

Bases de datos y metadatos en ecología: compartir para investigar en cambio global

B. Alonso, F. Valladares

Instituto de Recursos Naturales, Centro de Ciencias Medioambientales, CSIC, Serrano 115, 28006 Madrid.

➤ Recibido el 17 de febrero de 2006, aceptado el 29 de marzo de 2006.

Un avance significativo dentro del área de la ecología sólo se podrá lograr con la adopción generalizada de un sistema basado en el uso compartido de los datos entre científicos. Esta práctica, apoyada por el desarrollo de metadatos precisos que acompañen a los propios datos, conseguirá aumentar la escala temporal y espacial de los objetos de estudio, ventajas evidentes en un área en la que cuestiones de carácter global como el estudio de los efectos del cambio climático van adquiriendo cada vez mayor importancia. En la actualidad existen numerosos grupos de científicos que trabajan voluntariamente en el desarrollo de herramientas que faciliten a los científicos la documentación y el almacenaje de sus datos. Asimismo, Internet está demostrando ser un potente instrumento para compartirlos. Aprovechar estos recursos es decisión nuestra.

Palabras clave: Bases de datos, metadatos, cambio global, internet, sistemas de información geográfica

Data bases and metadata bases in ecology: sharing information to foster change global research. A significant advance in ecology will be achieved only if scientists decide to adopt a system based on sharing data. This practice, supported by the development of accurate metadata information enclosed with the data sets, will increase the temporal and spatial scale of a research topic. These are evident advantages in an area where global questions as the study of climate change effects are becoming more and more important. Nowadays, there are numerous groups of scientists who work as volunteers in the development of tools that facilitate the scientists the documentation and storage of the data. Internet is also demonstrating to be a strong instrument to share data. Making a good use of these resources is our decision.

Key words: Data bases, metadata, global change, internet, geographic information system

Campos como la genética o la biología molecular han experimentado un rápido avance en los últimos años fruto, en gran medida, del uso compartido de los datos que ejercen los investigadores en estas áreas. Si este principio de acceso público a los datos se extendiera a otros ámbitos de la biología, disciplinas relativamente menos adelantadas como la ecología podrían disfrutar de un desarrollo similar. Para ello, el primer paso es que los propios científicos reconozcan los beneficios derivados de esta práctica.

Compartir datos fortalece la investigación científica de diferentes maneras: alienta la diversidad de opinión, favorece la síntesis, promueve la formulación de hipótesis nuevas y permite la exploración de temas diferentes a los previstos inicialmente por los investigadores que generaron los datos. También evita la innecesaria replicación de datos. De hecho, en muchos casos, la extrema dificultad de una eventual replicación (a veces, incluso imposible, al desaparecer las condiciones iniciales del sistema que generaron las observaciones), confiere un carácter único a la información proporcionada.

La mayoría de la investigación en ecología está financiada con fondos públicos, por lo que se podría alegar que los datos pertenecen a todos los ciudadanos, producidos para el bien común y por lo cual, deberían ser de libre acceso (Arzberger *et al.*, 2004). Restricciones de disponibilidad sólo deberían estar justificadas en casos excepcionales (por ejemplo, contener información relacionada con especies en peligro de extinción, comprometer derechos de confidencialidad y privacidad, etc.). Al mismo tiempo, el uso compartido y el acceso público a los datos provenientes de investigación financiada con fondos públicos favorece la optimización de los recursos invertidos.

Existen otras implicaciones importantes de ámbito ético ¿Pueden los investigadores moralmente justificar el mantenimiento de los datos confidenciales si su utilización por terceros puede acelerar el hallazgo de soluciones a problemas medioambientales actuales? Los ecólogos tienen la responsabilidad de generar resultados cualitativa y cuantitativamente relevantes y conseguir publicarlos. Pero también tienen una responsabilidad ambiental y social desde el momento en que la difusión de esos datos puede contribuir a la mejora de la gestión del planeta (Zamora, 2005).

Intuitivamente se podría pensar en dos razones fundamentales por las cuales compartir datos no es todavía una práctica común en el campo de la ecología. Por un lado, está el deseo de los investigadores de guardar sus datos para un posible uso en trabajos posteriores. Por otro lado, la supuesta existencia de barreras logísticas.

Guardar los datos confidenciales para un posible uso futuro no pone en relieve toda la importancia de los mismos. En realidad, la valía de los datos incrementa con su utilización por terceras personas, hecho que además de aumentar el índice de citación del investigador que los generó, constituye un reconocimiento de la importancia de dicha investigación y va a ser considerado asimismo, como criterio para publicación, concesión de ayudas de financiación y reconocimiento profesional (Parr y Cummings, 2005).

Las barreras logísticas a la hora de compartir datos son también ilusorias. En la actualidad, existe una revolución en el desarrollo de herramientas y protocolos que permiten a los ecólogos documentar, compartir y almacenar sus datos de manera adecuada para su uso a largo plazo. La utilización sistemática de estas herramientas facilitará la generación de conocimiento ecológico y supondrá un enorme ahorro de tiempo a los ecólogos, tiempo que podrán dedicar a la investigación en sí misma (Palmer *et al.*, 2004). Concretamente en Estados Unidos de América, existen grupos muy activos de científicos que están trabajando de manera voluntaria en la creación de estos nuevos software que faciliten el uso de los datos ecológicos (ecoinformatics.org).

Para favorecer el uso compartido de los datos es imprescindible que los científicos tomen conciencia de la importancia de incorporar dentro de su actividad experimental como práctica generalizada el desarrollo de metadatos (información que describe el contenido, calidad, estructura y accesibilidad de una serie de datos concreta (Michener *et al.*, 1997)) y el correcto almacenamiento de los mismos junto a los datos. Éste es el único modo de hacer que los datos sean comprensibles para poder ser utilizados en el futuro por el investigador que los generó así como por otros investigadores no involucrados directamente en la investigación original. En otras palabras, el objetivo de los metadatos es eliminar el proceso de pérdida de información de los datos que se produce inevitablemente a través del tiempo (**Fig. 1**). Simultáneamente, la combinación de los metadatos y los propios datos va acompañada generalmente de un incremento del valor de los mismos (**Fig. 2**).

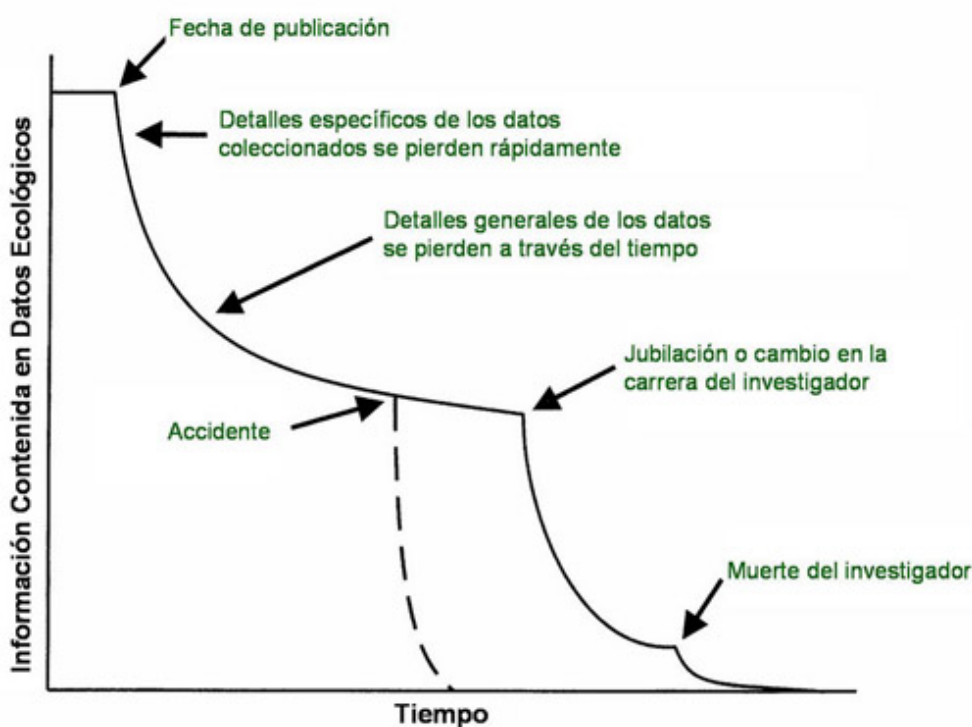


Figura 1. Evolución con el tiempo de la información de datos ecológicos en un sistema no basado en metadatos y sin acceso compartido a los datos (Michener, *et al.*, 1997).

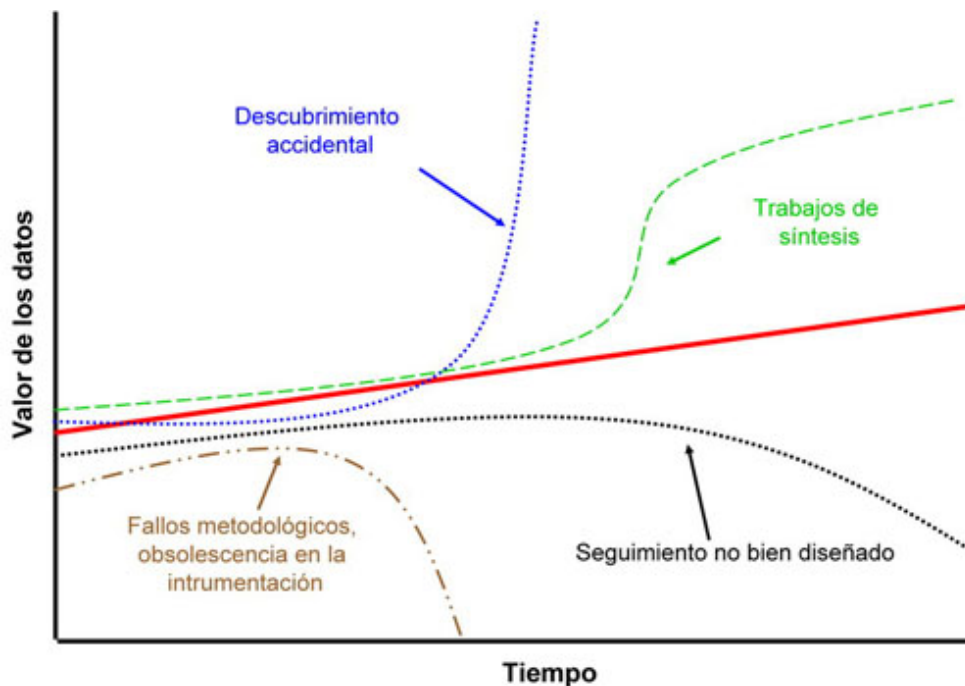


Figura 2. Evolución general del valor de los datos en un sistema basado en metadatos y acceso público a los datos.

El proceso de generación de metadatos lleva asociados múltiples beneficios, pero también conlleva una serie de costes que es necesario considerar:

- Tiempo tanto para el diseño y aprendizaje como para la ejecución en sí misma.
- 'Hardware' y 'software' relacionado con el desarrollo y almacenamiento de los metadatos (servidores, bases de datos, editores especiales, etc.).
- Costes derivados del mantenimiento a largo plazo de los metadatos.

Y sobre todo, para que la implementación de un sistema basado en datos y metadatos resulte exitoso, éste debe ser un proceso en el que exista una implicación por parte de todas las partes del equipo relacionado con la actividad experimental (técnicos de campo, investigadores, técnicos de laboratorio, etc.).

Conseguir la estandarización de los metadatos es muy importante porque permite definir una terminología común, permite llevar a cabo la entrada, validación, acceso, integración y síntesis de los datos de manera automatizada y asegura una documentación completa y precisa del contenido de los datos. Hoy en día existen diferentes estándares de metadatos (*Dublin Core*, *Content Standard for Digital Geospatial Metadata*, *ISO 19115 Geographic information metadata*, etc.). Dentro del área de la ecología hay que destacar el *Ecological Metadata Language*, un estándar muy robusto cuya importancia radica en haber sido creado por ecólogos para ecólogos y en consecuencia, tan completo que se puede aplicar a cualquier serie de datos obtenidos en esa área. El EML fue desarrollado en sus primeras versiones por el National Center for Ecological Analysis and Synthesis (NCEAS, Universidad de Santa Bárbara) basado en un trabajo previo de la Ecological Society of America (Michener *et al.*, 1997) y es el estándar que ha adoptado la red americana US Long Term Ecological Research Network.

Hoy en día, es posible compartir datos ecológicos a través de bases de datos institucionales. En Estados Unidos, redes como la [Knowledge Network for Biocomplexity](#) (KNB), o la [Biological Information Infrastructure](#) (NBII), proporcionan soporte a estaciones de seguimiento, laboratorios y científicos individuales para almacenar sus datos y metadatos y ponerlos a libre disposición de la comunidad científica para su consulta. A nivel nacional, la Red Española De Observaciones Temporales de Ecosistemas (REDOTE, www.redote.org) proporciona acceso a una base de metadatos de iniciativas de seguimiento a largo plazo desarrolladas en ecosistemas españoles (en la actualidad o en el pasado). Esta red, creada en 2002, pretende agrupar y coordinar estaciones de monitorización de ecosistemas con el objetivo de reunir la información necesaria para detectar tendencias y efectos del cambio global. Se trata de una red con vocación de apertura a toda la comunidad científica, que proyecta convertir el uso compartido de datos en una realidad siguiendo el ejemplo de redes como la US-LTER.

En estos momentos se puede ya acceder a través de Internet a algunos datos ecológicos de carácter muy heterogéneo: desde datos muy precisos de organizaciones y redes que utilizan tecnología y técnicas de manejo innovadoras, hasta datos menos completos a veces con carencias en cuanto a su documentación con metadatos (**Tabla 1**).

Tabla 1. Ejemplos de Instituciones que proporcionan acceso a datos a través de Internet. La información presentada corresponde a la Institución, los Datos disponibles y la dirección electrónica (URL).

<p>Long Term Ecological Research Network</p> <p>Series temporales de datos de biodiversidad, nutrientes, clima, ecofisiología, hidrología, sustrato, vegetación, fauna en distintos tipos de ecosistemas de Estados Unidos http://www.lternet.edu</p>
<p>Global Biodiversity Information Infrastructure</p> <p>Información sobre biodiversidad (colecciones botánicas y zoológicas) de todo el globo http://www.gbif.org</p>
<p>Information Center for the Environment's Biological Inventory Databases</p> <p>Información sobre especies vegetales y animales localizadas en áreas protegidas de todo el mundo http://www.ice.ucdavis.edu/bioinventory/bioinventory.html</p>
<p>The UK Environmental Change Network</p> <p>Series temporales de datos de clima, química atmosférica, hidrología, sustratos, vegetación, fauna, usos del suelo en ecosistemas terrestres y dulceacuícolas de Gran Bretaña http://www.ecn.ac.uk</p>
<p>Avibase-The World Bird Database</p> <p>Información de aves de todo el mundo http://www.bsc-eoc.org/avibase/avibase.jsp</p>
<p>Puertos del Estado</p> <p>Datos de oleaje y nivel del mar http://www.puertos.es/</p>
<p>Confederación Hidrográfica del Tajo</p> <p>Datos de aforo, volúmenes embalsados http://www.chtajo.es/</p>
<p>Confederación Hidrográfica del Júcar</p> <p>Datos de aforo, piezometría y calidad de aguas subterráneas http://www.chj.es/index2.HTM</p>
<p>Confederación Hidrográfica del Duero</p> <p>Calidad de aguas subterráneas http://www.chduero.es</p>
<p>Confederación Hidrográfica del Ebro</p> <p>Datos de aforo http://oph.chebro.es/</p>

Confederación Hidrográfica del Guadiana

Volúmenes embalsados, medidas piezométricas
<http://www.chguadiana.es/publica/index.htm>

Subsistema de Información Climatológica Ambiental (CLIMA) – Junta de Andalucía

Parámetros climatológicos de Andalucía
<http://www.juntadeandalucia.es/medioambiente/servtc5/sica/seleccionConsulta.jsp>

Reserva Biológica de Doñana

Clima, hidrología, sustrato, seguimiento de especies animales (mamíferos, aves, anfibios, etc.) y vegetales
<http://www-rbd.ebd.csic.es/>

World Conservation Monitoring Centre

Datos sobre especies, habitats y áreas protegidas y mapas interactivos basados en SIG (en desarrollo)
<http://www.unep-wcmc.org>

NOAA's National Geophysical Data Center (NGDC)

Datos y mapas interactivos sobre parámetros geofísicos
<http://www.ngdc.noaa.gov/ngdcinfo/onlineaccess.html>

Atlas Climático Digital de la Península Ibérica (Universidad Autónoma de Barcelona)

Servidor de mapas climáticos de la Península Ibérica
http://opengis.uab.es/wms/iberia/espanol/es_presentacio.htm

Pero en ambos casos, los datos ofrecidos por estas instituciones poseen tres características de gran relevancia: son gratuitos, su acceso y manejo es sencillo y tienen un elevado valor informativo, ya que en muchas ocasiones han sido registrados desde muchos años atrás, permitiendo así, a través de su análisis, observar la evolución de parámetros a lo largo del tiempo y los cambios repentinos sufridos durante la misma, y detectar tendencias y efectos del cambio global (**Fig. 3** y **Fig. 4**)

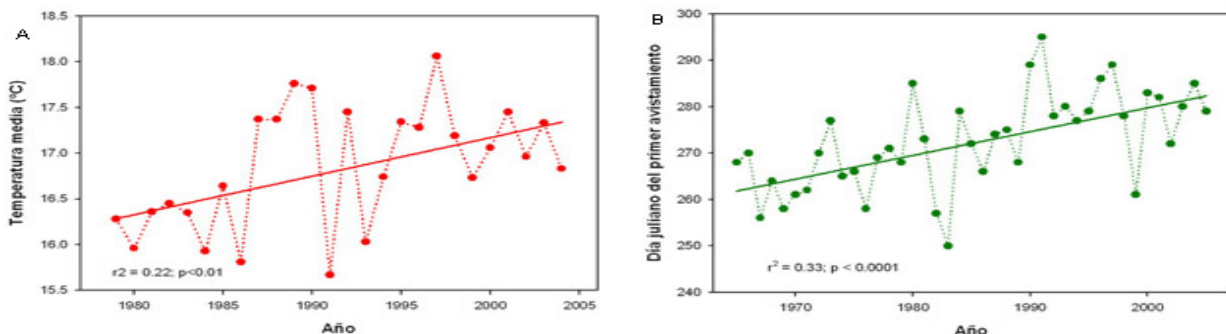


Figura 3. Ejemplos de series de datos disponibles de manera gratuita y sin restricciones en Internet. A la izquierda están representados datos de temperatura media en Doñana entre 1979-2004. A derecha está representada la fecha de la llegada del Ánsar común (*Anser anser*) a Doñana en el momento de su migración invernal hacia el sur (Chans et al., 2005) (Fuente datos: <http://www-rbd.ebd.csic.es/>). La representación gráfica muestra en ambos casos una tendencia significativa de cambio hacia el aumento de la temperatura y hacia el retraso del primer avistamiento del Ánsar común en Doñana.

Agrupamiento cronológico (chronological clustering)

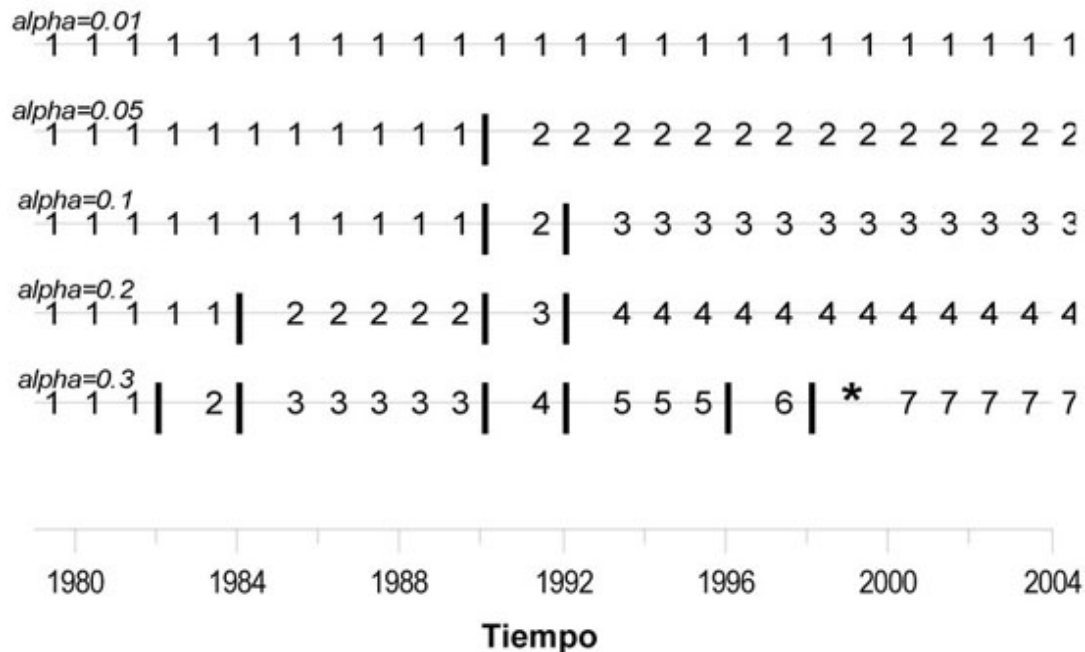


Figura 4. Resultados de aplicar un agrupamiento cronológico o *chronological clustering* a las series de datos de la Figura 3. Las líneas verticales corresponden a cambios repentinos. Con un *alpha* pequeño (0.01) se pueden ver los cambios más fuertes experimentados mientras que con *alphas* mayores se alcanza un grado de detalle mayor de manera que se pueden detectar cambios más leves. En nuestro caso podemos observar que tanto la temperatura media como la fecha de llegada del Ánsar común a Doñana experimentaron 2 cambios claros en su evolución: en 1990 y en 1992.

El creciente interés en los últimos años de la sociedad y los científicos por temas de importancia global como el cambio climático o la pérdida de biodiversidad, convierte en una necesidad el libre acceso a la información y el uso compartido de los datos. Es evidente que el estudio de estas cuestiones va más allá de jurisdicciones nacionales y tiene que verse asistido por una cooperación internacional y un compromiso entre científicos. Conocemos los beneficios y tenemos los medios. Aprovechar esta oportunidad está en nuestras manos.

Referencias

- Arzberger, P. Schroeder, A. Beaulieu, G. Bowker, K. Casey, L. Laaksonen, D. Moorman, P. Uhlir, P. Wouters. 2004. Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Science Journal* 3: 135-152.
- Chans, J. J., Máñez, M., Hiraldo, F., Calderón, J. and Ramo, C. Changes in distribution and arrival migration time in a Greylag Goose (*Anser anser*) population wintering in Spain. *Proceedings of the 9th Annual Meeting of Goose Specialist Group of Wetlands International*. 5-9. November 2005. Sopron, Hungary. Hungarian Waterfowl. Publicación en prensa
- Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B. and Stafford, S.G. 1997. Nongeospatial Metadata for the Ecological Sciences. *Ecological Applications*. 7: 330-342.
- Palmer, M.A. et al. 2004. Ecological Science and sustainability for a crowded planet. *Report from the Ecological Society of America* (www.esa.org/ecovisions).
- Parr, C.S. and Cummings, M.P. 2005. Data sharing in ecology and evolution. *Trends in Ecology and Evolution* 20: 362-363.
- Zamora, R. 2005. Aquí y ahora: una llamada al compromiso y la acción. [Ecosistemas. 2005/2](http://Ecosistemas.2005/2).