierta para los datos científicos con estándares comunes de metadatos, semántica y ontología. Algunos ejemplos son DataONE, el W3 (World Wide Web

Consortium), (ESIPFED) Earth Science Information Partners, el OGC (Open Geospatial) y el CSI (International Science Council). Además, en esta fase es preciso en muchas ocasiones invertir recursos en el pre-procesamiento de la información para eliminar o reducir el ruido, para reducir dimensiones mediante transformaciones y, por último, para el enriquecimiento mediante la combinación o enlace con otros conjuntos de datos complementarios (Fig. 4). En este monográfico Vega-Gorgojo et al. (2022) nos describen algunos de estos retos y sus soluciones al crear el Explorador Forestal, un navegador web de datos abiertos masivos.

Por otro lado, la computación en la nube y el acceso a una ingente capacidad de almacenaje de datos en la actualidad, está siendo esencial para lidiar con retos como el aumento en la velocidad de generación de datos y, el resultante volumen de información que almacenar y manejar. La computación y la disponibilidad generalizada de servicios públicos de computación en la nube a través de grandes proveedores como Amazon, Google y Microsoft, u otras como Docker han puesto a disposición de los científicos un importante medio para el almacenaje. La computación en nube proporciona acceso bajo demanda a recursos informáticos flexibles que en función de las necesidades permite obtener más espacio de almacenamiento o potencia de cálculo, y que supera la capacidad de cualquier individuo e institución (Farley et al. 2018). Otras soluciones programáticas que están surgiendo para facilitar el uso y el acceso al Big Data es el uso de interfaces de programación de aplicaciones (API) que hacen posible por ejemplo la disponibilidad de datos o aplicaciones sin tener que duplicar los mismos en ordenadores locales. Un buen ejemplo puede ser Google Earth Engine o Collect Earth Online que permite la visualización y análisis en línea de imágenes de teledetección, o Lifemapper, que ofrece varias herramientas en línea para poder modelizar la distribución de la biota terrestre (Fig. 4).

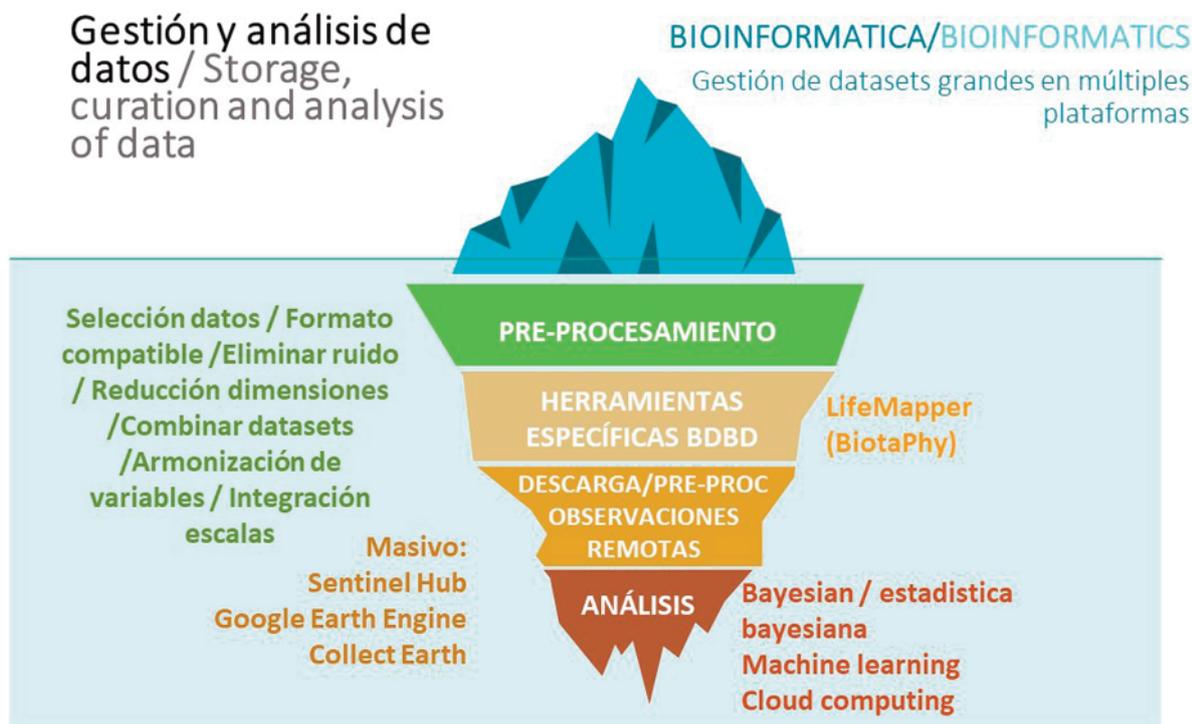


Figura 4. Esquema de algunos de los hitos más importantes en la etapa de gestión y análisis de grandes bases de datos sobre biodiversidad (BDBD).
Figure 4. Summary of some of the most important milestones in the storage, curation and analysis phase of large biodiversity databases (BDBD).

Con respecto al análisis de los datos, la heterogeneidad y desigual veracidad de los mismos, a veces no estructurados, requieren de enfoques analíticos muy flexibles. Además, el manejo del gran volumen de información y la velocidad con la que se genera hacen necesario que estas aproximaciones sean escalables. Farley et al. (2018) identifican la estadística bayesiana y las técnicas estadísticas de aprendizaje automático (machine learning, ML) como las dos aproximaciones de análisis más prometedoras en el marco del Big Data. Los modelos jerárquicos bayesianos permiten integrar de forma flexible múltiples capas de complejidad a diferentes escalas y resoluciones, sus interacciones, e incluso la incertidumbre. Sin embargo, la necesidad de estimar gran cantidad de parámetros es a menudo computacionalmente intensiva, limitando la escalabilidad, y en muchas ocasiones, la interpretación de resultados (Farley et al. 2018). El aprendizaje automático, sin embargo, suele ser escalable, pero tiene la limitación de que tiende únicamente a identificar patrones, más que a inferir factores subyacentes en los procesos ecológicos. Este último grupo de aproximaciones, muy relacionadas con la inteligencia artificial y los algoritmos automáticos, ha dado lugar a un interesante debate (Anderson 2008; Bowker 2014) sobre la deducción frente a la inducción o el humano frente a la máquina. Sin embargo, actualmente la mayoría de los autores coinciden en que ambas aproximaciones son necesarias y complementarias (Musvuugwa et al. 2021). No obstante, independientemente de la potencia analítica y de la cantidad de información, la calidad de los conocimientos que se derivan de cualquier análisis es una función de la idoneidad del conjunto de datos para la cuestión de interés y la capacidad de análisis e interpretación de los resultados. Está más que demostrado que más datos no se traduce necesariamente en más conocimiento (Bayraktarov et al. 2019). Uno de los grandes desafíos para sacar partido al Big Data, de hecho, está relacionado con la formación de colaboraciones y equipos multi-disciplinarios (bio-informáticos, estadísticos, técnicos y científicos de la biodiversidad) que permitan el manejo, actualización y mantenimiento de las ciberplataformas de información, a la vez que tengan capacidad analítica y conocimiento para interpretar de forma correcta los resultados, tanto aplicados como teóricos, sobre el seguimiento de la biodiversidad (Kühl et al. 2020). Este es uno de los objetivos de BigBioData, el nuevo grupo de trabajo cre-

ado en la Asociación Española de Ecología Terrestre (AEET) y que trata de servir de puente para el intercambio de conocimientos y experiencias entre diferentes profesionales o usuarios intermedios de la información espacial relacionada con el seguimiento de la Biodiversidad y que se describe en la nota de Hernández et al. 2022 en este número.

Oportunidades y retos generales

Tras esta revisión general del estado actual del seguimiento de la biodiversidad en la era del Big Data y la información pasamos a enumerar algunos de los retos y oportunidades identificados.

Oportunidades

- El acceso a distintos tipos de datos en el marco de la Big Data Science hace posible una investigación integradora sobre la biodiversidad, que ofrece un punto de partida para evaluar los efectos del Cambio Global sobre su estructura y función. Con la infraestructura adecuada, la mejora de los métodos analíticos y el conocimiento necesario es posible combinar datos a diferentes escalas, desde la biología molecular al dato de campo, clima o paisaje, que permite un seguimiento más completo en el espacio y en el tiempo.
- El incremento en los últimos años de la capacidad computacional y de almacenaje de información en la nube, promueve la generación de grandes conjuntos de datos sobre biodiversidad a través de agregación y armonización de registros históricos o mediante la colección masiva en ciertos campos. Las grandes plataformas de información suponen una nueva oportunidad para explorar de forma teórica la biodiversidad y para avanzar en aspectos técnicos asociados a metodologías de manejo, integración y análisis de datos masivos.
- Del mismo modo, el avance en técnicas flexibles para el análisis de diferentes fuentes de información mediante estadística bayesiana, ML e IA, supone nuevas oportunidades que, en sinergia con aproximaciones más deductivas, ayudarán a avanzar en el conocimiento y seguimiento de la biodiversidad global.

Retos

- Características intrínsecas al Big Data como la veracidad, el volumen, la variedad o heterogeneidad de los datos y la velocidad de recopilación de los mismos generan algunos de los mayores retos de gestión, análisis y generación de conocimiento sobre el seguimiento de la biodiversidad. Aunque algunas limitaciones asociadas a la velocidad y volumen de datos están bastante resueltas, estas características dificultan todavía la capacidad de análisis de datos (por ejemplo, sobreparametrización de modelos bayesianos). Además, aunque se están haciendo grandes avances en los procesos de armonización, estandarización, enlace y veracidad en las plataformas que unifican datos de diversas fuentes, aún queda camino por recorrer. Por último, si bien la información sobre biodiversidad es cada vez más accesible, en algunos casos el acceso a la totalidad de los datos sólo es posible bajo registro, petición o prepagado.
- Dada la existencia de ecosistemas poco representados y listas taxonómicas incompletas, es necesario aumentar el esfuerzo en recopilar información de aquellas formas de vida y hábitats menos carismáticos pero que ejercen un papel clave en la conservación de la biodiversidad. Asimismo, es importante mejorar la estandarización y verificación de nomenclaturas para facilitar el correcto uso de la información recopilada y evitar duplicidades e inconsistencias.
- A pesar del aumento y mayor consistencia durante los últimos años, las series temporales y espaciales de datos son incompletas. En la mayor parte de ecosistemas u organismos existe una mayor cantidad de información a partir de los años 80. De forma previa a esta fecha hay poca información, normalmente a escala local o regional, y no sistemática o armonizada. Además, existe disparidad en la cantidad y calidad de la representación en distintas regiones del mundo.
- Por último, pero no menos importante, en la actualidad, la continua y rápida evolución de plataformas y herramientas analíticas sobre el seguimiento de la biodiversidad dificulta a los usuarios finales e intermedios mantenerse al día de todos los cambios y novedades, y de las mejores herramientas disponibles. El uso efectivo del Big Data requiere de una continua formación en competencias técnicas para sacar partido a los datos, así como de colaboraciones y equipos de trabajo multidisciplinarios.

Conclusiones

El Big Data o Data Science es una herramienta innovadora y necesaria que permite cumplir objetivos de seguimiento de la biodiversidad cada vez más ambiciosos y globales. Con la actual capacidad de generación, almacenamiento y procesado de datos, las ramas científicas relacionadas con este seguimiento evolucionan rápidamente hacia una ciencia de “datos intensivos”, de acceso abierto, colaborativa e interdisciplinar. Las perspectivas futuras de esta nueva corriente parecen óptimas, y surgen fórmulas prometedoras para recabar información y generar conocimiento de forma colaborativa y bajo coste como la ciencia ciudadana o la i-Ecología. La necesidad de integrar, mantener, actualizar y analizar cantidades masivas de datos ha originado, incluso, una nueva rama científica, la ecoinformática o bioinformática.

Sin embargo, aunque la agregación de información de diversas fuentes en grandes ciberinfraestructuras y la generación masiva de datos satelitales supone una buena oportunidad para este seguimiento, no hay que olvidar que se han de seguir tomando datos bien planificados sobre el terreno para enriquecer, validar y actualizar el conocimiento a largo plazo. Además, se deben incrementar los esfuerzos en el seguimiento de organismos y ecosistemas poco representados en la actualidad. Algunos de los retos identificados, relacionados con la integración y gestión de plataformas de datos heterogéneos y su análisis posterior, también suponen desafíos y oportunidades futuros para muchas disciplinas relacionadas con el seguimiento de la biodiversidad en la era del Big Data.

Contribución de autores

Laura Hernández Mateo: conceptualización, redacción borrador inicial, edición; Jose Manuel Álvarez Martínez: conceptualización, redacción, edición; Cristina Gómez: conceptualización, redacción, edición; Rut Sánchez de Dios: conceptualización, edición; Borja Jimenez Alfaro: conceptualización, edición; Flor Álvarez-Taobada: conceptualización, edición.

Referencias

- Alberdi Asensio, I. 2015. *Metodología para la estimación de indicadores armonizados a partir de los inventarios forestales nacionales europeos con especial énfasis en la biodiversidad forestal*. Tesis de Doctorado, Universidad Politécnica de Madrid. ETSI Montes. Madrid, España.
- Álvarez-Martínez, J.M., Jiménez-Alfaro, B., Barquín, J., Ondiviela, B., Recio, M., Silió-Calzada, A., Juanes, J.A. 2018. Modeling the area of occupancy of habitat types with remote sensing. *Methods in Ecology and Evolution* 9(3), 580-593.
- Adame, P., Alonso, L., Cañellas, I., Hernandez, L., Pasalodos-Tato, M., Robla, E., Alberdi, I. 2022. Hacia un seguimiento más completo y armonizado de los daños en los bosques: Aplicación a la defoliación arbórea en España. *Ecosistemas* 31(3): 2387. <https://doi.org/10.7818/ECOS.2387>
- Anderson, C. 2008. The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine* 16: 7–16.
- Bayraktarov, E., Ehmke, G., O'connor, J., Burns, E.L., Nguyen, H.A., McRae, L., Possingham, H.P., et al. 2019. Do big unstructured biodiversity data mean more knowledge? *Frontiers in Ecology and Evolution* 6: 239.
- Bowker, G.C. 2000. Biodiversity datadiversity. *Social Studies of Science* 30: 643–683.
- Bowker, G.C. 2014. Big data, big questions| the theory/data thing. *International Journal of Communication* 8: 5.
- Callaghan, C.T., Poore, A.G., Major, R.E., Rowley, J.J., Cornwell, W.K. 2019. Optimizing future biodiversity sampling by citizen scientists. *Proceedings of the Royal Society B* 286(1912): 20191487.
- Callaghan, C.T., Poore, A.G., Mesaglio, T., Moles, A.T., Nakagawa, S., Roberts, C., Rowley, J.J., et al., 2021. Three frontiers for the future of biodiversity research using citizen science data. *BioScience* 71(1):55-63.
- CBD 2014. *Resourcing the Aichi Biodiversity Targets, An Assessment of Benefits, Investments and Resource needs for Implementing the Strategic Plan for Biodiversity 2011–2020*, Second Report of the High-Level Panel on Global Assessment of Resources for Implementing the Strategic Plan for Biodiversity 2011–2020. Secretariat of the Convention on Biological Diversity, Montréal, Canada.
- Devictor, V., Bensaude-Vincent, B. 2016. From ecological records to big data: The invention of global biodiversity. *History and Philosophy of the Life Sciences* 38: 1–23.
- Domínguez Lozano, F. 2019. *¿Qué sabes sobre biodiversidad?* Ediciones Mundi-Prensa, Madrid, España. 497 pp.
- European Commission, Directorate-General for Environment, Sundseth, K. 2008. *Natura 2000: protecting Europe's biodiversity*, Wegefelt, S.(editor), Mézard, N.(translator), European Commission. Disponible en: <https://data.europa.eu/doi/10.2779/45963>
- Farley, S.S., Dawson, A., Goring, S.J., Williams, J.W. 2018. Situating ecology as a big-data science: Current advances, challenges, and solutions. *BioScience* 68(8): 563-576.
- Gaiji, S., Chavan, V., Ariño, A.H., Otegui, J., Hobern, D., Sood, R., Robles, E. 2013. Content assessment of the primary biodiversity data published through GBIF network: status, challenges and potentials. *Biodiversity Informatics* 8(2) 94-172.
- Gómez, C., White, J.C., Wulder, M.A. 2016. Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 116, 55-72. <https://doi.org/10.1016/j.isprsjprs.2016.03.008>
- Gómez, C., Alejandro, P., Hermosilla, T., Montes, F., Pascual, C., Ruiz Fernández, L.Á., Valbuena, R. 2019. Remote sensing for the Spanish forests in the 21st century: A review of advances, needs, and opportunities. *Forest Systems* 28(1): 1-33.
- Hernandez, L. 2022. “Inventarios Forestales Nacionales de América Latina y el Caribe. Hacia la armonización de la información forestal” de Carla Ramírez, Iciar Alberdi, Carlos Bahamondez y Joberto Veloso de Freitas (coords.), 2021. *Ecosistemas* 31(3): 1449. <https://doi.org/10.7818/ECOS.2449>

- Hernández, L., Alberdi, I., McRoberts, R. E., Morales-Hidalgo, D., Redmond, J., Vidal, C. 2016. Wood resources assessment beyond Europe. En: Vidal, C., Alberdi, I., Hernández, L., Redmond, J. (eds.), *National Forest Inventories. Assessment of Wood Availability and Use*, pp. 105-118. Springer, Cham, Suiza.
- Hernandez, L., Álvarez-Martínez, J.M., Pescador, D.S. 2022. BigBioData, un nuevo grupo de trabajo en la AEET sobre el seguimiento de la Biodiversidad en la era del Big data. *Ecosistemas* 31(3), 2453. <https://doi.org/10.7818/ECOS.2453>
- Hey T, Tansley S, Tolle K. 2009. Jim Grey on eScience: A transformed scientific method. En: Hey, T., Tansley, S., Tolle, K. (eds.) *The Fourth Paradigm: Data-Intensive Scientific Discovery*, pp. xvii-xxxi. Microsoft Research, Redmond, WA 98052, Estados Unidos.
- IPBES 2019. *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. Díaz, S., Settele, J., Brondízio E.S., Ngo, H.T., Guèze, M., Agard, J., Arneth, A., et al. (eds.). IPBES secretariat, Bonn, Alemania. 56 pages. https://www.ipbes.net/sites/default/files/2020-02/ipbes_global_assessment_report_summary_for_policymakers_en.pdf.
- Jarić, I., Correia, R.A., Brook, B.W., Buettel, J.C., Courchamp, F., Di Minin, E., Firth, J.A., et al., 2020. iEcology: harnessing large online resources to generate ecological insights. *Trends in Ecology and Evolution* 35(7): 630-639.
- Kitchin, R. 2014. Big Data, new epistemologies and paradigm shifts. *Big Data and Society* 1: 2053951714528481.
- Kühl, H.S., Bowler, D.E., Bösch, L., Bruelheide, H., Dauber, J., Eichenberg, D., Eisenhauer, N., et al. 2020. Effective biodiversity monitoring needs a culture of integration. *One Earth* 3(4): pp.462-474.
- López Trullén, D., Álvarez-Martínez, J.M., Sánchez Labrador, J.D., Jiménez-Alfaro, B., Pérez-Silos, I., Hernández-Romero, G., Barquín, J. 2022. Espectrofenología con datos Sentinel 2: definición de curvas de referencia para la caracterización de ecosistemas forestales. *Ecosistemas* 31(3): 2411. <https://doi.org/10.7818/ECOS.2411>
- Louv, R., Fitzpatrick, J.W. 2012. *Citizen science: Public participation in environmental research*. Cornell University Press. Ithaca, NY, Estados Unidos.
- Matragkas, N., Williams, J.R., Kolovos, D.S., Paige, R.F. 2014. Analysing the 'biodiversity' of open source ecosystems: the GitHub case. En: *MSR 2014: Proceedings of the 11th Working Conference on Mining Software Repositories, May 2014*, pp. 356-359. Association for Computing Machinery, New York, NY, Estados Unidos. <https://doi.org/10.1145/2597073.2597119>
- McAfee, A., Brynjolfsson, E. 2012. Big Data. *Harvard Business Review* 90(10): 60-68.
- McRoberts, R.E., Tomppo, E.O., Schadauer, K., Ståhl, G. 2012. Harmonizing national forest inventories. *Forest Science* 58(3): 189-190.
- Michener, W.K., Jones, M.B. 2012. Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology and Evolution* 27: 85-93.
- Musvuugwa, T., Dlom, M.G., Adebowale, A. 2021. Big Data in Biodiversity Science: A Framework for Engagement. *Technologies* 9(3): 60.
- OECD 2007. *Principles and guidelines for access to research data from public funding*. OECD París, Francia. Disponible en: <http://www.oecd.org/dataoecd/9/61/38500813.pdf>
- Page, L.M., MacFadden, B.J., Fortes, J.A., Soltis, P.S., Riccardi, G. 2015. Digitization of biodiversity collections reveals biggest data on biodiversity. *BioScience* 65: 841-842.
- Pérez-Luque, A.J., Ros-Candeira, A. 2019. Compartiendo datos en Ecología: cómo añadir más valor a los datos. *Ecosistemas* 28(3):150-159. <https://doi.org/10.7818/ECOS.1838>
- Pereira, H.M., Cooper, H.D. 2006. Towards the global monitoring of biodiversity change. *Trends in ecology and evolution* 21(3): 123-129.
- Pescador, D.S., Vayreda, J., Escudero, A., Lloret, F. 2022. El potencial del Inventario Forestal Nacional para evaluar el estado de conservación de los tipos de Hábitat forestales de Interés Comunitario: nuevos retos para cumplir con las políticas de conservación de la biodiversidad. *Ecosistemas* 31(3):2384. <https://doi.org/10.7818/ECOS.2384>
- Pizarro, M., Mugica, A., Quintana-Buil, M., Miranda, H., Gómez, D., Fuentes, I., Martínez, J., et al. 2022. Aprendiendo de plantas y compartiendo observaciones con nuevas tecnologías. FLORAPYR y FLORAMON en iNaturalist. *Ecosistemas* 31(3): 2379. <https://doi.org/10.7818/ECOS.2379>
- Puerta-Piñero, C., Pérez-Luque, A.J., Rodríguez-Echeverría, S. 2020. Ecosistemas apuesta por la publicación de artículos de datos (Data Papers). *Ecosistemas* 29(3):2118. <https://doi.org/10.7818/ECOS.2118>
- Reichman, O. J, Jones, M.B., Schildhauer, M.P. 2011. Challenges and opportunities of open data in ecology. *Science* 331(6018): 703-705.
- Sánchez-Jardón, L., Ramos, B., Gómez Peral, E., Bunster, C., Acosta-Gallo, B. 2022a. Integración de datos de biodiversidad para la educación y el turismo: mamíferos marinos nativos en la región subantártica de Aysén, Chile. *Ecosistemas* 31(3): 2410. <https://doi.org/10.7818/ECOS.2410>
- Sánchez-Jardón, L., Uribe-Paredes, R., Álvarez-Saravia, D., Aldea, C., Raimilla, V., Velázquez, E., Millán, S., Águila, J. 2022b. Gestión local de la información en biodiversidad: fomentando la ciencia participativa en el sur de Chile. *Ecosistemas* 31(3): 2385. <https://doi.org/10.7818/ECOS.2385>
- Schimmel, D. S., Asner, G.P., Moorcroft, P. 2013. Observing changing ecological diversity in the Anthropocene. *Frontiers in Ecology and the Environment* 11(3): 129-137.
- Succi, S., Coveney, P.V. 2019. Big data: The end of the scientific method? *Philosophical Transactions of the Royal Society A* 377: 20180145.
- Troudet, J., Grandcolas, P., Bliin, A., Vignes-Lebbe, R., Legendre, F. 2017. Taxonomic bias in biodiversity data and societal preferences. *Scientific reports* 7(1): 1-14.
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steinger, M. 2003. Remote sensing for biodiversity science and conservation. *Trends in ecology and evolution* 18(6): 306-314.
- UN 2015. *Global Sustainable Development Report*. Department of Economic and Social Affairs. UNDESA, Nueva York, Estados Unidos.
- Vázquez de la Cueva, A., Aulló Maestro, I., Montes Pita, F. 2022. Cartografía de áreas cubiertas por *Acacia dealbata* en imágenes multispectrales invernales y estivales captadas por dron en la comarca de Verín (Ourense, Galicia, España). *Ecosistemas* 31(3): 2331. <https://doi.org/10.7818/ECOS.2331>
- Vega-Gorgojo, G., Ordóñez, C., Giménez-García, J.M., Bravo, F. 2022. Explorando datos abiertos forestales masivos con un navegador web. *Ecosistemas* 31(3): 2452. <https://doi.org/10.7818/ECOS.2452>
- Vidal, C., Alberdi, I., Hernández, L., Redmond, J.J. 2016. *National Forest Inventories. Assessment of Wood Availability and Use*. Springer, Cham, Suiza. <https://doi.org/10.1007/978-3-319-44015-6>
- Xu, L., Herold, M., Tsensbazar, N.E., Masiliūnas, D., Li, L., Lesiv, M., Fritz, S., Verbesselt, J. 2022. Time series analysis for global land cover change monitoring: A comparison across sensors. *Remote Sensing of Environment* 271, 112905. <https://doi.org/10.1016/j.rse.2022.112905>
- Zhu, Z. 2019. Science of Landsat Analysis Ready Data. *Remote Sensing* 11(18), 2166. <https://doi.org/10.3390/rs11182166>