

From Open Access to Open Data: Collaborative Work in the University Libraries of Catalonia

Mireia Alcalá Ponce de León

Àrea de Ciència Oberta, Consorci de Serveis Universitaris de Catalunya,
Barcelona, Spain

mireia.alcala@csuc.cat, orcid.org/0000-0003-2684-2110

Lluís Anglada i de Ferrer

Àrea de Ciència Oberta, Consorci de Serveis Universitaris de Catalunya,
Barcelona, Spain

lluis.anglada@csuc.cat, orcid.org/0000-0002-6384-4927

Abstract

In the last years, the scientific community and funding bodies have paid attention to data collected, generated or used in different research activities, because the dissemination of these data can be seen as a constituent of Open Science. For this reason, many funders are requiring or promoting the development of Data Management Plans, and depositing open data following the FAIR (Findable, Accessible, Interoperable and Reusable) data management principles.

Libraries and research offices of Catalan universities that work in a coordinated way within the Open Science Area of the University Services Consortium of Catalonia offer support services to research data management. The different activities carried out at the Consortium level, as well as the implementation of the service in each university are presented.

Keywords: research data management; university libraries; research support; FAIR data; open science.

1. Introduction

The traditional mechanisms for communicating research results have recently undergone profound changes as a result of the open access movement (Abadal, Ollé, & Redondo, 2018), which makes scientific information freely accessible and reusable. The movement began with research articles, but has recently been extended to other documents and now includes research data.

The scientific community and funding bodies are now focusing on the data gathered, generated or used during research activities, and public access to these data is one of the constituent elements of what is called open science (Anglada & Abadal, 2018). At the international level, the European Commission's Horizon 2020 (H2020) programme requires research data to be openly accessible and includes the obligation to draw up management plans in accordance with the FAIR principles: Findable, Accessible, Interoperable and Reusable (Wilkinson et al., 2016). In Spain, however, this requirement is only a recommendation of the new 2017–2020 State Plan (Ministerio de Economía, Industria y Competitividad, 2018) or a requirement to obtain the Severo Ochoa or María de Maeztu Excellence Awards (Ministerio de Ciencia, Innovación y Universidades, 2018).

Because of the strategic value now given to research data, more and more universities and research centres have decided to offer their researchers a data management service. The libraries and research offices of the Catalan universities, which work in coordination within the Open Science Area of the University Services Consortium of Catalonia (CSUC), offer such a service.

2. The CSUC and Research

The CSUC was set up in 2014 as a merger of the Centre for Scientific and Academic Services of Catalonia (CESCA, 2018) and the Consorci de Biblioteques Universitàries de Catalunya (CBUC, 2015), organizations related to information technology infrastructure and libraries, respectively. The CSUC's aim was to foster cooperation and sharing to make the Catalan university and research system more efficient.

The CBUC started to offer research support services in 1999, and this work has now been taken on by the CSUC's Open Science Department. The work has consisted of creating cooperative repositories of Catalan doctoral theses, books and scientific, cultural and scholarly journals, making open access mandates effective, managing article processing charges, implementing the unique ORCID identifier for each researcher, creating the Research Portal of Catalonia (Parusel & Reoyo, 2018) and supporting the management of research data.

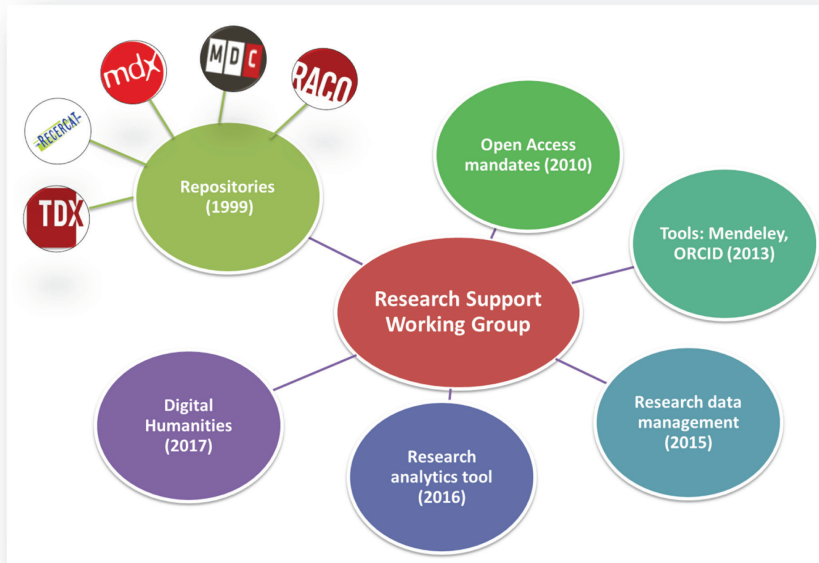
These activities, coupled with the changes arising from the digitalization of information, have made all the academic libraries of Catalonia adapt their organizational charts in one way or another to create structures to support research activities. In 2014, the new service needs of universities arising from the rapidly changing research scenario led the CBUC to add a new strategic line, support for research, to its two existing ones. Furthermore, in 2017 the CSUC decided to create the Open Science Area, which is dedicated exclusively to research services and undertakes the work of this type that had previously been done by the Libraries and Documentation Area.

With the creation of the new strategic line in 2014, it was decided to set up a Research Support Working Group (RSWG) in order to further alliances between the stakeholders of universities that were directly involved. The RSWG is made up of representatives of the University of Barcelona, the Universitat Politècnica de Catalunya, the Pompeu Fabra University, the University of Girona, the University of Lleida, the Universitat Rovira i Virgili, the Open University of Catalonia, the Vic-Central University of Catalonia, the Ramon Llull University and the Universitat Jaume I, in addition to the CSUC's Open Science and Information Technology areas. The activities of the group (Figure 1) are led by a committee formed by the vice-rectors for research of all the member universities of the CSUC.

3. Work Done with Research Data

One of the RSWG's objectives was to establish a framework of reference that would allow universities and research centres to establish a management policy for data generated by research activities. Starting in 2015, great efforts

Fig. 1: Areas of activity of the RSWG.



were made to agree on procedures that would allow libraries and research offices to collaborate with research groups to create data management plans (DMPs) and recommend repositories for depositing the data in open access.

3.1. Training and Exploration of Researchers' Needs

In order to align activities of the RSWG member institutions, training sessions were held at two levels: international experts were invited to explain their vision and experiences to the consortium, and in each institution the staff were trained to provide support to researchers.

In parallel to the training, a survey was designed for those responsible for the projects of universities of Catalonia with H2020 funding in order to determine researchers' needs with regard to data management.

Between December 2015 and January 2016 the survey was sent to a sample of 164 projects, and responses were obtained from 73 (45%). The

conclusions, presented in a report (CSUC, 2016a) and in the associated data (CSUC, 2016b), follow the general trends of other surveys, such as Tenopir et al. (2011) in the US and the Datasea project in Spain (Peset, Aleixandre-Benavent, Gonzalvo, & Ferrer-Sapena, 2017), and show the lack of knowledge of researchers in this field.

3.2. Data Management Plans

DMPs describe the life cycle of the research data and explain how they will be preserved, what vocabularies and standards will be used and how they will be shared. The working group considered that the most valuable contribution that libraries could make in the short term was to support the creation of DMPs. A guide was drawn up to help researchers create plans according to the FAIR requirements of H2020.

The guide, which was published in Catalan (CSUC, 2016c) and English (CSUC, 2016d), shows the fields required, with explanations and aspects to be taken into account, in addition to a selection of real examples that were highly valued by the researchers. When this work was completed, to maximize its dissemination and make it easier for researchers to use, the RSWG decided that the guide had to be converted into an online application or tool: the Research Data Management Plan (CSUC, 2018a) (Figure 2).

The CSUC's tool is an adaptation of the Digital Curation Centre's DMPOnline (DCC, 2016)¹, which has become the default software used by most institutions implementing a tool for setting up DMPs. DMPOnline and the tools derived from it allow researchers to draw up plans while consulting the guide and the specific specifications of each university, to share the plans with other researchers, and to export the plans in different file formats.

In order to disseminate the tool, infographics were published in Catalan and English (Figure 3) (CSUC, 2018b), offering a clear and easy explanation of the options of the tool and who to contact in each institution.

The tool is currently being reviewed to adapt it to new needs: the software version will be updated, and new templates will be implemented with the requirements of other funding bodies, such as the European Research Council.

Fig. 2: Screenshot of the Research Data Management Plan.

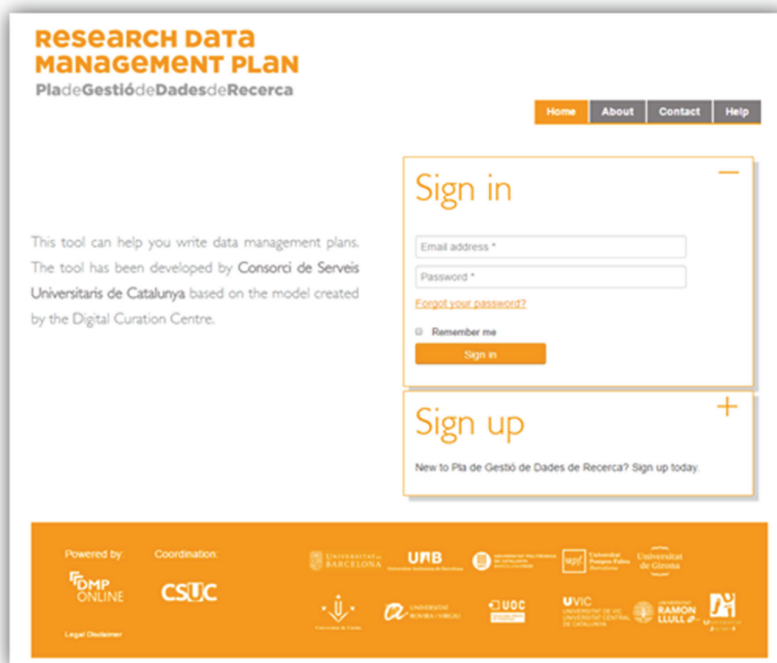








Fig. 3: Infographics to disseminate the DMP tool.



3.3. Recommendations for Data Repositories

When work began to support research data management, it seemed that the main objective was to build a data repository. After some discussion, however, the RSWG aligned with European trends to focus on offering advisory services rather than infrastructure (Tenopir et al., 2017). The various data repositories were studied and some recommendations (CSUC, 2017) were drawn up to support researchers in selecting a repository for their data. This document, which is updated regularly, includes the process and criteria for selecting a data repository following the OpenAIRE guidelines. It is accompanied by sources for selecting thematic and multidisciplinary repositories (directories, publishing recommendations, etc.) and a comparative table of the main multidisciplinary data repositories (Figure 4).

Fig. 4: Screenshot of the comparative table on data repositories.

	 ZENODO	 EUDAT	 figshare Digital Science (Macmillan)	 DRYAD SocioRxiv	 DataVerse JOSS (Harvard University)	 MENDELEY DATA Elsevier
1. Àmbit temàtic i geogràfic	Multidisciplinar Internacional	Multidisciplinar Internacional	Multidisciplinar Internacional	Amb èmfasi en ciència i medicina Internacional	Multidisciplinar Internacional	Multidisciplinar Internacional
2. Tipologia de dades (programari, imatges, raw data, etc.)	Tot tipus de formats i també documents publicats	Tot tipus de formats	Tot tipus de formats	Tot tipus de formats (amb exclusió de software) i amb preferència per formats de codi obert	Tot tipus de formats amb funcionalitats addicionals per alguns (RData, SPSS, STATA, CSV, slos, FITS)	Tot tipus de formats
3. Mida aproximada dels fitxers a incloure	Fins a 50 GB Per a mides superiors es pot contactar amb ells	Fins a 2 GB. Per a mides superiors, contactar amb ells	Fins a 5 GB	20 GB per un conjunt de dades. Per a mides superiors revisar les condicions	Fins a 2 GB.	10 GB
4. Dades obertes, emmagatzemades, restringides o tancades	Obertes Tancades Restringides Emmagatzemades (Pública)	Obertes Restringides Emmagatzemades	Obertes Tancades (per defecte fins que no es publiquen) Restringides Emmagatzemades	Obertes Emmagatzemades (per defecte fins el moment de la publicació)	Obertes Restringides	Obertes Tancades (per defecte fins que no es publiquen) Restringides Emmagatzemades
5. Llicències amb les que es volen difondre les dades	Permet varietat de llicències encara que recomanen llicències obertes. La descripció de les dades està sota CC0	Permet varietat de llicències per a dades i per a software	CC0 per a dades (<i>metadades & datasets</i>) CC-BY per a no dades (<i>audiovisuals, papers, etc.</i>) MITL per a codi i software Altres llicències: GNU, Apache 2.0, BSD 3 clause	CC0 PDM	Per defecte CC0 però es pot canviar	Permet varietat de llicències: CC-BY 4.0, CC-BY-NC, CC0 1.0, MIT, Apache 2.0, BSD 3 clause, BSD 2 clause, gplv3
6. Identificador permanent	DOI o URL	Handle	DOI	DOI	DOI	DOI
7. Condicions per retirar contingut	Es permet retirar les dades justificant-ne el perquè. El DOI i la URL quedaran restringits.	Per retirar contingut contactar amb EUDAT	Es pot sol·licitar la retirada de contingut públic justificant-ne el motiu	Cessió de dades <i>irrevocable</i> .	Es permet retirar les dades, tot i que les metadades quedaran sempre visibles	Per retirar el contingut contactar amb Mendeley .

3.4. Framework Agreement on Data Policy

To establish a research data management policy, the RSWG followed the same steps as those carried out in 2010 for creating mandates for open access to publications. A set of recommendations was prepared, establishing the elements that universities must commit to in the field of data management.

These recommendations follow the structure and content of the *Policy RECommendations for Open Access to Research Data in Europe* (RECODE, 2014), establishing that open access to data is the default position; that the competencies and responsibilities for the data, the place and the deposit period must be identified; that a data management plan is required; that responsibility for the costs must be assigned; and that the preservation and curation of the data must be determined.

Subsequently, these recommendations were drafted in an agreement and presented to the vice-rectors for research of the Catalan universities with the intention of helping the institutions adopt an internal policy for research data management. This proposal takes the form of a template for drawing up an institutional data policy and follows both the above recommendations and others proposed by LEARN (2017). The aim is for each institution to approve its own policy of open access to data in 2018.

3.5. Monitoring the Service

After all of this preliminary work, the universities started to offer the research data support service in September 2016. It was decided to monitor the universities through indicators collected by the RSWG every 6 months, which include H2020-funded projects, the dedicated staff, the number of activities and training courses related to research data management, the number of queries received, visits to the website, the number of users registered for the tool, and the number of plans created.

Since the start of the service, 142 users have registered for the tool and have produced 43 DMPs. In addition, 67 activities or training sessions related to this subject have been carried out at the universities and 142 queries have been received (including requests for help to draw up DMPs, to deposit data

or to resolve author or licence rights issues). The monitoring of the service shows a considerable gap between the services offered and the results in terms of actions carried out and plans created. However, the interest in and need for these services are growing.

4. Work to be Done with Research Data

The work of the RSWG currently focuses on two goals. The first is to improve the service and to introduce changes that meet the new needs of researchers. For this reason, it was decided to repeat the survey carried out at the beginning of the work and to update the software version of the DMP tool. The second—and now the most important task—is to create infrastructure for depositing open-access data in FAIR form.

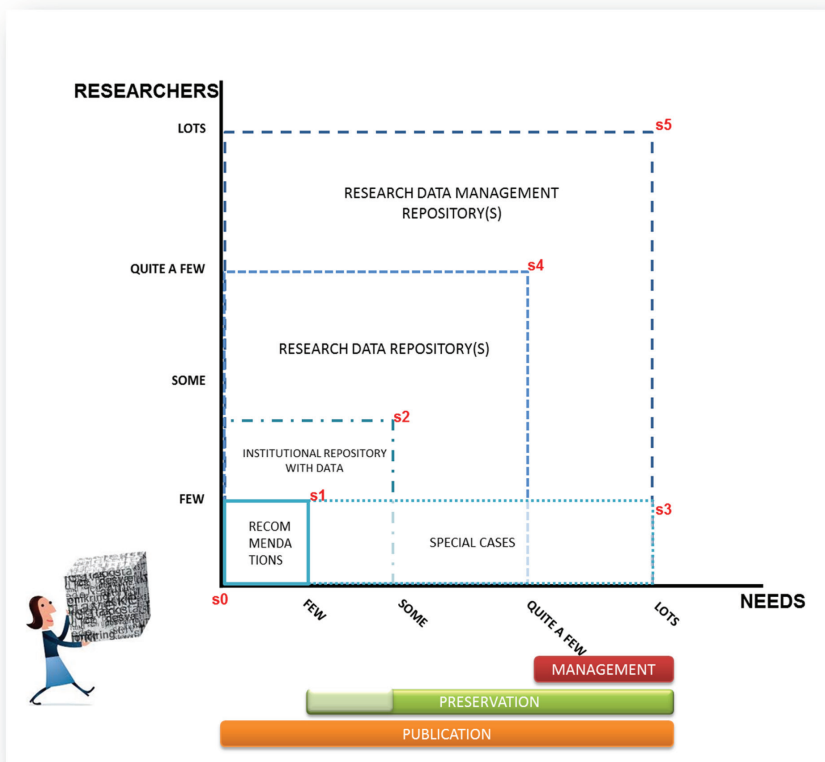
In the discussions of the RSWG it has been found that the research data are very diverse (in formats, dimensions, applications, etc.) and that there is no clear model of how to proceed. In addition to this, European experiences are scarce, embryonic and sometimes contradictory. For example, the Netherlands has a central national infrastructure, whereas the United Kingdom has a decentralized network of repositories at each university.

The group believes that there can be no single approach to solving this problem because there are different needs and groups. In order to respect the needs, it is necessary to distinguish between publishing the data accessibly, guaranteeing access to the data over a certain period of time, disseminating them in formats and protocols that make them interoperable and reusable, and, finally, managing them, i.e. having mechanisms for working with provisional data. The urgency of the needs varies according to disciplines and researchers. As an example, fast availability of research data is more important in domains like biomedicine or high-energy physics than in more slowly evolving domains like history or linguistics.

Five stages of work of growing complexity and exigency have been established for the CSUC's infrastructure (Figure 5).

First, some disciplines already have thematic repositories accepted by the research community for publishing their data in open access (for example,

Fig. 5: Stages of the research data repository.



Protein Data Bank for 3D protein structures or the National Oceanographic Data Centre for oceanographic data). For these cases, the group intends to continue revising and expanding the comparative table on data repositories. This stage does not require the creation of a dedicated infrastructure.

Second, some universities have expanded the services of their institutional repositories to include data files and immediately meet the requirements of H2020 and publisher policies. Continuing with the collaborative spirit of the group, the changes have been agreed by all the universities, although not all of them are applying them to their institutional repositories. This stage requires a low investment in resources but is necessarily provisional because

the current institutional repositories are not optimal for managing all types of data (e.g. large-size files).

Third, although many researchers do not even feel the need to publish data, for the groups that participate in European projects or have Excellence Awards it is intended to set up a pilot project in which the research groups will participate voluntarily. An existing repository will be used and shared experiences will be used to design a future infrastructure. This stage will have two requirements: first, it must identify the groups that form part of the programme and have them designate a person as an interlocutor; after this it must hire a data curator to collect and disseminate the experiences of the group.

Fourth, in the current stage, the RSWG is determining reasonable functional requirements for a consortial data repository. This is being done with the participation of the stakeholders (libraries, IT departments, research offices, etc.). Initially, it is believed that, to achieve economies of scale, the repository should be cooperative but not necessarily centralized and should not only allow publication but also guarantee the preservation, interoperability and reuse of data.

This repository will extend the provision of services and meet the needs of more projects than are currently covered by institutional repositories. To this end, the general categories of the functional requirements have been established:

- Stable identifiers, i.e. permanent identification codes that can be used to refer to data files uniquely over time, regardless of what part of the system the data are stored in.
- Large-scale storage capacity for most of the data produced by researchers in the Catalan Research System, though it is probably best to store some data (such as those of genetics and oceanographers) in thematic repositories.
- Storage capacity for most formats of the data files produced by researchers of the Catalan Research System; for reuse, this involves acquiring and updating the software.
- High performance data storage: i.e. several copies of the data files are stored, ideally hosted on remote servers, and frequent automatic checks of the data integrity are carried out.
- Interoperability between the different elements: i.e. there are mechanisms that allow the metadata of a data file to be offered in a research

portal or a Current Research Information System (CRIS) while the actual data are on another server; it must be possible to use the various elements transparently so that they act as a single system.

- Management of special features, such as different versions, the meta-data schemes for each discipline, the type of access allowed (open/closed/restricted), etc.

Fifth, in the not-too-distant future data management will be a generalized need that universities must satisfy by providing a service. This service should not be limited to the depositing of data safely over a period of time, but should also include management throughout their entire life cycle. Such a service requires considerable additional resources because, for example, it consumes much more storage space.

5. Conclusions

As the Australian Library and Information Association (ALIA, 2017) states, transformation is not a novelty for libraries, which have already overcome several barriers by taking advantage of opportunities that provide even greater access to information on a global scale. Through the internet and digital resources, users are given a previously unimaginable level of access to knowledge, and research data are just one part of this changing world.

However, facing this new scenario entails creating new services for which new knowledge and new infrastructure are necessary. We want to create a world in which research data management can be integrated into the researchers' own processes, and this will lead to an increase in the number of DMPs and in the amount of research data that is openly accessible. Achieving this is difficult because of the additional resources it entails and especially because it involves building a new scenario. The CSUC's RSWG believes that doing so jointly, in cooperation, saves many efforts, offers a better guarantee of success, and makes the task more enriching from a professional point of view.

References

Abadal, E., Olle, C., & Redondo, S. (2018). Open access monographs published by university presses in Spain. *El profesional de la información*, 27(2), 300–311.

Retrieved November 9, 2018, from <http://www.elprofesionaldelainformacion.com/contenidos/2018/mar/08.pdf>.

ALIA. (2017). *Australian libraries: The digital economy within everyone's reach*. https://www.alia.org.au/sites/default/files/Australian%20Libraries%20-%20the%20digital%20economy_website.pdf.

Anglada, L., & Abadal, E. (2018). ¿Qué es la ciencia abierta? *Anuario ThinkEPI*, 12, 292–298. Retrieved November 9, 2018, from <https://recyt.fecyt.es/index.php/ThinkEPI/article/view/thinkepi.2018.43>.

CBUC. (2015). *CBUC 2004–2013: Els últims deu anys d'activitats i de presència pública*. RECERCAT. <http://hdl.handle.net/2072/259600>.

CESCA. (2018). *Centre de Serveis Científics i Acadèmics de Catalunya*. https://ca.wikipedia.org/wiki/Centre_de_Serveis_Científics_i_Acadèmics_de_Catalunya.

CSUC. (2016a). *Gestió de les dades de recerca: Resultats de l'enquesta prospectiva a gener de 2016*. RECERCAT. <http://hdl.handle.net/2072/268186>.

CSUC. (2016b). *Gestió de les dades de recerca: Resultats de l'enquesta prospectiva a gener de 2016 [Dataset]*. Zenodo. <http://doi.org/10.5281/zenodo.183129>.

CSUC. (2016c). *Plans de gestió de dades: Versió 2, Desembre 2016*. RECERCAT. <http://hdl.handle.net/2072/273194>.

CSUC. (2016d). *Data management plans: Version 2, December 2016*. RECERCAT. <http://hdl.handle.net/2072/270395>.

CSUC. (2017). *Recomanacions per seleccionar un repositori per al dipòsit de dades de recerca: Versió 3, Maig 2017*. RECERCAT. <http://hdl.handle.net/2072/284974>.

CSUC. (2018a). *Research data management plan/Pla de gestió de dades de recerca*. <https://dmp.csuc.cat/>.

CSUC. (2018b). *Infografia research data management plan*. <http://www.csuc.cat/ca/recerca/gestio-de-dades-de-recerca>.

DCC. (2016). *DMPOnline*. <https://dmponline.dcc.ac.uk/>.

LEARN. (2017). *LEARN toolkit of best practice for research data management*. LEARN. <http://learn-rdm.eu/wp-content/uploads/RDMToolkit.pdf>.

Ministerio de Ciencia, Innovación y Universidades. (2018). *Apoyo a centros de excelencia "Severo Ochoa" y a unidades de excelencia "María de Maeztu," 2017*. Madrid. <http://www.ciencia.gob.es/portal/site/MICINN/menuitem.dbc68b34d11ccbd5d52f1feb801432ea0/?vgnnextoid=2da905f1b98ce510VgnVCM1000001d04140aRCRD>.

Ministerio de Economía, Industria y Competitividad. (2018). *Plan estatal de investigación científica y técnica de innovación 2017–2020*. Madrid. <http://www.ciencia.gob.es/stfls/MICINN/Prensa/FICHEROS/2018/PlanEstatatIDI.pdf>.

Parusel, I., & Reoyo, S. (2018). *Portal de la recerca de Catalunya: Agregant informació de procedència i institucions diverses*. RECERCAT. <http://hdl.handle.net/2072/308325>.

Peset, F., Aleixandre-Benavent, R., Gonzalvo, S., & Ferrer-Sapena, A. (2017). Experiencias y lecciones aprendidas en datos abiertos en investigación. El grupo Datasea. *evista PH*, 92, 224–225. www.iaph.es/revistaph/index.php/revistaph/article/view/3941.

RECODE. (2014). *Policy RECommendations for open access to research data in Europe*. <http://recodeproject.eu/>.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., ..., Frame, M. (2011). Data sharing by scientists: practices and perceptions. *PloS One*, 6(6), e21101 (n.p.). <https://doi.org/10.1371/journal.pone.0021101>.

Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., ..., Allard, S. (2017). Research Data Services in European Academic Research Libraries. *LIBER Quarterly*, 27(1), 23–44. <https://doi.org/10.18352/lq.10180>.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., ..., Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific data*, 3, 160018 (n.p.). <https://doi.org/10.1038/sdata.2016.18>.

Note

¹ Currently, in addition to the original version, DMPOnline is available in the Canadian, Danish, Finnish, Belgian and Spanish versions, among others. They are available at <https://github.com/DMPRoadmap/roadmap/wiki/Local-installations-inventory>.