Towards a Galician data commons

Lourdes Pérez González.
Biblioteca Universitaria, Universidade de Santiago de Compostela, Galicia, Spain

Introduction

Development in information and communication technologies is transforming the nature 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 users via science infrastructures means that institutions embrace 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 provide a framework for developing good practice, the authors will establish the National Science Foundation (NSF) and the Organisation for Economic Co-operation and Development (OECD), the commitment to long-term preservation through the data lifecycle, working close to this stage we define a policy framework and responsibility of the different stakeholders in the process.

Since our community is at different levels of work at local, national and international institutions, it is crucial that we balance the high cost resources digital preservation consumes. Based on the different levels of collection, 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 of Keeping Research Data Safe (JISC).

Metodology

Our study try to offer analysis of trends, best practices and key issues that could help us to construct robust infrastructures, content, and services in the new research knowledge and information.

Our first step was taking into account of international requirements and developments. There have been significant developments that indicate policy frameworks and proposal for a national and international data service.

It is recognised that different type of data created and managed across research discipline require specific approach to data management and sharing. In this way we gathered information on research activities and practices in the different research community in Galicia. Researchers create and collect different kinds of data, with different purposes and through different process. We must determine which data should be public available.

As an organisational level, we need to break with boundaries, libraries, data centers and researchers knowing that no single institution can be responsible. We assume that we need to work with key agents and stakeholders as national and international level in order to maximise the value of digital data for current and future researchers in Galicia. A stronger future for research data management and sharing is one in which multi-institutional collaborations available cyberinfrastructures and services.

We work from the lifecycle management of the digital data to ensure it is used as a resource for a foreseeable future.

Our project adopt two principles: 1) science system is based on openness and free exchange of data, information, and knowledge derived from publicly funded research should be made available and accessible for the public use; availability, access, and usability that are secured in public institutions. 2) digital research data of value arising from the actual data intensity science must be preserved and remain accessible for current and future generation.

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 found Galicia.

Research generated by public funded must be well-managed by 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 discipline 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.

Libraries and Research institutes archives as data providers are responsible for providing clear guidance to researchers and end users of their data management and sharing plans.

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 and implement 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.

3 Levels

Data creation

It’s a priority to promote and support good data management.

Data lifecycle, promoting appropriate skills.

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

The importance of standards on data collection.

Data entry or digitalization, file format, data documentation.

Data Curation

On the base of the research outputs.

Submission Information Package, SIPs 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 resources through data catalogues

• Locating aggregations to acknowledge rights

• Promoting the re-use of data

• Monitoring the secondary usage of data

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 in 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 defined by other organisations as the Data Seal of Approval by the Data Archiving and Networking Services (DANS) in The Netherlands, the different levels of collection established by the National Science Foundation (NSF) and the espacial recommendations of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, and the ongoing models of Keeping Research Data Safe (JISC).


Research Information Network(2008) . To Share or not to share publication and quality Assurance of Research Data Outputs. Research Information Network / Swan, A and Brown, S. Retrieved July 22, 2011 from http://eprints.ecs.soton.ac.uk/16742/. In this way we gathered information on research activities and practices in the different research community in Galicia. Researchers create and collect different kinds of data, with different purposes and through different process. We must determine which data should be public available.

As an organisational level, we need to break with boundaries, libraries, data centers and researchers knowing that no single institution can be responsible. We assume that we need to work with key agents and stakeholders as national and international level in order to maximise the value of digital data for current and future researchers in Galicia. A stronger future for research data management and sharing is one in which multi-institutional collaborations available cyberinfrastructures and services.

We work from the lifecycle management of the digital data to ensure it is used as a resource for a foreseeable future.

Our project adopt two principles: 1) science system is based on openness and free exchange of data, information, and knowledge derived from publicly funded research should be made available and accessible for the public use; availability, access, and usability that are secured in public institutions. 2) digital research data of value arising from the actual data intensity science must be preserved and remain accessible for current and future generation.

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 found Galicia.

Research generated by public funded must be well-managed by 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 discipline 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.

Libraries and Research institutes archives as data providers are responsible for providing clear guidance to researchers and end users of their data management and sharing plans.

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 and implement 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.

Long Term data sharing

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 specialist data centres are 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 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 conformity with explicit security protocols); copyright and data sharing.

Service provider must promote the politics of sharing research data on the base tal 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...