Routledge
Taylor & Francis Group

Records management for scientific data

Charlotte Maday* and Magalie Moysan

Charlotte Maday is Paris Diderot University's archivist and records manager for faculties and administrative services. With a Masters in History and Archives (2005), she first worked for the National Computing Centre for Higher Education and Research (CINES), where she contributed to the creation of the long-term preservation platform for higher education and research data. She is also President of the Section on Universities and Research Organisations of the French Association of Archivists (AAF/AURORE), and an ISO/AFNOR expert for records management.

Magalie Moysan is Paris Diderot University's archivist and records manager for laboratories and researchers. With a Masters in History and Archives (2009), she first worked in various administrations, including the French Department of Defence, a city hall and a private enterprise. She also coordinates the working group 'Scientific Records' of the French Association of Archivists.

Though this be madness, yet there is method in't. (*Hamlet*, William Shakespeare)

Venturing into research and science, for a records manager, is like walking into a den belonging to a very grumpy and very hungry lion. Firstly because we do not entirely comprehend 'scientific data': the OECD definition[1] is probably the most complete, but excludes some types of records that are more specific to research project management than research itself.

Also, considering the French definition of archives[2] we might assume that scientific data is, regardless of its shape, format or degree of evidence:

- that **produced** by researchers themselves, in order to prove ideas or theories;
- but also that **collected** by a researcher, a team or a laboratory, contributing to scientific arguments and aimed at providing conclusive evidence of ideas: from grey literature to raw data from satellites, this could be anything and everything.

Uncertainty about the nature and extent of those records is also complicated by the wide variety of formats, which raises practical questions related to appraisal and selection of information: do we count raw data in astronomy? Or computing products of endless algorithms? Or even 3D prints of stem cells entirely designed in laboratories?

There is also an increasing variety of actors dedicated to scientific projects: first, researchers have their own idea of what they need, and what they do and do not want to provide access to. The impacts of this on how data is identified, filed, described and indexed are immense. Records managers also have to deal with an oh-so-classic but always existing lack of comprehension of their area of expertise by researchers and/or academics who do not seem to understand the purpose of managing and preserving data

*Corresponding author. Email: charlotte.maday@univ-paris-diderot.fr

during and after the end of research projects. Preservation responsibilities are also often split, with libraries responsible for much preservation and access before archivists had the opportunity to become involved.

Large changes in the area of research data management are currently occurring and these could be used by the records and archives profession to help ensure better long-term management of research data. In recent years, editors of scientific journals have created a monopoly on the publication of scientific research. Scientists need to publish to gain validation of and recognition for their research.[3] The publication model levies fees from both the authors and readers of research. There is increasing reaction against this, with many research communities looking to establish free, open access publication models. To combat this, editors of research journals have tried to protect their business model by offering enhanced services, one of which is to collect and provide paid access to scientific data associated with published papers. But when it comes to publicly funded research, this amounts to privatised access to public records and endangers efficient open data and information policies. Through various mechanisms, therefore, the chain of proof for scientific research, represented by the publication system, is progressively collapsing, yet there is no reliable methodology for managing and preserving research, no data management requirements exist in national calls for projects and no credit line is kept in research projects even for basic data storage. All this leads to two fundamental problems:

(1) research data is subject to various competing forms of access;
(2) in the process of creating research data, trust, reliability and usability of the data are central, but it is challenging to deploy records management principles to support these critical requirements.

Records management or core archival principles such as respect for provenance are really put into question when applied to scientific data. But does that mean records management principles are irrevocably overwhelmed when it comes to science?

We would argue it is not the professional practice that is obsolete, it is the implementation of it, and when it comes to scientific data, the field progressively left by editors is wide open and a great opportunity. An experience we can consider to progressively use records management principles to support the management of research data could come from 'Horizon 2020', the new European framework program for research and innovation for 2014–2020.

According to the European Commission, 'a data management plan is a document outlining how the research data collected or generated will be handled during a research project, and after it is completed, describing what data will be collected/generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved'.[4] Data management plans are not part of French habits. In France, as in many countries, expressions such as 'open data' or 'big data' are very popular, but very few institutions have a policy of data management.[5]

Horizon 2020 represents a great opportunity to implement not only a policy but also a huge change in the habits and mentality of French researchers. A limited pilot action on open access to research data has been implemented and participating projects are required to develop a data management plan. For example, all project proposals submitted to 'Research and Innovation actions' or 'Innovation actions' under Horizon 2020 must include a section on research data management.[6] The scope of this program is still

small – projects may opt out of the pilot for a wide variety of reasons – but it reflects a timid move towards open data and data management. For the first time in France, researchers are obliged to consider the conservation of and access to their data at the beginning of their project.

Up until now, French archivists have classified scientific records on the retirement or the death of a researcher. This method was relevant when a researcher used to work on the same topics in the same institutions during his or her entire career. Today, research is based on projects, facilitating a much more flexible process of records management. The drafting of data management plans is an important evolution for French archivists, because researchers have to understand the data lifecycle and how this will impact on the project's specific data. Archivists can then provide help in establishing themselves as data management specialists. By defining how data is produced, they can control metadata and format and have an influence on interoperability. They can also make sure that records and data will be preserved on appropriate electronic preservation platforms.

For the first time, archivists might succeed in implementing appraisal for research data. Today, they fail in imposing selection criteria because they usually receive the archives long after creation and time has already made a 'natural selection'. Being associated from the beginning of a project will help to impose a strategic policy for data selection.

Above all, responsibilities for managing records and data will evolve. Nowadays, very few researchers feel concerned about records. They usually keep their data until they no longer need it, then they transfer responsibility for it to the archivist. Thanks to data management plans, responsibilities will be more balanced. The archivist will have to ensure data preservation and the researcher will have to make sure that his or her scientific choices are compatible with preservation and access.

Endnotes

1. OECD, *Principles and Guidelines for Access to Research Data from Public Funding*, 2007.
2. In the French law on archives (Loi 2008-696 relative aux archives, codifiée au Code du Patrimoine), the word *archives* refers to both records and archives.
3. The publication model was created to avoid theft of ideas and ensure authenticity and reliability of research, and ease of access to facilitate peer review.
4. European Commission, 'Guidelines on Open Access to Scientific Publications and Research Data', in *Horizon 2020*, Version 1.0, December 2013, p. 10.
5. The French National Institute of Agronomic Research (INRA) was one of the first national institutions to write a data share policy: Institut national de recherche agronomique, 'Rapport du groupe de travail sur la gestion et le partage des données', June 2012, available at <*http://www.pfl-cepia.inra.fr/uploads/gdp_docs/Rapport-GestionDonnees-web.pdf*>, accessed March 2014.
6. European Commission, 'Guidelines on Data Management', in *Horizon 2020*, p. 2.