

Curso “Open Science for Synthesis 2014”

A.J. Pérez-Luque^{1,*}, S. Varela²

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

(2) Department of Ecology, Faculty of Science, Charles University, Vinicná 7, 128 44 Praha 2, Czech Republic.

* Autor de correspondencia: A.J. Pérez-Luque [ajperez@ugr.es]

> Recibido el 09 de octubre de 2014 - Aceptado el 13 de octubre de 2014

Pérez-Luque, A.J., Varela, S. 2014. Curso “Open Science for Synthesis 2014”. *Ecosistemas* 23(2): 117-118. Doi.: 10.7818/ECOS.2014.23-3.18

Durante los pasados días 21 de Julio a 8 de Agosto tuvo lugar el taller **Open Science for Synthesis** organizado por el *National Center for Ecological Analysis and Synthesis (NCEAS, University of California Santa Bárbara)* y el *Renaissance Computing Institute (RENCI, University of North Carolina)*, que se celebró simultáneamente en Santa Barbara (California) y en Chapel Hill (Carolina del Norte).

Un total de 45 participantes de 35 instituciones procedentes de 11 países (**Fig. 1**) tuvimos la oportunidad de disfrutar de este intensivo taller que giró en torno a dos conceptos principales: **Ecoinformática** y **Open Science**, esto es, un marco de trabajo abierto basado en la utilización de herramientas de desarrollo de software, que permite a los investigadores tanto generar, gestionar y almacenar su producción científica de manera óptima, como difundir, compartir y colaborar con otros investigadores (**Michener y Jones 2012**).

Durante tres semanas un equipo multidisciplinar de **profesores** impartieron sesiones teórico-prácticas sobre diferentes temáticas, entre las que destacan:

- Modelos de colaboración entre investigadores
- Gestión y preservación de datos (Bases de datos y Metadatos)
- Técnicas para compartir datos de forma eficiente (data-sharing)
- Manipulación, exploración e integración de datos ecológicos
- Flujos de trabajo científicos
- Reproducibilidad en Ciencia
- Modelización y análisis de datos
- Técnicas de comunicación y divulgación de resultados científicos
- Documentación de modelos ecológicos

Uno de los aspectos importantes tratados fue el concepto de **Reproducibilidad** (*Reproducible Science*). La reproducibilidad incrementa la transparencia en la ciencia (**Ram 2013**), permitiendo, entre otros una evaluación objetiva de los resultados publicados. Sin embargo, la práctica de depositar los datos utilizados en las publicaciones científicas en bases de datos abiertas (e.g. **KNB**) y el código utilizado para generar los resultados obtenidos en repositorios públicos (e.g. **github**) no es muy frecuente entre los ecólogos, aunque poco a poco las instituciones están incentivando este tipo

de prácticas (**Wolkovich et al. 2012**). En este sentido los participantes aprendimos durante el taller diferentes herramientas y procedimientos que fomentan la reproducibilidad y la colaboración entre investigadores, como el uso de **github** para compartir y desarrollar software en equipo (**Ram 2013; Wilson et al. 2014**).

Cada vez mas los ecólogos trabajamos con grandes volúmenes de datos, siendo necesario implementar una cultura transparente de intercambio y agregación de datos (**Hampton et al. 2013**). Para ello es importante mejorar la forma en la que documentamos y almacenamos los datos ecológicos (**Jones et al. 2006**). Durante el taller aprendimos técnicas para **metadatar** datos y procedimientos para depositarlos en repositorios públicos (e.g. **DataONE**). De igual modo la utilización de flujos de trabajo científicos (**Reichman et al. 2011**) permite mejorar la forma en la que gestionamos los modelos que aplicamos a nuestros datos, sean sencillos o complejos. En este sentido, para aumentar la reproducibilidad, un aspecto clave es documentar los diferentes procedimientos que aplicamos a nuestros datos para obtener un resultado, esto es, documentar los modelos que aplicamos a los datos.

Durante el taller también trabajamos técnicas para elegir el mejor modelo que se aplica a cada conjunto de datos (**Bolker 2008**), y otras para llevar a cabo síntesis de conjuntos de datos voluminosos y heterogéneos, como por ejemplo, el meta-análisis (**Lortie 2013**).

No menos importante fue el aprendizaje de procedimientos para mejorar la forma en que los ecólogos comunicamos los resultados de la investigación, tanto en la comunidad científica como hacia el público general o los medios (**Baron 2010**). Por otro lado se trabajaron técnicas para estimular la colaboración entre científicos en equipos de trabajo multidisciplinarios (**Hampton y Parker 2011**)

Además de las sesiones teórico-prácticas y los foros de discusión, los participantes formamos grupos de trabajo de síntesis, donde aplicamos las habilidades aprendidas (técnicas de colaboración, reproducibilidad, análisis, etc.) a diferentes temas ecológicos. Entre las temáticas elegidas encontramos: efectos de la sequía de California sobre la productividad de los ecosistemas (todos los materiales generados en este proyecto están disponibles en un **repositorio en github**); comparación de patrones ecológicos obtenidos con datos procedentes de ciencia ciudadana frente a los obtenidos

por datos tomados por los científicos; respuestas ecohidrológicas de los ecosistemas a perturbaciones; utilización de grandes conjuntos de datos experimentales para evaluar patrones de diversidad; interacción multiescala entre usos del suelo, biodiversidad y servicios ecosistémicos; etc.

Este es el segundo año que se imparte el curso. Desde aquí queremos animar a los ecólogos interesados en aprender técnicas y herramientas para optimizar los análisis, los flujos de trabajo y la reproductibilidad en la ciencia a que soliciten participar en la próxima convocatoria que presumiblemente será en Julio de 2015.

Para predicar con el ejemplo, esta nota se ha generado usando *markdown*, un lenguaje de edición de texto que permite un control fácil de versiones usando *Git*. Además hemos colaborado para su edición a través de *github*. El texto completo (así como material adicional) se puede consultar en el siguiente [repositorio](#).

Referencias

Baron, N. 2010. *Escape from the ivory tower: a guide to making your science matter*. Island Press. Washington, DC. Estados Unidos.

Bolker, B.M. 2008. *Ecological models and data in R*. Princeton and Oxford, Princeton University Press. Princeton, NJ, Estados Unidos.

Hampton, S.E., Parker, J.N. 2011. Collaboration and productivity in scientific synthesis. *BioScience* 61(11): 900-910, doi: 10.1525/bio.2011.61.11.9

Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., Duke, C.S., Porter, J.H. 2013. Big data and the future of ecology. *Frontiers in Ecology and the Environment* 11: 156-162. doi: 10.1890/120103

Jones, M.B., Schildhauer, M.P., Reichman, O.J., y Bowers, S. 2006. The new bioinformatics: integrating ecological data from the gene to the biosphere. *Annual Review of Ecology, Evolution, and Systematics* 37: 519-44, doi: 10.1146/annurev.ecolsys.37.091305.110031

Lortie, C.J. 2013. Formalized synthesis opportunities for ecology: systematic reviews and meta-analyses. *PeerJ PrePrints* 1:e39v1 doi: 10.7287/peerj.preprints.39v1

Michener, W.K., Jones, M.B. 2012. Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology and Evolution* 27 (2): 85-93, doi: 10.1016/j.tree.2011.11.016

Ram, K. 2013. Git can facilitate greater reproducibility and increased transparency in science. *Source Code for Biology and Medicine* 8: 1-7 doi:10.1186/1751-0473-8-7

Reichman, O.J., Matthew, B.J., Schildhauer, M.P. 2011. Challenges and opportunities of open data in ecology. *Science* 331: 703-705, doi: 10.1126/science.1197962

Wilson, G., Aruliah, D.A., Brown, C.T., Chue Hong, N.P., Davis, M., Guy, R.T., Haddock, S., Huff, K.D., Mitchell, I.M., Plumbley, M.D., Waugh, B., White, E.P., Wilson, P. 2014. Best Practices for scientific computing. *PLoS Biology* 12(1): e1001745. doi: 10.1371/journal.pbio.1001745

Wolkovich, E.M., Regetz, J., O'Connor, M.I. 2012. Advances in global change research require open science by individual researchers. *Global Change Biology* 18(7): 2102-2110, doi: 10.1111/j.1365-2486.2012.02693.x



Figura 1. Foto de los participantes en la sede del National Center for Ecological Analysis and Synthesis. Los autores en la fila inferior (segundo y cuarta a la izquierda). Foto: Ginger Gillquist.