

Open Access movement in Spain: RECOLECTA Project.

Scientific data management Report for the repository community.

Mercedes de Miguel Estévez

SPANISH FOUNDATION FOR SCIENCE AND TECHNOLOGY (FECYT)
Department for Scientific Information Management
Repositories and Open Access Unit

COAR Annual Meeting
Istanbul, May 7 & 8, 2013



FECYT: The Spanish Foundation for Science and Technology



Is a public foundation, founded in 2001, dependent on the **Spanish Ministry of Finance and Competitiveness**.

Principles:

- Rationalisation
- Transparency
- Efficiency

Aims:

- › To develop social participation instruments that favour science.
- › To be the right tool for science **dissemination** and boosting **science culture**.
- › To act as a communication channel for the Spanish scientific community abroad.
- › To become a **metric** reference for Spanish R&D&i.
- › To support information and **scientific resource management** structures.

OPEN ACCESS IN SPAIN



FECYT participation in Open Access projects and initiatives:

- National project:
 - ✓ **RECOLECTA:** the “*National Open Science Harvester*”
- European projects under the FP7 of the EC (FECYT is the Spanish partner):
 - ✓ **MedOANet:** addresses the necessity for coordinated strategies and policies in OA, to scientific information, in six Mediterranean countries.
 - ✓ **OpenAIRE:** to support the application of OA in Europe. It provides the resources to promote and realise the widespread adoption of OA policies.
 - ✓ **OpenAIREplus:** the continuation of OpenAIRE for the incorporation of research data sets linked to scientific papers from the FP7. In the short term, OpenAIREplus will issue a series of basic directives for datasets.
- International participation:
 - ✓ **COAR** member since 2010.

Nationwide **infrastructure** of OA scientific repositories.

Platform that gathers all the national **scientific repositories** together in place.

National reference for the **OA movement** in Spain.



The **aims**....

- To promote and coordinate the national **infrastructure of OA** digital scientific repositories.
- To foster, support and facilitate the **adoption of OA policies** by Spanish universities and R&D organizations.
- To give a greater **visibility and impact** to the Spanish national research outcomes.

Services to stakeholders:

- Repository managers
- Researchers
- Decision-makers

...in conclusion RECOLECTA facilitates:

- **Open and free access** to Spanish scientific production.
- Support **services** for users and strengthens the national OA community.
- Promotes **activities** to spread information and fosters collaboration.



SOME RECENT MILESTONES...



- New **Website** has been launched
<http://recolecta.fecyt.es/>
(English and Spanish version)

Home

RECOLECTA or *Recolector de Ciencia Abierta* [Open Science Harvester] is a platform that gathers all the national scientific repositories together in one place and provides services to repository managers, researchers and decision-makers.

RECOLECTA is the result of the collaboration since 2007 between the Spanish Foundation for Science and Technology (FECYT) and the Network of Spanish University Libraries (REBIUN) run by the Conference of Vice-Chancellors of Spanish Universities (CRUE). Their work is aimed at creating a nationwide infrastructure of open access scientific repositories.

NEWS

11 January 2013
[La ciencia abierta se expande en internet](#)

28 October 2012
[La comunidad internacional, unida ante la necesidad de crear indicadores de calidad para las revistas en abierto](#)

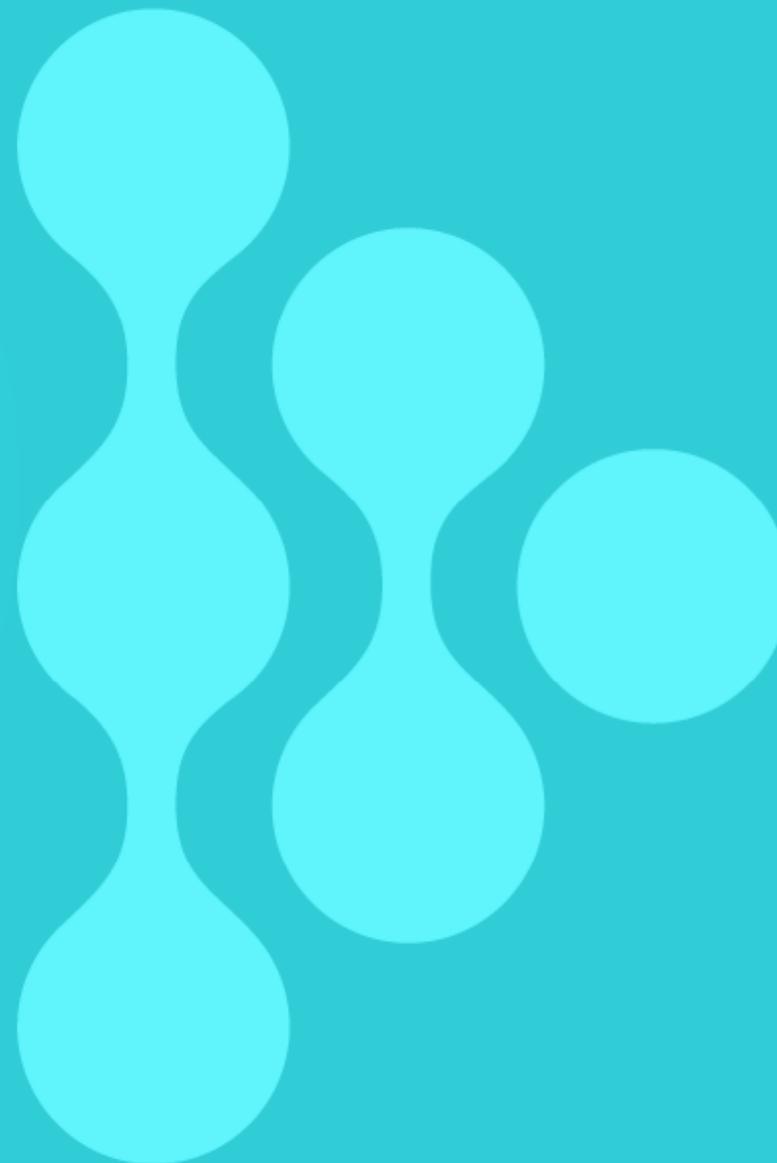
MORE



- Designed a national repositories usage-data for getting **statistics** (implementation process)

- Coordination of a research data **working group**

Scientific data management report for the repository community



“The preservation and reuse of scientific data in Spain. Report of the good practices working group”



Working Group on “*The depositing and management of data in Open Access*” as part of the RECOLECTA project.

Madrid: Spanish Foundation for Science and Technology, FECYT (2012)



The report is under a [Creative Commons License: Attribution-NonCommercial-NoDerivatives 3.0 Unported](#)

THE NATIONAL REPORT ON RESEARCH DATA



Who: participating institutions

- Polytechnic University of Catalonia (UPC)
- Carlos III University of Madrid (UC3M)
- Complutense University of Madrid (UCM)
- Spanish High Council for Scientific Research (CSIC)
- University of Alicante (UA)
- Centre for Scientific and Academic Services of Catalonia (CESCA)
- Juan March Institute
- Polytechnic University of Valencia (UPV)
- Spanish Foundation for Science and Technology (FECYT)

When: working group launched in 2011. Publication date: December 2012

Where: available at www.fecyt.es and www.recolecta.fecyt.es (English and Spanish version)

Why:

- As a response to the challenge about how to include **research data** in repositories.
- Reinforce the application of **Spanish Science Law** (regarding article 37 of OA dissemination).
- Analyze the **state of the art** of the management of scientific research data and its use in the field of repositories in Spain.

How: illustrates the Spanish **landscape in data management**, as well as the long way ahead to be covered in the **challenge** of building data repositories.

What:



RESEARCH DATA



Data are facts, observations or experiences on which an argument, theory or test is based. Data may be numerical, descriptive or visual. Data may be raw or analysed, experimental or observational. Data includes: laboratory notebooks; field notebooks; primary research data (including research data in hardcopy or in computer readable form); questionnaires; audiotapes; videotapes; models; photographs; films; test responses. Research collections may include slides; artefacts; specimens; samples. Provenance information about the data might also be included: the how, when, where it was collected and with what (for example, instrument). The software code used to generate, annotate or analyse the data may also be included.

Definition from Melbourne University (Australia)

Types of data (based on its origin):

- Observational:** historical records which can only be obtained in one place and at one moment in time.
- Experimental:** accompanies experiments from the planning and preparation stage until results are obtained.
- Computational:** accompanies simulations which tend to include input data, certain programmes and results.



ACTORS INVOLVED

- ❖ **Researchers:** they are data producers, authors and users of research data.
- ❖ **Universities & Research Centres:** to lay down **internal policy** for scientific data management (establish standards).
- ❖ **Institutional repositories:** basic role in **short-term** data storage. The use of standards is fundamental to facilitate interoperability between repositories and data centres.
- ❖ **Data centres:** they establish guides of good practices and the selection of data which must be preserved in the **long-term**. They protect the property rights.
- ❖ **Data managers:** to manage and promote the use of data once it is created **to ensure its use and availability** to be located and reused.
- ❖ **Users who reuse data:** they must **comply** with the **license conditions and permissions** of use of the researchers who have produced the data.
- ❖ **Funding agencies:** implement data **policies** with the actors involved, they set preservation dates, resolve problems of confidentiality, data protection and use of licenses.
- ❖ **Scientific publications:** publishers have started **linking journal articles** with their **research data**, in order to share this data with readers and researchers.



ASPECTS FOR AN ADEQUATE DATA MANAGEMENT



- Definition of the roles and responsibilities of the different parties → **Policies** (at the level of funding and institutional agencies).
- Data preservation in long term → **Financial resources**
- Data generation, use and preservation → Specialised **human resources**
- Equipment, high capacity data storage, high capacity networks → Coordinated **infrastructures** to guarantee its interoperability.
- Among the actors involved → **Cultural change**



Research data management should be carried out during the entire research process:
before the creation of data
during its creation and use
during its life cycle

DATA MANAGEMENT PLAN



Any data management plan, which should be **included** in every **funding proposal**, should consider the following:

- **Organisation and documentation** of the data according to standards. Data **storage, back-up, security, and sharing** mechanisms.
Datasets are stored and managed in interoperable network repositories.
- **Ethical and legal points:** legal rights, intellectual property, confidentiality, data protection (public, restricted, or private), agreements, licences should all be taken into account.
Data collection and databases are protected by intellectual property.
- **Data preservation plan** according to international standards. A final copy of the data should be archived in specialized data centres with different types of support.

CASE STUDIES IN SPAIN



Several **Spanish initiatives in scientific data management** have been studied:

- **Academic and professional literature**
- **Seminars and conferences**
- **Projects**

Case study: **ODISEA**



DiSEA

International registry on
Research Data

<http://odisea.ciepi.org/>

International registry of databases on research datasets.
It allows finding a specific data bank, or suggesting another one.

In addition it has other purposes:

- to provide basic data regarding the **banks** registered.
- to be the starting point for a future meta search engine for data groups, **opendatascience**.
- to create **knowledge** in management and production of research data.

ODiSEA has several deposits, including specialized banks, data libraries, repositories, and image banks.



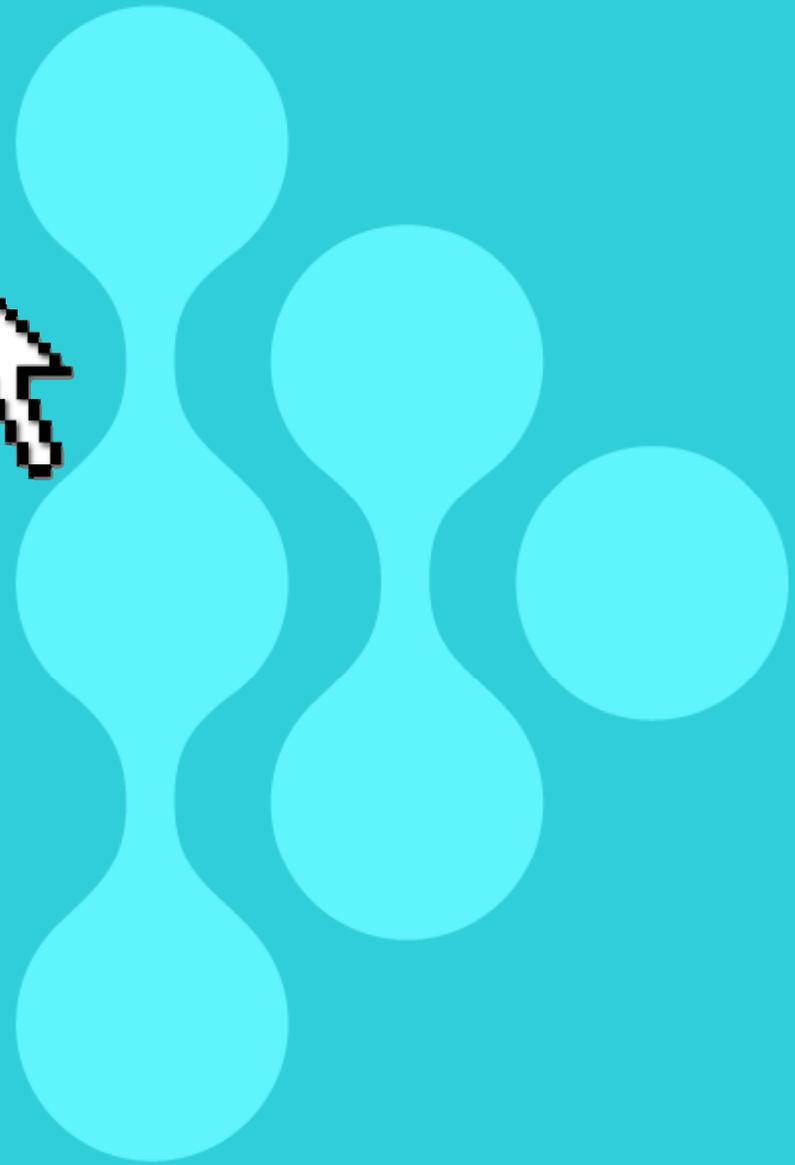
CONCLUSIONS

- This is the beginning...
- The time is right for a **national strategy**, at the highest level of authority and with all agents involved.
- It is required a **structure** which allows for the coordination and integration of information for scientific community.
- ...Spain has a long way ahead to be covered in the **challenge** of building data repositories.
- Research data should be directed, to the general tendency, by the handling of **large quantities of data** (BigData).

Thank you

recolecta@fecyt.es

www.recolecta.fecyt.es



SPANISH FOUNDATION FOR SCIENCE AND TECHNOLOGY

