

# Plan de gestión de datos de investigación: una propuesta argentina del CAICYT-CONICET

**FERNANDO ARIEL LÓPEZ**

*Centro Argentino de Información Científica y Tecnológica (CAICYT)*

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)



## Resumen

Estudio de caso del Centro Argentino de Información Científica y Tecnológica del Consejo Nacional de Investigaciones Científicas y Técnicas (CAICYT-CONICET) en investigación, desarrollo e implementación de un Plan De Gestión de Datos de Investigación para el Observatorio Nacional de la Degradación de Tierras y Desertificación (ONDTyD) y para el CONICET. La primera de las tareas fue adquirir conocimiento de la disciplina y conocer las prácticas de investigación, documentación producida y flujos de trabajo del grupo del Observatorio Nacional de la Degradación de Tierras y Desertificación (ONDTyD). Mantuvimos reuniones frecuentes con los coordinadores del grupo, algunas puntuales con algunos investigadores principales y reuniones generales con todos el grupo. Dichas reuniones nos sirvieron para entender, determinar y acordar entre todos el flujo y ciclo de vida de los datos de investigación. Continuamos, por realizar un relevamiento, análisis y comparación de planes de gestión de datos de investigación requeridos por Digital Curation Center (DCC, UK), Horizon2020 (Unión Europea), National Science Fundation (NSF, EEUU) y Australian Research Council (ARC, Australia) que detallamos en el documento de trabajo “Crosswalk de Data Management Plan” del Laboratorio de Información del CAICYT-CONICET. El desarrollo de un Plan de Gestión de Datos de Investigación del ONDTyD, donde desarrollamos un *diccionario de datos* (detallando la información requerida sobre los datos, definiciones y opciones). El PGD-ONDTyD incluye información sobre los bloques: (a) Datos administrativos; (b) Recolección de datos; (c) Documentación y metadatos; (d) Almacenamiento y copias de seguridad; (e) Selección y preservación; y (f) Re-

uso de Datos. Además, incluimos una sección de *Buenas prácticas* respecto a: (a) Formato de Datos, (b) Estructura de Carpetas y Archivos, (c) Control de Versiones, y (d) Esquemas de Metadatos.

### Palabras clave

Ciencia abierta; datos abiertos; datos de investigación; plan de gestión de datos; CAICYT-CONICET.

### Abstract

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, Argentinean Center of Science and Technology Information of the National Council of Science and Technology Research) in the research, development and implementation of a Research Data Management Plan for the Observatorio Nacional de la Degradación de Tierras y Desertificación (ONDTyD, National Observatory of Soil Degradation and Desertification) and for CONICET. The first task was to acquire knowledge in the field of data management and identify research practices, documentation generated and group workflows at ONDTyD. We sustained regular meetings with the group coordinators, some of them with specific researchers and other of a more general nature with the whole group. The meetings allowed us to understand, determine and reach consensus among participants about research data lifecycle and workflows. We continued with the identification, analysis and comparison of research data management plans required by Digital Curation Center (DCC, UK), Horizon2020 (European Union), National Science Foundation (NSF, USA) and the Australian Research Council (ARC, Australia), as specified in the working paper “Crosswalk de Data Management Plan” of the Information Laboratory of CAICYT-CONICET. The following action was to develop a Research Data Management Plan of ONDTyD, in which a *data dictionary* was 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.

### Keywords

Open science; open data; research data; data management plan; CAICYT-CONICET.

## Estudio de caso: CAICYT-CONICET (Argentina)

---

En el siguiente estudio de caso contaremos la experiencia del Centro Argentino de Información Científica y Tecnológica del Consejo Nacional de Investigaciones Científicas y Técnicas (CAICYT-CONICET) en investigación, desarrollo e implementación de un Plan De Gestión de Datos de Investigación para el Observatorio Nacional de la Degradación de Tierras y Desertificación (ONDTyD) y para el CONICET.

### Plan de gestión de datos de investigación del CAICYT-CONICET

---

Diversos organismos internacionales relacionados con el ámbito científico-tecnológico (Agencias Nacionales de Investigación, Financiadores, Consorcios de Universidades, etc.) comenzaron a solicitar junto a la solicitud del proyecto a financiar, que los mismos sean acompañados por un Plan de Gestión de Datos de Investigación (PGD, Data Management Plan – DMP) elaborados por el investigador principal y/o grupo de investigación solicitante.

El PGD permite, por un lado, una organización de la gestión de los datos para el Investigador, y por otro, la capacidad de diagnóstico, caracterización y predicción acerca de la información que el PGD contiene, siendo un insumo de valor para una Institución que gestiona Ciencia y Tecnología. Además, el PGD se vuelve una herramienta fundamental para evaluar y medir el potencial impacto (social, económico, cultural, etc.) que implica el desarrollo de dicha investigación.

En Argentina, ya contamos con legislación y normativa que contextualiza y formaliza la solicitud de los Planes de Gestión de Datos (PGD):

- Es una exigencia de la Ley 26.899 [Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos](#), sancionada (2013) y reglamentada (2016).
- Resolución CONICET 2705/15 y Políticas del Repositorio Institucional [CONICET Digital](#).
- Política Datos de CONICET [en elaboración].

## ¿Qué es un Plan de Gestión de Datos científicos (PGD)?

---

Un plan de gestión de datos científicos (Data Management Plan - DMP) es un documento elaborado por el investigador o grupo de investigación donde se define:

- Qué datos van a ser creados y cómo.
- Cómo se van a describir, organizar, almacenar y gestionar los datos.
- Quién o quiénes serán los responsables de realizar cada una de estas actividades.
- De qué forma van a ser compartidos, explicando cualquier restricción de uso que pueda ser aplicada.

El plan de gestión de datos de investigación (PGD) es un documento vivo, va evolucionando hasta el final de la investigación y su posterior publicación. Un PGD se suele pedir en los siguientes momentos: (1) Al solicitar financiamiento, acompañando al proyecto de investigación. (2) Iniciado el proyecto. (3) A mitad del proyecto. (4) Finalizado el proyecto.

## Problemas con datos de investigación

---

El Observatorio Nacional de Degradación de Tierras y Desertificación (ONDTyD) es un sistema nacional de evaluación y monitoreo de tierras a diferentes escalas (nacional, regional y de sitios piloto), basado en un abordaje integral, interdisciplinario y participativo. Está sustentado en una red de organizaciones científico-tecnológicas y políticas que proveen datos y conocimientos y al mismo tiempo son usuarios de la información. Para su visualización se desarrollan mapas interactivos, publicaciones y un repositorio de datos geoespaciales en línea. El objetivo del ONDTyD es identificar las causas de la desertificación, prever los riesgos y cooperar en la remediación de los ecosistemas afectados.

En su metodología desarrollada, el Observatorio trabaja con indicadores con dimensiones biofísicas y socioeconómicas. Pero sus investigadores no tenían en claro el ciclo de vida de sus datos, ni como gestionarlos, ni la documentación que acompaña el uso, re-uso, licencias, exposición o preservación a largo plazo (múltiples versiones, fuentes varias, falta de normalización, etc.).

Desde la ONDTyD nos convocaron para colaborar a mejorar estos aspectos del proyecto de investigación comenzado y finalizado en el caso de algunos indicadores.

### **Desarrollo del Plan de Gestión de Datos de investigación**

---

La primera de las tareas fue adquirir conocimiento de la disciplina y conocer las prácticas de investigación, documentación producida y flujos de trabajo del grupo del Observatorio Nacional de la Degradación de Tierras y Desertificación (ONDTyD). Mantuvimos reuniones frecuentes con los coordinadores del grupo, algunas puntuales con algunos investigadores principales y reuniones generales con todos el grupo. Dichas reuniones nos sirvieron para entender, determinar y acordar entre todos el flujo y ciclo de vida de los datos de investigación.

Continuamos, por realizar un relevamiento, análisis y comparación de planes de gestión de datos de investigación requeridos por Digital Curation Center (DCC, UK), Horizon2020 (Unión Europea), National Science Foundation (NSF, EEUU) y Australian Research Council (ARC, Australia) que detallamos en el documento de trabajo *Crosswalk de Data Management Plan* del Laboratorio de Información del CAICYT-CONICET.

El siguiente paso fue desarrollar un Plan de Gestión de Datos de Investigación del ONDTyD, donde desarrollamos un *diccionario de datos* (detallando la información requerida sobre los datos, definiciones y opciones). Además, incluimos una sección de *Buenas prácticas* respecto a: (a) Formato de Datos, (b) Estructura de Carpetas y Archivos, (c) Control de Versiones, y (d) Esquemas de Metadatos.

El PGD-ONDTyD incluye información sobre los bloques: (a) Datos administrativos; (b) Recolección de datos; (c) Documentación y metadatos; (d) Almacenamiento y copias de seguridad; (e) Selección y preservación; y (f) Re-uso de Datos.

### **Plataforma de gestión de PGD, formación y soporte**

---

La siguiente etapa, fue desarrollar e implementar una herramienta digital donde el grupo de investigación (ubicado en distintas provincias y ciudades a

los largo de Argentina) pueda cargar, editar, almacenar y publicar de manera remota un Plan de Gestión de Datos (PGD-ONDTyD).

También realizamos un relevamiento y análisis de diferentes plataformas en línea para la Gestión de PGD. Por diversos motivos, la herramienta seleccionada fue el [DMPonline](#) desarrollada por la Digital Curation Center (DCC, UK). Encaramos la localización y traducción de la plataforma para la ONDTyD.

Para asegurar la implementación y el uso correcto por parte de todos los miembros del grupo del Observatorio, el paso restante fue encarar la formación y soporte:

- Desarrollo de un workshop: Datos Científicos: calidad, normalización y visualización.
- Desarrollo de un curso virtual sobre el PGD-ONDTyD donde se incluye la información sobre los bloques requeridos y las buenas prácticas (antes descriptas).
- Línea de Soporte, para preguntas surgidas durante el proceso de llenado del PGD-ONDTyD.

## Impacto

---

A partir de las reuniones e intercambios con el ONDTyD, el grupo consideró necesario reconsiderar algunas decisiones metodológicas, resultando en la mejora de los datos, la documentación de los mismos y gestión de los datos de investigación ya generados y los aún por generar. El grupo de investigadores del ONDTyD mejoraron su comprensión y competencias en la gestión de datos científicos.

La Fundación Williams, financiadora del proyecto, se demostró interesada en incorporar esta herramienta como insumo fundamental al momento de recibir solicitudes de financiamiento de proyectos.

En base a la experiencia y trabajo realizado con el ONDTyD, a pedido de la Gerencia de Desarrollo Científico del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina):

- Desarrollamos un Plan de Gestión de Datos de Investigación para el CONICET. Dicho PGD es general y cuenta con 3 niveles de detalle de presentación de un proyecto: (1) Global: aspectos generales

imprescindibles que permiten conocer quiénes son los responsables de los datos, sus características básicas y los aspectos legales ligados a ellos. (2) Gestión: contempla aspectos concretos de la administración y toma de decisiones para documentación y re-uso. (3) Set de datos: todo lo referido a aspectos específicos de los datos científicos generados en proyectos de investigación financiados por el CONICET.

- Lanzamiento de una Encuesta Piloto PGD de CONICET en la convocatoria de Proyectos Estratégicos de CONICET, que tiene como objetivos: (a) Conocer el tratamiento de los datos de los investigadores; y (c) Relevar el interés y las necesidades de los investigadores, agencias de investigación y financiadores.
- Recepción de invitación a integrar un Grupo consultor sobre Gestión de Datos Científicos de CONICET para la elaboración: (a) de una Política de Datos en CONICET; y (b) de una Hoja de ruta.

## Conclusiones

---

Es fundamental adquirir conocimiento de la disciplina, conocer las prácticas de investigación y flujos de trabajo de los grupos de investigación especializados en el o las áreas temáticas. Asimismo, tener feedback constante de grupos de investigación y/o investigadores por área temática para acordar y consensuar el ciclo de vida de datos, plan de gestión de datos, metadatos, etc.

EL PGD permite a los investigadores planificar la creación y recolección, además de organizar la gestión de los datos. Un buen PGD multiplica la posibilidad de uso, re-uso e impacto de la investigación en la comunidad científica y en la sociedad en general.

La exigencia de un PGD por parte de las instituciones que gestionan y financian la investigación en Ciencia y Tecnología constituye un insumo de valor para el diagnóstico y predicción para el desarrollo de infraestructuras y para evaluar y medir el potencial y/o real impacto (social, económico, cultural, etc.) que implica el desarrollo de dicha investigación y su financiación.

Desarrollo e implementación de una plataforma digital para Gestionar los PGD del ONDTyD, dicha herramienta debe ser flexible, modular e interoperable con Repositorios de datos, de publicaciones, etc.

La formación y soporte a los investigadores del ONDTyD son elementos fundamentales para lograr el éxito de la implementación y desarrollo de los PGD, lo que facilitará el posterior uso y re-uso de los datos.

## Bibliografía

- COUTO CORREA, FABIANO. (2016). *Gestión de datos de investigación*. Barcelona: UOC. 156 p. (EPI Scholar; no. 6). ISBN 978-84-9116-430-2
- GONZÁLEZ, LUIS-MILLÁN; SAORÍN, TOMÁS; FERRER-SAPENA, ANTONIA; ALEIXANDRE-BENAVENT, RAFAEL; PESET, FERNANDA (2013). Gestión de datos de investigación: infraestructuras para su difusión. *El Profesional de la Información*, 22(5), 415-423. Disponible en <<http://eprints.rclis.org/20912/>>
- GRUPO DE TRABAJO DE «DEPÓSITO Y GESTIÓN DE DATOS EN ACCESO ABIERTO» DEL PROYECTO RECOLECTA. (2012). La conservación y reutilización de los datos científicos en España. Informe del grupo de trabajo de buenas prácticas. Madrid: Fundación Española para la Ciencia y la Tecnología, FECYT. Disponible en <[http://www.recolecta.net/buscador/documentos/informe\\_datos\\_cientificos\\_en\\_esp.pdf](http://www.recolecta.net/buscador/documentos/informe_datos_cientificos_en_esp.pdf)>
- LERU RESEARCH DATA WORKING GROUP. (2013). LERU Roadmap for Research Data. Leuven: LERU. Disponible en <[http://www.leru.org/files/publications/API4\\_LERU\\_Roadmap\\_for\\_Research\\_data\\_final.pdf](http://www.leru.org/files/publications/API4_LERU_Roadmap_for_Research_data_final.pdf)>
- NINA-ALCOCER, VICTOR; BLASCO-GIL, YOLANDA; PESET, FERNANDA (2013). Los inicios del datasharing: guía práctica para compartir datos de investigación. *El Profesional de la Información*, noviembre-diciembre, 22(6), 562-568. Disponible en <<http://eprints.rclis.org/20907/1/datasharing.pdf>>
- PESET, F. ; R. ALEIXANDRE-BENAVENT; BLASCO-GIL, Y.; FERRER-SAPENA, A. (2017). Datos abiertos de investigación. Camino recorrido y cuestiones pendientes. *Anales de Documentación*, 20(1). Disponible en <<http://revistas.um.es/analesdoc/article/view/272101/210391>>