

Sobre el intercambio de datos de vegetación: el estándar 'Veg-X' y el paquete de R 'VegX'

M. De Cáceres^{1,2,*}

(1) Centro Tecnológico Forestal de Cataluña, Ctra. antigua St. Llorenç de Morunys km. 2; 25280 Solsona, Catalunya, España.

(2) CREAM, Cerdanyola del Vallès 08193, España.

* Autor de correspondencia: M. De Cáceres [miquelcaceres@gmail.com]

> Recibido el 20 de marzo de 2018 - Aceptado el 20 de marzo de 2018

De Cáceres, M. 2018. Sobre el intercambio de datos de vegetación: el estándar 'Veg-X' y el paquete de R 'VegX'. *Ecosistemas* 27(1): 128-129. Doi.: 10.7818/ECOS.1570

Actualmente existe una cantidad ingente de datos sobre la vegetación de los ecosistemas terrestres del mundo. En muchos casos se trata de datos de parcelas de muestreo, establecidas para estudios concretos o que forman parte de esfuerzos colectivos de observación a gran escala, como son los inventarios florísticos o los inventarios forestales. A menudo dicha información se encuentra archivada en bases de datos relacionales. El *Global Index of Vegetation-Plot Databases* (GIVD; www.givd.info) (Dengler et al. 2011) contiene información, en marzo de 2018, sobre 260 bases de datos de vegetación. Sin embargo, la mayoría de ellas (y otras de uso personal no incluidas en GIVD) almacenan sus datos en estructuras idiosincráticas, es decir con definiciones de campos y relaciones específicas. Esto no supone un problema para estudios que utilicen bases de datos grandes, tales como los inventarios forestales, dado que solo es necesario comprender una sola estructura de datos (Wiser 2016). Sin embargo, cuando uno desea integrar distintas fuentes de datos para realizar nuevos análisis es necesario establecer mapas de campos y conceptos, tarea que ralentiza en gran medida el proceso y puede ser fuente de errores. En el caso de redes de parcelas a escala mundial que compartan los mismos protocolos de muestreo de campo, el problema se puede aliviar estableciendo **estándares de bases de datos** (p.ej. Condit et al. 2014). Sin embargo, esta solución no es aplicable para armonizar e integrar datos provenientes de distintos estudios (p.ej. distintas formas de las parcelas y/o subparcelas, distinta definición de estratos, etc.). Para datos observacionales de biodiversidad (de tipo *especie-lugar-fecha-observador*) existe desde hace años el **estándar de intercambio de datos Darwin Core** (<http://rs.tdwg.org/dwc/>), siendo una de sus claves de éxito su relativa sencillez. Los análisis requiriendo la integración de datos de vegetación de varias fuentes se ven dificultados porque no existe un estándar de intercambio de datos reconocido internacionalmente (análogo a Darwin Core) y las correspondientes herramientas para usarlo con facilidad.

En 2003, el grupo de trabajo de ecoinformática de la *International Association for Vegetation Science* (IAVS; www.iavs.org) decidió promover el desarrollo de un **estándar de intercambio de datos de parcelas de vegetación**. A partir del trabajo en distintos talleres

internacionales se llegó a la publicación de *Veg-X*, un estándar de intercambio de datos implementado en un esquema XML (*eXtensible Markup Language*) y diseñado para ser compatible con las bases de datos de vegetación más utilizadas (Wiser et al. 2011). El estándar distingue entre la entidad observada (p.ej. un árbol) y el acto de observación (es decir, las mediciones), y permite almacenar tanto datos de plantas individuales (p.ej. diámetros, alturas, etc.) como datos 'agregados' (p.ej. porcentaje de cobertura de una especie), incluyendo su posición en estratos verticales. El estándar también soporta el muestreo repetido de organismos y parcelas, así como la agrupación de observaciones por criterios no temporales (p.ej. a partir de definir sub-parcelas). Todos los elementos del esquema XML están claramente definidos para facilitar la interoperabilidad (es decir, el mapeo de los campos desde otras estructuras de datos).

Desde su publicación, uno de los impedimentos a la adopción de *Veg-X* para el intercambio de datos ha sido su relativa complejidad, por otro lado, necesaria para acomodarse a la gran variedad de estructuras existentes. Un caso análogo sería el desarrollo del *Ecological Metadata Language* (<https://knb.ecoinformatics.org/>), cuya elevada complejidad supone una barrera para su uso. Para conseguir que el estándar *Veg-X* sea adoptado por un conjunto amplio de usuarios es importante desarrollar herramientas ecoinformáticas para facilitar la interoperabilidad, así como la integración de documentos, armonización de unidades, etc. Desde 2018, en el grupo de trabajo de ecoinformática de la IAVS nos hemos puesto a desarrollar un paquete de R que sirva para realizar las tareas mencionadas. El desarrollo del paquete, llamado *VegX*, se ha realizado en paralelo a una revisión importante del estándar mismo enfocada a simplificarlo. Tanto el esquema en XML del estándar (ver. 2.0) como el paquete de R, la descripción detallada del esquema, y un manual de usuario se encuentran disponibles en un repositorio de **GitHub** (<https://github.com/miquelcaceres/VegX>).

Aunque se puede considerar funcional, el paquete *VegX* se encuentra actualmente en un estado de desarrollo en el que es necesario comprobar su utilidad para importar e integrar distintos tipos de datos fuente. Además, uno de nuestros objetivos es el de pro-

mover que los datos de vegetación de un estudio científico que sean archivados en repositorios públicos (p.ej. DRYAD; <https://datadryad.org/>) utilicen el estándar Veg-X, para facilitar su reutilización en estudios posteriores. Para ello, sería interesante disponer de varios ejemplos de importación de datos (escritos p.ej. en Rmarkdown). Es por ese motivo que quisiera aprovechar esta Nota Ecoinformática para hacer un llamamiento a investigadores (ecólogos terrestres, ingenieros forestales, botánicos, etc.) con datos de vegetación propios (p.ej. en tablas de Excel o ficheros ASCII de texto) interesados en probar la importación de sus datos y reportarnos los problemas o dudas que hayan surgido, así como el código usado en el proceso. No es necesario que dichos datos sean públicos, dado que solamente nos interesa su estructura, no su contenido. Las personas interesadas en ayudarnos pueden hacernos llegar sus datos de ejemplo y scripts de R o documentos Rmarkdown directamente por correo electrónico o a través de GitHub. Esta información nos ayudará enormemente a mejorar la herramienta que proponemos. A largo plazo, también nos planteamos el desarrollo de aplicaciones interactivas en shiny (<https://shiny.rstudio.com/>) que llamen internamente a funciones del paquete VegX. Dichas aplicaciones facilitarían la transformación de datos al estándar para usuarios con dificultades para manejar R. Cuando alcancen una madurez suficiente, creemos que el conjunto de las herramientas mencionadas (estándar, paquete de R, documentación y aplicacio-

nes shiny) pueden contribuir de modo sustancial a mejorar los procesos de intercambio y armonización de datos de vegetación y facilitar la reutilización de los mismos.

Agradecimientos

Gracias al grupo de ecoinformática de la AEET por animarme a escribir esta nota. Miquel De Cáceres cuenta con una ayuda Ramón y Cajal del MINECO (RYC-2012-11109).

Referencias

- Condit, R., Lao, S., Singh, A., Esufali, S., Dolins, S. 2014. Data and database standards for permanent forest plots in a global network. *Forest Ecology and Management* 316: 21-31.
- Dengler, J., Jansen, F., Glöckler, F., Peet, R.K., De Cáceres, M., Chytrý, M., Ewald, J. et al. 2011. The Global Index of Vegetation-Plot Databases (GIVD): a new resource for vegetation science. *Journal of Vegetation Science* 22: 582-597.
- Wiser, S.K. 2016. Achievements and challenges in the integration, reuse and synthesis of vegetation plot data Chiarucci, A. (ed.), *Journal of Vegetation Science* 27: 868-879.
- Wiser, S.K., Spencer, N., De Cáceres, M.D., Kleikamp, M., Boyle, B., Peet, R.K. 2011. Veg-X - an exchange standard for plot-based vegetation data. *Journal of Vegetation Science* 22: 598-609.

