

La Gestión de Datos de Investigación

Marisa Pérez Aliende

Universidad Autónoma de Madrid

mp.aliende@uam.es

Sevilla, 13 de Junio 2017





La gestión de datos: Planificación

- El Plan de Gestión de Datos
- H2020: Directrices para la redacción de un PGD
- Herramientas: PGDonline / DMPonline
- Tendencias futuras: maDMP

¿Qué es el Plan de Gestión de Datos?

“El Plan de Gestión de Datos es un documento formal que describe los datos producidos en el curso de un proyecto de investigación, tanto en la fase activa, como después de que haya terminado.”

(Sarah Jones, DCC)

“Documento en el que se describe cómo se gestionan los datos de investigación durante y después del proyecto de investigación”
(UE. Directrices para Gestión de Datos en H2020. 2013)

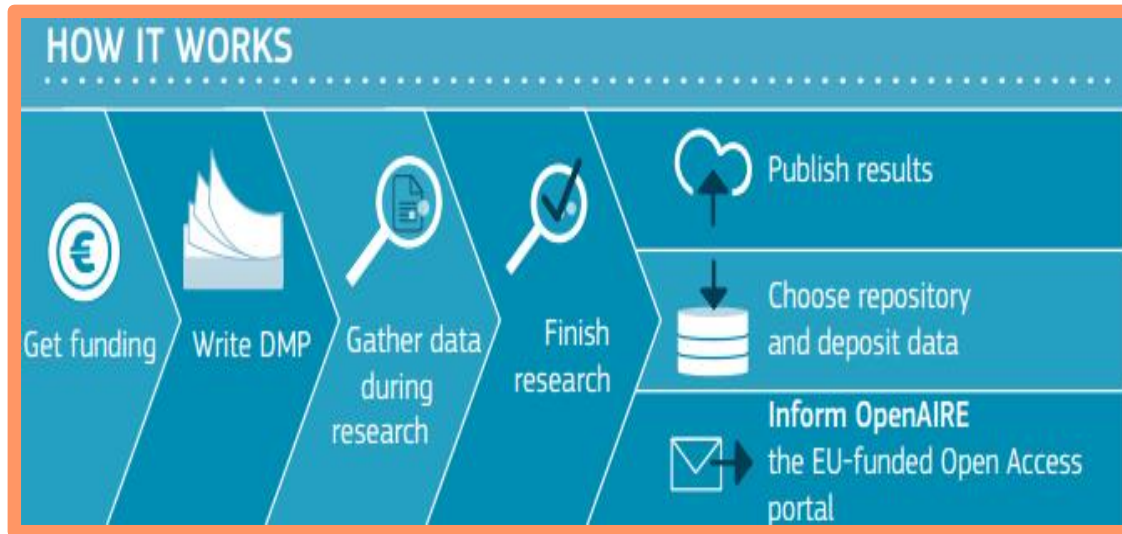
“EL PGD es un elemento clave de una buena gestión de datos. Describe el ciclo de vida de la gestión de datos que se van a recopilar, procesar o generar por un proyecto de H2020. Como parte de hacer los datos de investigación *findable, accessible, interoperable y reusable* (FAIR). Debiera de incluir información sobre.”:

- El manejo de los datos de investigación durante y después del proyecto.
- Qué datos serán recopilados, procesados y/o generados.
- Qué metodologías y normas/estándares se les van a aplicar.
- Cómo los datos se van preservar, incluso después de finalizado el proyecto.

El DMP se requiere para todos los proyectos que participan en el piloto ORD (una primera versión a los 6 meses), a no ser que adopten el *opt out*, pero incluso estos pueden optar por suministrar el DMP de forma voluntaria.

(UE. Guidelines on FAIR Data Management in H2020. 2016)

EI PGD



(Open Reserch Data in H2020
https://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf)

El PGD: ¿Por qué se gestionan los datos?

A la hora de escribir un PGD hay que pensar detenidamente sobre la gestión de datos.

¿Por qué se gestionan los datos?:

- Permiten verificar o refinar los resultados de investigación publicados
- Reduce el posible fraude científico
- Promueve nueva investigación
- Proporciona nuevos recursos para la formación de investigadores
- Desalienta la redundancia involuntaria (existencia de datos que son adicionales a las actuales datos, pudiendo ser una copia completa de los datos, y permite la corrección de errores de datos almacenados o transmitidos)

(Edinburgh University Data Library Research Data Management Handbook)

¿Por qué realizar un PGD?

- Requerimiento obligatorio de la agencia financiadora y/o de la institución
- Parte de unas buenas prácticas de investigación: los datos necesitan de una gestión a lo largo de todo el ciclo de investigación

El PGD: Beneficios

Entre los beneficios que aporta la redacción de un plan de gestión de datos podemos mencionar:

- Encontrar los datos cuando se necesitan
- Garantizar la continuidad del proyecto, independientemente de la participación de los investigadores
- Garantizar la integridad de la investigación y reproductibilidad
- Evitar duplicaciones y tareas innecesarias
- Aumentar la eficiencia y la calidad de la investigación
- Asegurar que los datos sean precisos, completos y fidedignos
- Mantener el conjunto de datos generados que permita la validación de los resultados
- Compartir los datos, permitiendo un alto nivel de colaboración y de avance en la investigación
- Ahorrar tiempo y recursos a largo plazo
- Mejorar la protección de datos y minimizar el riesgo de pérdida
- Garantizar su conservación y preservación
- Si los datos están en abierto, tendrán visibilidad
- Otros investigadores podrán citarlos y la investigación obtendrá más prestigio

El PGD: Cuestiones básicas I

Antes de empezar a crear su PGD es conveniente que previamente se consideren una serie de cuestiones relacionadas con la gestión de datos en el proyecto:

- ¿Qué datos se van a recoger?
 - Tipo de datos, volumen, almacenamiento
- ¿Cómo se van a recoger o crear?
 - Métodos de colección, estándares, organización de ficheros, control, protocolos de garantía de calidad durante y después de la colecta
- ¿Qué documentación y metadatos acompañarán a los datos?
 - Describir la documentación para interpretar y usar: descripción de la metodología; metadatos; libros de códigos (codebooks); cuestionarios o instrumentos; análisis de los procedimientos
- ¿Cómo gestionar cualquier tema ético?
 - Proteger confidencialidad: manejo y almacenamiento de datos sensibles, restricciones de acceso
- ¿Cómo gestionar el copyright y la propiedad intelectual?
 - Identificar al propietario de los datos y las condiciones de uso, discutir permisos con los productores de datos
- ¿Cómo manejar el acceso y la seguridad?
 - Definir el acceso restringido, describiendo medidas de seguridad y el cumplimiento de las normas de esas medidas

El PGD: Cuestiones básicas II

- ¿Qué datos debieran de retenerse, compartir y/o preservarse?
 - Anotar el valor potencial de los datos, los esfuerzos que se requieren para preparar los datos para preservar y acceder
- ¿Cuáles son los plazos de preservación a largo plazo del dataset?
 - Identificar un repositorio para archivar los datos, planes para preparar y documentar los datos y asegurarse de que serán preservados y usables en el futuro
- ¿Cómo compartir los datos?
 - Identificar los mecanismos necesarios para compartir los datos, describir como otros pueden encontrar los datos y como los ficheros se entregarán, especificar el reconocimiento y requerimientos de citación
- ¿Se requieren restricciones a compartir los datos?
 - Discutir preocupaciones de privacidad, habilitar la fecha en que se compartirán, anonimización de los datos, acuerdos de uso, explicar períodos de embargo
- ¿Quién será el responsable de la gestión de los datos?
 - Identificar al encargado de supervisar e implementar el PGD
- ¿Qué recursos necesitarás para implementar tu plan?
 - En función del tipo de dato (tamaño, geográficos...): identificar personal experto o equipo especializado, considerar los costes de la gestión durante y después del proyecto

El PGD: Cuestiones básicas III

En este punto es interesante consultar la lista de control del Digital Curation Center que presenta los principales temas y preguntas que los investigadores pueden querer cubrir a la hora de escribir un PGD.



Checklist for a Data Management Plan, v4.0

Please cite as: DCC. (2013). *Checklist for a Data Management Plan. v.4.0*. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/data-management-plans>

DCC Checklist	DCC Guidance and questions to consider
Administrative Data	
ID	A pertinent ID as determined by the funder and/or institution.
Funder	State research funder if relevant
Grant Reference Number	Enter grant reference number if applicable [POST-AWARD DMPs ONLY]
Project Name	If applying for funding, state the name exactly as in the grant proposal.
Project Description	<p>Questions to consider:</p> <ul style="list-style-type: none"> - What is the nature of your research project? - What research questions are you addressing? - For what purpose are the data being collected or created? <p>Guidance:</p> <p>Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created.</p>
PI / Researcher	Name of Principal Investigator(s) or main researcher(s) on the project.
PI / Researcher ID	E.g ORCID http://orcid.org/
Project Data Contact	Name (if different to above), telephone and email contact details
Date of First Version	Date the first version of the DMP was completed
Date of Last Update	Date the DMP was last changed
Related Policies	<p>Questions to consider:</p> <ul style="list-style-type: none"> - Are there any existing procedures that you will base your approach on? - Does your department/group have data management guidelines? - Does your institution have a data protection or security policy that you will follow? - Does your institution have a Research Data Management (RDM) policy? - Does your funder have a Research Data Management policy? - Are there any formal standards that you will adopt? <p>Guidance:</p> <p>List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security. Some of the information you give in the remainder of the DMP will be determined by the content of other policies. If so, point/link to them here.</p>

1. Administrative Data
2. Data collection
3. Documentation and metadata
4. Ethics and legal compliance
5. Storage and backup
6. Selection and preservation
7. Data sharing
8. Responsibilities and resources

DCC. Checklist for a Data Management Plan v4.0. Edinburgh: Digital Curarion Centre. 2013.
 Available online: http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf

El PGD: Directrices en H2020 I

Los proyectos que participan en el Piloto (ORD Pilot) tienen que proporcionar una primera versión del PGD en los 6 primeros meses que se actualizará a lo largo del proyecto siempre que se produzcan cambios significativos, y/o por lo menos a la mitad y final del proyecto.

(OpenAire Research Data Management briefing paper. Abril 2017)

H2020 proporciona una plantilla con los puntos que debe cumplir el PGD, que se actualizará a medida que la política evolucione.

Serie de preguntas:

1. **Resumen de datos** (origen (generados o recolectados), tipos, formatos, volumen)
2. **Datos FAIR. Hacer los datos:**
 - a. **Y metadatos localizables** (asignación de metadatos, identificadores persistentes, palabras clave)
 - b. **Accesibles en abierto** (qué datos, cómo, restricciones, documentación y softwares de acceso)
 - c. **Interoperables**, permitiendo el intercambio y la reutilización entre los investigadores, instituciones, etc. (vocabularios estándar, metodologías)
 - d. **Incrementar la reutilización de los datos** (a través de licencias)
3. **Localización y recursos** (costes de hacer los datos FAIR)
4. **Seguridad de los datos** (recuperación, almacenamiento y transferencia de datos sensibles)
5. **Aspectos éticos** (impacto al compartir los datos, consentimiento, datos personales)
6. **Otros** (referencia a otros procedimientos si los hubiera sobre la gestión de datos a nivel nacional/financiador/departamento)

El PGD: Directrices en H2020 II

SUMMARY TABLE 1

FAIR Data Management at a glance: issues to cover in your Horizon 2020 DMP

This table provides a summary of the Data Management Plan (DMP) issues to be addressed, as outlined in Annex I. You should refer to the annex and the main text of the guidelines for further guidance.

DMP component	Issues to be addressed
1. Data summary	<ul style="list-style-type: none"> State the purpose of the data collection/generation Explain the relation to the objectives of the project Specify the types and formats of data generated/collected Specify if existing data is being re-used (if any) Specify the origin of the data State the expected size of the data (if known) Outline the data utility: to whom will it be useful
2. FAIR Data	
2.1. Making data findable, including provisions for metadata	<ul style="list-style-type: none"> Outline the discoverability of data (metadata provision) Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? Outline naming conventions used Outline the approach towards search keyword Outline the approach for clear versioning Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how

2.2 Making data openly accessible	<ul style="list-style-type: none"> Specify which data will be made openly available? If some data is kept closed provide rationale for doing so Specify how the data will be made available Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Specify where the data and associated metadata, documentation and code are deposited Specify how access will be provided in case there are any restrictions
2.3. Making data interoperable	<ul style="list-style-type: none"> Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
2.4. Increase data re-use (through clarifying licences)	<ul style="list-style-type: none"> Specify how the data will be licenced to permit the widest reuse possible Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why Describe data quality assurance processes Specify the length of time for which the data will remain re-usable
3. Allocation of resources	<ul style="list-style-type: none"> Estimate the costs for making your data FAIR. Describe how you intend to cover these costs Clearly identify responsibilities for data management in your project Describe costs and potential value of long term preservation
4. Data security	<ul style="list-style-type: none"> Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	<ul style="list-style-type: none"> To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	<ul style="list-style-type: none"> Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

(EC. "H2020 Programme: Guidelines on FAIR Data Management in H2020. Julio 2016. v3)

El PGD: Directrices H2020 III

La Comisión Europea hace referencia a una serie de Herramientas que pueden ser útiles a la hora de elaborar un PGD:

- El **Metadata Standards Directory** elaborado por un GT de la Research Data Alliance
 - Contiene los estándares y herramientas para aplicar metadatos en función de la materia de los datos
- La herramienta **EUDAT B2SHARE** que incluye un asistente para construir licencias y selección de la adecuada
 - EUDAT (Research Data Services & Technology Solutions). Es una plataforma de asistencia a los investigadores europeos de cualquier disciplina a lo largo del ciclo de vida de los datos:
 - **B2DROP**: almacenamiento e intercambio de datos entre investigadores colegas o del equipo
 - **B2SHARE**: almacenar y publicar datos de investigación
 - **B2SAFE**: replica los datos de manera segura
 - **B2STAGE**: transferencia de datasets
 - **B2FIND**: encontrar datos de investigación
 - **B2HANDLE**: registrar los datos de investigación
 - **B2ACCESS**: identidad y autorización
- Elección de repositorio:
 - **Registro de repositorio de datos** (re3data.org)
 - Propuesta de **Zenodo**
 - Herramienta para la elaboración del PGD: **DMPonline**

EI PGD:



EUDAT

EUDAT B2 service suite

- B2DROP**
Sync and Exchange Research Data
- B2SHARE**
Store and Share Research Data
- B2SAFE**
Replicate Research Data Safely
- B2STAGE**
Get Data to Computation
- B2FIND**
Find Research Data

Covering both access and deposit, from informal data sharing to long-term archiving, and addressing identification, discoverability and computability of both long-tail and big data, EUDAT's services **address the full lifecycle of research data**

<https://www.eudat.eu/>

GROOTVELD, M. y ROSS-HELLAWER, T. "Open Research Data in H2020". 2017.
https://www.slideshare.net/OpenAIRE_eu/20170530open-research-data-in-horizon-2020

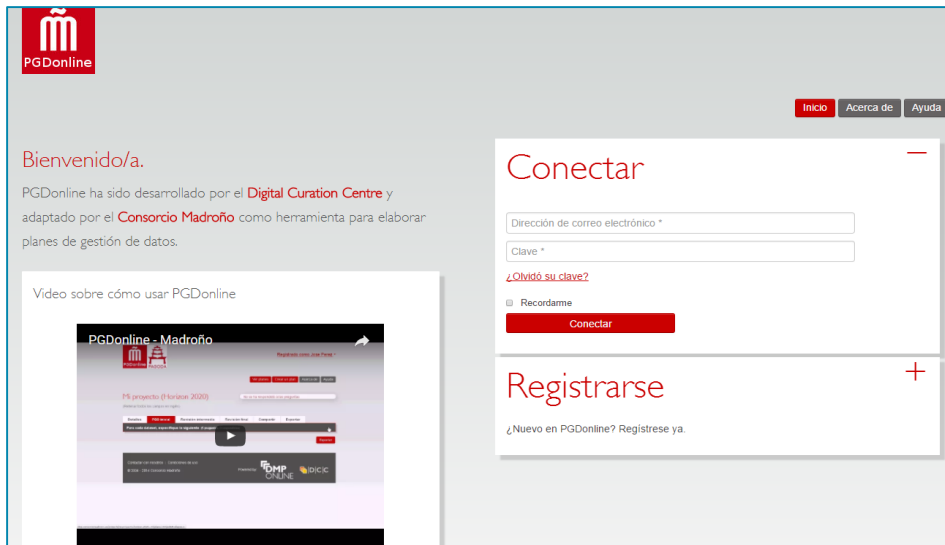
El PGD: PGDOnline

DMPonline es la herramienta para la elaboración de Planes de Gestión de Data del DCC, guía para ayudar a los investigadores. Incluye un gran número de plantillas con los requisitos de las agencias de financiación que afectan al Reino Unido.

PGDonline es la versión española, la aplicación está traducida y adaptada al castellano por el Consorcio Madroño. Al igual que Madroño existen otros DMPonline nacionales como DMP OPIDoR (Francia) o DMP TUULI (Finlandia).

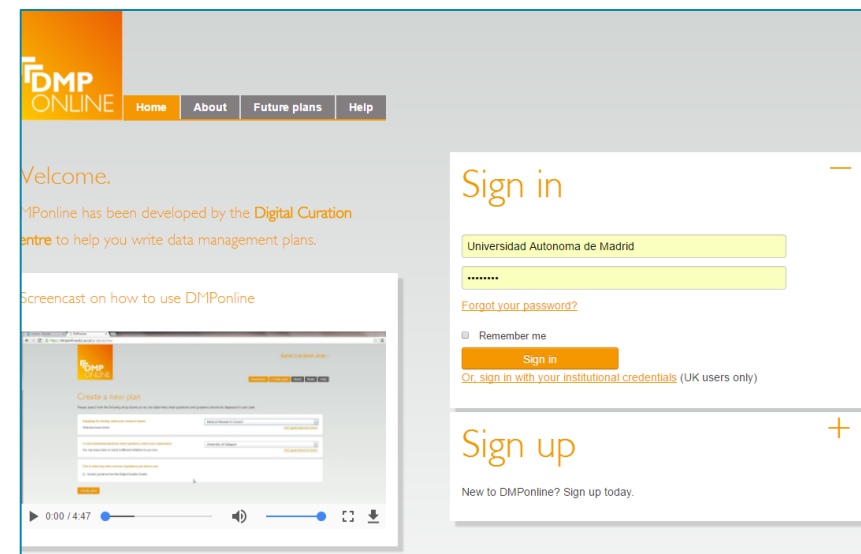
Está disponible en el Portal del Investigador del Consorcio Madroño InvestigaM, en la plataforma **PaGoDa** encontramos la herramienta, así como material de apoyo para la elaboración de un PGD.

Vídeo sobre como usar PGDonline



<http://http://pgd.consorciomadrono.es/>

Vídeo sobre como usar DMPonline



<https://dmponline.dcc.ac.uk/>

El PGD: PGDonline

Bienvenido/a.

PGDonline ha sido desarrollado por el **Digital Curation Centre** y adaptado por el **Consorcio Madroño** como herramienta para elaborar planes de gestión de datos.

Video sobre cómo usar PGDonline

Conectar

Dirección de correo electrónico *

Clave *

[¿Olvidó su clave?](#)

Recordarme

Conectar

Registrarse

¿Nuevo en PGDonline? Regístrese ya.

Plan creado con éxito

Mi proyecto (Horizon 2020)

(Rellenar todos los campos en inglés)

Detalles | PGD inicial | Revisión intermedia | Revisión final | Compartir | Exportar

Por favor, rellene los detalles básicos del proyecto abajo y haga clic en 'Actualizar' para guardarlos

Nombre de proyecto:

ID:

Número de la subvención:

Investigador principal:

ID del Investigador Principal:

Datos del contacto del proyecto:

Descripción:

Cancelar

EI PGD: PGDonline

DMP checklist 2013.' The modal has 'Cancel' and 'Create plan' buttons. Below the modal, a dropdown menu for 'Funder' is open, listing various funding sources such as BBSRC, CRUK, ESRC, EPSRC, Horizon 2020, MRC, NSF, and NERC. The ESRC option is currently selected." data-bbox="71 188 890 895"/>

Es seguro | <https://dmponline.dcc.ac.uk/projects/new> | Signed in as Marisa Perez

DMP ONLINE | View plans | **Create plan** | About | Future plans | Help

Create a new plan

Please select from the following drop-downs so we can determine what questions and guidance should be displayed in your plan.

If you aren't responding to specific requirements from a funder or an institution, [select here to write a generic DMP](#) based on the most common themes.


If applying for funding, select your research funder.
Otherwise leave blank.

Funder

- Biotechnology and Biological Sciences Research Council (BBSRC)
- Cancer Research UK (CRUK)
- Economic and Social Research Council (ESRC)**
- Engineering and Physical Sciences Research Council (EPSRC)
- European Commission (Horizon 2020)
- Medical Research Council (MRC)
- National Science Foundation (USA)
- Natural Environment Research Council (NERC)

Contact us | Terms of use | DMPonline previous version
© 2004 - 2017 Digital Curation Centre (DCC)

EI PGD: PGDonline

View plans Create plan About Future plans Help

Create a new plan

Please select from the following drop-downs so we can determine what questions and guidance should be displayed in your plan.

If you aren't responding to specific requirements from a funder or an institution, [select here to write a generic DMP](#) based on the most common themes.

If applying for funding, select your research funder. European Commission (Horizon 2020) [Not applicable/not listed.](#)
Otherwise leave blank.

To see institutional questions and/or guidance, select your organisation. University of Edinburgh
You may leave blank or select a different organisation to your own.

Tick to select any other sources of guidance you wish to see.

- DCC guidance
- Roslin Institute

Create plan

Confirm plan details

Where your funder or institution doesn't have specific requirements (or if you left these options blank), you will see the DCC Checklist. This offers a generic set of DMP questions and guidance. For more details see: [DMP checklist 2013](#).

Funder: European Commission (Horizon 2020)
Institution: University of Edinburgh
Template: Horizon 2020 DMP
Other guidance:
DCC guidance

Cancel Yes, create plan

EI PGD: PGDonline

DMP ONLINE View plans Create plan About Future plans Help

Plan was successfully created.

My plan (Horizon 2020 DMP)

Plan details Initial DMP Detailed DMP Final review DMP Share

Please fill in the basic project details below and click 'Update' to save

Plan name:

ID:

Grant number:

Principal Investigator/Researcher:

Principal Investigator/Researcher ID:

Plan data contact:

Description:

DMP ONLINE View plans Create plan About Future plans Help

My plan (Horizon 2020 DMP)

Plan details Initial DMP Detailed DMP Final review DMP

- 1. Data summary** (1 question, 0 answered)
- 2. FAIR data** (4 questions, 0 answered)
- 3. Allocation of resources** (1 question, 0 answered)
- 4. Data security** (1 question, 0 answered)
- 5. Ethical aspects** (1 question, 0 answered)
- 6. Other** (1 question, 0 answered)

DMP ONLINE View plans Create plan About Future plans Help

My plan (Horizon 2020 DMP)

Signed in as Marisa Perez

0/9 questions answered
approx. 18% of available space used

Plan details Initial DMP Detailed DMP Final review DMP Share Export

- 1. Data summary** (1 question, 0 answered) +
- 2. FAIR data** (4 questions, 0 answered) -

In general terms, your research data should be 'FAIR' that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard or implementation-solution.

2.1 Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Guidance Share note

EC Guidance

The Research Data Alliance provides a [Metadata Standards Directory](#) that can be searched for discipline-specific standards and associated tools.

B / I / [List] [List] [Link] [Grid]

EI PGD: PGDonline

My plan (Horizon 2020 DMP)

Plan details | Initial DMP | Detailed DMP | Final review DMP | Share | **Export**

From here you can download your plan in various formats. This may be useful if you need to submit your plan as part of a grant application. Select what format you wish to use and click to 'Export'.

Initial DMP +

Detailed DMP -

Format


pdf

- csv
- html
- json
- pdf **(formatting values)** -
- text
- xml
- docx

Final review DMP +

Contact us | Terms of use | DMPonline previous version

© 2004 - 2017 Digital Curation Centre (DCC)



El PGD: Ejemplos

DMP title

Project Name Synthetic Chemistry / EPSRC example

Description This research project involves the development of a new chemical reaction for incorporating [your atom of choice] into [your molecule / compound of interest]. An experimental procedure will be developed that will allow the preparation of a range of compounds.

Funder Engineering and Physical Sciences Research Council

Institution University of Glasgow

Data Collection

What data will you collect or create?

The data produced from this work will fall into two categories:

1. The various reaction parameters required for optimisation of the chemical transformation.
2. The spectroscopic and general characterisation data of all compounds produced during the work.

I anticipate that the data produced in category 1 will amount to approximately 2 MB and the data produced in category 2 will be in the range of X - Y GB.

How will the data be collected or created?

The reaction conditions will be recorded and collated using Excel spreadsheets and named according to each generation of reaction.

The various experimental procedures and associated compound characterisation will be written up using the Royal Society of Chemistry standard formatting in a Word document. The associated NMR spectra will be collated in chronological order in a .pdf document.

These are standard practices for synthetic methodology projects.

Documentation and Metadata

What documentation and metadata will accompany the data?

The data will be accompanied by the following contextual documentation, according to standard practice for synthetic methodology projects:

1. spreadsheet documents which detail the reaction conditions.
2. text files which detail the experimental procedures and compound characterisation.

Files and folders will be named according to a pre-agreed convention.

The final dataset as deposited in the institutional data repository will also be accompanied by a README file listing the contents of the other files and outlining the file-naming convention used.

Ethics and Legal Compliance

How will you manage any ethical issues?

This document was generated by DMPonline (<http://dmponline.doc.ac.uk>)

1 of 3

There are no ethical issues in the generation of results from a synthetic methodology project. There are no human subject or samples involved.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

This project is being carried out in collaboration with an industrial partner. The intellectual property rights are set out in the collaboration agreement. The intellectual property generated from this project will be fully exploited with help from the University of Glasgow's IP and Commercialisation Office.

The aim is to patent the final procedure and then publish the work in a research journal.

Storage and Backup

How will the data be stored and backed up during the research?

The data will be stored on hard-drives belonging to the researchers involved in the work. These hard-drives are backed up onto the School of Chemistry's local servers.

How will you manage access and security?

Files created during this project will be encrypted so that only the PI and researcher will be able to access them. Data will be transferred between the PI and researcher on memory storage devices rather than by email.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Any data from this research which underpin or contribute to our patent application or subsequent research publications will be considered to be of long-term value and will be retained and preserved. These data would be suitable for sharing only once the intellectual property is protected by a patent.

What is the long-term preservation plan for the dataset?

Data which underpin our patent application and research publications will be stored on the School of Chemistry's server.

The dataset will also be deposited in Enlighten: Research Data, the University of Glasgow's institutional data repository. Data in the repository will be stored in accordance with funder and University data policies. Files deposited in Enlighten: Research Data will be given a Digital Object Identifier (DOI) and the associated metadata will be listed in the University of Glasgow Research Data Registry and the DataCite metadata store. The retention schedule for data in Enlighten: Research Data will be 10 years from date of deposition in the first instance, with extensions applied to datasets which are subsequently accessed. This complies with both University of Glasgow guidance and EPSRC policy.

Enlighten: Research Data is backed by commercial digital storage with is audited on a twice-yearly basis for compliance with the ISO27001 Information Security Management standard.

Data Sharing

This document was generated by DMPonline (<http://dmponline.doc.ac.uk>)

2 of 3

How will you share the data?

If the research is successful, the research will be protected by the filing of a patent. Following this, the research will be disseminated by the publication of an open-access manuscript in a chemical journal. The manuscript will be deposited in our institutional publication repository, Enlighten: Publications. The manuscript will contain a data citation indicating where and on what terms the data can be accessed.

The data which underpins the publication and patent will be made available for sharing via Enlighten: Research Data, the University of Glasgow's Data Repository. This will be available at the time of publication of the corresponding manuscript. Data in the repository will be issued with a Digital Object Identifier (DOI). This can be included as part of a data citation in publications, allowing the datasets underpinning a publication to be identified and accessed. DOIs will also be linked with appropriate records in Enlighten: Publications, the University's publication repository, to enhance visibility of datasets.

Metadata about datasets held in the Institutional repository will be publicly searchable and discoverable and will indicate how and on what terms the dataset can be accessed.

Information about datasets from the repository will be displayed on researcher profile pages on the University of Glasgow webpages which will also increase the visibility of the datasets.

Are any restrictions on data sharing required?

After the patent is filed to protect the intellectual property, there will be no restrictions on data sharing. The PI will actively disseminate the results as widely as possible.

Responsibilities and Resources

Who will be responsible for data management?

The PI will be ultimately responsible for data management of this project. The researcher will be responsible for organisation and storage of the data as it is produced during the project. The School of Chemistry's IT staff will manage the school's server, where the data will be stored. The University's library staff will be responsible for management of the institutional repositories, Enlighten: Publications and Enlighten: Research Data.

What resources will you require to deliver your plan?

The researcher already has the required software to implement the data collection plan. Funds to cover final deposit of the dataset in the institutional repository have been costed into the grant application as advised by the Research Data Management Service. Funds to support open access publication of the research are available from the institutional RCUK fund for this purpose.

This document was generated by DMPonline (<http://dmponline.doc.ac.uk>)

3 of 3

PGD de la Universidad de Glasgow realizado con DPM según requisitos de EPSRC:
http://www.gla.ac.uk/media/media_418166_en.pdf

El PGD: Ejemplos

<http://doi.org/10.5281/zenodo.48171>

HELIX NEBULA - THE SCIENCE CLOUD Grant Agreement: 687614

HELIX NEBULA
THE SCIENCE CLOUD

Helix Nebula – The Science Cloud

Deliverable Title: Data Management Plan
Partner Responsible: CERN
Work Package: 1
Submission Due Date: 26 January 2016
Actual Submission Date: 23 March 2016
Distribution: Public
Nature: Report

Abstract: This document describes the initial Data Management Plan (DMP) for the HNSciCloud project. It is based on the Guidelines on Data Management in Horizon 2020 document version 2.0 of 30 October 2015. "Participating projects will be required to develop a Data Management Plan (DMP), in which they will specify what data will be open."

Initial Data Management Plan 1

HELIX NEBULA - THE SCIENCE CLOUD Grant Agreement: 687614

Document Information Summary

Deliverable number: 1.1
Deliverable title: *Initial Data Management Plan*
Editor: *Jamie Shiers, CERN*
Contributing Authors: *Bob Jones, Rachida Amsaghrou, CERN*
Reviewer(s): *David Foster, CERN*
Work package no.: *WP1*
Work package title: *Consortium Management*
Work package leader: *Bob Jones, CERN*
Work package participants: *Rachida Amsaghrou, David Foster, Helge Meinhard, Jamie Shiers*

Distribution: *Public*
Nature: *Report*
Version/Revision: *1.0*
Draft/Final: *Final*
Keywords: *Data Management Plan, Data Preservation*

Disclaimer

Helix Nebula – The Science Cloud (HNSciCloud) with Grant Agreement number 687614 is a Pre-Commercial Procurement Action funded by the EU Framework Programme for Research and Innovation Horizon 2020.

This document contains information on the HNSciCloud core activities, findings and outcomes and it may also contain contributions from distinguished experts who contribute to HNSciCloud. Any reference to content in this document should clearly indicate the authors, source, organisation and publication date. This document has been produced with co-funding from the European Commission. The content of this publication is the sole responsibility of the HNSciCloud consortium and cannot be considered to reflect the views of the European Commission.

Grant Agreement Number: 687614
Start Date: 01 January 2016
Duration: 30 Months

Initial Data Management Plan 2

Más ejemplos en:

<http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples>

<https://doi.org/10.3897/rio.2.e11065>

RIO Articles About About Pensoft Books E-Books Blog Journals

Data Management Plan Research Ideas and Outcomes 2: e11065 (09 Nov 2016)
<https://doi.org/10.3897/rio.2.e11065> Reviewable v1

A Data Management Plan for Effects of particle size on physical and chemical properties of mine wastes

▼ Hong Xu, Mayu Ishida, Minglu Wang

Abstract ▼

data management plan; mine wastes; Mojave Desert; Souther California; NSF

Keywords ▲

Types of data produced ▲

Physical samples

- Bulk samples of different media including mine tailings, waste rock, streambed sediments, and background materials. All sample stored in borosilicate glass jars with Teflon lids.
- Streambed water samples stored in HDPE bottles.
- Field notes about information of bulk samples and water samples

Data of samples

- Data about the bulk samples in XLS/CSV
- Data about the water samples in XLS/CSV
- Scanned images of field notes in TIFF

Contents Article info Citation Cited

Article metadata
Types of data produced
Data and metadata standards
Policies for Access and Sharing
Policies for re-use and distribution
Plans for archiving and preservation
Acknowledgements

El PGD: Tendencias futuras

OpenAIRE tiene planes con **DMPonline**:

- Añadir una opción que permita a los proyectos depositar el DMP en Zenodo como medio para publicar un plan y obtener un DOI.
- Utilizar la API de OpenAIRE que permita seleccionar el PI, identificador de proyecto H2020 automáticamente, con el ID field del proyecto y enlazar el DMP con todos los outputs.

Otra herramienta para elaborar Planes de Gestión de Datos es **DMPTool** del UC3 of the California Digital Library. Ambos se obtienen online gratuitamente.

DMPTool + DMPonline = **DMPRoadmap**

DMPRoadmap: convergencia basada en DMPonline añadiendo funcionalidades de DMPTool. Disponible en GitHub. Foco en:



Machine-Actionable Data Management Plans (maDMP)

Cubren los mismos temas que los DMPs estandar pero secciones determinadas se rellenan con información obtenida de herramientas existentes. Se les puede considerar como metadatos extraídos automáticamente.

Habitualmente los investigadores escriben los PGD de manera manual, apoyados por una *checklist* y por una herramienta, pudiendo producirse errores tanto por quien lo escribe, o el momento en que se hace (al final de proyecto puede que datos importantes no estén disponibles nunca más).

EI PGD: maDMP

Ejemplos:

- **Data Stewardship Wizard** (cuestionario dirigido en función de las respuestas se abren nuevas ventanas, puede resultar largo pero lo consideran eficiente “*Actionable and computer readable*”)

DATA STEWARDSHIP WIZARD v0.3, KM: JAN 19, 2017

Questionnaire generated from the precompiled JSON of [Data Stewardship Knowledge Model](#). This is a proof-of-concept work not intended for serious use, yet.

Right now, saving the data is implemented for the ELIXIR All-hands workshop. Please register to receive updates. The functionality coming soon:

- Loading of stored data :-)
- Profile editing
- Password recovery (right now, send us an email)
- Multiple plans management
- Linking with [the book](#)

Privacy notice: Your personal and all other data are considered private and will not be shared with any third parties. Statistical information based on your data may be shared.

Technical contacts
[Robert Pergal](#)
[Marek Suchánek](#)

Data stewardship action team

Logos: elixir, ICB AS CR, FIT, elixir NETHERLANDS, DTL, and a network logo.

Introduction: In the data design and planning phase, we will make sure that we know what data comes when, that we have enough storage space and compute power to deal with it, and that all the responsibilities have been taken care of.

• Have you identified types of data that you will use that are used by others too? No, I am not using any common data types Yes, I will make sure to use common formats for common data types

• Will you be using new types of data? No, all of my data will fit in common formats Yes, I will need to use custom formats for some of my data

• Will you need to add fields in your data format to a data type registry? No, all of my data types are described in a data type registry already Yes, I will add new types to an existing data type registry Yes, I will create my own data type registry

• Do you need to create vocabularies or ontologies for any of your data items? No, suitable public controlled vocabularies or ontologies exist for all of my data types Yes, I will make and publish a vocabulary or ontology for some of my data

• Which data type registries will you use?

Sometimes the type of data you collect can not be stored in a commonly used data format. In such cases you may need to make your own, keeping interoperability as high as possible.

La Gestión de Datos de Investigación

Marisa Pérez Aliende

Universidad Autónoma de Madrid

mp.aliende@uam.es

Sevilla, 13 de Junio 2017

