

Documentación de modelos y flujos de trabajo: el siguiente reto en el manejo de información ecológica

F.J. Bonet-García^{1,2,*}, R. Pérez-Pérez^{1,2}, B.M. Benito^{1,2}, A.J. Pérez-Luque^{1,2}, R.J. Zamora^{1,2}.

(1) Laboratorio de Ecología (iEcolab), Instituto Interuniversitario Sistema Tierra, Universidad de Granada, Avda. del Mediterráneo s/n, Granada 18006 España.

(2) Departamento de Ecología, Facultad de Ciencias, Universidad de Granada, Avda. Fuentenueva s/n, Granada 18071 España.

* Autor de correspondencia: F.J. Bonet [fjbonet@ugr.es]

> Recibido el 12 de noviembre de 2013, aceptado el 10 de diciembre de 2013.

Bonet-García, F.J., Pérez-Pérez, R., Benito, B.M., Pérez-Luque, A.J., Zamora, R.J. 2013. Documentación de modelos y flujos de trabajo: el siguiente reto en el manejo de información ecológica. *Ecosistemas* 22(3):37-45. Doi.: 10.7818/ECOS.2013.22-3.06

Documentación de modelos y flujos de trabajo: el siguiente reto en el manejo de la información ecológica. Los modelos ecológicos se han convertido en una pieza clave de esta ciencia. La generación de conocimiento se consigue en buena medida mediante procesos analíticos más o menos complejos aplicados sobre conjuntos de datos diversos. Pero buena parte del conocimiento necesario para diseñar e implementar esos modelos no está accesible a la comunidad científica. Proponemos la creación de herramientas informáticas para documentar, almacenar y ejecutar modelos ecológicos y flujos de trabajo. Estas herramientas (repositorios de modelos) están siendo desarrolladas por otras disciplinas como la biología molecular o las ciencias de la Tierra. Presentamos un repositorio de modelos (ModeleR) desarrollado en el contexto del Observatorio de seguimiento del cambio global de Sierra Nevada (Granada-Almería). Creemos que los repositorios de modelos fomentarán la cooperación entre científicos, mejorando la creación de conocimiento relevante que podría ser transferido a los tomadores de decisiones.

Palabras clave: metadatos; modelos ecológicos; repositorio; flujo de trabajo.

Bonet-García, F.J., Pérez-Pérez, R., Benito, B.M., Pérez-Luque, A.J., Zamora, R.J. 2013. Documenting models and workflows: the next challenge in the field of ecological data management. *Ecosistemas* 22(3):37-45. Doi.: 10.7818/ECOS.2013.22-3.06

Documenting models and workflows: the next challenge in the field of ecological data management. Ecological models have become a key part of this scientific discipline. Most of the knowledge created by ecologists is obtained by applying analytical processes to primary data. But most of the information underlying how to create models or use analytic techniques already published in the scientific literature is not readily available to scientists. We are proposing the creation of computer tools that help to document, store and execute ecological models and scientific workflows. These tools (called model repositories) are being developed by other disciplines such as molecular biology and earth science. We are presenting a model repository (called ModeleR) that has been developed in the context of the Sierra Nevada Global Change Observatory (Granada-Almería, Spain). We believe that model repositories will foster cooperation among scientists, enhancing the creation of relevant knowledge that could be transferred to environmental managers.

Key words: metadata; ecological model; repository; workflow

Introducción: el reto de almacenar, documentar y gestionar algoritmos y modelos en ecología

Los modelos son una pieza clave de la ciencia ecológica del siglo XXI. Desde los modelos estadísticos que nos permiten conocer las relaciones causa-efecto entre factores abióticos y el funcionamiento de los ecosistemas, hasta modelos complejos que explican el funcionamiento de un ecosistema mediante la simulación de procesos fisiológicos, el concepto de modelo se ha convertido en un elemento imprescindible para la ciencia ecológica actual. Este gran peso que ha adquirido la modelización en las últimas décadas, se debe, en nuestra opinión, a tres factores que se complementan.

En primer lugar contamos con una enorme cantidad de información primaria sobre la estructura y funcionamiento de los ecosistemas. Buena parte de esta información está georreferenciada, lo que facilita la creación de modelos ecológicos espacialmente explícitos. Esta gran cantidad de información es fruto en buena medida del gran esfuerzo descriptivo realizado por los ecólogos en

décadas pasadas. Grandes iniciativas como GBIF (Global Biodiversity Information Facility), LTER (Long Term Ecological Research), NEON (National Ecological Observatory Network), etc. han contribuido a esta situación (Whitlock 2011). Gracias a todos estos esfuerzos contamos con grandes repositorios de información primaria (distribución de especies, funcionalidad del ecosistema, etc.), que han fomentado la investigación ecológica y su transformación en la ciencia que conocemos hoy día (Michener y Jones 2012).

En segundo lugar, las técnicas analíticas han experimentado un auge exponencial en las últimas décadas (Crowley 1992, Green et al. 2005, Metzger et al. 2010). Tanto los métodos de análisis estadístico como los procedimientos de simulación de procesos ecológicos han mejorado considerablemente su funcionamiento y aplicabilidad. Esto hace que los ecólogos dispongan de una enorme batería de procedimientos de modelización aplicables a los grandes conjuntos de datos comentados anteriormente.

Por último, las ciencias de la computación proporcionan el marco físico (hardware) y lógico (software) necesario para simular

sistemas complejos y abordar la resolución de modelos con requerimientos de computación avanzados (Hobbie 2003, Fegraus et al. 2005, Plaszczak y Wellner 2005)

La ciencia ecológica se enfrenta también a nuevos retos (Miche-ner y Jones 2012). En este trabajo nos centraremos en uno de ellos, que consideramos de los más relevantes: ¿Cómo gestionamos los modelos y algoritmos con los que procesamos los datos primarios para obtener conocimiento útil? O dicho de otra manera, ¿Cómo almacenamos las recetas con las que cocinamos los datos?.

La realidad es que la inmensa mayoría de la información ecológica existente (datos de campo en bruto, datos procesados mediante técnicas analíticas y/o modelos diversos) está almacenada sólo en las memorias (biológicas y digitales) de los investigadores. Buena parte de esa información no aparece publicada en las revistas científicas, por lo que corre el riesgo de perderse cuando el investigador se jubila o deja el trabajo académico. Creemos que la recopilación de todo ese conocimiento es básico si queremos que realmente la ecología: a) responda a los retos de mejorar la comprensión del sistema terrestre (Voinov y Cerco 2010) y los impactos de la actividad humana y b) permita la generación de instrumentos que mejoren la forma en la que administramos y gestionamos dichos recursos naturales (Crutzen 2002, Chapin et al. 2010).

Creemos que al igual que hace 30 años fue necesaria la creación de repositorios de datos primarios en ecología, ha llegado el momento de que abordemos el almacenamiento y gestión de los modelos ecológicos. La creación de repositorios de modelos puede ser considerado como el siguiente paso que la ecología debe dar en materia de gestión de información. La creación de herramientas para conservar y gestionar modelos ecológicos podría contribuir a que la Ecología ocupara su lugar entre las llamadas “grandes ciencias”, caracterizadas por promover la creación de repositorios de datos, documentación de datos y procedimientos analíticos y la creación de infraestructuras para promover la colaboración entre científicos (Borgman et al. 2007). En este sentido, entendemos un repositorio de datos como una herramienta informática capaz de gestionar, documentar y ejecutar modelos y flujos de trabajo creados por ecólogos.

Este reto no es exclusivo de la ecología, sino que también afecta a otras disciplinas científicas que han seguido una evolución similar. En este trabajo describiremos cómo otras disciplinas relacionadas con las ciencias de la tierra y con la biología básica, han abordado o están abordando retos similares. Estas disciplinas han seguido una evolución similar en las últimas décadas: de la recopilación de datos y su catalogación a la creación de modelos complejos para generar conocimiento útil, para a continuación abordar la cuestión de cómo almacenar y documentar los modelos anteriores. En definitiva, han pasado de dedicar muchos esfuerzos a describir la realidad (capturar datos en bruto) a tratar de usar dichos datos para explicar su funcionamiento (modelización).

Por último describiremos ModeleR, que es un prototipo funcional en el que hemos tratado de implementar alguna de las funciones que consideramos ha de tener un repositorio de modelos ecológico. ModeleR es el núcleo básico del sistema de información asociado al programa de seguimiento de los efectos del cambio global de la reserva de la biosfera (y sitio LTER) de Sierra Nevada. Describiremos alguna de las funciones principales de este repositorio de modelos local que ha sido diseñado para conectarse con otros repositorios a escala internacional.

Documentación de modelos en distintas disciplinas científicas

Distintas ramas de la ciencia y de la tecnología han abordado o están abordando el problema de documentar, reutilizar y compartir los modelos o procedimientos analíticos que aplican a los datos primarios. Se trata, en nuestra opinión, de una cuestión más propia de la “era de la información” que de una disciplina científica concreta. Hemos identificado algunos patrones comunes que nos ayudarán a diseñar un repositorio de modelos útil para los ecólogos.

La biología molecular es quizás la disciplina que más ha avanzado en el diseño e implementación de repositorios de modelos. El gran rendimiento científico que para los investigadores ha supuesto la centralización de la información primaria en repositorios de libre acceso como GenBank (Benson et al. 2006), ha hecho que el siguiente paso (centralización de los modelos que usan la información anterior) haya sido considerado como algo casi obvio. Según Buckingham (2007) los repositorios de modelos deben de permitir la documentación de modelos, han de estar conectados a las bases de datos de información primaria (para facilitar la ejecución del modelo en cuestión) y deben de estar diseñados usando el concepto de servicio web. Estos conceptos se han implementado en diversas iniciativas (Hunter y Nielsen, 2005, Snoep et al. 2006), entre los que destacamos Biomodels (Li et al. 2010) que es un repositorio de modelos sobre ciencias biomédicas que han sido publicados en revistas científicas. Este repositorio permite a los científicos el almacenamiento, búsqueda y acceso a modelos matemáticos complejos. Cuando se crea un modelo con Biomodels, es posible controlar las distintas versiones del mismo y cambiar parámetros en las distintas ejecuciones. Incluso es posible descargar el modelo en distintos formatos para poder ejecutarlo en local. Para ello se han desarrollado estándares específicos como SBML (System Biology Markup Language) (Hucka et al. 2003) o MatML (Mathematical Markup Language) (Miner 2005). Una estrategia interesante para promover el uso de estos repositorios consiste en animar a los autores a que suban su modelo a una plataforma web una vez que han sido publicados. Este relativo avance del concepto de repositorio de modelos en el ámbito de la biología molecular se explica parcialmente por la homogeneidad de los datos primarios que usa esta disciplina (Jones et al. 2006). Esto también ha fomentado que dichos repositorios aspiren a tener una escala global al igual que las grandes bases de datos primarios.

Las llamadas ciencias de la Tierra tienen su razón de ser en el desarrollo de modelos que integren consistentemente la interacción entre la atmósfera, la hidrosfera, la litosfera, la biosfera e incluso la heliosfera. Esta necesidad de integración también ha fomentado que surjan conceptos asimilables al de repositorio de modelos. Pero en esta ocasión la idea de repositorio de modelos se orienta más hacia el desarrollo de software específico que sea capaz de acoplar o interconectar diversos modelos adecuadamente documentados (Bulatewicz et al. 2009, Castronova et al. 2013) que además tengan en cuenta la existencia de versiones en los modelos (Thornton et al. 2005). No se trata tanto de tener una base de datos donde los investigadores busquen, ejecuten o descarguen modelos concretos, sino de una aplicación que permita encadenar la ejecución de distintos modelos modulares gracias a la existencia de los conectores adecuados. Esta aproximación implica que los modelos a interconectar sean autodescriptibles. Es decir, que los resultados de la ejecución de un módulo determinado contengan los metadatos tanto del modelo como de los datos de entrada (Turuncoglu et al. 2013). Además de esta modularidad, se promueve que la creación de modelos estén basados en el esfuerzo colectivo y colaborativo de un conjunto de expertos en distintas materias (científicos de distintas disciplinas, ingenieros informáticos, etc.). Esto es lo que se denomina “community modeling systems” (Voinov et al. 2010). Hay multitud de ejemplos que ponen en práctica las ideas anteriores. Entre ellos destacamos OpenMI (Gregersen et al. 2007) que suministra un interfaz estandarizado para definir, documentar, acoplar y ejecutar modelos hidrológicos. ESMF (Earth System Modeling Framework) (Hill et al. 2004) comparte la misma filosofía que el proyecto anterior, pero centrado en modelos climáticos y meteorológicos. Por último CSDMS (Community Surface Dynamics Modeling System) (Peckham et al. 2013) es un excelente ejemplo de repositorio de modelos que pretende integrar una gran variedad de procesos naturales teniendo en cuenta distintas escalas temporales y espaciales. En definitiva, los científicos de esta disciplina han avanzado en el desarrollo de técnicas que facilitan la ejecución sincronizada y modular de los modelos que hay en un repositorio.

También se han producido iniciativas interesantes en el ámbito de la ecología. Benz (1997) escribió una carta al editor de la revista

“Ecological Modelling” en la que proponía la creación de un protocolo común para documentar modelos ecológicos. Partiendo de la constatación de que cada científico ha de ser responsable de documentar el trabajo que realiza y del gran número de modelos ecológicos existentes, pero no disponibles, Benz propone un protocolo para documentar los modelos ecológicos. Este protocolo consta de tres posibles niveles de documentación: a) Descripción general del proyecto (objetivos, metodología, bibliografía asociada, autores, etc.) b) Lo anterior más una descripción general de los algoritmos matemáticos subyacentes al modelo en cuestión y c) Lo anterior más la caracterización detallada de los algoritmos matemáticos que se usan para ejecutar el modelo documentado. El protocolo de documentación debería de estar estandarizado al igual que en la actualidad contamos con estándares para documentar datos ecológicos. Estas ideas fueron parcialmente implementadas en un repositorio llamado ECOBAS (Hoch et al. 1998, Benz et al. 2001, Strube et al. 2008) que actualmente se haya alojado en el ORNL (Oak Ridge National Laboratory) y permite tanto la documentación como la ejecución en línea de los modelos. Otra iniciativa interesante es el llamado “Forest Model Archive”. Este proyecto fue iniciado en la Universidad de Greenwich (Reino Unido) y tiene como objetivo reutilizar los modelos forestales y plantear la inclusión de los mismos en un archivo como una “nueva forma de publicar los modelos”. Definen una serie de características para documentar los distintos modelos: Descripción de la teoría subyacente al modelo, despliegue del código del modelo, algoritmos de ajuste, datos de validación, etc. (<http://goo.gl/oigcbi>). Otros proyectos como el “Forest Growth Models”, (<http://models.etiennethomassen.com/>) auspiciado por la Universidad de Wageningen, persiguen la misma filosofía de recopilar modelos pero en este caso sin ningún criterio de normalización de los metadatos.

En resumen, los ecólogos que se han aproximado a la cuestión de los repositorios de modelos se han centrado en la documentación escalable de los mismos en función de su complejidad y tipo. Esto es una consecuencia más de la gran heterogeneidad de tipos de modelos que se han desarrollado en esta ciencia.

Definición operativa de modelo ecológico

Crear un modelo ecológico es una tarea compleja. Es necesario disponer de datos heterogéneos sobre distintas variables del territorio sobre el que se desarrolla el modelo. Estos datos pueden tener una dimensión espacial (datos georreferenciados), una dimensión temporal (cuando cubren un intervalo de tiempo) o ambas simultáneamente, cuando el análisis tiene una dimensión espacio-temporal. Además, los datos también pueden tener distintos grados de precisión y exactitud, según las fuentes que los proporcionan, o distintas escalas espaciales o temporales.

Un procedimiento de análisis de datos está descrito mediante un lenguaje informático, y su código puede ejecutarse en cada ocasión que sea necesaria la información que puede proporcionar. Un procedimiento de estas características homogeniza criterios entre las distintas unidades de gestión de un territorio, acelera la obtención de información, y facilita esta información rápidamente en formatos que aceleran el proceso de interpretación, como gráficos, tablas o material multimedia, como animaciones tridimensionales.

Observemos por un momento un ejemplo de un proceso de análisis territorial sencillo y automatizado, en su mínima expresión: el cálculo de la pendiente del terreno (ángulo de la línea de máxima pendiente respecto a la horizontal de referencia). Tomando como datos de entrada los valores de altitud proporcionados por un modelo digital de elevaciones, un algoritmo transforma las diferencias en elevación entre celdas adyacentes en valores de pendiente ofrecidos en grados. La salida de este proceso es un mapa digital con los valores de pendiente correspondientes asignados a cada una de las celdas en las que se divide el territorio (ver Fig. 1).

Analizando la estructura conceptual de este conjunto, podemos observar a primera vista, varios componentes elementales (ver Fig. 2): unos datos de entrada, que son los valores de elevación del mo-

delo de elevaciones; un núcleo de procesamiento, representado por el algoritmo que realiza la transformación; una información de salida, que se ofrece como un mapa digital de pendientes. Sin embargo, un análisis más profundo de los componentes nos revelaría nuevos elementos: El núcleo de procesamiento tiene, al menos, un punto de entrada, a través del que recibe los datos (los importa), y un punto de salida, a través del que emite la información resultante (la exporta). Por esta analogía importación-exportación, a los puntos de entrada y salida se los denomina “puertos”. Los puertos reciben o emiten datos e información a través de “canales”, que son las líneas de flujo que unen los elementos, dando consistencia física a la estructura. Además, hay una jerarquía intrínseca, a la que denominaremos “método de ejecución”, que determina el comportamiento y orden de ejecución del conjunto: en primer lugar los valores de elevación se introducen en un canal, que llega a un puerto de entrada del núcleo de procesamiento. Éste realiza la transformación de la elevación en pendiente, y solo después, emite la información resultante a través un puerto de salida que la envía a un canal, que la hace llegar a su destino (en este caso, la pantalla de un PC, por ejemplo).

A la estructura descrita, la denominaremos, de ahora en adelante, modelo, y, para concretar, su estructura formal está compuesta por los siguientes elementos (ver Fig. 2), en número variable según la naturaleza del modelo: 1) datos de entrada. 2) núcleos de procesamiento: 2.1) puertos de entrada. 2.2) algoritmo. 2.3) puertos de salida. 3) información de salida. 4) canales. 5) método de ejecución.

En el campo de los modelos ecológicos, es difícil encontrar estructuras tan simples como la del modelo de pendiente del terreno. En este tipo de modelos es muy común la utilización de distintas fuentes de información (como mapas temáticos, imágenes satélite o tablas con datos alfanuméricos), que debe ser procesada mediante distintos métodos y algoritmos (por ejemplo, clasificación automática de coberturas, análisis estadístico o lógica difusa), y con salidas variadas (nuevos mapas, gráficas o tablas resumen). Esta combinación de elementos da lugar a modelos con estructuras más complejas, con varios datos de entradas, núcleos de procesamiento con varios puertos de entrada y salida, canales ramificados y distintos tipos de información de salida. Estas estructuras complejas, además, pueden necesitar métodos de ejecución diferentes. Por ejemplo, en una secuencia lineal de núcleos de procesamiento, el método de ejecución será secuencial, porque un núcleo, para ejecutarse, requiere que su predecesor le haya proporcionado datos (un buen ejemplo de esto podrían ser los modelos de distribución de especies: Benito et al, en prensa). Sin embargo, en redes de análisis más imbricadas, el método de ejecución puede tener dis-

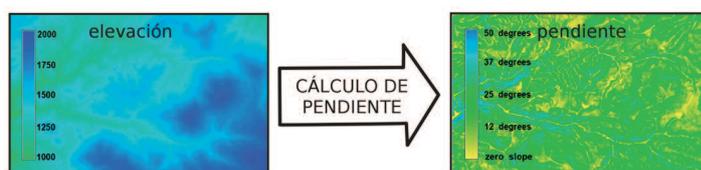


Figura 1. Representación gráfica de un modelo de pendiente del terreno, mostrando los datos de origen (elevación), el algoritmo de cálculo representado por la flecha, y el resultado (pendiente).

método de ejecución

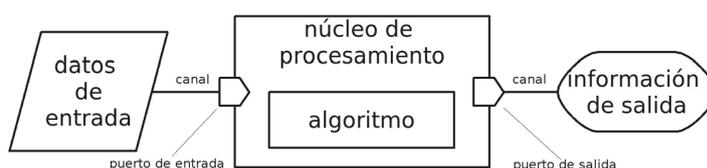


Figura 2. Estructura formal de un modelo. Este concepto de modelo ha sido utilizado para diseñar el repositorio de modelos descrito en este trabajo.

tintos requerimientos, como ejecución concurrente (en paralelo) de segmentos del modelo, ejecución condicional de una rama u otra, o iteraciones (los modelos de reintroducciones de especies son un buen ejemplo de iteración ya que los resultados condicionan las futuras intervenciones: Johnson et al. 2010). En ocasiones, incluso la propia naturaleza del modelo puede requerir, por ejemplo, un método de ejecución dependiente del tiempo real, como podría ser el caso de un sistema de ecuaciones diferenciales para analizar la evolución temporal de variables en un modelo hidrológico (Herrero et al. 2009).

Esta estructura es necesariamente escalable, cuando unos modelos necesitan como entrada la información de salida ofrecida por otros modelos. En este contexto, un modelo, tanto puede ser uno compuesto por un solo núcleo de procesamiento, como una red de modelos, compuesta por “n” modelos con sus respectivos núcleos de procesamiento. En esta nueva jerarquía compuesta por varios modelos interdependientes, cada modelo de la estructura puede anidarse como un único núcleo de procesamiento. Cada uno de estos modelos anidados, además, puede poseer su propio método de ejecución, según sus requerimientos.

Como se puede ver, usamos una acepción bastante abierta y versátil de modelo. Se trata de conseguir un alto grado de abstracción que nos permita avanzar en el desarrollo de herramientas que documenten, almacenen y ejecuten distintos tipos de modelos y técnicas de procesamiento de datos.

ModeleR, nuestra propuesta de repositorio de modelos

Teniendo en cuenta las ideas anteriores hemos creado una primera versión de un repositorio de modelos ecológicos. El desarrollo de esta herramienta fue encargada a nuestro grupo por la Red de Información Ambiental de Andalucía (REDIAM), que es el sistema de información ambiental de la Junta de Andalucía. Nuestra propuesta de repositorio de modelos ecológicos se llama ModeleR (Pérez-Pérez et al. 2012, Bonet et al. 2014) y actualmente funciona en el contexto del Observatorio de seguimiento del cambio global de Sierra Nevada (Granada-Almería) (Aspizua et al. 2010, Bonet et al.

2010). ModeleR puede documentar cualquier modelo ecológico utilizando distintos niveles de profundidad en el proceso de documentación. También permite la conexión de un modelo determinado con cualquier fuente de datos previamente documentada usando el estándar EML (Ecological Metadata Language) (Michener 2006). Además, el sistema es capaz de crear un prototipo de un flujo de trabajo ejecutable usando el software Kepler (Altintas et al. 2004). Este prototipo puede ejecutarse tanto en local como en un servidor. También es posible conectar dos modelos mediante sus entradas/salidas. Una fuente de datos generada por un modelo puede ser utilizada como entrada por otro modelo diferente. En la actualidad ModeleR gestiona más de 200 modelos y procesos analíticos del sitio LTER de Sierra Nevada. Esta herramienta, que es accesible vía web previo registro gratuito (<http://modeler.obsnev.es/>), se ha convertido en el núcleo del sistema de información de este sitio LTER.

La Figura 3 muestra la estructura general del ModeleR. Para su diseño e implementación se han tenido en cuenta los avances realizados en este ámbito por las disciplinas científicas mencionadas anteriormente. Describimos a continuación y de manera resumida las principales funciones de ModeleR:

Documentación de modelos

La definición más simple de metadatos se puede resumir en la siguiente frase: “datos estructurados sobre los datos.” Más formalmente se trata de la información de nivel superior que describe el contenido, calidad, estructura y accesibilidad de un conjunto específico de datos (Michener et al. 1997, Michener 2006). Esta información descriptiva puede ser de un objeto o recurso, tanto si es físico como electrónico (Michener 2006).

Permiten conocer el contenido de los datos a los que se está accediendo, sus posibilidades y limitaciones, proporcionando de esta forma lo necesario para una correcta interpretación de los datos. Se trata de información del contexto de los datos necesaria para comprenderlos y usarlos (Jones et al. 2006, Michener y Jones 2012), ya que describen **quién, qué, cuándo, por qué, dónde y cómo** para cada aspecto de los datos (Jones et al. 2001, Fegeaus et al. 2005, Michener 2006).

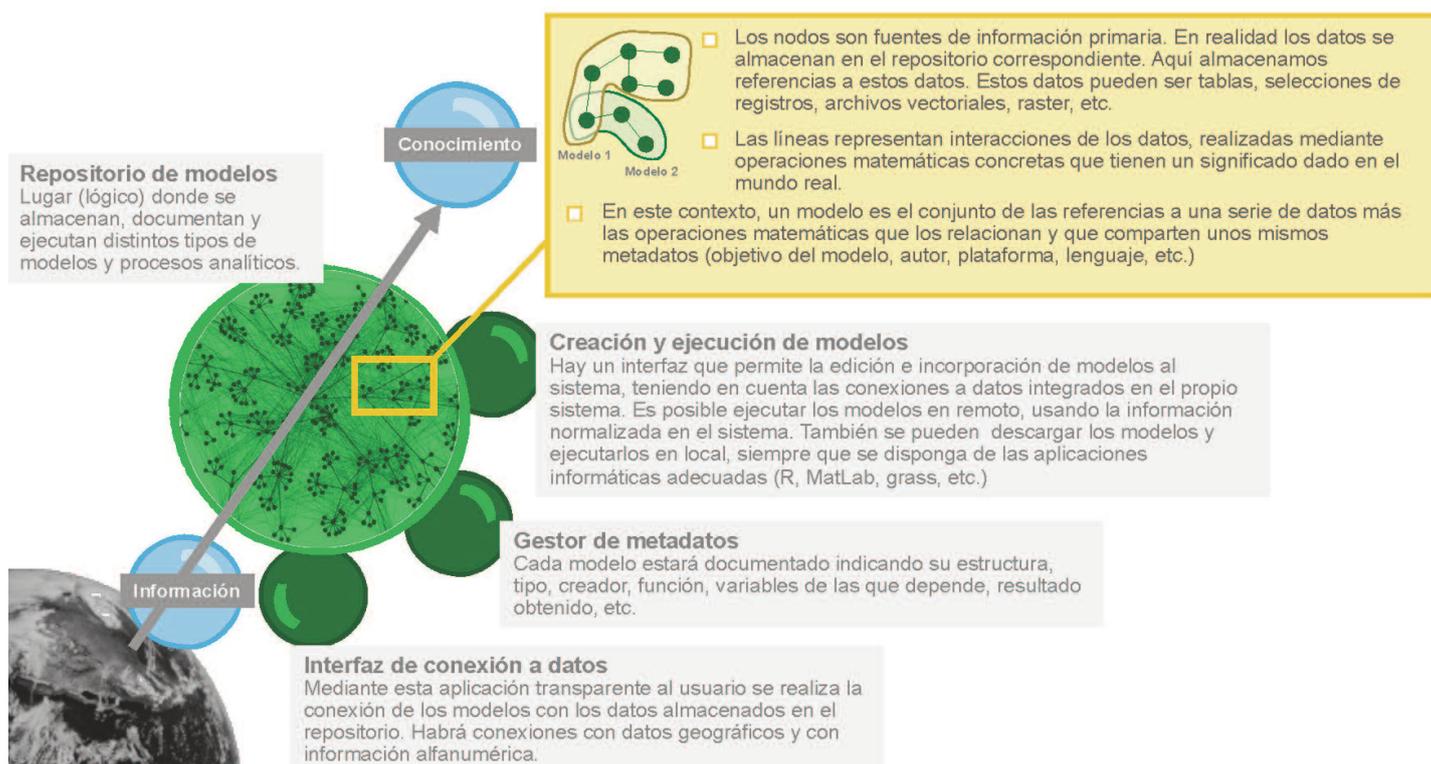


Figura 3. Estructura teórica general propuesta para un repositorio de modelos ecológicos. Esta estructura ha sido adoptada para implementar ModeleR en el contexto del Observatorio de seguimiento del cambio global de Sierra Nevada. Se observan los distintos módulos de los que consta el sistema, así como sus principales funciones.

La importancia de los metadatos ha sido destacada tanto en Ecología como en otras disciplinas (ver [Jones et al. 2006](#)), ya que proporcionan la documentación necesaria para entender el contenido, el formato y el contexto de un conjunto de datos ([Michener y Jones 2012](#)). De hecho, se considera un componente crítico de la infraestructura de gestión de datos en ecología ([Michener 2006](#)). En el caso de la documentación de modelos ecológicos destaca la iniciativa CREM (Council for Regulatory Environmental Modeling) ([CREM 2009](#)), auspiciada por la EPA (Environmental Protection Agency) para estandarizar la forma en la que se describen los modelos ambientales.

Buena parte del éxito que tienen los grandes repositorios de datos primarios en los últimos años (GBIF, LTER, etc.) se debe a que han dedicado muchos esfuerzos a documentar con detalle sus

paquetes de información ([Michener et al. 2011](#)). De esta forma se consigue que los paquetes de datos estén más accesibles y permite una reutilización de los datos por parte de los científicos y gestores.

En el caso de un repositorio de modelos la clave es documentar la forma en la que se “combinan” los datos para generar conocimiento útil. Si bien en el caso de los datos primarios hay diferentes estándares de documentación (EML, Darwin Core, INSPIRE, etc.) (ver [San Gil et al. 2011](#) para una revisión), no ocurre lo mismo con la forma de documentar los modelos y procesos analíticos. Debido a la importancia de la documentación de los modelos, para poder hacerlos accesibles y reutilizables por otros científicos, hemos desarrollado un protocolo de documentación de los modelos gestionados por ModeleR ([Pérez-Pérez et al. 2012](#)). Se trata de una propuesta compatible con otros estándares de metadatos ecológicos ([Tabla 1](#)).

Tabla 1. Propuesta de metadatos necesarios para la documentación de modelos ambientales mediante modeleR (Modificado a partir de [Pérez-Pérez et al. 2012](#))

Apartados	Metadatos				
<p>Identificación Contiene información relativa a aspectos básicos del modelo que permiten ser identificado por sus usuarios potenciales</p>	<p>Nombre del modelo Nombre común que identifica al modelo</p> <p>Acrónimo</p> <p>Autor Identificación de los autores del modelo</p> <p>Palabras clave</p> <p>Bibliografía asociada Listado de documentos que son importantes para entender el funcionamiento del modelo en cuestión</p>				
<p>Lógica Interna Conjunto de elementos que permiten conocer cómo funciona el modelo en términos generales</p>	<p>Tipo de modelo Definición de la naturaleza del modelo desde un punto de vista matemático. Se basa en un listado de algoritmos y otras técnicas matemáticas y estadísticas.</p> <p>Versión Extendida Información detallada sobre la lógica interna, extraída de las fuentes de información detalladas en la bibliografía.</p> <p>Mapa conceptual Mapa mental o diagrama de la lógica interna del modelo</p> <p>Limitaciones Factores que limitan la aplicabilidad de los resultados del modelo, es importante tenerlas en cuenta para la interpretación de los resultados</p>				
<p>Implementación y ejecución Detalles técnicos sobre la implementación del modelo en un programa informático. En esta sección deben tener cabida tanto los programas “empaquetados” en un único ejecutable como los formados por un conjunto de scripts y librerías</p>	<table border="1"> <tr> <td data-bbox="596 1218 639 1711" style="writing-mode: vertical-rl; transform: rotate(180deg);">Implementación</td> <td data-bbox="639 1218 1508 1711"> <p>Nombre Nombre completo del programa</p> <p>Enlace de descarga</p> <p>Versión</p> <p>Código fuente</p> <p>Lenguaje Para realizar implementaciones propias de los modelos es necesario conocer el lenguaje en que está programado el modelo</p> <p>Diagrama de bloques Descripción de los bloques o módulos que forman el modelo y las funciones y relaciones que se establecen entre ellos</p> <p>Plataforma Sistemas operativos que soportan el programa</p> <p>Requerimientos de software: Dependencias, incompatibilidades</p> <p>Requerimientos de hardware Prestaciones mínimas necesarias</p> </td> </tr> <tr> <td data-bbox="596 1711 639 2103" style="writing-mode: vertical-rl; transform: rotate(180deg);">Ejecución</td> <td data-bbox="639 1711 1508 2103"> <p>Datos de entrada Enumera los datos (capas, tablas, etc.) necesarios para ejecutar el modelo, junto a los modelos de datos y formatos requeridos en cada caso. Se trata de uno de los aspectos más importantes de todo el repositorio. Para que un modelo almacenado en el mismo pueda ejecutarse, debe de estar conectado con las fuentes de datos que necesita para su funcionamiento.</p> <p>Parámetros de configuración Se enumeran los parámetros que deben introducirse en el modelo, y sus valores típicos según la situación (si son conocidos). Para cada parámetro se registra su nombre y su valor típico.</p> <p>Información de salida Describe la información resultante de la ejecución del modelo. Los productos del modelo se tratan como las fuentes de datos necesarias para ejecutar el modelo (entradas). De esta forma cada salida de un modelo puede usarse como entrada de otro diferente. Esto promueve la conectividad entre los modelos del repositorio.</p> </td> </tr> </table>	Implementación	<p>Nombre Nombre completo del programa</p> <p>Enlace de descarga</p> <p>Versión</p> <p>Código fuente</p> <p>Lenguaje Para realizar implementaciones propias de los modelos es necesario conocer el lenguaje en que está programado el modelo</p> <p>Diagrama de bloques Descripción de los bloques o módulos que forman el modelo y las funciones y relaciones que se establecen entre ellos</p> <p>Plataforma Sistemas operativos que soportan el programa</p> <p>Requerimientos de software: Dependencias, incompatibilidades</p> <p>Requerimientos de hardware Prestaciones mínimas necesarias</p>	Ejecución	<p>Datos de entrada Enumera los datos (capas, tablas, etc.) necesarios para ejecutar el modelo, junto a los modelos de datos y formatos requeridos en cada caso. Se trata de uno de los aspectos más importantes de todo el repositorio. Para que un modelo almacenado en el mismo pueda ejecutarse, debe de estar conectado con las fuentes de datos que necesita para su funcionamiento.</p> <p>Parámetros de configuración Se enumeran los parámetros que deben introducirse en el modelo, y sus valores típicos según la situación (si son conocidos). Para cada parámetro se registra su nombre y su valor típico.</p> <p>Información de salida Describe la información resultante de la ejecución del modelo. Los productos del modelo se tratan como las fuentes de datos necesarias para ejecutar el modelo (entradas). De esta forma cada salida de un modelo puede usarse como entrada de otro diferente. Esto promueve la conectividad entre los modelos del repositorio.</p>
Implementación	<p>Nombre Nombre completo del programa</p> <p>Enlace de descarga</p> <p>Versión</p> <p>Código fuente</p> <p>Lenguaje Para realizar implementaciones propias de los modelos es necesario conocer el lenguaje en que está programado el modelo</p> <p>Diagrama de bloques Descripción de los bloques o módulos que forman el modelo y las funciones y relaciones que se establecen entre ellos</p> <p>Plataforma Sistemas operativos que soportan el programa</p> <p>Requerimientos de software: Dependencias, incompatibilidades</p> <p>Requerimientos de hardware Prestaciones mínimas necesarias</p>				
Ejecución	<p>Datos de entrada Enumera los datos (capas, tablas, etc.) necesarios para ejecutar el modelo, junto a los modelos de datos y formatos requeridos en cada caso. Se trata de uno de los aspectos más importantes de todo el repositorio. Para que un modelo almacenado en el mismo pueda ejecutarse, debe de estar conectado con las fuentes de datos que necesita para su funcionamiento.</p> <p>Parámetros de configuración Se enumeran los parámetros que deben introducirse en el modelo, y sus valores típicos según la situación (si son conocidos). Para cada parámetro se registra su nombre y su valor típico.</p> <p>Información de salida Describe la información resultante de la ejecución del modelo. Los productos del modelo se tratan como las fuentes de datos necesarias para ejecutar el modelo (entradas). De esta forma cada salida de un modelo puede usarse como entrada de otro diferente. Esto promueve la conectividad entre los modelos del repositorio.</p>				

Ejecución de modelos

Suponiendo un modelo compuesto por múltiples elementos (varios modelos anidados, con distintos modos de ejecución, distintos núcleos de procesamiento y cierta variedad en cuanto a entradas y salidas) a priori puede resultar sencillo establecer la estructura y relaciones que definen el modelo, pero en absoluto lo es diseñar la misma estructura en un entorno informático capaz de ejecutarla eficientemente.

Un modelo, tal y como se ha definido en secciones anteriores, puede implementarse mediante un programa informático clásico, como un listado de órdenes que guían al ordenador para que ejecute todos los procesos implicados. Sin embargo, dicho programa carece de posibilidades de ejecutar cualquier otra tarea que no sea la asignada. Para ejecutar otra tarea distinta, será necesario construir un nuevo programa, y así, para cada nuevo tipo de modelo que sea necesario llevar a cabo. En un entorno de investigación en el que se usan distintos tipos de modelos como herramientas de uso diario, es importante disponer de herramientas flexibles, que faciliten el desarrollo de nuevos modelos en un plazo razonable y con un coste contenido.

En los últimos años están proliferando los llamados entornos basados en “flujos de trabajo” (workflow) (Ogasawara et al. 2009). Como flujo de trabajo puede entenderse una red de procesos analíticos, que puede ser simple y lineal, o muy compleja y no lineal, diseñada para trabajar sobre un conjunto heterogéneo de datos, en el que tanto el flujo de datos como los componentes que los procesan son representados según un lenguaje formal específico. Consideramos que esta aproximación es muy adecuada para implementar la funcionalidad de ejecución de modelos en nuestro repositorio. El flujo de trabajo está descrito en un lenguaje informático, pero tiene una representación visual de la estructura de nodos y canales, que puede ser igual, o muy parecida, al esquema conceptual del modelo que ejecuta. La representación gráfica de estos elementos facilita la comprensión de las funciones del flujo de trabajo, una localización y corrección rápida de los errores, y ofrece la posibilidad de modificar la estructura física del flujo de trabajo para obtener nuevas funcionalidades.

ModeleR utiliza este concepto de flujo de trabajo como elemento básico para realizar la ejecución de los modelos que alberga. Si un modelo está adecuadamente documentado, ModeleR genera automáticamente un prototipo de flujo de trabajo. Para ello se usa un entorno denominado Kepler (Altintas et al. 2004). Este prototipo de flujo de trabajo contiene las fuentes de datos de entrada del modelo previamente documentado, así como sus conexiones a través de diversos tipos de algoritmos. El usuario puede descargar el prototipo a su ordenador personal, completarlo y ejecutarlo para obtener los resultados deseados. Gracias al uso de Kepler (o de otros programas de gestión de flujos de trabajo) es posible incluir en un modelo casi cualquier algoritmo generado por un lenguaje de programación scriptable (que se pueda ejecutar usando un macro). De esta manera es posible incluir en ModeleR modelos creados por software como Vensim, Stella, Matlab u otros. Contemplamos así el concepto de meta-modelo (Blanco 2012) en el que cada modelo consta de diversos “módulos” cada uno de los cuales se alimenta de los resultados del anterior.

El proceso de ejecución del modelo conlleva una serie de requisitos importantes además de la mera aplicación de uno o varios algoritmos a un conjunto de datos. Es muy importante, por ejemplo, realizar un seguimiento de las acciones específicas de dichos algoritmos. Este seguimiento de la ejecución de un modelo (tracking provenance) (Frew et al. 2008, Dozier y Frew 2009), permite la optimización de su ejecución.

Ejemplos de uso de ModeleR

Como ya hemos comentado, ModeleR maneja en torno a 200 modelos y flujos de trabajo en el contexto del Observatorio de seguimiento del cambio global de Sierra Nevada. Con objeto de ilus-

trar de manera sencilla las potencialidades de los repositorios de modelos, incluimos en este apartado tres ejemplos concretos del uso de ModeleR en la “vida real”. Los tres ejemplos tienen que ver con tres ámbitos temáticos diferentes y ponen en juego metodologías analíticas y de procesamiento completamente diferentes. La Figura 4 muestra de forma esquemática los flujos de trabajo necesarios para ejecutar cada uno de estos tres modelos.

El primer ejemplo que describimos tiene que ver con la gestión de la información creada en la toma de datos en un programa de seguimiento. El Observatorio de Sierra Nevada cuenta con más de 40 metodologías de seguimiento diferentes (Aspizua et al. 2012). En todas ellas se ha definido un flujo de trabajo que nos permite almacenar de manera automática la información generada. El ejemplo (Fig. 4A) representa concretamente la captura de información sobre aves dispersantes. En primer lugar se ha creado una secuencia usando el software Cybertracker para capturar los datos de campo en un dispositivo móvil (PDA o Smartphone). La información generada se almacena en una base de datos relacional (PostgreSQL). Los datos allí registrados son consultados mediante un cubo OLAP (Online Analytical Processing) (McGuire et al. 2008). Estos cubos permiten consultar los datos en virtud de diferentes variables de agregación. En el caso que nos ocupa se podrían mostrar gráficas de abundancia de distintas especies en función de su hábitat, ecología trófica, ecosistema en el que está el transecto, etc. Todos estos valores son personalizables por el usuario. Todo el proceso desde los datos digitales suministrados por cybertracker hasta las gráficas generadas por el cubo OLAP están documentadas en ModeleR.

El segundo ejemplo (Fig. 4B) también aborda la cuestión de la gestión y procesamiento de datos procedentes de métodos de seguimiento. Pero en este caso el “agente” que toma los datos del campo es un sensor alojado en un satélite. Concretamente se trata del sensor MODIS que está en el satélite Terra de la NASA. Este sensor suministra información sobre distintas variables biofísicas a escala planetaria. En este caso presentamos un flujo de trabajo que extrae información sobre la extensión de la cobertura de nieve en Sierra Nevada. Se han diseñado e implementado varios flujos de trabajo que descargan imágenes de los servidores de la NASA con periodicidad diaria y semanal. Estas imágenes son almacenadas en una base de datos. A continuación se calculan varios indicadores de estado de la cubierta de nieve: fecha de inicio de la temporada de nieve, último día con nieve, duración de la nieve por año hidrológico y número de ciclos de fusión (Wang et al. 2008). Una vez calculados los índices se ofrecen al usuario gráficas y mapas explicativos a través de un portal de datos (<http://linaria.obsnev.es> Accesible previo registro gratuito). Todos los procesos de descarga, extracción y análisis de datos están documentados usando ModeleR.

El último ejemplo (Fig. 4C) ilustra la creación de un modelo de distribución de especies. Los datos de presencia se obtienen del portal de GBIF. Las variables ambientales necesarias para ejecutar el modelo están almacenadas en un ordenador local en este caso. Los algoritmos necesarios para ejecutar el modelo se almacenan en distintos scripts que recogen las diferentes aproximaciones metodológicas disponibles (GARP, Maxent, MARS, etc.). Cada algoritmo genera un modelo de distribución. Todos ellos se ensamblan mediante otro algoritmo también documentado en ModeleR. El modelo resultante se proyecta hacia el futuro mediante un conjunto de variables ambientales predichas en el futuro y en el presente con las variables ambientales actuales. Los dos resultados son también documentados y registrados en ModeleR.

Conclusiones: Beneficios de los repositorios de modelos en ecología

En nuestra opinión, la creación de repositorios de modelos contribuirá al avance de la Ecología como ciencia en dos aspectos importantes: promoverán la capacidad de crear conocimiento relevante y también aumentarán la capacidad de la Ecología de transferir ese conocimiento a los tomadores de decisiones.

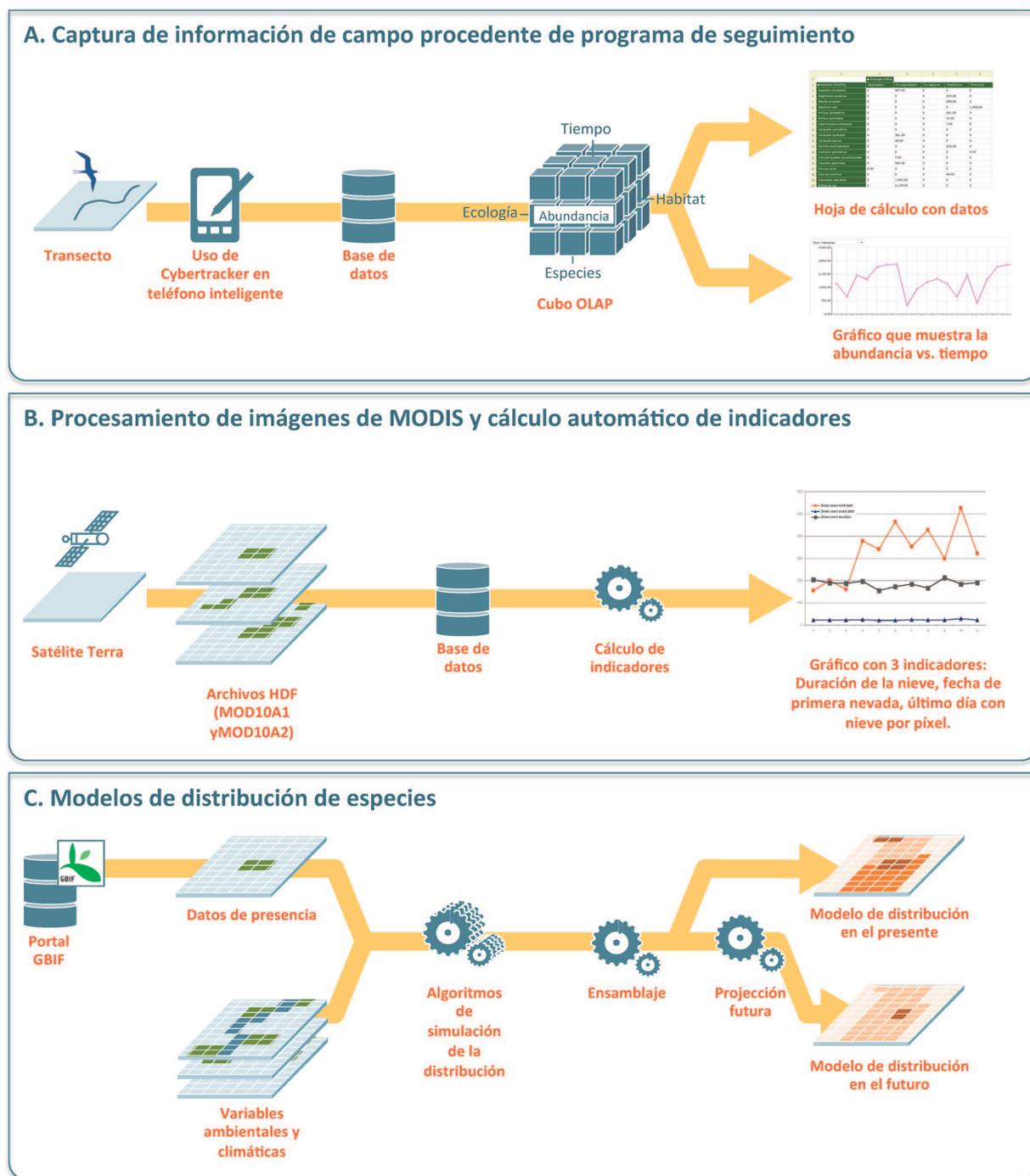


Figura 4. Descripción gráfica de tres ejemplos de modelos documentados y ejecutados por ModeleR. A) Flujo de trabajo para registrar información sobre censos de aves dispersantes en una base de datos. B) Flujo de trabajo para descargar y procesar automáticamente información procedente de imágenes de satélite. C) Modelo de distribución de especies según quedaría implementado en ModeleR.

La documentación de modelos es una buena práctica en la disciplina de la modelización (Scholten 1999). Esta tarea facilita el proceso de conceptualización, descripción matemática e implementación real de un modelo determinado (Keller y Dungan 1999). Un repositorio de modelos también contribuye a la difusión de un modelo entre diferentes científicos. Creemos que el hecho de compartir un modelo a través de estos repositorios permitirá a los ecólogos avanzar en la reproducibilidad de sus análisis (Casey 2006). Asimismo, la reutilización del código y de los modelos fortalecerá la transferencia del conocimiento obtenido a los gestores (Scholten 1999).

A pesar de las ventajas que aporta la documentación de modelos y los repositorios de modelos, estas herramientas no están siendo adoptadas con mucho éxito en las ciencias naturales. Algunos científicos argumentan que usar estas herramientas consume mucho tiempo. No estamos de acuerdo con esa afirmación. Hemos

calculado que para describir un modelo complejo (parecido a los ejemplos suministrados) se necesitan unas dos horas aproximadamente. El proceso es más rápido si el usuario ha descrito previamente los distintos pasos del flujo de trabajo de su modelo. También es posible que algunos autores consideren que publicando sus modelos en estas herramientas pierdan el control sobre los productos de su trabajo. Parte de la respuesta podría estar también en que la publicación de modelos en esta plataformas no está acompañada de ningún incentivo hacia los investigadores.

En definitiva, los beneficios de este tipo de herramientas informáticas se pueden resumir en una simple palabra: sinergia. Cuanto más fácil sea compartir cómo se analizan datos ecológicos para generar conocimiento, más fuerte será la colaboración entre científicos. Tenemos la tecnología y el conocimiento necesarios para crear estas herramientas, así que el éxito de las mismas depende (en nuestra opinión) de la implicación de los ecólogos.

Agradecimientos

El desarrollo de ModeleR ha sido financiado por la Consejería de Medio Ambiente y Ordenación del Territorio de la Junta de Andalucía a través de la Red de Información Ambiental (REDIAM), gracias a un convenio llamado "Diseño y creación de un repositorio de modelos para la red de información ambiental de Andalucía". A.J. Pérez-Luque agradece al MICINN por el contrato PTA 2011-6322-I.

Referencias

- Altintas, I., Berkley, C., Jaeger, E., Jones, M., Ludascher, B., Mock, S. 2004. Kepler: an extensible system for design and execution of scientific workflows. *Proceedings of the 16th International Conference on Scientific and Statistical Database Management (SSDBM 2004)*, 21-23 June 2004, Santorini Island, Greece. IEEE Computer Society. Doi:10.1109/SSDM.2004.1311241
- Aspizua, R., Bonet, F.J., Zamora, R., Sánchez, F.J., Cano-Manuel, F.J., Henares, I. 2010. El observatorio de cambio global de Sierra Nevada: hacia la gestión adaptativa de los espacios naturales. *Ecosistemas* 19(2):56-68.
- Aspizua, R., Barea-Azcón, J.M., Bonet, F.J., Pérez-Luque, A.J., Zamora, R. 2012. *Observatorio de Cambio Global Sierra Nevada: metodologías de seguimiento*. Consejería de Medio Ambiente, Junta de Andalucía. 112 pp. Disponible en: <http://es.scribd.com/doc/81687592/Observatorio-de-Cambio-Global-Sierra-Nevada-Metodologias-de-seguimiento>
- Benito, B., Lorite, J., Pérez-Pérez, R., Gómez-Aparicio, L., Peñas, J. (en prensa). Forecasting plant range collapse in a mediterranean hotspot: when dispersal uncertainties matter. *Diversity and Distributions* 20: 72-83
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L. 2006. GenBank. *Nucleic Acids Research* 34: D16-20. Doi:10.1093/nar/gkj157
- Benz, J. 1997. Call for a common model documentation etiquette. *Ecological Modelling* 97(1-2): 141-143.
- Benz, J., Hoch, R., Legovic, T. 2001. ECOBAS - modelling and documentation. *Ecological Modelling* 138: 3-15.
- Blanco, J.A., 2012. Más allá de los modelos de crecimiento: modelos ecológicos híbridos en el contexto del manejo forestal sostenible. *Cuadernos de la Sociedad Española de Ciencias Forestales* 25: 11-25.
- Bonet, F.J., Pérez-Luque, A.J., Moreno, R., Zamora, R. 2010. Observatorio de Cambio Global en Sierra Nevada. Estructura y contenidos básicos. Consejería de Medio Ambiente, Junta de Andalucía - Universidad de Granada. Granada, España. Disponible en: <http://es.scribd.com/doc/32239398/Observatorio-Sierra-Nevada-Datos-basicos>
- Bonet, F.J., Pérez-Pérez, R., Benito, B. M., Albuquerque, F. S., Zamora, R. 2014. Documenting, storing, and executing models in ecology: A conceptual framework and real implementation in a Global Change Monitoring program. *Environmental Modelling and Software* 52: 192-199.
- Borgman, C.L., Wallis, J.C., Enyedy, N. 2007. Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries* 7(1-2): 17-30.
- Buckingham, S. 2007. To build a better model. *Nature Methods* 4(4): 367-373.
- Bulatewicz, T., Yang, X., Peterson, J.M., Staggenborg, S., Welch, S.M., Steward, D.R. 2009. Accessible integration of agriculture, groundwater, and economic models using the Open Modeling Interface (OpenMI): methodology and initial results. *Hydrology and Earth System Sciences Discussions* 6(6): 7213-7246.
- Cassey, P. 2006. Reproducibility and repeatability in ecology. *BioScience* 56(12): 958-959.
- Castronova, A.M., Goodall, J.L., Ercan, M.B. 2013. Integrated modeling within a Hydrologic Information System: An OpenMI based approach. *Environmental Modelling and Software* 39: 263-273.
- Chapin, F.S., Carpenter, S.R., Kofinas, G.P., Folke, C., Abel, N., Clark, W.C., Olsson, P., Smith, D.M.S., Walker, B., Young, O.R., Berkes, F., Biggs, R., Grove, J.M., Naylor, R.L., Pinkerton, E., Steffen, W., Swanson, F.J. 2010. Ecosystem stewardship: sustainability strategies for a rapidly changing planet. *Trends in Ecology and Evolution* 25(4): 241-249.
- Crowley, P.H. 1992. Resampling methods for computation-intensive data analysis in Ecology and Evolution. *Annual Review of Ecology and Systematics* 23: 405-447.
- Council for Regulatory Environmental Modeling (CREM) 2009. 'Guidance on the Development, Evaluation, and Application of Environmental Models', Publication EPA/100/K-09/003, U.S. Environmental Protection Agency, Washington, D.C.
- Crutzen, P. 2002. Geology of mankind. *Nature* 415: 23.
- Dozier, J., Frew, J. 2009. Computational provenance in hydrologic science: a snow mapping example. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 367: 1021-1033.
- Fegraus, E., Andelman, S., Jones, M.B., Schildhauer, M. 2005. Maximizing the value of ecological data with structured metadata: an introduction to ecological metadata language (EML) and principles for metadata creation. *Bulletin of the Ecological Society of America* 86: 158-168.
- Frew, J., Metzger, D., Slaughter, P. 2008. Automatic capture and reconstruction of computational provenance. *Concurrency and Computation: Practice and Experience* 20(5): 485-496.
- Green, J.L., Hastings, A., Arzberger, P., Ayala, F.J., Cottingham, K.L., Cuddington, K., Davis, F., Dunne, J.A., Fortin, M.J., Gerber, L., Neubert, M. 2005. Complexity in ecology and conservation: mathematical, statistical, and computational challenges. *BioScience* 55(6): 501-510.
- Gregersen, J.B., Gijbbers, P.J.A., Westen, S.J.P. 2007. OpenMI: Open modelling interface. *Journal of Hydroinformatics* 9(3): 175-191.
- Herrero, J., Polo, M., Moñino, A., Losada, M. 2009. An energy balance snowmelt model in a Mediterranean site. *Journal of Hydrology* 371(1): 98-107.
- Hill, C., DeLuca, C., Suarez, M. 2004. The architecture of the earth system modeling framework. *Computing in Science and Engineering* 6(1): 18-28.
- Hobbie, J.E. 2003. Scientific accomplishments of the Long Term Ecological Research program: an introduction. *BioScience* 53(1): 17-20.
- Hoch, R., Gabele, T., Benz, J. 1998. Towards a standard for documentation of mathematical models in ecology. *Ecological Modelling* 113(1-3): 3-12.
- Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., et al. 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19(4): 524-531.
- Hunter, P., Nielsen, P. 2005. A strategy for integrative computational physiology. *Physiology (Bethesda)* 20: 316-325.
- Johnson, S., Mergersen, K., de Waal, A., Marnewick, K., Cilliers, D., Houser, A.M., Boast, L. 2010. Modelling cheetah relocation success in southern Africa using an Iterative Bayesian Network Development Cycle. *Ecological Modelling* 221: 641-651.
- Jones, M.B., Berkley, C., Bojilova, J., Schildhauer, M. 2001. Managing scientific metadata. *Internet Computing, IEEE* 5(5): 59-68
- Jones, M.B., Schildhauer, M.P., Reichman, O.J., Bowers, S. 2006. The new Bioinformatics: integrating ecological data from the gene to the biosphere. *Annual Review of Ecology, Evolution, and Systematics* 37(1): 519-544.
- Keller, R.M., Dungan, J.L. 1999. Meta-modeling: a knowledge-based approach to facilitating process model construction and reuse. *Ecological Modelling* 119(2-3): 89-116.
- Li, C., Donizelli, M., Rodriguez, N., Dharuri, H., Endler, L., Chelliah, V., Li, L., He, E., Henry, A., Stefan, M.I., Snoep, J.L., Hucka, M., Le Novère, N., Laibe, C. 2010. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology* 4(1): 92. doi:10.1186/1752-0509-4-92
- McGuire, M., Gangopadhyay, A., Komlodi, A., Swan, C. 2008. A user-centered design for a spatial data warehouse for data exploration in environmental research. *Ecological Informatics* 3(4-5): 273-285.
- Metzger, K.J., Klaper, R., Thomas, M.A. 2010. Implications of informatics approaches in ecological research. *Ecological Informatics* 6(1): 4-12.
- Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B., Stafford, S.G. 1997. Nongeospatial metadata for the ecological sciences. *Ecological Applications* 7(1): 330-342
- Michener, W.K. 2006. Meta-information concepts for ecological data management. *Ecological Informatics* 1(1): 3-7.
- Michener, W.K., Porter, J., Servilla, M., Vanderbilt, K. 2011. Long term ecological research and information management. *Ecological Informatics* 6(1): 13-24
- Michener, W.K., Jones, M.B. 2012. Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology and Evolution* 27(2): 85-93.

- Miner, R. 2005. The importance of MathML to communication. *Notices of the American Mathematical Society* 52(5): 532-538.
- Ogasawara, E., Rangel, P., Murta, L., Werner, C., Mattoso, M. 2009. Comparison and versioning of scientific workflows. *2009 ICSE Workshop on Comparison and Versioning of Software Models*, 17, 2009, Vancouver, Canada. pp. 25-30. Doi:10.1109/CVSM.2009.5071718
- Peckham, S., Hutton, E., Norris, B. 2013. A component-based approach to integrated modeling in the geosciences: The design of CSDMS. *Computers and Geosciences* 53: 3-12.
- Pérez-Pérez, R., Benito, B.M., Bonet, F.J. 2012. ModeleR: An environmental model repository as knowledge base for experts. *Expert Systems with Applications* 39(9): 8396-8411.
- Plaszczak, P., Wellner, R. 2005. *Grid Computing. The Savvy Manager's Guide*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.
- San Gil, I., Vanderbilt, K., Harrington, S.A. 2011. Examples of ecological data synthesis driven by rich metadata, and practical guidelines to use the Ecological Metadata Language specification to this end. *International Journal of Metadata, Semantics and Ontologies* 6(1): 46-55
- Scholten, H. 1999. Good modelling practice. En: Makowski, M. (Ed.), *Abstracts of the 13th JISR-IIASA Workshop on Methodologies and Tools for Complex System Modeling and Integrated Policy Assessment*, IIASA, Laxenburg, Austria, pp. 57-59. Disponible en: <http://www.iiasa.ac.at/~marek/ftppub/Pubs/csm99/abst.pdf>
- Snoep, J. L., Bruggeman, F., Olivier, B.G., Westerhoff, H.V. 2006. Towards building the silicon cell: a modular approach. *BioSystems* 83(2-3): 207-16.
- Strube, T., Benz, J., Kardaetz, S., Brüggemann, R. 2008. ECOBAS - A tool to develop ecosystem models exemplified by the shallow lake model EMMO. *Ecological Informatics* 3(2) 154-169.
- Thornton, P., Cook, R., Braswell, B., Law, B.E., Corvallis, W.M., Shugart, H.H., Rhyne, B.T., Hook, L.A. 2005. Archiving numerical models of biogeochemical dynamics. *EOS Transactions American Geophysical Union* 86(44): 6-7.
- Turuncoglu, U.U., Dalfes, N., Murphy, S., DeLuca, C. 2013. Toward self-describing and workflow integrated Earth system models: A coupled atmosphere-ocean modeling system application. *Environmental Modelling and Software* 39: 247-262.
- Voinov, A.A., DeLuca, C., Hood, R., Peckham, S., Sherwood, C.R., Syvitski, J.P.M. 2010. A community approach to earth systems modeling. *Eos, Transactions American Geophysical Union* 91(13): 117-118.
- Voinov, A., Cerco, C. 2010. Model integration and the role of data. *Environmental Modelling and Software* 25(8): 965-969.
- Wang, X., Xie, H., Liang, T. 2008. Evaluation of MODIS snow cover and cloud mask and its application in Northern Xinjiang, China. *Remote Sensing of Environment* 112(4): 1497-1513.
- Whitlock, M.C. 2011. Data archiving in ecology and evolution: best practices. *Trends in Ecology and Evolution* 26(2): 61-65.