

K.5. Hacia la gestión de datos de investigación en las universidades: la *Data asset framework*

Por **Daniel Torres-Salinas**

20 noviembre 2009

Torres-Salinas, Daniel. "Hacia la gestión de datos de investigación en las universidades: la *Data asset framework*". *Anuario ThinkEPI*, 2010, v. 4, pp. 262-265



Resumen: Las bibliotecas y otros organismos están empezando a valorar los datos de investigación como un producto académico. Esta situación está haciendo emerger nuevas competencias entre los profesionales de la información y especialmente en las bibliotecas universitarias. Ya se están lanzando las primeras propuestas, experimentando soluciones y explorando metodologías para iniciar la protección de tales datos. En esta nota explicaremos una de estas iniciativas llevadas a cabo por las bibliotecas universitarias del Reino Unido: la *Data asset framework*. Su objetivo es conocer mediante una auditoría las características y las condiciones de los datos generados por profesores e investigadores, y en función de los resultados la creación de un marco de trabajo que permita la puesta en marcha de políticas de sensibilización y preservación de los mismos.

Palabras clave: Datos científicos, Resultados de la investigación, Bibliotecas universitarias, Preservación digital, Iniciativa DAF.

Title: *Toward data management in university research: the Data asset framework initiative*

Abstract: Libraries and various agencies are beginning to value research data as an academic product. This situation is bringing out new skills among information professionals, especially in university libraries that are beginning to launch the first proposals, testing methodologies and exploring solutions to begin protecting our users' research data. We explain one of these initiatives undertaken by the university libraries in the UK, the *Data Asset Framework (DAF) initiative*. The objective of this project is to identify the characteristics and conditions of the data generated by professors and researchers and use the results to create a framework allowing the implementation of policies for the conservation and preservation of research data.

Keywords: Research data, Research outputs, Academic libraries, Digital preservation, DAF initiative, Data audit framework, Data asset framework.

1. Abriendo los cajones de los investigadores

EN EL TEXTO ANTERIOR hablamos del *data sharing* (Torres-Salinas, 2010) ofreciendo un brevísimo estado de la cuestión sobre el asunto y apuntando la preocupación común de agencias, instituciones, investigadores y documentalistas por el destino final de los datos en bruto utilizados en los trabajos de investigación.

Como comenta Borgman (2007), los datos están empezando a ser valorados como un producto académico. Podríamos decir que estamos en un momento de puesta en valor de los datos científicos. Concluíamos aquel texto apuntando que quizás las bibliotecas universitarias tendrían algo que decir y, en ese punto, me gustaría retomar el tema.

La custodia a cargo de las bibliotecas de los datos de investigación generados por nuestros centros es una cuestión que ya introdujo y desarrolló con acierto Martínez-Urbe (2008) del *Oxford e-Research Centre*. Uno de los aspectos que se deduce de este trabajo es el carácter emergente de estas nuevas competencias por lo que aún son escasas las bibliotecas que se hagan cargo de las mismas. Quizás el problema de muchas de ellas sea la falta de una hoja de ruta que ayude a sus bibliotecarios a afrontar o siquiera plantearse el problema. Sin embargo, para todos aquellos bibliotecarios que se digan "vale, los datos de mis usuarios/investigadores son importantes pero, ¿por dónde empiezo?" he de animarles diciéndoles que es una duda común en la mayor parte de las bibliotecas, incluso en las más punteras.

Justamente en el momento en el que nos

Data audit framework, <http://www.data-audit.eu/>

encontramos se están lanzando las primeras propuestas, experimentando soluciones y explorando métodos para iniciar este tipo de servicios.

“¿Qué está pasando con los datos de los investigadores?”

En octubre de 2009 concluyó una iniciativa que supone ese primer paso hacia el cambio (Jones, 2008a), la *Data asset framework (DAF)*, llamada al principio *Data audit framework*. En ella no se plantean aún empezar a crear repositorios de datos sino algo previo y de momento más fácil: ver qué guardan los investigadores en sus cajones, y después con tranquilidad, ver cómo se puede ayudarles a poner orden en los mismos.

2. Auditorías de datos y bibliotecas universitarias

No voy a contar aquí en detalle el proyecto *DAF* ya que toda la información sobre el mismo está perfectamente organizada en su web¹, pero sí esbozaré de qué va la historia.

Nació de una necesidad apuntada por Liz Lyon (2007) en un informe encargado por el *Join*

Information Systems Committee (Jisc), del Reino Unido, y era la creación de un marco de trabajo que permitiera a las universidades conocer qué datos son generados por sus departamentos y poner en marcha políticas de sensibilización y preservación de los mismos.

Ante esta necesidad el mismo *Jisc* financia el proyecto *DAF*, liderado por la *Universidad de Glasgow*, para desarrollar dicha metodología entre 2008 y 2009; el pasado octubre se presentó la guía final (*Jisc*, 2009). En sus inicios se plantearon los siguientes objetivos básicos:

- Ver dónde se generan los datos, de qué tipo son, cómo se almacenan, gestionan y comparten.
- Identificar problemas: pérdidas, mal uso, localizar datos en riesgo, formatos desfasados.
- Conocer las actitudes de los investigadores a la hora de crear y compartir datos.
- Sugerir mejoras para la gestión de datos.

Para llevar a cabo estos objetivos de partida diseñaron una auditoría que consta de cuatro pasos (Jones, 2008b): planificar la auditoría, identificar y clasificar los datos, evaluar la importancia de los datos localizados, y finalmente ofrecer unas recomendaciones para implementar una gestión adecuada.

Como en toda auditoría, la información se obtiene de gestores e investigadores a través de encuestas, entrevistas o incluso revisando los *papers* producidos por los centros.

Para testear esta metodología se realizaron diferentes auditorías en algunas universidades (Edimburgo, Bath, Glasgow, King's College, Southampton y Oxford) en centros muy concretos con líneas temáticas diversas (geociencias, arqueología, humanidades, etc.). Al final cada universidad presentó un informe con los resultados. En ellos se explica cómo seguir mejorando la metodología pero quizás lo más interesante son algunas de las conclusiones que se apuntan y que responden a preguntas como ¿qué está pasando con los datos de los investigadores? y ¿qué podemos hacer desde la biblioteca para empezar a ayudarles?

cómo conservar sus datos y que éstos suelen estar dispersos por cds, dvds, discos duros externos, pcs (trabajo y domésticos). Esto provoca pérdidas irrecuperables y en ocasiones poner en marcha investigaciones paralelas para localizarlos, identificar la última versión o encontrar a la persona responsable (que a veces ya no trabaja en el mismo centro).

“Los investigadores gestionan sus datos de forma casera y fragmentada”

3. Respuesta a las preguntas y apuntes finales

Cualquiera que esté un poco introducido en el mundo de la investigación sabe qué ocurre con los datos, pero es mejor que los bibliotecarios que han abierto esos cajones con sus auditorías nos lo cuenten (Jones et al., 2008c).

En primer lugar, revelan que no existe una política clara entre los investigadores acerca de

Además los investigadores no utilizan ninguna convención para denominar a sus archivos, ni crean catálogos de las colecciones de datos y por supuesto tampoco tienen en cuenta su preservación a largo plazo; y aunque algunos se lo plantean no tienen las herramientas y conocimientos para atajar el problema.

Se puede decir que en la actualidad los investigadores gestionan sus datos de forma casera

Could you get more from your research data?

Data Audit Framework

What is it?
The Data Audit Framework is a mechanism to establish an overview of data holdings and the policies and practices in place to manage them to help organisations improve on existing data management.

Who is it for?
DAF can help anyone who wants to be in greater control of their data. Have you ever struggled to find data, had difficulties accessing and understanding other people's work, encountered version control issues, or suffered from data loss or irretrievability? If so a data audit could help.

How do I use it?
DAF is a self-audit methodology, designed to be used without specialist skills. Four incremental steps are followed to collect information on the organisation's data assets and management practices.

1. Complete the Audit
2. Assess the Current State
3. Develop a Plan
4. Review and Implement

The audit generates two key outputs: an inventory of data assets, divided into groups according to their value for the organisation (Stage 2); and a final report with recommendations on how data management could be improved (Stage 4).

Why is it useful?
Auditing data brings several benefits. Data assets can be identified and monitored to avoid loss and corruption. Duplication of effort becomes apparent so resources can be better utilised. Audits facilitate capacity planning, compliance, and help organisations realise the value of their data assets through improved access and reuse. Knowledge is the cornerstone of effective data management. DAF audits place organisations in a position of knowledge.

Where do I find out more?
www.data-audit.eu

Four steps to effective data management

JISC | **DCC** Digital Curation Centre | **University of Glasgow** Humanities Advanced Technology & Information Institute

The Data Audit Framework is funded under the JISC Repositories Programme. Development has been led by HATI at the University of Glasgow with support from partners at the Universities of Edinburgh and Bath (UKOLN), King's College, Imperial College and University College London. The collaboration was made possible through the Digital Curation Centre which has acted as an umbrella for this work.

y fragmentada. Por ello los informes de la DAF plantean algunas acciones básicas:

1. Implementar políticas institucionales para la creación y conservación de datos y acompañar éstas de guías para favorecer buenas prácticas.

Éstas versarían sobre cómo gestionar los diferentes tipos de datos, cómo realizar copias a largo plazo o cómo crear planes de gestión de datos; en fin, toda una serie de cuestiones que están bien ejemplificadas en la guía realizada por la *Australian National University (ANU, 2008)* dentro de su programa de alfabetización informacional y que nos da una idea bastante clara de por dónde van los tiros.

2. Formar a los nuevos investigadores en técnicas de gestión de datos, es decir crear pequeños bibliotecarios *amateur*. Todo esto no es sino el paso inicial hacia la creación de repositorios institucionales, no necesariamente abiertos, para la conservación de los datos de nuestros investigadores.

“Junto a una investigación de excelencia debe fomentarse también una gestión de los datos de excelencia”

Está claro que esta auditoría, de muy bajo coste para las bibliotecas porque no requiere personal especializado, está aún en un estado muy embrionario pero he querido compartirla aquí porque me parece muy ilustrativa y bastante realista de ese primer paso hacia la custodia de los datos por parte de las bibliotecas.

Y efectivamente, después llegarán los repositorios y familiarizarnos más profundamente con términos de la e-ciencia y de las e-infraestructuras, pero primero hay que conocer qué ocurre en los despachos de nuestros centros y aprovechar la oportunidad de participar más activamente en los problemas diarios de los investigadores.

Además, justamente ahora que se quiere y se busca la excelencia, hay que ser conscientes de que junto a una investigación de excelencia debe

fomentarse también una gestión de los datos de excelencia, que permita su acceso óptimo y garantice su reutilización futura. Creemos que la iniciativa DAF es un buen ejemplo a imitar en la universidad española.

Nota

1. <http://www.data-audit.eu/index.html>

Referencias

Australian National University (ANU). *Data management manual: managing digital research data at the Australian National University*, 2008. http://ilp.anu.edu.au/dm/ANU_DM_Manual_v1.03.pdf

Borgman, Christine L. *Scholarship in the digital age: information, infrastructure, and the Internet*. Cambridge, Mass: MIT Press, 2007.

Jisc. *Data asset framework: Implementation guide*. 2009. http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf

Jones, Sarah. “The data audit framework: A first step in the data management challenge”. *The international journal of data curation*, 2008, v. 2, n. 3, pp. 112-120.

Jones, Sarah; Ross, Seamus; Raivo, Ruusalepp. *The data audit framework: a toolkit to identify research assets and improve data management in research led institutions*. 2008a. <http://eprints.gla.ac.uk/6240/>

Jones, Sarah; Ekmekcioglu, Cuna; Ball, Alex; Grace, Stephen. *Data audit framework lessons learned*. 2008b. http://www.data-audit.eu/docs/DAF_lessons_learned.pdf

Lyon, Liz. *Dealing with data: roles, rights, responsibilities and relationships, Consultancy report*, 2007. http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing_with_data_report-final.pdf

Martínez-Urbe, Luis; Macdonald, Stuart. “Un nuevo cometido para los bibliotecarios académicos: data curation”. *El profesional de la información*, 2008, v. 17, n. 3, mayo-junio, pp. 273-280.

Torres-Salinas, Daniel. “Compartir datos (data sharing) en ciencia: el contexto de una oportunidad”. *Anuario ThinkEPI*, 2010, vol. 4, pp. 258-261. <http://www.thinkepi.net/compartir-datos-data-sharing-en-ciencia-el-contexto-de-una-oportunidad>