

## The time has come for managing and sharing research data in universities

---

**Ignasi Labastida**

### **Abstract**

During the last decade universities have developed policies and infrastructures to support open access to publications but now it is time to move a step forward. There is an increasing demand for accessing data supporting the research results to validate and reproduce them. Therefore universities have to be prepared for this new challenge that goes beyond dissemination because it requires a strategy for managing research data within institutions. In this paper I will try to give some hints on how to deal with this challenge that can be framed in the new open science movement aimed at providing openness in all the whole cycle of research.

### **Keywords**

Scholarly communication

---

### **Introduction**

Currently we can state that the dissemination of the research outputs has changed radically from the last century. It has not been just the shift of support facilitated by the technology, but a set of different changes. On one hand we see a change on the typology of outputs researchers publish and share. Besides the text explaining the results achieved during a research, now we can access to data or code created or used. Moreover, in some case, this access is provided during the research process and not at the end, once the results are achieved. This new way of sharing research activities receives the name of open science<sup>1</sup> and it helps to change the idea of researchers locked in laboratories without sharing any discovery unless it is published in a paywall publication not affordable or accessible to all. Open science is an opportunity to engage research and society by offering more transparency and accessibility but it also represents a challenge for a research institution, as universities, to give an answer to researchers that want to engaged in those activities. It is more an issue of facilitating and rewarding open science than an issue of requiring to do research openly.

---

<sup>1</sup>Between July and September 2014 the European Commission conducted a public consultation on Science 2.0. The validation results are available at: [https://scienceintransition.files.wordpress.com/2014/10/science\\_2\\_0\\_final\\_report.pdf](https://scienceintransition.files.wordpress.com/2014/10/science_2_0_final_report.pdf). The main page is available at: <http://scienceintransition.eu/>.

And here there is the key term: open. Openness has been the key drive to change many of the current scholarly activities. Although there could be many meanings for the term openness I would like to follow the one used by the open definition<sup>2</sup>

“Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).”

Therefore open is not just a question of gratuity but a question of reusability.

## Sharing content in the academia

The concept of openness, as defined above, seems to match to any scholar activity. However, not all the contents produced in the academia are open in that way. For instance, a huge amount of scientific literature is still behind a paywall and some teaching materials are public but they are not published under a license allowing reuse. Nevertheless, the number of materials that are provided openly is growing everyday.

If we look back, probably the first area where academia released open material was the software. Scholars share code among colleagues and build applications out of it. The use of licenses to allow software reusability is well spread in the academia and there are licenses named after a research institution (for instance BSD or MIT licenses). Universities not only develop and create free or open source software, they are users of it when they adopt tools for their activities: virtual campuses, institutional repositories, databases. . .

From software we moved to publications. At the beginning of this century it appeared a movement advocating for a change in the scholarly publication. This movement received the name of open access because it was aimed at facilitating unrestricted online access to the peer-reviewed journal literature.<sup>3</sup> By “open access” the movement meant free online availability to this literature without any legal, technological and economic barriers. It is not just a question of gratuity but a question of reusability.

To reach that final goal, the proponents of this initiative offered two strategies that nowadays have become the two categories of the open access. The first strategy was to build open electronic archives to host copies of any paper published by scholars. Those archives are known as repositories and this strategy has received the name of green or gratis open access because there is no paywall but in many cases there are still restrictions to reuse those copies. The second strategy, known as gold or libre open access, advocated for the creation of a new generation of journals providing unrestricted access and reusability subject to proper attribution of authorship.

In general, universities have adopted both strategies. On one hand, universities have built their own institutional repositories and offer services to support the first strategy, and on the other hand many universities host open access journals or support them by providing funds for publishing or by joining membership

<sup>2</sup>Open definition is available in different languages at: <http://opendefinition.org/>.

<sup>3</sup>Budapests Open Access Initiative (BOAI), <http://www.budapestopenaccessinitiative.org/>.

programs.<sup>4</sup> According to a study funded by the European Commission and undertaken by Science-Metrix, a research evaluation consultancy, in 2013<sup>5</sup> open access has reached a tipping point: more than 50% of papers published in 2011 are now available for free. Those papers are available in repositories, or in their own journals.

Obviously, the concept of openness reached other activities within the academia. At the same time the open access was growing, a new movement appeared on education and teaching with the rise of the open educational resources (OER) and the open courses mainly led by the MIT with its Open Courseware (OCW) that were adopted by many universities around the world.<sup>6</sup> As an evolution of the OCW we have now the current Massive Open Online Courses (MOOC) [‘From OCW to MOOC: Deployment of OERs in a Massive Open Online Course. The Experience of Universidad Carlos III de Madrid (UC3M)’] where the term open has different meanings and it is mainly used to state that there is a free enrolment to the course and not to indicate an unrestricted reuse of the educational materials provided.

Therefore scholars have been sharing papers, code, and teaching materials not just with colleagues but with the society in general, and universities have been providing them with the infrastructure and the services needed. But why do they share all those contents? There are many answers for this question, but it is important to take into account that initially it was done voluntarily. Currently, in some cases, sharing has become mandatory. This requirement is especially strong for publications derived from research funded by a public administration, as we will see in the next section.

## Policies requiring openness

As a consequence of the open access movement, research funders started to ask to their grantees to share their results beyond the traditional paywall journals. Initially, some funders recommended to post copies of published papers in repositories but in 2006 Wellcome Trust adopted a stronger policy by requiring it [Walport and R., 2006]. That policy was an example that has been followed by other research funders around the world becoming a common practice among them. It is important to notice that public research funders have adopted this mandatory approach because beneath this kind of policy there is the idea to return to the public what is funded with public money.

Besides funders, many research institutions have also adopted its own policy following the Queensland University of Technology (QUT) that was the first university to establish an institutional self-archiving policy in 2004.<sup>7</sup> Although there are differences among these institutional policies because some are

<sup>4</sup>There are some lists of funds available at the following addresses:

[http://oad.simmons.edu/oadwiki/OA\\_journal\\_funds](http://oad.simmons.edu/oadwiki/OA_journal_funds), <http://www.sparc.arl.org/sites/default/files/OA%20Funds%20in%20Action%20attachment%202014%20%281%29.pdf>.

<sup>5</sup>Press release EC Open access to research publications reaching ‘tipping point’, August 2013, [http://europa.eu/rapid/press-release\\_IP-13-786\\_en.htm](http://europa.eu/rapid/press-release_IP-13-786_en.htm).

<sup>6</sup>The list of Open Education Consortium members is available at: <http://www.oeconsortium.org/members/>.

<sup>7</sup>There is an international registry of open access mandates and policies adopted by universities, research institutions and research funders called ROARMAP available at: <http://roarmap.eprints.org/>.

mandatory and others are just a recommendation, all of them are aimed at populating institutional repositories.

Therefore researchers are currently required to follow one or more self-archiving policy and as a consequence to post a copy of their peer-reviewed papers in a repository accessible to the public.

However, research outputs cannot be reduced to publications and for that reason funders, institutions and even publishers are modifying their policies. Researchers are now asked to publish and share other outputs than papers, for instance data. This new requirement is aimed at facilitating the validation and reproducibility of the results.

Currently many journals have adopted a data sharing policy.<sup>8</sup> In certain cases, researchers are not required to deposit data in a specific repository and therefore they ask their own institution to provide this service. When the amount of data is small, universities offer their institutional repositories but they can reach a point where they could not meet the demands.

Moreover, in some countries like the UK, the requirement to share data comes from research funders like the RCUK<sup>9</sup> and it also includes the need to develop a data management plan. This plan allows researchers to think how they are going to manage the data they will work with during their research. The requirement of a data management plan is also included in the new pilot developed by the European Commission in its current Horizon 2020 research program [European Commission, 2013]. The pilot is addressed to seven areas although any project can voluntarily join it. Obviously there are some cases where researchers can opt out, for instance if there are conflicts with security or with personal data. The first requirement of the pilot is to create a data management plan that has to be delivered during the first six months of the project. This plan can be changed along the project if it is needed. In the management plan, researchers must state how they are going to disseminate the data they will obtain by experiments, observations or simulations. The pilot advise them to release them under a non restrictive license and to public them in suitable repositories. In fact, the main target of the pilot is the data needed to validate scientific publications.

All those policies requires that institutions provide researchers with a set of services to help them in managing data and moreover institutions must value their own data. Nevertheless institutions have to overcome a first barrier: researchers' concerns about sharing their data.

## Concerns about sharing data

As it has been said, data is a valuable asset in research. For that reason, researchers are really concerned about sharing data. There are some fields where data sharing is a common activity, for instance genetics or astronomy, but there are other fields where there is a strong reluctance to do so.

---

<sup>8</sup>The library of the MIT provides a list of those journals, <http://libraries.mit.edu/data-management/share/journal-requirements/>.

<sup>9</sup>The RCUK Common Principles on Data Policy are available at: <http://www.rcuk.ac.uk/research/datapolicy/>.

Policies help to get over this reluctance but it is important also to convince researchers that sharing will help in the discovering of knowledge. Policies must protect the acknowledgement of the first who obtains or observes data and therefore they must give a period of exclusivity for them. Probably it does not make sense to require the release of raw data immediately. Currently most of the policies are asking to publish data related to a publication. It means that data must be released with the publication and not before.

Besides policies there is a need for an institution to establish a culture of sharing but overcoming all those fears and concerns. Some of them were gathered in the “Open Data Excuse” Bingo.<sup>10</sup>

Therefore an institution must start to face how to deal with research data once some of those concerns are addressed. They must provide more than a simple place to publish data, they have to acknowledge the value of research data and therefore develop a strategy to secure the data in all the whole research cycle. In this next section I will try to give some hints.

### Proposed action

Many institutions have already established a policy or a strategy on research data, especially in the UK. However, there are still a lot of institutions without it. It is not just a decision about where to place research data to be shared with society but to acknowledge the value of data. Many institutions are offering its institutional repositories to researchers to publish data but there is a need to start a broader debate about what it must be done with the data produced within a research institution.

A good way to start is to identify the different stakeholders within the institution. Generally, the library can play the leading role due to its experience with open access but it needs to engage other strategical units within the university as, for instance the research and technical offices. Once the different stakeholders are identified we can start the conversation, as suggested by an interesting document published in 2013 by the OCLC [Erway, 2013]. This document emphasizes the importance of involving all the key actors in the research data management. There are other documents that give some recommendations to universities in order to establish a policy or a strategy. Among them we can mention the LERU Roadmap for Research Data [LERU, 2013], published at the end of 2013, and the set of policy recommendations issued this year by RECODE, a project funded by the European Commission [Tsoukala et al., 2015]. Both documents can be used as a guide to follow and they show also some best practices.

Therefore the first step to be taken by the institution, once the different stakeholders are identified, is to establish a working group or steering committee to draw a roadmap aimed to have a strategy on research data management. This leading group can then follow some of the recommendations from the different documents mentioned before.

The set of recommendations can also be used to make an initial picture of where the institution is placed. Initial questions to be asked are which services are we offering

---

<sup>10</sup>As an example someone can check: <http://ves.cat/mf58>. The template is available at: <http://data.dev8d.org/devbingo/>.

to researchers, who can answer questions about this topic, which is the infrastructure we provide, or even legal questions. From this initial picture the group or committee can establish the final goal to achieve for the institution.

For instance, a policy can be useful to establish the different roles of each individual in the institution defining duties and rights and at the same time assuming all the different responsibilities. As an example, we can see the research data policy at UCL where you can find all the roles from students to the provost [Ayriss, 2013].

Another important issue to be tackled is the selection of data and the time of preservation. Universities must establish a criteria on those issues. Obviously not all the data must be preserved and some selection has to be done. During the research cycle universities can offer to researchers some facilities to use, to store and to share data with colleagues but at one point some of that data must be preserved and some has to be erased.

Another important thing to be taken into account in the selection process is which data must be published or offered to the public. Especially now when many journals are asking researchers to provide links to the data they reference in the papers submitted for publication. All that data must be identified and described, and again some services of the university, as the library can play an important role here supporting researchers.

And of course we cannot forget the issue of the infrastructure. Institutions must decide which will be their model when providing researchers with a solution. Basically, institutions can choose between building their own infrastructure for data, including specific repositories, or use the services that are offering externally. In the first group we can see universities as UCL<sup>11</sup> or Edinburgh,<sup>12</sup> and in the second group universities like Monash<sup>13</sup> or Loughborough.<sup>14</sup> Another choice is to build alliances with other universities and work together in a common infrastructure. This later model is the one used in the Netherlands through the Data and Archiving Networked Services (DANS).<sup>15</sup>

Another important topic that an institution cannot forget is the legal status of data. Although in many cases data cannot be protected by copyright, databases can be and institutions must have a clear policy in this issue. To whom belongs data, if they belong to someone? [Carroll, 2015].

And finally, at the same time a university is deciding about which is the best infrastructure, and it is shaping a suitable policy, it must train people to go along with researchers in all the process of management. It is important to build new roles

---

<sup>11</sup>UCL Research Data Services,  
<https://www.ucl.ac.uk/research-it-services/about/research-data-service>.

<sup>12</sup>Data Library and consultancy,  
<http://www.ed.ac.uk/information-services/research-support/data-library>.

<sup>13</sup>Monash University and figshare partner to combine cloud management and discoverability with institutional storage, [http://figshare.com/blog/Monash\\_University\\_and\\_figshare\\_partner\\_to\\_combine\\_cloud\\_management\\_and\\_discoverability/137](http://figshare.com/blog/Monash_University_and_figshare_partner_to_combine_cloud_management_and_discoverability/137).

<sup>14</sup>Loughborough University, figshare, Arkivum and Symplectic announce pioneering research data management solution, [http://figshare.com/blog/Loughborough\\_University\\_figshare\\_Arkivum\\_and\\_Symplectic\\_announce\\_pioneering\\_rese/136](http://figshare.com/blog/Loughborough_University_figshare_Arkivum_and_Symplectic_announce_pioneering_rese/136).

<sup>15</sup><http://www.dans.knaw.nl/>

within different units to deal with it. And once those roles are created it is important to train researchers, too: from doctoral students to senior researchers. The training process can go along with a marketing campaign by means of specific webpages or any other format to disseminate the new services to manage research data.

## Conclusions

There are no more excuses to start working on developing a roadmap to implement a global research data management plan within your institution. Not all the data has to be stored and preserved, neither all data have to be shared but we have to be ready to answer many questions and to decide which data we will keep, for how long, and how it will be shared not just with colleagues but with all the society. Let's start working!

## References

- Ayris, P. (2013). *UCL Research Data Policy*. URL: <https://www.ucl.ac.uk/isd/services/research-it/documents/uclresearchdatapolicy.pdf>.
- Carroll, M. W. (2015). 'Sharing Research Data and Intellectual Property Law: A Primer'. *PLoS Biol* 13 (8), e1002235. DOI: [10.1371/journal.pbio.1002235](https://doi.org/10.1371/journal.pbio.1002235).
- Erway, R. (2013). *Starting the Conversation: University-wide Research Data Management Policy*. URL: <http://www.oclc.org/content/dam/research/publications/library/2013/2013-08.pdf>.
- European Commission (December 2013). *Guidelines on Open Access to Scientific Publications and Research*. URL: [http://ec.europa.eu/research/participants/data/ref/h2020/grants/\\_manual/hi/oa/\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants/_manual/hi/oa/_pilot/h2020-hi-oa-pilot-guide_en.pdf).
- League of European Research Universities (2013). *LERU Roadmap for Research Data*. URL: [http://www.leru.org/files/publications/AP14\\_LERU\\_Roadmap\\_for\\_Research\\_data\\_final.pdf](http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf).
- Tsoukala, V., Angelaki, M., Kalaitzi, V., Wessels, B., Price, L., Taylor, M. J., Smallwood, R., Linde, P., Sondervan, J., Reilly, S., Noorman, M., Wyatt, S., Bigagli, L., Finn, R., Sveinsdottir, T. and Wadhwa, K. (2015). *RECODE Recommendations*. URL: [http://recodeproject.eu/wp-content/uploads/2015/02/RECODE-D5.1-POLICY-RECOMMENDATIONS\\_FINAL.pdf](http://recodeproject.eu/wp-content/uploads/2015/02/RECODE-D5.1-POLICY-RECOMMENDATIONS_FINAL.pdf).
- Vida Fernández, J. and Webster, S. 'From OCW to MOOC: Deployment of OERs in a Massive Open Online Course. The Experience of Universidad Carlos III de Madrid (UC3M)'. *Open Praxis* 6 (2), pp. 145–158. URL: <http://openpraxis.org/index.php/OpenPraxis/article/view/137>.
- Walport, M. and R., K. (2006). 'Open access, UK PubMed Central and the Wellcome Trust'. *J. R. Soc. Med.* 99 (9), pp. 438–439. DOI: [10.1258/jrsm.99.9.438](https://doi.org/10.1258/jrsm.99.9.438).

## Author

Ignasi Labastida, Ph.D. in Physics from Universitat de Barcelona (UB). Head of the Office for the Dissemination of Knowledge and the Research Unit at the CRAI (Library) in the UB. Member of the Steering Committee of the LERU (League of European Research Universities) Community of Chief Information Officers and member of its working groups on Open Access and Research Data. Member of the Working Group on copyright at LIBER (Association of European Research Libraries). Public Leader of Creative Commons in Spain. He has publications on copyright, open knowledge, and research dissemination. He has trained library staff, researchers and doctoral students in copyright, open access, and research data. E-mail: [ilabastida@ub.edu](mailto:ilabastida@ub.edu).

## How to cite

Labastida, I. (2015). 'The time has come for managing and sharing research data in universities'. *JCOM* 14 (04), C03.



This article is licensed under the terms of the Creative Commons Attribution - NonCommercial - NoDerivativeWorks 4.0 License.  
ISSN 1824 – 2049. Published by SISSA Medialab. <http://jcom.sissa.it/>.