

***Electronic laboratory notebooks (ELN)*, gestores del ciclo de vida de objetos de investigación**

Electronic laboratory notebooks (ELN): Managers of the life cycle of research objects

Fernanda Peset y Luis-Millán González-Moreno

Peset, Fernanda; González-Moreno, Luis-Millán (2016). "*Electronic laboratory notebooks (ELN)*, gestores del ciclo de vida de objetos de investigación". *Anuario ThinkEPI*, v. 10, pp. ¿¿-¿?

<http://dx.doi.org/10.3145/thinkepi.2016.59>

Publicado en *IweTel* el 11 de enero de 2016



Resumen: Se reflexiona sobre la naturaleza de la producción de conocimiento científico. Al tratarse de un proceso humano, sus características son especialmente complejas y difíciles de procedimentar. No obstante sus diferencias por disciplinas, tiene como común denominador la aplicación del método científico y la obtención de resultados en forma de datos y publicaciones. Se revisa el modelo de ciclo de vida de los objetos de investigación. Se discute que para entender los datos éstos tienen que estar contextualizados, algo extremadamente difícil si no se añade información en el momento de ser producidos. Para facilitar la gestión del conocimiento de forma integral en entornos académicos

nacen los *electronic laboratory notebooks* como *LabArchives*, *LabGuru* y *ROHub*. Se concluye la necesidad de implantar estos sistemas como forma de garantizar la preservación de los datos de investigación.

Palabras clave: Datos de investigación; Gestión del conocimiento; Software para laboratorios; Cuadernos electrónicos de laboratorio; Preservación; Ciclo de vida.

Abstract: This paper is a reflection on the nature of scientific knowledge production. Being a human process, its characteristics are particularly complex and difficult to establish with standard procedures. Despite the differences between disciplines, the research processes always includes the common application of the scientific method to obtain results in the form of data and publications. In this article the life cycle model of research objects is reviewed. To be understood, data need to be contextualized, something that is difficult if the necessary information is not added at the time they are produced. Electronic laboratory notebooks (ELN), such as *LabArchives*, *LabGuru*, and *ROHub*, have begun to appear in academic settings to facilitate comprehensive knowledge management. It is recommended that ELN systems are used to guarantee research data preservation.

Keywords: Research data; Knowledge management; Laboratory tools; Laboratory software; Electronic laboratory notebooks; Preservation; Life cycle.

Introducción

La investigación busca responder preguntas cuya solución se desconoce, es decir, procura reducir la incertidumbre, lo que la relaciona con el manejo intensivo y metódico de información. La capacidad de formular preguntas que pueden

ser contestadas nos diferencia de los animales y las máquinas. De hecho, "el conocimiento, o el proceso de buscar conocimiento, es una forma de juego; así ocurre por cierto con todos los hombres de ciencia e inventores que valen y que en verdad logran resultados valederos" (**Yutang, 1944**). La

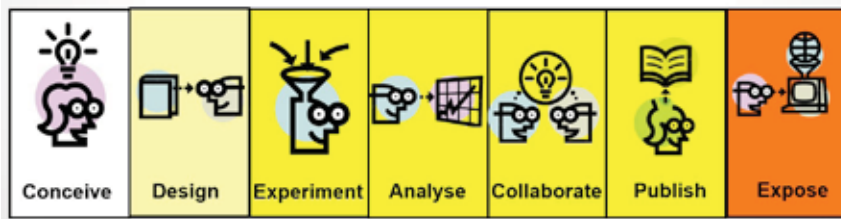


Figura 1. Proceso investigador (imagen de Anthony Beitz)
<http://www.slideshare.net/lrcrice/rice-recoderice2014>

investigación, así como la innovación, contienen un alto grado de creatividad, de curiosidad y de esfuerzo por interrelacionar diversos aspectos de la realidad.

“Si la atención a los datos de investigación se produce en su etapa final, la de publicación, su apertura será menos fluida y generalizada”

El proceso científico

La investigación tiene múltiples maneras de realizarse según las disciplinas, los paradigmas metodológicos escogidos, la cantidad de recursos necesarios o los sectores a los que se dirija. La ciencia está tan especializada que los métodos son tan infinitos como las cuestiones a contestar. Sin embargo, lo que se entiende por método científico no varía excesivamente. Consistiría a grandes rasgos en plantear un objeto de estudio, contar con el material y método adecuado para abordarlo, ejecutar la investigación y extraer conclusiones basadas en evidencias documentales, experimentales... Esto supone:

- proceso de creación (preguntas a responder);
- acopio de materiales (con sus variables a estudiar);
- producción de documentos en los que se registran las observaciones y datos que evidencian la investigación (datos de investigación);
- documentos donde se hace pública esa investigación (publicaciones o apariciones públicas).

De este proceso de registro

se producen los datos generados por las observaciones, que pueden ser de naturaleza cualitativa o derivadas del investigador, o cuantitativa, que registra él o un aparato. El tipo de documentos que genera pueden ser cuadernos de laboratorio, códigos metodológicos, manuales de estilo..., además de los resultados

finales que serían las publicaciones y conjuntos de datos (*datasets*).

Gray (2009) refrenda esta visión: «Obtenemos datos a partir de instrumentos, luego los calibramos y limpiamos, incluyendo llenar los vacíos según sea necesario. Reformateamos la información y, finalmente, la ponemos en una base de datos que publicamos en internet».

Pero “los datos crudos desvinculados de su contexto no pueden ser usados por cualquier investigador; simplemente son series numéricas sin sentido” (**Peset; González, 2016**). Para comprenderlos, tal y como apuntaba la *Royal Society (2012)*, se necesita información sobre las variables, la forma de obtención de datos, el software utilizado para el análisis y almacenamiento... Por tanto, aparatos, protocolos y métodos son los tres ejes que fijan una investigación y el valor de los datos que se derivan de ella. Documentar estos datos o *research objects (RO* en adelante)

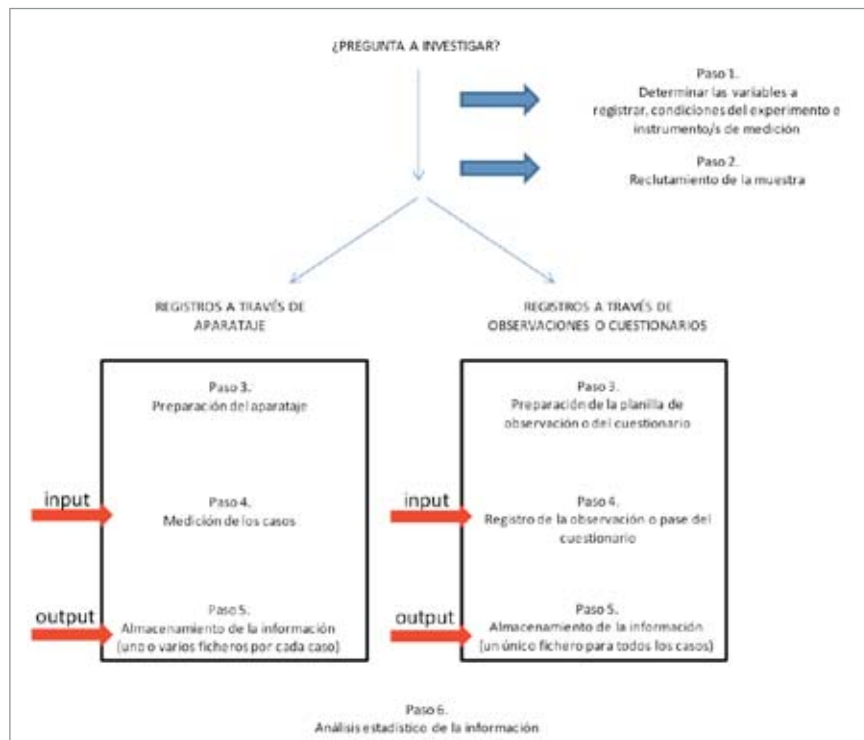


Figura 2. Esquema general de los pasos para conseguir datos crudos (*raw*) en un proyecto de investigación (**Peset; González, 2016**)

una vez finalizada la investigación es una tarea ardua, razón por la cual los laboratorios de investigación comienzan a utilizar programas de gestión del conocimiento. Éstos resultan de ayuda durante todo el ciclo de vida de los RO y quizá sean la única manera de “manejar lo inmanejable”. Estos sistemas integrales de control de toda la información suelen llamarse en este contexto *electronic laboratory notebooks (ELN en adelante)*.

Programas de gestión del conocimiento

Para gestionar correctamente todos los RO que se generan durante el transcurso de un proyecto de investigación existen numerosas aplicaciones para controlar y facilitar los procesos. En el caso de no utilizar un sistema integral de gestión, encontraremos diversas utilidades por la Red que permiten trabajar en colaboración.

Por ejemplo, para estructurar o anotar un proyecto contamos con *PBworks*, *Microsoft OneNote*

o *Evernote*. Sin embargo estas soluciones equivalen sólo a la etapa de inicio del trabajo de investigación, a modo de un cuaderno de laboratorio.

No son capaces de reflejar todos los objetos que se producen a lo largo de un proyecto (resultados de experimentos, *datasets*, etc.), los flujos o etapas por los que pasan los objetos (prototipos o publicaciones) ni su almacenamiento a largo plazo. Es decir, no gestionan todo el ciclo de vida de los RO.

Los *ELN* por tanto incluyen el almacenamiento, gestión del ciclo de vida y la preservación de los objetos de investigación (**Besemer, 2014**).

LabArchives y *LabGuru* son programas proporcionados por empresas privadas de servicios para laboratorios en la nube: *LabArchives LCC* y *Biodata* de *Digital Science* respectivamente. Permiten crear, almacenar, compartir y gestionar los datos de investigación de forma flexible y escalable.

<http://www.labarchives.com>
<http://www.labguru.com>

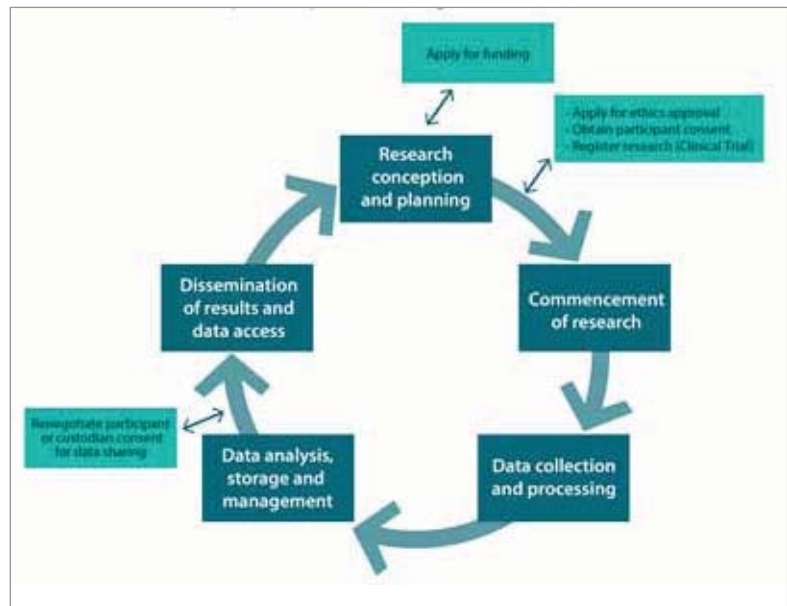


Figura 3. Datos producidos durante el ciclo de vida de una investigación
<http://www.nhmrc.gov.au/grants-funding/policy/nhmrc-statement-data-sharing>



Figura 4. *LabArchives* y *LabGuru*
<http://www.labarchives.com>
<http://www.labguru.com>



Figura 5. ROHub
<http://www.rohub.org/portallhome>

Otro ejemplo, *ROHub* se define como un prototipo de biblioteca digital para objetos de investigación. Especialmente facilita el trabajo de los científicos (creación de objetos predefinidos, *snapshots* y resúmenes para compartir en cualquier fase del desarrollo del objeto) y su preservación en el futuro. El modelo se ha construido en torno a una ontología que proporciona la estructura básica para la descripción de los recursos y sus anotaciones, además de unas extensiones para describir la evolución de los objetos (Palma et al., 2014).

<http://www.rohub.org/portallhome>

“Los ELN incluyen el almacenamiento, gestión del ciclo de vida y la preservación de los objetos de investigación”

Conclusión

Creemos que si la atención a los datos de investigación se produce en su etapa final, la de publicación, su apertura será menos fluida y generalizada. Sólo si los datos se gestionan desde su producción, estandarizando la forma en que se van a recopilar, producir y manejar durante los proyectos, será fácil abrirlos, cumpliendo con su vertiente de rendición de cuentas, ya que al fin y al cabo forman parte de la información del sector público.

Dado que los datos son dependientes de la disciplina en la que se producen tanto en su génesis como en su publicación y reutilización, involucrar a quienes los crean es la base del *Open Research Data Pilot* de H2020. Esto determina el peso de los planes de gestión de datos en la fase de diseño del experimento. Pero en este ámbito de trabajo un solo perfil profesional no puede garantizar el éxito que se augura a los datos como nuevo petróleo de la economía. Las organizaciones necesitan implantar sistemas capaces de gestionar el ciclo de vida completo de los RO para, en definitiva, facilitar la preservación y publicación de los datos que subyacen a las investigaciones.

Bibliografía

Besemer, Hugo (2014). Data: legal issues. *WGS data management planning course*.

<http://datamanagementplancourse.pbworks.com/w/page/52243666/Program20150512>

Gray, Jim (2009). “Jim Gray on eScience: A transformed scientific method”. En: Hey, Tony; Tansley, Stewart; Tolle, Kristin (eds.). *The fourth paradigm. Data-intensive scientific discovery*. Washington: Microsoft Research. ISBN: 978 0 9825442 0 4

http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf

Peset, Fernanda; González, Luis-Millán (2016). *Ciencia abierta y gestión de datos de investigación*. Gijón: TREA. ISBN: 978 84 9704 907 8

Palma, Raúl (2014). *ROHub demo*. <https://youtu.be/TxW2wvreyoQ>

Palma, Raúl; Holubowicz, Piotr; Corcho, Óscar; Gómez-Pérez, José-Manuel; Mazurek, Cezary (2014). “ROHub - A digital library of research objects supporting scientists towards reproducible science. En: *Semantic web evaluation challenge. Communications in computer and information science*, v. 475, n. 2014, pp. 77-82. http://dx.doi.org/10.1007/978-3-319-12024-9_9

Royal Society (2012). *Science as an open enterprise*. <http://royalsociety.org/policy/projects/science-public-enterprisereport>

Yutang, Lin (1944). *La importancia de vivir*. Apostrofel Pokhara. ISBN: 978 8435016674

Fernanda Peset
Universitat Politècnica de València
mpeset@upv.es

Luis-Millán González-Moreno
Universitat de València