Towards a Galician data commons

Lourdes Pérez González.

Biblioteca Universitaria, Universidade de Santiago de Compostela, Galicia, Spain

Metodology

Introduction

Development in information and communications technologies is transforming the natures and scale of research and education in the way that they have become increasingly data-intensive. The ease with which digital data can be stored, disseminated and accessible to secondary user via science infrastructures means that institutions embrace in the benefits of sharing research data to increase its impact and visibility.

The research community of Galicia needs to establish a policy that ensures that data and research derived from publicly-funded research should be available and accessible for public use. To treat that, a framework for developing good praxis to all research communities and institutions is necessary. Researchers and researchers founders may propose some principles reinforced with a number of programs to disseminate it, taking account that some issues related to science integrity, transparency, professional responsibility, interoperability, protection of intellectual property and ultimately understand that long term preservation of digital material is the central problem for e-science.

In this work we move on several fronts: our acceptance of the principles of access and preservation of research data established by the Organization for Economic Cooperation and Development (OECD), the commitment to long-term preservation thought the use and implementation of the Standard Reference Model for an Open Archival Information System (OAIS), we insist in a good data management policy according with the data lifecycle, working close to the steps we define roles and responsibilities of the different stakeholders in the process.

Since our community is at different levels of work at local, national and international level, it is clear that the principle of sharing costs models that balance the high cost resources digital preservation consumes. Based on the different levels of collection elaborated by the National Science Foundation (NSF) we work with the recommendations and best practices of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, and the ongoing models Keeping Research Data Safe (KDDS).

3 Levels

Principles and policy: International, European and National policy (governments and agencies across the world encourage effective dissemination and sharing of research outputs and promote the preservation of them).

Principles: Public-funded research data are public good, produced in public interest; public-funded research data should be openly available to the maximum extent possible. Policy statement.

Open access to all research outputs, wholly or partially funded by the public fund in Galicia; Research generated by public funded must be well-managed and the different stakeholders during the research process, in this way research data must be accompanied by high-quality metadata enable further re-use and in case long-term preservation.

Management: Project managers and program managers in multi-institutional collaborative approach.

It should be desirable a mandatory data management and sharing plan in the different disciplines according the international standards where it exists.

Any data management and sharing plan should include any cost for its implementation.

All the process requires appropriate funding for data management and different institutions must provide their policy.

Different stakeholders in process of data lifecycle: Data creation, data analysis, research outputs, data curation, services.

We must incur in two strategically points: data creation and end of research, in this sense we try to implement a new research culture that ensure that all funded projects develop a data management plan to ensure that data are well managed through the duration of a research project. The goal is to create or develop a research resource for the benefit of the research community.

Steps in data lifecycle

Data creation

Its a priority to promote and support good data management.

Data literacy, promoting appropriate skills.

Start in the scientific process, it should become part of standard research practice.

The importance of standards on data collection.

Data entry or digitization: file format, data documentation.

Data Curation

On the base of the research outputs

Submission Information Package (SIP) are sent to the archive archive by producers.

SIP should be negotiated between the research community and the Archive.

Specifying criteria like file formats, subject matter, ingest schedule, access restrictions, and verification protocols.

Service onset research data:

Offer online resource through data catalogues.

Licensing agreements to acknowledge data rights.

Promoting the re-use of data.

Monitoring the secondary usage of data.

Metadata

Data sharing and re-use: Data re-use / Data Management Principles.

In this work we move on several fronts: our acceptance of the principles of access and preservation of research data established by the Organization for Economic Cooperation and Development (OECD), the commitment to long-term preservation thought the use and implementation of the Standard Reference Model for an Open Archival Information System (OAIS), we insist in a good data management policy according with the data lifecycle, working close to the steps we define roles and responsibilities of the different stakeholders in the process.

Since our community is at different levels of work at local, national and international level, it is clear that the principle of sharing costs models that balance the high cost resources digital preservation consumes. Based on the different levels of collection elaborated by the National Science Foundation (NSF) we work with the recommendations and best practices of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, and the ongoing models Keeping Research Data Safe (KDDS).

For data to be shared in the research community and public in general, data must be properly curated and stored. It must be on time, preferably on the early phases of a research project increasing the value, validity, impact and durability of data and decreasing the cost of storage, curation and preservation.

Archives and research community in Galicia must work closely adopting policies, and签署 by other organisations as the Data Seal of Approval by the Data Archiving and Networked Services (DANS) in The Netherlands, the different levels of collection established by the National Science Foundation (NSF) and the essential recommendations of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, and the ongoing models Keeping Research Data Safe (KDDS).

Acknowledgments: We want to thank the intellectual guidance to the Digital Curation Centre, DCC UK and UK Data Archive.

Conclusion

Data sharing is a core component in research data management and sharing plan for effective implementation through all the research life-cycle.

Good practice: Data Seal of Approval from Data Archiving and Networked services, DANS.

Data long-term preservation and data management are central issues surrounding data access, data security and ethical use. Data management occurs in the lifecycle of the research process, data creation and digital preservation it should be central to data archive and specialised staff.

The value of data depends on the quality of data. Data creator and data managers in Galicia must pay attention to ensure compliance with quality standards in different international institutions and research institutions.

Data sharing starts at the end of research on the base of research outputs whith the Submission Information Package, SIPs are sent to the archive archive by producers and it should be negotiated between the research community and the Archive. Specifying criteria like file formats, subject matter, ingest schedule, access restrictions, verification protocols.

The goal of data curation is the log-term data sharing. The value of data collections to be preserved depends on the data policy and data management during the research process.

Metadata are the core:

Should provides standardised structured information explaining the origin, purpose, creator, access conditions.

Considering re-use of data in long term, we work with the most appropriate and usable software and data format to use. Guarantee long term data access and usability imply use open standard formats. Most of times when data collection is at end of research, the research team should convert data to preferred data preservation format.

Data quality control, stakeholder must assign clear roles and responsibilities for quality assurance at all stage of research.

Value must be added to data through annotation, addition of additional databases by researchers and by curation aggregation and enhancement.

Research outputs: According to level of collections of NSF, there is different approach on depositing and shared data: centralised and specialised data centres are useful in expertise and resource in data curation, but that centers do not accept all. We promote distributed, local data storage in research collections but in Galicia we must work our expertise to data management at regional and local level.

Service onset research data

Research libraries and archives as data service provider in Galicia should be responsible for ensuring long-term access to data that has been placed in their care; ensuring that these data are usable, reliable and available.

Should be responsible for guidance and advice for data creator on issues related to data management, confidentiality (especially in sensitive and confidential data), security (ensuring the protection of data from unauthorised use, change, disclosure or destruction in a generally secure manner), ethical use (ensuring that data is treated with the necessary respect and sensitivity), and copyright and data sharing.

The service provider must promote the politics of sharing research data on the base that they are valuable resource.

Promoting share data should be accompanied by: data catalogues, licensing agreements to acknowledge data ownership, monitoring of the secondary usage of data, safe keeping of research data in a secure environment, management of access, etc.

Bibliography


Acknowledgments: We want to thank the intellectual guidance to the Digital Curation Centre, DCC UK and UK Data Archive.