DOI: https://doi.org/10.14324/000.learn.23

Case Study 22

Developing a Data Management Plan: a case study from Argentina.

Author: Fernando-Ariel López, Chief of Institutional Communications and Training Sectors (CAICYT – CONICET)
Email: flopez@conicet.gov.ar

22.1. INTRODUCTION

This case study will describe the experience of the *Centro Argentino de Información Científica y Tecnológica del Consejo Nacional de Investigaciones Científicas y Técnicas* (CAICYT-CONICET)[1] in the research, development and implementation of a Research Data Management Plan for the *Observatorio Nacional de la Degradación de Tierras y Desertificación* (ONDTyD)[2] and for CONICET.

22.2. A RESEARCH DATA MANAGEMENT PLAN BY CAICYT-CONICET

Several international organisations related to the field of science and technology (National Research Agencies, Funders, University Consortia, etc.) have started to require that research project funding applications be accompanied by a Research Data Management Plan (DMP) elaborated by the lead researcher and/or the group of researchers applying for funds.

The DMP allows for, on the one hand, the organisation of research data for researchers and, on the other, the ability to diagnose, characterise and predict, based on the information contained in the DMP, thus making it a valuable instrument for institutions managing Science and Technology. Furthermore, the DMP becomes a fundamental tool to assess and evaluate the potential impact (social, economic, cultural, etc.) implied in the development of research projects.

In Argentina there exists legislation and regulations that provide a framework and formalise the requirement for Data Management Plans (DMP):

- Data Management Plans are required by the law 26899 "Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos[3]", enacted in November 2013 and revised in November 2016;[4]
- Resolution CONICET 2705/15 and Institutional Repository Policies "CONICET Digital" require open access to publications and data funded by CONICET to researchers and institutes affiliated to CONICET;[5]
- The CONICET Data Policy [in development] will be aligned with the law 26899 and the Resolution CONICET 2705/15, requiring and regulating Data Management Plans and other aspects of data sharing.

1 English translation: Argentinean Centre of Science and Technology Information of the National Council of Science and Technology Research, http://www.caicyt-conicet.gov.ar/, last accessed 02/02/2017.
2 English translation: National Observatory of Soil Degradation and Desertification, http://www.desertificacion.gob.ar/, last accessed 02/02/2017.
3 English translation: "Creation of Institutional Open Access Repositories, Own or Shared".
4 SNRD: http://repositorios.mincyt.gob.ar/recursos.php; last accessed 02/02/17.
5 CONICET: http://ri.conicet.gov.ar/themes/Mirage/RD%2020150710-2705.pdf, last accessed 02/02/2017.

22.3. WHAT IS A DATA MANAGEMENT PLAN (DMP)?

A research data management plan (DMP) is a document elaborated by a researcher or a group of researchers, where the following is defined:

- What data will be created and how;
- How data will be described, organised, stored and managed;
- Who will be responsible for each of these activities;
- How data will be shared, explaining any use restriction that could apply.

The data management plan (DMP) is a live document, which evolves until the end of the research and its subsequent publication. Usually, a DMP is required at the following points in time: (1) at the time of requesting funding, accompanying the research project proposal; (2) once the project has started; (3) half way through the project; (4) at the end of the research project.

22.4. PROBLEMS WITH RESEARCH DATA

The National Observatory of Soil Degradation and Desertification (ONDTyD) is a national system for the evaluation and monitoring of soil across different scales (national, regional and pilot sites), based on an integral, interdisciplinary and participatory approach. It is sustained by a network of science and technology, and political organisations that provide data and knowledge and, at the same time, are also users of that information. Interactive maps, publications and an online geospatial data repository are being developed for their visualisation. The goal of ONDTyD is to identify the causes of desertification, to anticipate environmental risks and to collaborate in the restoration of affected ecosystems.

In the methodology developed, ONDTyD uses indicators of biophysical and socioeconomic vectors. However, the researchers were not aware of the lifecycle of their data, data management practices, documentation of their use, re-use, licenses or long-term preservation. The result was multiple versions of data from various sources and a lack of standardisation.

ONDTyD invited CAICYT-CONICET to collaborate in the improvement of these areas of their ongoing research project, whose indicators have varying levels of progress in terms of data collection.

22.5. DEVELOPMENT OF THE RESEARCH DATA MANAGEMENT PLAN

The first task was to discover the level of awareness of the field of data management amongst the researchers and to identify the research practices, documentation generated and group workflows at ONDTyD. We established regular meetings with the group coordinators, with specific researchers, as well as other meetings of a more general nature with the whole group. These meetings allowed us to understand, determine and reach consensus among participants about research data lifecycles and workflows.

We continued with the identification, analysis and comparison of research data management plans required by the Digital Curation Centre (DCC, UK), Horizon 2020 (European Union), the National Science Foundation (NSF, USA) and the Australian Research Council (ARC, Australia), as specified in the Information Laboratory of CAICYT – CONICET's working paper "Analysis of Data Management Plans".

The following action was to develop a Research Data Management Plan for ONDTyD, incorporating a *data dictionary* which was also developed (the dictionary specifies what information is required and incorporates definitions and alternative answers to the questions of the DMP). Furthermore, a section on *Best Practices* was included, referring to: (a) Data formats, (b) Folders and files structure, (c) Version control, and (d) Metadata schemas.

The ONDTyD-DMP includes the sections: (a) Administrative data; (b) Data collection; (c) Documentation and metadata; (d) Storage and security copies; (e) Selection and preservation; and (f) Data re-use.

22.6. PLATFORM FOR DMP MANAGEMENT, TRAINING AND SUPPORT

The next phase was to develop and to implement a digital tool to enable the research group (located across different provinces and cities in Argentina) to load, edit, and store and publish remotely a Data Management Plan (ONDTyD-DMP).

We identified and compared different online platforms for the management of a DMP. For diverse reasons, the tool selected was *DMPonline*[6] developed by the Digital Curation Centre (DCC, UK). Following acquisition, we then undertook the customisation and translation of the platform for use by the ONDTyD.

To ensure the implementation and correct use by all members of the Observatory, the next step was to deal with training and support:

- Development of a workshop entitled "Scientific Data: quality, normalisation and visualisation"
- Development of a virtual course about the ONDTyD-DMP, which incorporated information on the required sections and best practices (mentioned above).
- Establishment of a support helpline, to answer questions emerging in the process of filling out the ONDTyD-DMP

22.7. IMPACT

After meeting and exchanging information with ONDTyD, the combined workgroup deemed it necessary to reconsider some methodological decisions, resulting in the enhancement of data, their documentation and the management of research data created and to be generated in the future. In this way the group of researchers of ONDTyD improved their understanding and skills in the management of research data.

The Fundación Williams[7] - the project funder - make clear its interest in incorporating the DMP as an integral element in the process of receiving future research funding applications.

Based on the previous experience and the work carried out with ONDTyD, at the request of the Gerencia de Desarrollo Científico[8] of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina), we:

- Developed a Research Data Management Plan for CONICET. The DMP is of a generic nature and has 3 levels of information detail for the presentation of a project:

6 For further information about DMPonline, see DCC, UK: https://dmponline.dcc.ac.uk/about_us, last accessed 02/02/2017.
7 English translation: Williams Foundation: http://www.fundacionwilliams.org.ar/, last accessed 02/02/2017.
8 English translation: Scientific Development Division of the National Council of Science and Technology Research.

1. Global: Necessary general aspects that provide information about who is responsible for data, their basic characteristics and related legal aspects;
2. Management: Consideration of concrete aspects of management and decision-making for data documentation and re-use;
3. Data Set: Reference to specific aspects of scientific data generated in research projects funded by CONICET.

- Launched a CONICET DMP Pilot Survey, as part of the call for Strategic Projects of CONICET, with the following objectives: (a) to know the treatment of data generated by researchers, and (b) to draw attention to the interests and needs of researchers, research agencies and funders.
- Accepted an invitation to participate in the Consultant group on Scientific Data Management of CONICET, for the establishment of: (a) a Data Policy for CONICET, and (b) a Roadmap for the Management of Scientific Data at CONICET.

22.8. CONCLUSIONS

It is fundamental to acquire an appreciation of the discipline and to know research practices and workflows of specialised research groups in the thematic area. It is also important to allow for constant feedback from research groups and/or researchers in each thematic area to reach consensus in regard to data lifecycle, data management plans, metadata, etc.

The DMP enables researchers to plan the creation and collection, as well as the organisation, of data. A good DMP will multiply the possibilities for data use, re-use, and the impact of research in the scientific community and in society at large.

The requirement of a DMP by institutions that manage and fund research in Science and Technology constitutes an important input for diagnosis and prediction, necessary for the development of infrastructure and for the evaluation and measurement of potential and/or real impact (social, economic cultural, etc.) that a piece of research and its funding imply.

ONDTyD's digital platform to manage their DMPs was developed and implemented. The platform should be flexible, modular and interoperable with repositories of data, publications, etc.

Training and support of the researchers at ONDTyD have proved vital elements to success with the implementation and development of DMPs, the implementation of which will facilitate future use and re-use of data.

The Research Data Management Plan of the Observatorio Nacional de la Degradación de Tierras y Desertificación (PGD – ONDTyD) and the Digital Platform for DMP Management were developed by the Centro Argentino de Información Científica y Tecnológica (CAICYT) of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) with the support and funding of Fundación Williams.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.