MareData

Red Española sobre Datos de Investigación en Abierto

Recommendations on managing research data

Addressed to researchers

Summary

Introduction 3

The importance of sharing 7

FAIR data ecosystem 8

Recommendations from MareData 10

Bibliography 15

MareData Spanish Open Research Data Network.
Author: Remedios Melero
Collaboration: Ernest Abadal, Rafael Aleixandre, Agustí Canals, Antonia Ferrer, Tony Hernández, Alexandre López-Borrull, Jose Norberto Mazón, Eva Méndez, Candela Ollé and Fernanda Peset.

December 2018.

Project CSO2015-71867-REDT

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Introduction

MareData is a thematic network funded by the Spanish Ministry of Economy and Competitivity under the "Networks of Excellence" aid scheme of the Programme for Fostering Excellence in Scientific and Technical Research. This initiative forms part of the National Sub-Programme for Generation of Knowledge, in the framework of the Spanish National Plan for Scientific and Technical Research and Innovation 2013-2016 (CSO2015-71867-REDT). The network was created and is managed by seven research groups from various institutions (Institute of Agrochemistry and Food Technology-CSIC, Institute of Management of Innovation and Knowledge-CSIC, Universidad de Alicante, Universitat de Barcelona, Universidad Carlos III de Madrid, Universitat Oberta de Catalunya and Universitat Politècnica de València) whose lines of investigation are related to the management of research data: interoperability, publication, access, localisation, preservation and impact metrics.

The general objective of the **MareData** network is to bring together and consolidate collaboration among Spanish research groups for which the production, analysis and handling of data represents a significant burden. Differences in the habits and practices of researchers and in data types and formats also tend to hinder the possibility of adopting a single strategy for data handling and management without considering the diversity of data and their thematic field (Science Europe, 2018).

Consequently, drawing up recommendations on the management of research data should be understood as providing guidelines, bases or fundamentals for investigators to create their own system, adapted to the corresponding scientific community. There are currently several organisations engaged in this activity. The **DataONE** network, for instance, collects and shares Earth observational data and also offers tools and training in subjects related to data management, adapted to the characteristics of the scientific speciality.

Astronomy, climatology, particle physics and genetics are areas that clearly share data by default, because collaboration among research groups of these disciplines is crucial to the advancement of knowledge.

The recommendations, which form part of the specific aims of **MareData**, are **addressed to researchers** and based on the experience of network members and documentation relating to data management. The bibliography developed in recent years around this subject has been abundant, in particular the guidelines of research-funding agencies such as the European Commission, or of associations of university and research institutions, among them the European University Association (EUA), the League of European Research Universities (LERU), the Young European Research Universities Network (YERUN) or the Ligue des Bibliothèques Européennes de Recherche (LIBER), and projects like OpenAIRE, FOSTER, OpenUp, EDISON, LEARN and RECODE.

In Spain, the European Commission has been the major driver of promotion and support for open access to science and especially research data. The pilot project for open access to publications deriving from EU-funded projects in certain thematic areas was implemented in the 7th Framework Programme for Research and consolidated as a requirement for all disciplines in the Horizon 2020 programme. This programme also launched the pilot project for research data, in which certain disciplines were required to submit a data management plan and provide open access to the data underlying published papers. As from 1 January 2017 (European Commission, 2016a), this pilot project was extended to all areas of the Framework Programme and all funded projects must currently provide a data management plan specifying the data that will be produced, where it will be shared and preserved or whether, for some justifiable reason, the described data is exempt from open sharing by the project having opted out.

The value of research data is no longer solely scientific; it is now considered key to economic progress and innovation. This was highlighted by the European Commissioner for research, Carlos Moedas, in his key speech entitled “Open Innovation, Open Science, Open to the World” at the inauguration

of the "Opening up to an ERA of Innovation" conference held in June 2015 (Moedas, 2015). During this event, the Commission announced its intention to work for a data-intensive economy (*open science for a knowledge and data-driven economy*).

In the Draft European Open Science Agenda, the Commission initially identified five lines of potential policy actions to support the development of Open Science (European Commission. RTD, 2016):

1. Fostering and creating incentives for Open Science
2. Removing barriers to Open Science
3. Developing research infrastructures for Open Science
4. Mainstreaming Open Access to research results
5. Embedding Open Science in society

Subsequently, the European Commissioner for Research divided this agenda into eight priorities or goals for open science in Europe: rewards and incentives; research indicators and next-generation metrics; the future of scholarly communication; the European Open Science Cloud (EOSC); **FAIR** data (findable, accessible, interoperable and reusable); research integrity; skills and education in open science; and citizen science. The European Open Science Policy Platform (OSPP) was created during the Competitiveness Council of May 2016, comprising a group of experts that advise the Commission on how to develop and implement open science policies in Europe. The OSPP also recommends strategic actions for effective implementation by the different interest groups (funding agencies, research-performing organizations such as universities, publishers and those responsible for research infrastructure, among others). OSPP recommendations respond to the Commissioner's eight priorities, which include **FAIR** research data and may be consulted at **OSPP-REC**.

The European Open Science Cloud (EOSC) was officially launched in April 2016. The EOSC is a common virtual environment where European researchers can store, manage, analyse and reuse data for research, innovation

and education. The idea that underlies the EOSC is to create a trusted open environment in which the scientific community can share and reuse research data and results (European Commission, 2016b). The EOSC is a kind of meta-infrastructure where resources and capabilities of existing infrastructures in Europe can be pooled to manage and handle research data on a cross-disciplinary basis. If the EOSC is to be an effective reality, a catalogue of data and services must be created together with a system of governance, and Member States must collaborate closely and be fully committed to its implementation.

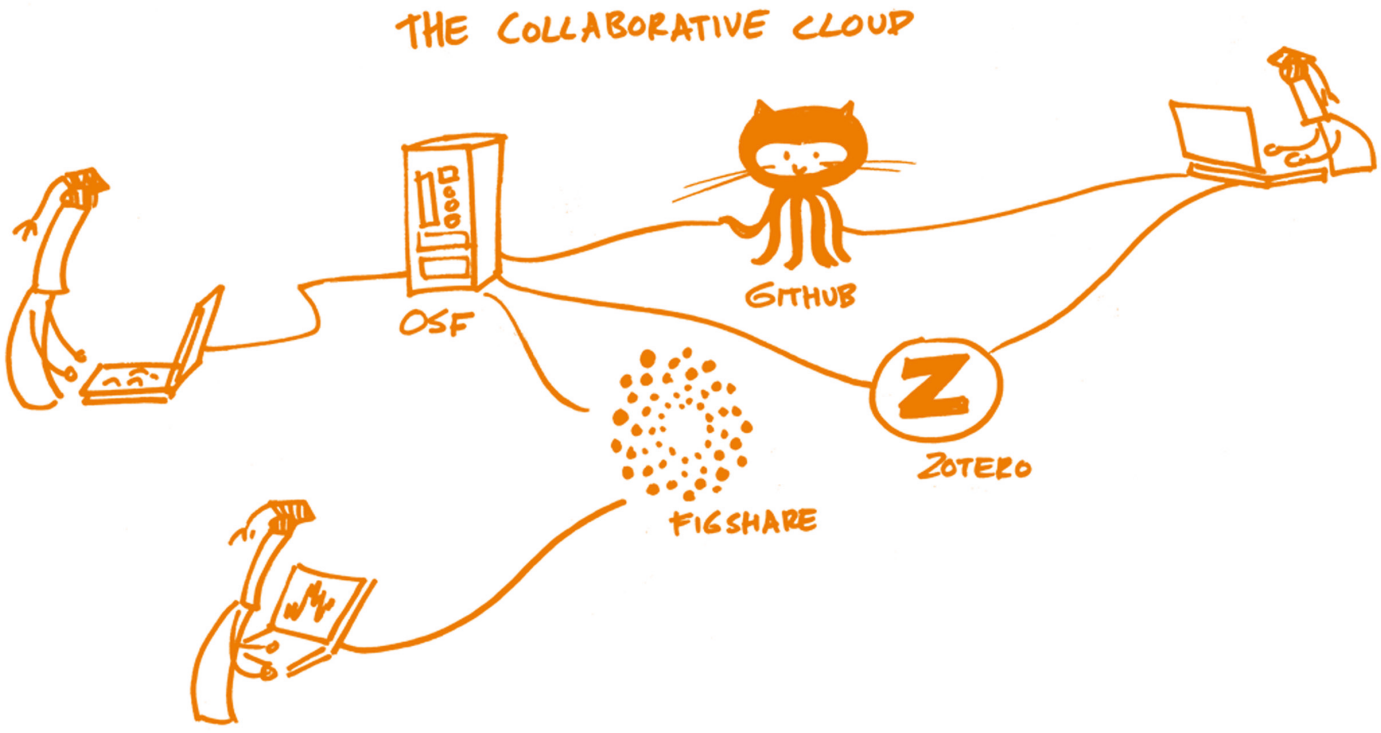
All policies, mandates and initiatives **at the European level** reflect the value of research data and the importance of being able to share and reuse them. Ultimately, data lend meaning to any institutional coordination of initiatives and actions undertaken by researchers.

Consequently, initiatives that avoid redundancy and consolidate coordinated projects among research groups and other stakeholders involved in managing and opening up research data must be strengthened and brought together. The **MareData** network sets out to facilitate this and contribute to attaining these objectives. It seeks to ensure that data which are the fruit of Spanish research fulfil the four criteria defined as "intelligent openness", meaning that they are "accessible, intelligible, assessable and usable" (*Science as an open enterprise*, 2012, p. 14); or in other words, using the universal acronym for data, meaning that they are FAIR (findable, accessible, interoperable and reusable) (Wilkinson *et al*., 2016).

The importance of sharing

Above all, sharing data

- promotes innovation and the reuse of data which could have new applications;
- enables collaboration among data users, creators and reusers;
- maximises data transparency and reliability;
- facilitates the reproducibility of experimental trials;
- allows research results to be verified;
- reduces costs by avoiding data duplication;
- increases the impact and visibility of research;
- promotes the research projects from which the data and publications originate;
- generates direct recognition of the researchers who produce the data, as occurs with any other research result.



FAIR data ecosystem

According to the model suggested by the European Commission Expert Group on Turning **FAIR** data into reality, the **FAIR** data ecosystem is composed of: policies that regulate and define the data, the researchers who produce or use them, data management plans, identifiers, standards, trusted digital repositories and the cloud services where the data are stored. At the same time, to make the ecosystem possible these components should be developed in a proactive framework of four key elements: skills, metrics, investment and rewards (Figure 1).

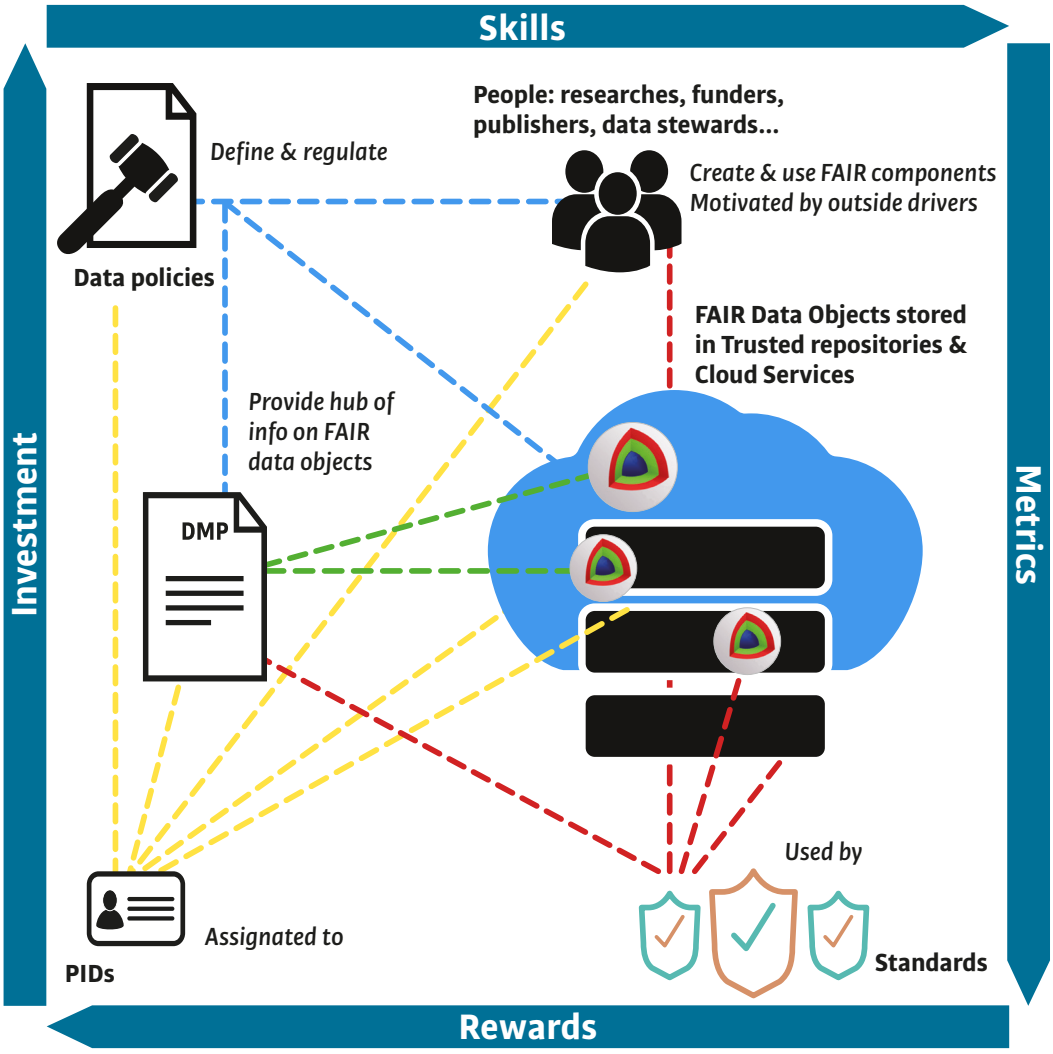


Figure 1. Components of a FAIR data ecosystem (Hodson, *et al.*, 2018a)

In this ecosystem, the **FAIR** data objects comprise their data file, a persistent identifier, a description of the standards and formats used to obtain, represent and visualise the data and the metadata describing datasets that enable their interpretation and reuse (Figure 2).

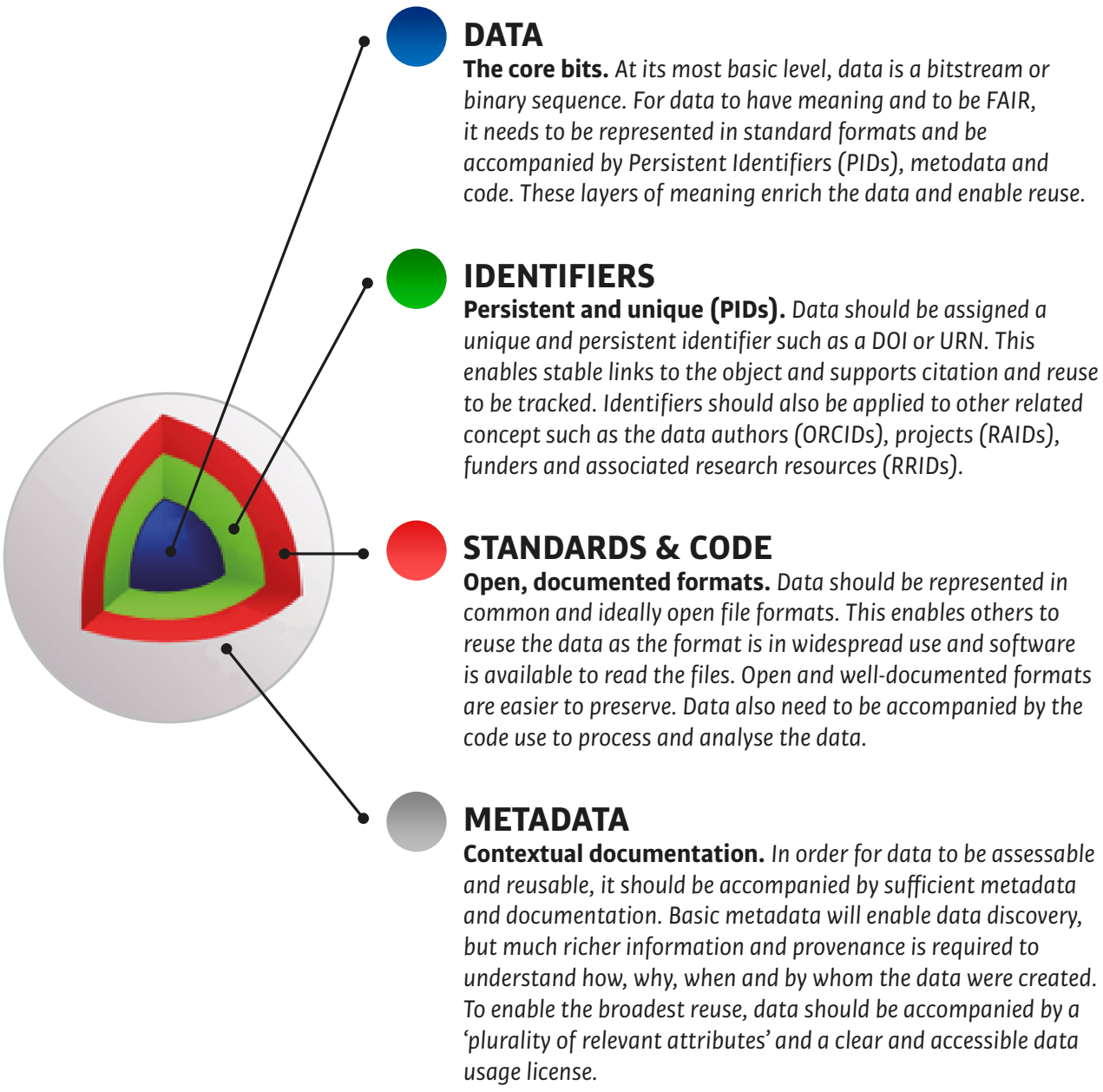
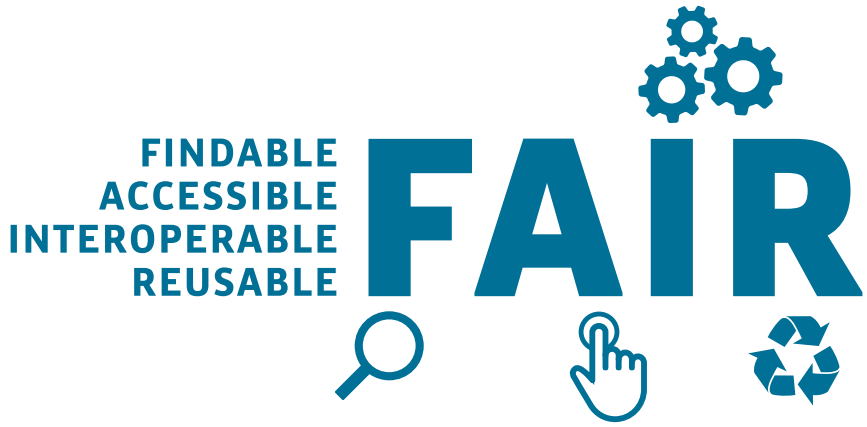


Figure 2. Model for FAIR data objects (Hodson, *et al.*, 2018b)

Recommendations from MareData

The **MareData** srecommendations on research data, which are addressed to all researchers and research groups, are based on the **FAIR** data ecosystem and focus on **how to make your data FAIR and how to better ensure the retrievability and preservation of research data**.



REC1. Adopt the FAIR Guiding Principles and facilitate open, legal and reusable scientific production. Adopting the FAIR Guiding Principles entails the following requirements (Data Citation Synthesis Group, 2014):

To be Findable:

- **F1.** assign a globally unique and persistent identifier to the (meta)data.
- **F2.** describe the data with rich metadata (i.e. adding semantic descriptions, annotations, etc. to the data they describe, which improve the quality of the metadata)
- **F3.** register or index the (meta)data in a searchable resource (repository, service, etc.)
- **F4.** clearly and explicitly include an element of metadata for the persistent identifier of the data it describes

To be Accessible:

- **A1.** use a standardised communications protocol to retrieve (meta)data by their identifier
- **A1.1** the protocols must be open, free, and universally implementable
- **A1.2** the protocols must allow for an authentication and authorization procedure, where necessary
- **A2.** metadata must be accessible, even when the data are no longer available.

To be Interoperable:

- **I1.** use formal, accessible, shared, and broadly applicable languages for (meta) data representation
- **I2.** use vocabularies (diagrams, ontologies, etc.) that follow **FAIR** principles to describe (meta)data
- **I3.** include qualified references and links from (meta)data to other (meta)data

To be Reusable:

- **R1.** ensure meta(data) are richly described with a plurality of accurate and relevant attributes
- **R1.1.** release (meta)data with a clear and accessible data usage license
- **R1.2.** use detailed provenance criteria (creation, attribution and record of versions) to associate (meta)data to data throughout their life cycle
- **R1.3.** ensure that the standards of (meta)data used meet domain-relevant community standards of the field of knowledge to which the data refer

- *If you want further information about how to make your data FAIR or check whether they are FAIR, we recommend:*

  GO-FAIR: https://www.go-fair.org/fair-principles

  FAIR assessment tool: https://www.ands-nectar-rds.org.au/fair-tool.
- *If you don't fully understand which standards, metadata or identifiers these principles refer to, ask in your university library.*

REC2. Choose those data that are interesting for future uses.

- *Sometimes it is not necessary to preserve all the data generated during a research project: "clean your data" and carefully choose those that may prove useful.*

REC3. Apply **FAIR** principles to the metadata and the data management plan.

- *The metadata describing your data and the data management plan documenting the management process of the data project must also be Findable, Accessible, Interoperable and Reusable.*

REC4. In the project evaluation reports, describe the benefits and advantages of having made the research data **FAIR**.

- *In your deliverables and reports, reflect the advantages inherent in making the data contained in your project* **FAIR**.

REC5. Draw up a data management plan which reflects what your **FAIR** data will be like.

- *The data management plan (DMP) facilitates planning, organising and documenting how the data will be retrieved and managed in the framework of the research project. The data management plan is a live document which is modified and updated in line with development of the research.*
- *Use the DMP templates offered by one of the following tools:*
  DMPtool: https://dmptool.org
  DMPonline: https://dmponline.dcc.ac.uk
  PGDOnline (in Spanish): http://dmp.consorciomadrono.es.
- *The data management plan is usually a project deliverable (generally in the sixth month of its preparation). Make sure you include it in your proposal and that you complete it if the project is funded.*

REC6. Indicate in the data management plan whether existing data have been consulted or reused instead of finding them again (avoid redundancy).

- *If your project is based on existing data you should record this in your data management plan.*

REC7. Use interoperable standards with the **FAIR** data ecosystem to enable interoperability among systems.

- *There are many standards involved when creating FAIR data (identifiers, metadata, vocabularies, etc.); make sure you use the most appropriate. Check the relevant RDA (Research Data Alliance) information:*
  PID (persistent identifiers): https://rd-alliance.org/group/pid-information-types-wg/outcomes/pid-information-types.
  Metadata standards directory: https://rd-alliance.org/group/metadata-standards-catalog-wg/outcomes/metadata-standards-directory-wg-recommedations.html.
- *If you have any doubts about which standards to use, ask in your library.*

REC8. Make research data available in line with the standards recognised by the particular scientific community.

- *Data and metadata standards are usually linked to the scientific domain. Make sure you employ the ones that are most used in your discipline and that have a level of maturity, normalisation and use which enhances their interoperability. You can consult:*
  Fairsharing: https://fairsharing.org
  DCC disciplinary metadata: http://www.dcc.ac.uk/resources/metadata-standards
  RDA: http://rd-alliance.github.io/metadata-directory/.
- *If you have any doubts about which standards to use, ask in your library.*

REC9. Use open licenses as far as possible to facilitate reuse of the research data.

- *It is advisable to clearly assign the type of research-data-use license; **preferably the most open possible**.*

A classification of license types for datasets from the most open to the most restrictive can be found at:
https://help.data.world/hc/en-us/articles/115006114287-Common-license-types-for-datasets.
Further information on open licenses is available in the guide at Open Definition (OKF):
https://opendefinition.org/guide.

- *Open Knowledge International recommends using licenses in accordance with their open definition (attribution-share alike). Examples of these licenses, such as Open Data Commons and Creative Commons, may be found on their website.*
- *If you have any doubts about which type of license to assign to your data, ask in your library.*

REC10. Provide a detailed description of the conditions for reproducing experimental trials.

- *Don't forget to include negative results: they are also relevant in avoiding unnecessary repetitions.*

REC11. Identify the repository that is most suitable for depositing your data. Analyse discipline-specific repositories as a source of data for consultation and for help on how to make the deposit.

- *There are numerous thematic data repositories you can use, both for consultation and for help on how to deposit your data. If you want further information about these, see re3data.org, the registry of discipline-specific data repositories. Bear in mind the key data infrastructures in your discipline.*
- *If you have any doubts about which repositories to use, ask in your library.*

REC12. Check the policy terms or requirements regarding your research data. Funding agencies, research centres and universities and, increasingly, scientific journals, have their own policies on research data.

- *Although these days all research-funding institutions have similar research data management requirements (create a FAIR data management plan, share open datasets, etc.), specific demands may vary. Check these in the terms of the call for the funding of your research, and review the policies of publications if you intend to publish your results in a journal.*
- *FAIRsharing compiles a catalogue of data preservation, management and sharing policies of international funding agencies, regulators and leading journals. See: https://fairsharing.org/policies.*

REC13. Promote the citation of datasets as any other bibliographic resource.

- *Cite the data of other researchers you use in your work and provide the citation of your own data.*
- *To cite recommendations of the RDA (Research Data Alliance):*
  https://rd-alliance.org/group/data-citation-wg/outcomes/data-citation-recommendation.html.
- *Make sure your data have a permanent identifier to enable their citation.*
- *We recommend you follow the guidelines of the Data Citation Principles, based on the relevance of data as a source of knowledge, persistence, and the interoperability of datasets (https://www.force11.org/group/joint-declaration-data-citation-principles-final).*
- *DataCite enables the search of datasets that have a DOI (digital object identifier) assigned as a unique identifier, and shows how to cite them.*

REC14. Follow appropriate and internationally recognised codes of conduct, above all in those disciplines that involve working with living beings.

- *If your data refer to people, safeguard their privacy in accordance with the terms established in European Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data:*
  https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32016R0679.

REC15. Contribute to the dissemination and sharing of good practices among your scientific community with respect to data management and lessons learned.

- *Share and disseminate good practices on research data management with other researchers in your domain and with the data managers.*

REC16. Foster collaboration among professionals in the field and data and information managers in preparing and taking part in training programmes.

- *Actively participate in the design of training courses for young researchers and continuous training for researchers in relation to the management of scientific data.*

REC17. Final recommendation: encourage researchers and the centres they work for to endorse the paradigm of open science.

- *Contribute to implementing the Open Science paradigm in your institution or research centre from a holistic perspective, involving all stakeholders and processes that affect the production, communication and preservation of research results financed with public funds.*

Bibliography

Australian National Data Service (2018). *Data citation*. [Melbourne; Canberra]: Australian National Data Service. <https://www.ands.org.au/working-with-data/citation-and-identifiers/data-citation>.

Ayris, Paul; López de San Román, Alea; Maes, Katrien; Labastida, Ignasi (2018). «Open science and its role in universities: a roadmap for cultural change». Advice paper, no. 24 (May). <https://www.leru.org/publications/open-science-and-its-role-in-universities-a-roadmap-for-cultural-change>.

Dai, Qian; Shin, Eunjung; Smith, Carthage (2018). «Open and inclusive collaboration in science: a framework». *OECD science, technology and industry working papers*, no. 7, 29 p. <https://doi.org/10.1787/2dbff737-en>.

Data Citation Synthesis Group (2014). *Joint Declaration of Data Citation Principles: final*. Martone, M. (ed.). La Jolla, CA: FORCE11. <https://doi.org/10.25490/a97f-egyk>.

European Commission (2018). «Commission recommendation (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information». *Official journal of the European Union*, vol. 61, L 134, p. 12-18. <https://ec.europa.eu/digital-single-market/en/news/recommendation-access-and-preservation-scientific-information>.

European Commission (2016a). *H2020 programme: guidelines on FAIR data management in Horizon 2020*. [Brussels]: European Commission, Directorate-General for Research & Innovation. 12 p. <http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf>.

European Commission (2016b). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: European Cloud Initiative: building a competitive data and knowledge economy in Europe*. Brussels: European Commission. 13 p. <http://ec.europa.eu/transparency/regdoc/rep/1/2016/EN/1-2016-178-EN-F1-1.PDF>.

The FAIR data principles (2017). La Jolla, CA: FORCE11. <https://www.force11.org/group/fairgroup/fairprinciples>.

Hodson, Simon; Jones, Sarah; Collins, Sandra; [et al.] (2018a). *FAIR Data Action Plan: interim recommendations and actions from the European Commission Expert Group on FAIR data*. June. 21 p. <https://zenodo.org/record/1285290#.WyOWGSB9jt4>.

Hodson, Simon; Jones, Sarah; Collins, Sandra; [et al.] (2018b). *Turning FAIR data into reality: interim report from the European Commission Expert Group on FAIR data*. June. <https://zenodo.org/record/1285272#.W7uasPmYSCg>.

Kroes, Neelie (2013). «The economic and social benefits of big data». [Brussels]: European Commission. <http://europa.eu/rapid/press-release_SPEECH-13-450_en.htm>.

Melero, Remedios; Hernández-San-Miguel, Javier (2014). «Acceso abierto a los datos de investigación, una vía hacia la colaboración científica». *Revista española de documentación científica*, vol. 37, n.º 4. <http://dx.doi.org/10.3989/redc.2014.4.1154>.

Moedas, Carlos (2015). «Open innovation, open science, open to the world». En: *A new start for Europe: opening up to an ERA of innovation Conference*. [Brussels]: European Commission. <http://europa.eu/rapid/press-release_SPEECH-15-5243_en.htm>.

Open Science Policy Platform (2018). *Integrated advice of the Open Science Policy Platform on 8 prioritised open science ambitions*. [Brussels]: European Commision. <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform>.

Peset, Fernanda; González, Luis-Millán (2017). *Ciencia abierta y gestión de datos de investigación (RDM)*. Gijón: Trea. 191 p. (Biblioteconomía y administración cultural; 301). ISBN 978-84-9704-907-8.

RECODE Project Consortium (2014). *Policy recommendations for open access to research data*. [London]: RECODE. 40 p. <http://web.archive.org/web/20180205153933/http://recodeproject.eu/wp-content/uploads/2015/01/recode_guideline_en_web_version_full_FINAL.pdf>.

Royal Society. Science Policy Centre (2012). *Science as an open enterprise*. London: The Royal Society. 104 p. (The Royal Society Science Policy Centre report ; 2). <https://royalsociety.org/topics-policy/projects/science-public-enterprise/Report/>.

Science Europe (2018). *Science Europe guidance document: presenting a framework for discipline-specific research data management: [D/2018/13.324/1]*. Brussels: Science Europe. 46 p. <http://www.snf.ch/SiteCollectionDocuments/SE_Guidance_Document_RDMPs.pdf>.

Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; [et al.] (2016). «The FAIR guiding principles for scientific data management and stewardship». *Scientific data*, vol. 3, no. 160018. <http://dx.doi.org/10.1038/sdata.2016.18>.

Red Española sobre Datos de Investigación en Abierto

Project CSO2015-71867-REDT