|  |
| --- |
| El propósito del Plan de Gestión de Datos (PGD) es proveer un análisis de los principales elementos de la política de gestión de datos que se usarán por los solicitantes en relación con todos los datasets que serán generados por el proyecto. El PGD no es un documento fijo, sino que va evolucionando durante la vida del proyecto.  Ver [Ejemplos](#_EJEMPLO_1:) |
| Project title and brief description:  |  |  | | --- | --- | | Plan name |  | | Project identifier | - | | Project coordinator | - | | Contact details of the coordinator |  | | Contact details of the author/s of the Data Management Plan |  |  Título del proyecto , datos de contacto-identificación y breve descripción: |
| What data will be produced?/¿Qué datos se producirán? (Tipos)Origen de los datos[Ayuda](#_What_data_will) |
| How will data be documented and described?/¿Cómo se documentarán y describirán los datos? [Ayuda](#_How_will_data) |
| How will data be structured and stored?/¿Cómo se estructurarán y almacenarán los datos? Estándares usados. [Ayuda](#_How_will_data_1) |

|  |
| --- |
| Are there any ‘special’ requirements for your data?/¿Existe algún requisito “especial” para sus datos? [Ayuda](#_Are_there_any) |
| What are the plans for data sharing and access?/¿Cuáles son los planes para el intercambio de datos y para el acceso a los mismos? [Ayuda](#_What_are_the) |
| What are your main data challenges? Who can help? /¿Cuáles son sus principales retos en lo que concierne a los datos? ¿Quién puede ayudar? [Ayuda](#_What_are_your) |
| Who is responsible for managing the data?What resources (will you need?  1. Posibles Grupos de interés respecto a sus datos   **¿Quién se encarga de la gestión de los datos? ¿Qué recursos necesitará (ECONÓMICOS Y TÉCNICOS?** [**Ayuda**](#_Who_is_responsible) |

### No olvide revisar y actualizar su plan de gestión de datos de forma regular.

**Plan Gestión de Datos – Hoja de ayuda**  
Todo puede convertirse en datos de investigación si se utiliza con fines de investigación - los datos no son sólo números en una hoja de cálculo. Piense creativamente sobre los materiales que está utilizando y produciendo: lo que podría compartirse con otros investigadores que estén interesados en su trabajo; ¿Qué podría reutilizarse para producir nuevos conocimientos? Cualquier evidencia o material que sustente o arroje luz sobre sus hallazgos, sus publicaciones académicas, su tesis o su proyecto pueden ser considerados datos de investigación

|  |
| --- |
| What data will be produced?/ ¿Qué datos se producirán? **¿Qué datos físicos se estudiarán? (ej. Artefactos, muestras, archivos de papel, etc.)**  **¿Qué datos digitales se generarán? (ej. Notas de campo, imágenes, hojas de cálculo, entrevistas en audio, datos de sondeo, bibliografía comentada, etc.)**  **¿Qué formatos de archivo y qué software utilizará?** |
| How will data be documented and described? / ¿Cómo se documentarán y describirán los datos? **¿Serán sus datos entendibles por terceras personas? Documéntelos. Asegúrese de indicar claramente los valores de las tablas y hojas de cálculo.**  **¿Qué información se recogerá sobre la metodología de acopio de los datos?**  **¿Es importante para la investigación el ser reproducible? por qué/por qué no si lo fuera, ¿Qué documentación adicional o qué indicadores se requerirán?**  **¿Creará programas? ¿Escribirá código? ¿Dónde estarán documentados y almacenados para uso futuro?** |
| How will data be structured and stored? / ¿Cómo se estructurarán y almacenarán los datos? **Calcule qué cantidad de datos producirá a lo largo del tiempo - ¿Tiene suficiente almacenamiento?**  **Está Vd. haciendo pleno uso del almacenamiento protegido proporcionado por la Universidad?¿Cómo se guardarán los datos generados en el almacenamiento seguro de la Universidad?**  **¿Tiene alguna normalización lógica de la denominación de archivos y estructura de directorios?**  **¿Cómo gestionará el versionado para que se pueda identificar la versión actual de los documentos/datos?** |
| Are there any ‘special’ requirements for your data? / ¿Existe algún requisito “especial” para sus datos? **¿Son datos confidenciales? ¿Están debidamente almacenados y encriptados?**  **¿Despersonalizará sus datos?**  **¿ Establece el financiador de la investigación requisitos concretos para la gestión e intercambio de datos?**  **¿Habrá que destruir algunos datos? ¿Cuándo y cómo?** |
| What are the plans for data sharing and access in the short and long term?/ ¿Cuáles son los planes para el intercambio de datos y para el acceso? **¿Ha sopesado el intercambio de datos con sus colaboradores/supervisor de investigación?**  **Si su investigación involucra a personas, ¿ha obtenido el consentimiento debido para compartir datos?**  **¿Se pueden liberar sus datos de forma inmediata, o están sujetos a embargo (retrasar el acceso )?**  **¿Qué datos conservará? ¿Quién lo decide?**  **¿Estarán abiertos los datos para todo el mundo o habrá restricciones de acceso?**  **¿Durante cuánto tiempo estarán/deberán estar los datos disponibles?**  **¿Utilizará un repositorio de datos?¿Cuál?** |
| What are your main data challenges? Who can help? ¿Cuáles son sus principales retos en lo que concierne a los datos? ¿Quién puede ayudar? **¿Necesita formación o apoyo? ¿Dispone de ello?**  **¿Qué políticas universitarias o de otro ámbito son aplicables a su proyecto? ¿Las ha leido y comprendido?** |
| Who is responsible for managing the data? What resources will you need? ¿Quién se encarga de la gestión de los datos? ¿Qué recursos necesitará? **¿Quién es el responsable de los datos en las distintas etapas de su ciclo vital?**  **¿Hay suficientes recursos disponibles (habilidades, personas, almacenamiento, tecnología) para lanzar su plan?** |

## EJEMPLO 1:

**Project Name** Drosophila Genetics - BBSRC Example

**Description** This project will investigate the role of Polo kinase in metaphase to anaphase

transition in Drosophila melanogaster.

**Funder** Biotechnology and Biological Sciences Research Council

**Institution** University of Glasgow

**Data areas and data types**

**Outline the volume, type and content of data that will be generated e.g. experimental measurements, models, records and images**

This project will generate three main types of raw data.

1. Images from transmitted-light microscopy of giemsa-stained squashed larval brains.

2. Images from confocal microscopy of immunostained whole-mounted larval brains.

3. Western blot data.

Measurements and quantification of the images will then be recorded in spreadsheets.

Micrograph data is expected to total between 100GB and 1TB over the course of the project.

Scanned images of western blots are expected to total around 1GB over the course of the

project.

Other derived data (measurements and quantifications) are not expected to exceed 10MB.

**Standards and metadata**

**Outline the standards and methodologies that will be adopted for data collection and management, and why these have been selected**

All samples on which data are collected will be prepared according to published standard protocols in the field. All microscopes used for sample examination are serviced and recalibrated regularly. All Drosophila lines used in experiments are checked periodically for phenotypic markers. Drosophila are maintained in live culture according to standard methods in the field.

Files will be named according to a pre-agreed convention. The dataset will be accompanied by a README file which will describe the directory hierarchy and filenaming convention.

Each directory will contain an INFO.txt file describing the experimental protocol used in that experiment. It will also record any deviations from the protocol and other useful contextual information.

Microscope images capture and store a range of metadata (field size, magnification, lens

phase, zoom, gain, pinhole diameter etc) with each image.

This should allow the data to be understood by other members of our research group and add contextual value to the dataset should it be reused in the future.

**Relationship to other data**

**State the relationship to other data available in public repositories**

This dataset will provide a novel characterisation of Drosophila Polo kinase mutants

documented in the Flybase database. To the best of my knowledge, no other study has

perturbed the metaphase to anaphase transition in these mutants, then examined the

phenotypes seen in mitosis.

**Secondary Use**

**Outline the further intended and/or foreseeable research uses for the completed**

**dataset(s)**

The confocal and transmitted light images generated in this work may well be of use in the future. It is entirely possible that another study would want to measure a different aspect of mitosis in Drosophila (both the wild-type controls and the mutants) treated as per the protocols in this study.

I cannot see the western blot data being of future use.

**Methods for data sharing**

**Outline the planned mechanisms for making these data available, e.g. through**

**deposition in existing public databases or on request, including access Mechanisms where appropriate**

Datasets from this work which underpin a publication will be deposited in Enlighten: Research Data, the University of Glasgow’s institutional data repository, and made public at the time of publication. 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 funder policies.

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

The DOI issued to datasets in the repository 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 University Registry will be publicly searchable and

discoverable and will indicate how and on what terms the dataset can be accessed.

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

**Proprietary data**

**Outline any restrictions on data sharing due to the need to protect proprietary or**

**patentable data**

It is not anticipated that this study will generate any patentable data or proprietary data which would have to be protected.

**Timeframes**

**State the timescales for public release of data**

Data will be made available at the point of publication of the associated paper or publication.

**Formats**

**State the format of the final dataset**

Images will be stored as .tif

Data in spreadsheets will be stored as .csv

Data in freetext documents will be stored as .txt.

These formats are platform agnostic and should support future access and reuse.

Any data which has to be stored in a proprietary format will have the necessary software

(including version number) noted in the associated INFO.txt file.

Plan

**EJEMPLO 2: Data Management Plan: Opening access to economic data to prevent tobacco related diseases in Africa**

**Data Collection**

**What data will you collect or create?****How will the data be collected or created?**

**Data Collection methods**

**1. Desk-based search of official websites of project countries**

This will involve searching websites of government departments These will include data

on:

1. Tobacco production (Departments of Agriculture)

2. Prevalence of tobacco-related diseases, and tobacco-related morbidity and

mortality (Departments of Health)

3. Tobacco taxation (Internal Revenue Services)

4. Tobacco products manufacturing, tobacco imports/exports (Departments of Trade

and Industry)

5. Tobacco usage, from Surveys by National Statistics Agencies (NSAs). In South

Africa unit record administrative data from government departments, repackaged as

research datasets, are also shared by the NSA. If data collection instruments

(administrative forms) used to collect the data are available on these sites they can

provide useful information on the data.

**2. Desk-based search of websites of International Development Organisations**

This desk-based study will allow us to discover tobacco data on project countries that has

been collected by international organisations. The will include the websites of the

international DHS Program and UN bodies such as the World Health Organisation

(WHO). From this a "question bank" will be created of useful variables and the datasets

where these can be found.

**2. Desk-based search of industry websites**

The third component of our desk-based research will involve examing online records of the

tobacco industry. From these we hope to obtain data on: Cost of tobacco production, and

profits, in the industry, prices of raw tobacco and tobacco products, salaries, capital and

foreign investment, mergers and acquisitions, advertising spend, and regulations in the

industry

**3. Approaches to data holders**

We will create metadata on our platform for surveys already shared by others online. Our

desk search may also reveal the existence of datasets with a tobacco data components but

which are not in the public domain. In these cases we will approach the relevant research

projects in the project countries to release this data and allow the Project to host this on

their Open Data portal. This may be a fraught process but any challenges and successes

can be written up to inform our future work.

**4. Own surveys**

The project has already crowd-sourced data on current prices of tobacco products in two

project countries. This may be expanded during the course of the project to all project

countries. We will upload and share the metadata and data from these surveys.

**Documentation and Metadata**

**What documentation and metadata will accompany the data?**

**Supporting documents**

Data collection and data analysis documents will be shared along with the data files, where

available. Forms used for collecting administrative data will be shared with administrative

datasets. Data collection instruments (questionnaires, diaries) will be made available with

the survey data. Code lists used in collecting the data will also be provided. Final reports

from data collection projects will also be shared, where available.

**Metadata**

Each dataset will have a metadata record to help data users analyse the data. This

metadata record will be created during examination of the data and data collection

instruments. It will include information gathered on the dataset during the data collection

process. This documentation can be an invaluable source of provenance and usage

information for those analysing the data. Notes on data quality will form part of the

metadata record. Metadata will be created according to the Data Documentation Initiative

(DDI) international metadata standard, using Nesstar Publisher, which is free data markup

software for the creation of XML-compliant metadata according to the DDI standard.

**Ethics and Legal Compliance**

**How will you manage any ethical issues?**

The administrative data we will collect will mostly be in the public domain, in the form of

reports and other records from government departments. The survey data we will collect

will be anonymised data already shared with researchers, although not always online. The

industry data will be data made available to shareholders and the public. We are adding

value by bringing these sources together and providing a means for researchers to easily

discover and download these data.

However, we will endeavour to make data available that is not yet in the public domain. In

these cases, we will ensure that:

1. We have the necessary permissions from data owners to make these data open.

2. The data is suitably anonymised, to protect respondent confidentiality and privacy

3. We take national laws on sharing data across borders into account. Where such

restrictions exist, we will be unable to host this data.

4. We work with all stakeholders to ensure agreement on what will be shared, how,

and with whom.

4 Woolfrey L

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

The government data we will collect is not subject to IPR. The tobacco industry data we will

collect will be public records. We will therefore not publish information that would

compromise any IP rights. However, we will check each traunch of data we obtain, to

ensure we have permission to pass the data on to third parties.

**Storage and Backup**

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

The data would be stored on a server managed and backed up by the University of Cape

Town's Commerce IT Department. Curation of the data will be the responsibility of

DataFirst's Research Data Service. DataFirst is a technical partner on the Project. Each

preservation dataset will consist of data files, document files, metadata files, and any

programme files used in creating the data files. Data Service staff will be responsible for

adding data updates to datasets. We will also handle version control to ensure the most

recent and accurate data files are published, and provide tombstone citations to earlier

versions for verification or replication of research which may cite these supercede versions

of the data.

**How will you manage access and security?**

Access to the server hosting the preservation datasets will be password controlled.

Passwords will be allocated by the Commerce IT manager only to Data Service staff.

Server software will monitor data security and integrity.

**Selection and Preservation**

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

**preserved?**

Criteria for preservation will be:

1. Data is tobacco-related

2. Data covers project countries

3. Data is accurate and reliable (we will undertake quality audits to determine this)

4. Data is unit record data (not aggregated but available at the level at which it was

collected)

5. Data is not readily available from another repository

Retention:

It is difficult to predict what data has long-term value. Our policy will be to store unit record

tobacco data indefinitely. As these datasets grow, so will their value over time. Time-series

data will continue to be useful for economic and health policy research in the long term.

Sharing:

Because we aim to establish an Open Data portal, all data retained/preserved will also be

shared. The Project's policy is aligned to DataFirst's policy: We do not archive data which

cannot be shared with researchers in some form and at some access level.

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

There will be numerous datasets. Our long term preservation plan for the Project's data

holdings depends on the sustainability of DataFirst's Research Data Service. The service

was established in 2001 and is a unit at the University of Cape Town, a well- funded and

well-established university in South Africa. Our sustainability prospects are therefore good.

**Data Sharing**

**How will you share the data?**

The data will be shared as discrete datasets (by country, year, data source). DataFirst

hosts and shares data via an online dissemination platform, based on the National Data

Archive Open Source software developed by the World Bank's Development Data Group.

The platform provides a number of data access options. The Project's data will be shared

as Public Use data. That is, researchers will need to register on our site and say for what

purpose they will use the data, but access will be immediate and automatic, with no vetting

of use. The usage information we collect will be to support service improvements.

The data will be shared in a number of formats, includind excel spreadsheets, and data

files in the commonly used statistical analysis programmes (SPSS, Stata). We will also

make the data available as .csv files, in line with Open Data requirements.

**Are any restrictions on data sharing required?**

We aim to share the tobacco data we collect as public access data. We do not aim to

support research-use only requirements, as this is counter to Open Data principles. Policy

research, academic research, business analysis, and private sector innovation all need

good data, and countries benefit from informed decision-making in all these spheres.

6 Woolfrey L

**Responsibilities and Resources**

**Who will be responsible for data management?**

The Manager of DataFirst will be responsible for curating the Project's data. This is in line

with the project proposal for DataFirst to be funded to provide technical support. DataFirst's

Manager has 25 years' experience in managing research data and working with data users.

Skills learned from undertaking data rescue projects in South Africa will also be useful in

assisting with data collection activities.

**What resources will you require to deliver your plan?**

Funding for data collection as been budgeted for in the Project. This may need to include

funding to travel to project countries and negotiate with data collectors in government and

academia to release their data, and allow its reuse. Funding has been provided for a

Project Manager. The Project Manager is responsible for conducting online data audits and

downloading data, and populating the database which DataFirst's Manager will curate. This

will be a time- and labour- intensive task and more staff hours may need funded for this.